Computational Approaches to Stock Price Prediction

Jiachen Liu

Smeal College, Management Information Systems, Penn State University, State College, PA, United States of America
jonathan0330.liu@gmail.com

Abstract. Stock price prediction is one of the most significant and difficult problems of computational finance. As the volume of data and computing capabilities have rapidly increased, forecasting models have shifted toward a data-driven forecasting method based on machine learning and deep learning. In this paper, a comparative review of three key groups of stock prediction methods is presented. The initial part is about the conventional methods that are defined by technical and fundamental indicators wherein the interpretability of the technique and its limitation is discussed. The second section will discuss machine learning techniques of regression models, support vector machines, and a tree-based algorithm using an empirical example of Apple Inc. and Google to explain performance and errors. The third segment presents the deep learning algorithms, such as the Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Convolutional Neural Network (CNN) that are the current leading edge in the field of financial forecasting. This paper finds that the deep learning models tend to be superior in nonlinearity and long-term modeling nonlinear and long-term dependencies compared to classical methods, with the tradeoff that such models do need large datasets and are more expensive to compute. The results indicate the way in which the transformation between traditional analysis and the deep learning reflects the broader tramsformation in global financial data analysis and decision-making.

Keywords: Stock Price Prediction; Machine Learning; Deep Learning; Financial Forecasting; Computational Finance.

1 Introduction

Stock markets play an important role in economic systems around the world as they determine investment, risk, and the stability of the financial systems. Stock prices are thus of great importance in prediction to the investor, institutions, and the policymakers. Nevertheless, with the nonlinear, volatile, and highly dynamic characteristics of financial markets, predicting future prices is a challenging task. The price movements are complex and noisy since stock returns are conditioned by macroeconomic conditions, corporate fundamentals, and investor sentiment. This has motivated scientists to constantly come up with different methods of computation to achieve more accurate analysis of these trends [1, 2].

K. Subramanian (ed.), *Proceedings of the International Workshop on Advances in Deep Learning for Image Analysis and Computer Vision (IWADIC 2025)*, Advances in Computer Science Research 128,
https://doi.org/10.2991/978-94-6239-648-7_57

In the past, most investors were guided either by intuition, chart analysis, or simple statistical guidelines to determine the trend in the market. Such conventional techniques may be valuable in offering insights, even though they are generally ineffective when it comes to fast market change or an abnormal volatility. The recent decades saw the rise of machine learning and deep learning as the most fundamental paradigms of financial prediction. Such techniques can learn complex relationships in the huge volumes of data automatically and is therefore applicable to the problems with nonlinear time series like stock prices [3, 4].

This paper aims at summarizing and comparing the key techniques of stock price prediction. It speaks about three methods of analysis, such as traditional statistical and rule-based techniques, machine learning designs, like regression and SVM, and deep learning designs, including LSTM, GRU, and CNN. It consists of both conceptual argumentation and empirical investigation on the large-cap stocks in the U.S. (Apple, Microsoft, Amazon, and Google). It aims at showing the way prediction methods have developed and what the benefits of current algorithms are in financial decision-making.

2 Essentials of Stock Market Data.

In addition to scholarly attention, price forecasting of stocks is an important issue in the contemporary financial systems. These algorithmic trading systems operate on predictive signals to trade at a certain frequency excuting trades within milliseconds to execute high-frequency trades, whereas institutional investors operate on forecasting models to invest capital and deal with risks in a portfolio. The macroeconomic policy and financial stability are also supported by precise predictions, identifying the possible bubbles or market anomalies at an earlier stage. Moreover, the development of artificial intelligence has made it possible to include in an automated decision-making platform the stock prediction models that track sentiment, volatility, and liquidity in real time. This convergence of technology and finance explains the reason behind why predictive modeling has emerged as an increasingly active field of applied research in the world of computational economics that explicitly connects theory, data, and practice in an increasingly automated market space.

