Dean&Francis
ISSN 2959-6157

Application of Machine Learning in Stock Price Prediction

Zeyu Chen[1]

[1]Software Engineering, Macau University of Science and Technology, Macau, China
1220024771@student.must.edu.mo

Abstract:
In the financial world, it is always difficult to predict stock prices due to the high volatility of stock prices and the noise inherent in market data. And because of some of the benefits of machine learning techniques, researchers and practitioners are increasingly adopting these techniques to increase the accuracy of predictions and support investment decisions. Therefore, this paper provides classification model (random forest, logical regression), linear regression model (linear regression, XGBoost regression), time series model (LSTM, Prophet) and other complex machine learning models. It explains the theoretical basis, key functions, and practical applications in finance and stock markets. Furthermore, it also explores the deployment of these models in specific contexts such as short-term trading guidance, portfolio rebalancing, and quantitative trading strategy development. It also addresses critical challenges associated with ML-based prediction, including data quality, model overfitting, and the need for robust risk management frameworks. Finally, this paper critically examines the essential advantages and limitations of ML in this sphere and suggests a positive outlook for future research, including the integration of unstructured data and the development of hybrid models that combine statistical and deep learning approaches.

Keywords: machine learning; stock market; classification model; linear regression model; time series model

1. Introduction

Stock markets are a key barometer of economic health and a fundamental platform for global capital allocations. Accurate prediction of stock prices is constantly challenged by the complex, non-linear and often chaotic nature of market dynamics. However, traditional methods such as technical and fundamental analysis have limitations in capturing the dependencies of complex nonlinear stock market models and large financial data sets.

The advent of machine learning has introduced powerful computer tools that can autonomously learn from past data sets, identify potential patterns, and generate data-driven predictions.

The purpose of this paper is to systematically sum-

ISSN 2959-6157

marize the application of machine learning models in stock price forecasting and explain a series of algorithms including regression models, classification models and time series models. This paper introduces them by checking the effectiveness of each transaction and investment scene. The purpose of this study is to provide a structured understanding for researchers and practitioners, and to explore the potential applications of machine learning techniques and the limitations of financial forecasting.

The following part of this paper is organized as follows: Chapter 2 illustrates the main machine learning models will use in the stock price prediction, discussing their theoretical underpinnings. Chapter 3 examines the practical applications of these models in short-term trading, portfolio adjustment, and quantitative strategy development. Chapter 4 presents a discursive analysis of the advantages and disadvantages identified. Finally, Chapter 5 concludes the paper and outlines avenues for future research.

2. Main Models of Machine Learning

2.1 Regression Models

This section mainly illustrates some regression models used in the analysis of stock price prediction. The main role and core formula will be illustrated.

2.1.1 Linear Regression

Establishes a linear relationship between features and the target. It is simple, interpretable, and coefficients indicate feature influence. However, it is sensitive to outliers and assumes normally distributed residuals.

Main Function:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n \quad (1)$$

ŷ: Predicted value; β_0: Intercept β_i:Regression coefficients x_i : Input features

The model's primary value lies in the economic interpretability of its coefficients: the coefficient β_i quantifies the marginal effect of feature x_i on the target variable $\hat{y}$. A positive β_i indicates that the factor is positively correlated with the expected price movement, and vice versa. The magnitude of its absolute value reflects the strength of the influence. This allows researchers to clearly identify and evaluate the driving effects of different risk factors on asset prices.

2.1.2 EXtreme Gradient Boosting Regressor

XGBoost is an efficient implementation of Gradient Boosting Decision Trees (GBDT), iteratively training weak learners and minimizing an additive objective function comprising loss and regularization terms. This design effectively controls model complexity and enhances generalization.

Main function:

$$Obj = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t)}) + \sum_{k=1}^{t} \Omega(f_k), \Omega(f) = \gamma T + \frac{1}{2}\lambda \| w \|^2 \quad (2)$$

$\sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t)})$:the empirical loss, i.e., the discrepancy between the true value y_i and prediction value $\hat{y}_i^{(t)}$; $\sum_{k=1}^{t} \Omega(f_k)$: a regularization function, where $\Omega(f) = \gamma T + \frac{1}{2}\lambda \| w \|^2$

In financial applications, XGBoost is particularly useful for credit risk assessment, stock trend classification, fraud detection, etc., but the structure of financial data enables tree-based methods.

This model is highly efficient because it is resistant to missing values and supports parallel computation. It's a bit complicated and easy to predict, so users will adjust the parameters.