After presented the purpose and structure of this paper, it should be noted that the basic peculiarities of those data employed in stock market prediction should be outlined. Time series data of stock prices in the form of daily trading and price data are commonly used to forecast stock prices. The standard line of structure contains the opening price, highest price, lowest price, closing price, and trading volume which is also known as OHLCV. These variables would represent both the short-term and long-term market trends. Other features like moving averages relative strength index (RSI), and momentum indicators are also applied in the empirical studies with the aim of improving the predictive capability [5, 6].

In this paper, four large technology stocks will be included, these are, Apple, Microsoft, Amazon, and Google. The analyzed data is between the years of early 2022 and late 2025; these data were read at Yahoo Finance. All four stocks are associated with characteristic behavior and high comovement as they were all exposed to the U.S.

technology industry. A standardized change of their adjusted close prices indicates that Microsoft and Google registered the highest growth rates, whereas the recovery of Amazon trailed behind marginally since 2023. This is indicative of larger sector workings and divergent company underpinnings.

The moving averages of Apple show apparent medium and long-term trends. Short comparison MA 20 and MA 50 react promptly to changes in the prices whereas MA 200 shows the long-term trend. This is the same case with Google, with all the three moving averages coinciding well during stable periods and diverging in the event of a market shock as an indicator of change in momentum. Overall, these indicators assist in detecting crossover points, which may be applied in rule-based trading or used as input resources in the machine learning model.

The four stocks correlation analysis shows that the four stock are highly correlated with respectively positive coefficients (mostly higher than 0.75). This interdependence indicates that industry-movements have a strong impact on the price of stocks. The results of such findings provide justification to use multivariate learning techniques that consider the cross-stock dependencies, as opposed to using individual price series.

3 Prediction Methods

Once the general outline of stock market data has been done, the second thing to consider is to learn how scholars have gone about the problem of price prediction with various methodologies. The stock prediction analysis was conducted under two broad areas of focus, namely, traditional methods and machine learning-based methods.

To begin with, conventional methods. The stock forecasting technologies used traditionally are constructed based on two analytical philosophies, the technical analysis, and fundamental analysis.

Technical analysis assumes that all the information that is relevant is contained in prices and that past patterns can be used to forecast future patterns. Examples of these common tools are trend lines, moving averages, oscillators, and candlestick formations. There are also signals, like so-called golden cross or death cross, most used by traders when short and long-term moving averages meet each other that can be used to guess a possible reversal of the trend. These methods are simple and intuitive but are based on a significant amount of visual interpretation and subjective judgment, which makes them less consistent and scalable.

Fundamental analysis however examines the underlying value of a firm through the scrutiny of financial statements, cash flows and the economic indicators. Price-to-earnings ratios or debt-to-equity are some of the ratios which can be used to determine the undervaluation or overvaluation of a stock. Nevertheless, short-term price fluctuations can be forecasted using strong fundamentals not all times since markets tend to over respond to news or investor mood.

Both traditional approaches are limited in responding to the nonstationary and high frequency character of the contemporary financial data. Unpredictable events in the traditional analysis are human bias, emotional trading, and events around the world, which are difficult to explain by means of a traditional approach. These limitations encouraged the overall shift towards quantitative methods using data-driven methods and computational algorithms.

Secondly, machine learning algorithms. Machine learning allows finding nonlinear dependencies between input variables and future prices without making any direct assumptions about the data-generating process. The popular applications of machine learning in stock forecasting are the use of linear and polynomial regression, support vector machines, decision trees, random forests and gradient boosting models. All of them have their own advantages regarding the volume of data, noise, and complexity of the features.

Random Forest and Extreme Gradient Boosting are the two ensemble-based methods that have received a lot of attention due to their compromise between accuracy and interpretability. Random Forest uses a combination of decision trees and averages the output of these trees to mitigate variance and overfitting, so it is appropriate with financial data where outliers and noise are widespread [7]. This makes it be valuable in estimating feature importance enabling its analysts to determine which variables. The latter include such as, closing price, volume, or moving averages, which are most contributory to prediction accuracy. Random Forest models can however turn to be less effective once the feature space is extensive or when the data are sequentially dependent. Gradient boosting algorithms such as Extreme Gradient Boosting are more appropriate in such situations in the sense they optimize the weak learners in a repetitive scheme, to reduce the prediction error [6]. Research has indicated that the default objectives are characterized by the fact that the performance of Extreme Gradient Boosting has been observed to be superior to other regressors in both classification and regression in stock forecasting if aspect of parameter tuning, and cross-validation are properly handled. Although they are powerful, these models are completely static and unable to learn the temporal relationships directly, and this weakness can be partly addressed by deep learning architectures that process time-dependent dependencies in a cascade fashion.