2.2 Classification Models

This section is about classification models. This is a model that needs to be tagged like a known input and the correct class corresponding to it. The output is a classification model.

2.2.1 Random Forest

Random forest is a holistic learning method based on a bagging strategy. Its basic idea is to use boot sampling and functional subspace selection to form several independent decision trees and aggregate those predictions by a large number of votes.

Main function:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^{T} h_t(x) \quad (3)$$

T is the total number of decision trees and $h_t(x)$ is prediction of the t-th decision tree.

The main advantage of this generic mechanism is overadaptive tolerance, which improves the model's robustness and generalization ability. In the stock market, random forests are used for credit risk prediction, fraud detection, and stock market trend classification. For example, by predicting stock price movements, structural features such as historical prices and trading volumes can be exploited.

Overall, random forest is important in sampling and feature selection of data, even noisy or unbalanced data sets that are common in financial markets, and can enhance the performance of generalization. Can produce stable forecasts.

2.2.2 Logistic Regression

It is a widely used classification algorithm, applicable in binary decision-making contexts. This method performs well when there exists a linear relationship between the features and the class label and is characterized by excellent interpretability and computational efficiency.
Main function:

$$P(y=1|x)=\frac{1}{1+e^{-(\beta_0+\beta_1x_1+\beta_2x_2+\ldots+\beta_nx_n)}} \tag{4}$$

Here, $P(y=1|x)$ denotes the probability of a positive outcome, while β_i represents the coefficients learned from the data. This function maps any real numeric input to sections (0,1)(0,1) (0,1) allowing probabilistic interpretation. The LR algorithm, used in finance, is frequently used to classify market direction. This algorithm predicts whether stock prices will rise or fall based on technical indicators and macroeconomic characteristics. Its low computational costs and compatibility with large data sets make it widely used in the stock market.

2.3 Time Series Model

Time series is a series of data points arranged in time series over a specific period of time and can be analyzed and processed as discrete time.

2.3.1 LSTM model(Long-Short Time Model)

LSTM is an advanced RNN architecture that can efficiently capture long-term dependencies. This concept is particularly well-suited for solving the zero-slope problem and modeling the complex temporal dependencies inherent in stock price sequences. Recent studies have confirmed its effectiveness.
Bathla demonstrated that LSTM supports vector regression and provides high accuracy in stock price prediction [1]. Ashtagi et al. also conducted a comparative study of several machine learning models and confirmed that LSTM outperforms traditional regression methods in dynamic financial environments [2].
Main function:

$$h_t=f\left(x_t,h_{t-1},c_{t-1}\right) \tag{5}$$

h_t is the hidden state. x_t is the current input. c_t represents the memory cell that stores long-term information.
This mechanism allows the model to retain relevant information over time, making it particularly effective for forecasting financial time series.

2.3.2 Prophet Model

Prophet is a time series model developed by Facebook that decomposes data into trends, seasonality, and holiday effects. It provides a flexible framework that handles missing data and outliers, and is widely used in business forecasting tasks such as stock price and demand forecasting. Pant et al. highlighted Prophet‘s interpretability, particularly in decomposing complex seasonality in financial data [3]. Similarly, Susindran et al. demonstrated a practical application of Prophet in a stock market prediction web application, combining it with LSTMs to achieve robust real-time predictions [4].
Main functions:

$$y(t)=g(t)+s(t)+h(t)+\epsilon_t \tag{6}$$

g(t) is the long-term trend, s(t) is seasonality, h(t) is the holiday or event effects, and ϵ_t captures the random error.
This decomposition makes the model interpretable and practical in financial forecasting.

3. The Application of Machine Learning in Stock Market

3.1 Short-Term Trading Guidance

Short-Term trading is a kind of trading that continue a few days or weeks, which means investors need to predict the stock market in these a few days, so invenstors need to use some linear regression models to illustrate the stock market fluctuation [5][6]. As mentioned in part 2, XGBoost and Simple Linear Regression Model can help to predict the trend with a little noise point. XGBoost is a kind of model that can capture non-linear relationships and improve prediction accuracy. And investors can also use Time Series model LSTM. LSTM is particularly effective in modeling sequential dependencies in intraday time-series data, enabling traders to capture sudden market fluctuations, the guidance is also combined with SVM [7]. In 2019, Ramos-Pérez et al proposed “a model based on a set of Machine Learning approaches namely Gradient Descent Boosting (GDB), Random Forest (RF), Support Vector Machine (SVM) and Artificial Neural Network (ANN) to forecast S&P500 volatility. The model relied solely on market data and used predictions from these algorithms as inputs for a final artificial neural network [7].