Support Vector Regression (SVR) model was appllied to Apple's stock in this research. The technical model was referred to as Open, High, Low, Close, and Volume where forecasts on the following day price were made based on past such. The forecasted curve administered the overall market trend but underestimated very handsome upward moves afterward 2024. The error distribution showed that the model was effective when dealing with stable periods but not when dealing with high volatility. In the case of Google, slight success was noted by the support vector regression where the trends were much smoother. Root mean square error (RMSE) values were around 55 in the case of Google to 132 in the case of Microsoft consistent with prior findings showing comparable performance of boosting and support vector models on moderately volatile assets [6].

Besides regression, support vectors machine classifier was also tested to determine the directional changes. Within the four stocks, accuracy was 46-49 or slightly above chance. Majority of misclassifications were done in sideways markets in which the changes that occurred in a day were small. These findings are in line with the available literature that indicates that although support vector machine models are computationally efficient and robust, they do not have temporal memory hence, they do not capture sequential dependencies effectively [5].

However, machine learning algorithms such as the support vector machine are still useful in providing benchmarks. They need fewer data, they can be trained rapidly and they can provide feature importance that can be interpreted. As an illustration, the concern of the feature ranking in terms of correlation revealed that the following attributes had the strongest effect on the next-day predictions: close and adjusted closing price, then moving averages, volume. In general, machine learning is an essential intermediary phase between classic analysis and current deep learning, that is both more automatable and statistically rigorous yet not too costly to compute.

Machine learning methods offered a significant gap between old models and the new artificial intelligence. Based on this background, deep learning has come up as more advanced computational framework, which automatically discovers hierarchical patterns in raw market data. Deep learning architectures like Long Short-Term Memory, Gated Recurrent Unit and Convolutional Neural Network, can directly learn temporal and nonlinear relationships on stock time series without any manual design of indicators.

The Long Short-Term Memory (LSTM) model can be capturing long-term dependencies that store and forget information over time steps using gated cells using long-term memory model introduced by Hochreiter and Schmid at [3]. Fischer and Krauss [1] used an LSTM network to forecast the movement of S&P 500 stock in the market in the period between 1992 and 2015. Their model beat out logistic regression and random forests and provided a directional improvement of approximately 7 per cent and led to positive cumulative returns when in backtests. On the same note, Nelson et al. applied LSTM to predict price changes of significant stocks in the United States and established that it performed better than multilayer perceptrons and SVMs [2].

The gated recurrent unit is a simplified version of LSTM with fewer parameters and could be trained faster, though it also has a similar accuracy. Chen et al. were able to suggest a better gated recurrent unit model to reconstruct missing data and used industry details, which were more effective at generalizing on smaller datasets [4]. The gated recurrent unit has these properties, rendering it interesting in financial forecasting where the data to work with is usually incomplete or noisy.

Even though the convolutional neural network was originally created to be used with image recognition, it has shown its efficiency in analyzing time-series. As illustrated by Hao [8], a hybrid model of convolutional neural network-long short-term memory had the ability of capturing the local patterns through convolution and the long-term correlations through the recurrent layers. The hybrid architecture had a lowered root mean square error than either CNN or LSTM alone, indicating the usefulness of architecture mixture. On the same note, Sonani et al. combined LSTM and Graph Neural Networks to simulate the relationships between multiple stocks, which demonstrate better results in multi-asset forecasting problems [9].