3.2 Portfolio Adjustment

Portfolio adjustment is a kind of method that considers different kinds of situations, then combine them together and conduct a comprehensive financial analysis of the stock market. According to this , Random Forest can

ISSN 2959-6157

classify different situations into outperforming, neutral, or underperforming categories, then help investors in rebalancing their portfolios [8]. Also Logistic Regression is employed to estimate the probability of price increase or decline, supporting binary investment decisions [7]. Different models can also be used for comprehensive consideration of Portfolio Adjustment, such as stock market analysis based on neural networks [9].

Table 1. Results for different algorithms

Models	RMS	R2	MAE
Moving Average	65.19206217	-2.036683033	58.03018002
Linear Regression\[{R}^{2}\]	249.4082088	-43.43054073	247.0054433
KNN	212.50602332	-31.28301299	170.4863214
Auto Arima	143.9723997	-13.81057611	121.9446558
Facebook prophet	131.428607	-11.34212199	109.8763046
LSTM	9.424258267	0.8194021263	8.625745379

In Table 1, the rows represent different techniques used like Moving Average, Linear Regression etc. and columns represent different evaluating parameters including RMS, R2, and MAE [9].

3.3 Quantitative Trading Strategy Development

Quantitative trading strategies rely on rule-based signals derived from predictive models. LSTM is a special type of repetitive neural network that learns long-term dependencies of data. In addition, Prophet is a predictive tool based on a decomposable addition model. This model is particularly suitable for corporate time series data, which is highly seasonal and affected by holidays, spanning multiple periods. It is also suitable for datasets covering multiple historical periods and effectively handles outliers and missing values [1]. Therefore, this time series model can reflect a variety of situations that can occur in the stock market.

3.4 Risk Management And Market Anomaly Detection

Risk management and market anomaly detection are important aspects of stock market applications. Traditional models fail to recognize sudden fluctuations and increasing exposure to financial risk. Recent researches have shown that machine learning offers a more adaptive and robust approach in this field. As Lumoring et al. noted, a holistic approach is effective in detecting short-term volatility peaks and identifying anomalies in financial data sets [10]. Patil et al. have developed a real-time forecasting system that dynamically adjusts portfolio strategies in response to new risk signals [11]. Kumar et al. noted that a hybrid framework combining deep learning and traditional methods improves noise tolerance and responsibility for anomaly detection results at the same time [12]. It is important to note that machine learning plays a key role in providing early warning mechanisms, mitigating risk for investors and ensuring more sustainable investment results.

4. Discussion

4.1 Advantages

LSTM and Prophet models make it easy to predict long-term or short-term stock market movements. LSTM automatically provides short-term trading predictions, while the Prophet model helps analyze and predict future stock returns and the appropriate volatility range. Using Machine Learning to predict stock markets, it is possible to distinguish between types of stocks as well. And it can work fully automatically, using very little human resources to achieve stock market trading and trading.

4.2 Disadvantages

The stock market is inherently volatile, characterized by substantial risks and fluctuations—an aspect that poses significant challenges to stock market prediction using machine learning. Owing to such volatility, traders are compelled to take a broader range of factors into account. When employing predictive models, corresponding noise points should also be incorporated; yet, even with this adjustment, non-negligible errors persist. Consequently, large-scale application of machine learning for stock market prediction remains unstable. These models often only identify relevant linear relationships, thereby overlooking errors that exist in real-world scenarios.

5. Conclusion

5.1 Summary in Findings

All above, this paper has provided a systematic investigation into application of machine learning techniques for stock price prediction, with particular emphasis on Regression, Classification, Time-Series based approaches. The comparative analysis of representative models such as Linear Regression and XGBoost in regression tasks, Random Forest and Logistic Regression in classification, and advanced temporal architectures such as LSTM and Prophet, they illustrate that machine learning offers promising alternatives to traditional forecasting methodologies by capturing hidden structures and nonlinear dependencies within financial data.

5.2 Practical Implications

The analysis indicates that machine learning models can effectively capture both linear and non-linear relationships within financial data, providing valuable insights for short-term trading signals, portfolio management strategies, and the development of automated quantitative trading systems. Nevertheless, the results also underscore that stock markets remain highly complex and inherently volatile systems, where external shocks, market sentiment, and macroeconomic factors often generate unpredictable noise that cannot be fully explained by data-driven models. So, issues such as model interpretability, robustness under distributional shifts, and overfitting risks limit their widespread adoption in real-world financial environments.