Deep learning models are more capable of dealing with the nonlinear and high-dimensional interdependency of financial data than classical machine learning approaches. Automatic feature extraction decreases manual preprocessing since they can do it by themselves. Yet, these benefits have their own price, to implement deep learning, it is necessary to deal with large datasets and massive computational resources and be careful not to overfit the network. Furthermore, the interpretability is a problem since the inner representations acquired by these models are likely to be incomprehensible to the human researchers [10].

These issues notwithstanding, deep learning is now being promoted as a stock price predictor of choice both in academia and industry. According to the general tendency of the research, long short-term memory, and gated recurrent unit perform best in the case of univariate or multi-variate sequential forecasting, whereas convolution neural network-based hybrids perform better in high-frequency or in multi-asset contexts [11].

4 Conclusion

In this paper, the development of the stock price prediction approaches through the classical statistics analysis to the recent deep learning-based ones was reviewed. Conventional techniques, such as technical and fundamental analysis, are a great source of intuition, yet not very resistant to noise and nonlinear changes. Machine learning models such as SVM and Random Forest introduced automation and the ability to model nonlinearity and provided good base lines on small to medium datasets. But they are constrained by the fact that they do not have the ability to encode temporal dependencies between sequences.

Deep learning-based models, in particular, LSTM and GRU, and CNN models have transformed financial prediction by building more complex correlations in time series. Both the previous literature and the data utilized in this study provide empirical evidence that deep architectures are much more likely to enhance accuracy of prediction, especially in volatile situations. Simultaneously, they require more data and computational resources, which may be inaccessible to smaller organizations.

Multimodal and interpretable learning will probably be featured in research in financial forecasting. Modern markets generate enormous volumes of unstructured data, news headlines and social media posts, investor sentiment, and macroeconomic indicators that can be added to numerical data on prices. By incorporating such heterogeneous sources, it is possible to make the models aware of not only the changes that occur in the market, but also the reasons that cause such changes. An additional opportunity is Explainable Artificial Intelligence that strives to make the deep models more open and credible by pointing out the important characteristics upon which predictions are based. This would ensure regulators; fund managers and retail investors are more comfortable adopting AI tools. Lastly, as the computing resources and data keep on increasing, greater expectations are raised to develop energy-efficient and sustainable algorithms that can maintain prediction ability but consume less energy. The development of this nature will help to make sure that not only accuracy but also the larger objectives of ethical and sustainable technological advancement can be achieved in the sphere of AI application in the field of finance.

References

1. Fischer, T., Krauss, C.: Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research 270(2), 654–669, 2018
2. Nelson, D.M., Pereira, A.C., de Oliveira, R.A.: Stock market's price movement prediction with LSTM neural networks. In: Proc. Int. Joint Conf. Neural Netw. (IJCNN), 1419–1426, 2017
3. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Computation 9(8), 1735–1780, 1997
4. Chen, C., Li, J., Xu, H.: Improved GRU-based stock price prediction using reconstructed datasets. Applied Sciences 13(15), Art. 8813, 2023
5. Patel, A., Shah, S., Thakkar, P., Kotecha, K.: Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. Expert Systems with Applications 42(1), 259–268, 2015
6. Zheng, Y.: A comparative study of XGBoost and LSTM in stock price prediction. Procedia Computer Science 226, 842–850, 2024
7. Vijh, R., Chandola, R., Saurabh, A.: Stock closing price prediction using machine learning techniques. Procedia Computer Science 167, 599–606, 2020
8. Hao, Y., Gao, Q.: Predicting the trend of stock market index using the hybrid neural network based on multiple time scale feature learning. Applied Sciences 10(11), Art. 3961, 2020
9. Sonani, M.S., Badii, A., Moin, A.: Stock price prediction using a hybrid LSTM-GNN model. arXiv preprint, arXiv:2502.15813, 2025
10. Bhandari, H.N., Ibba, S.N.: Predicting stock market index using LSTM. Machine Learning with Applications 10, 100–115, 2022
11. Rahmadeyan, A., Khotijah, N., Permana, A.C.: Long short-term memory and GRU for stock price prediction. Procedia Computer Science 240, 201–210, 2024