5.3 Future Work

And for future research, they should prioritize the development of more robust and interpretable model architectures. First, researchers can hybrid frameworks that combine machine learning algorithms with traditional econometric or domain-specific knowledge may enhance both predictive accuracy and explanatory power. Second, the integration of alternative data sources, such as textual sentiment, macroeconomic indicators, and real-time transaction-level data may further enrich model inputs and provide a more holistic representation of market dynamics. Finally, based on the real-time situation of the market, establish and improve a sound risk assessment mechanism and allocate the most suitable models for market types, and robust evaluation frameworks that incorporate risk-adjusted performance metrics and stress-testing methodologies are essential for ensuring practical applicability in financial decision-making.

5.4 Future direction

In conclusion, while machine learning provides a potent and increasingly sophisticated toolkit for stock price prediction, its application necessitates prudence, because machine learning is not a panacea for financial forecasting but represents a transformative tool when carefully integrated with financial theory and domain expertise. The future development direction should focus on building models that are precise, interpretable, robust and adaptable, so as to better cope with the rapidly evolving market environment. Moreover, through a collaborative approach that combines these data-driven technologies with solid financial theories, a sound risk management framework and indispensable human expertise, It can better apply machine learning to the financial market.

References

[1] G. Bathla, Stock Price prediction using LSTM and SVR, 2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC), Waknaghat, India, 2020:211-214.

[2] R. Ashtagi, A. Deshmukh, S. Dhomse, G. Dhumal, S. Dhurve, A Comparative Study of Machine Learning Models for Stock Price Prediction, 2024 2nd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS), Erode, India, 2024:768-773.

[3] J. Pant, H. K. Pant, P. Kotlia, D. Singh, V. K. Sharma and J. Bhatt, Utilizing the Prophet Model for Time Series Based Stock Price Forecasting, 2024 5th International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2024:13-17.

[4] S. Susindran, P. Prabu, S. Nagarajan and R. D. Bharathi, Stock Market Prediction Web App Using LSTM Model and Facebook Prophet, 2024 International Conference on Smart Technologies for Sustainable Development Goals (ICSTSDG), Chennai - 600077, Tamil Nadu, India, 2024:1-5.

[5] S. Ravikumar and P. Saraf, Prediction of Stock Prices using Machine Learning (Regression, Classification) Algorithms, 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 2020:1-5.

[6] A. Behera and A. Chinmay, Stock Price Prediction using Machine Learning, 2022 International Conference on Machine Learning, Computer Systems and Security (MLCSS), Bhubaneswar, India, 2022:3-5.

[7] P. P. Kadu and G. R. Bamnote, Comparative Study of Stock Price Prediction using Machine Learning, 2021 6th International Conference on Communication and Electronics Systems (ICCES), Coimbatre, India, 2021:1200-1204.

[8] A. Ruke, S. Gaikwad, G. Yadav, A. Buchade, S. Nimbarkar and A. Sonawane, Predictive Analysis of Stock Market Trends: A Machine Learning Approach, 2024 4th International Conference

ISSN 2959-6157

on Data Engineering and Communication Systems (ICDECS), Bangalore, India, 2024:1-6.
[9] S. Kumar, R. Gupta, P. Kumar and N. Aggarwal, A Survey on Artificial Neural Network based Stock Price Prediction Using Various Methods, 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2021:1866-1872.
[10] N. Lumoring, D. Chandra and A. A. S. Gunawan, A Systematic Literature Review: Forecasting Stock Price Using Machine Learning Approach, 2023 International Conference on Data Science and Its Applications (ICoDSA), Bandung, Indonesia, 2023:129-133.
[11] A. Patil, K. Malap, H. Manna, S. Mahto and Y. Raut, „Real-Time Future Stock Price Prediction Using Machine Learning Algorithms," *2023 International Conference on Advanced Computing Technologies and Applications (ICACTA)*, Mumbai, India, 2023, pp. 1-5.
[12] A. Kumar, M. Trivedi, Vikas and S. K. Upadhyay, „A Machine Learning-Based Analysis of Stock Market Forecasting: A Review," *2024 1st International Conference on Advanced Computing and Emerging Technologies (ACET)*, Ghaziabad, India, 2024, pp. 1-5.