Dean&Francis
ISSN 2959-6157

# The Investigation Related to the Role of Machine Learning in Predicting Stock Price During the COVID-19 Pandemic

**Yiou Ding**

School of Mathematical Science, University of Nottingham, Nottingham, United Kingdom

pmyyd1@nottingham.ac.uk

**Abstract:**
The COVID-19 pandemic has caused turbulence in the world's financial markets on an unprecedented scale, testing the validity of conventional models that forecast the stock price. Having unrivaled capabilities for processing big volumes of data and complicated relationships, machine learning algorithms have emerged as one of the powerful tools available for forecasting stock prices in the COVID-19 period. The paper aims to provide a review and discuss various models developed using machine learning, specifically Linear Regression, Support Vector Machine (SVM), Decision Trees, Random Forests, Long Short-Term Memory (LSTM), and Reinforcement Learning, and their individual effectiveness regarding stock price prediction in the face of pandemic uncertainty. The performance of the models has ranged between good performance to successful handling of market fluctuations, yet issues with interpretable results and integration of external factors like news and policies have persisted. Expert systems, explainable tools like SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME), Automated Machine Learning (AutoML), and transfer learning can be considered for overcoming the challenges mentioned above and improving the models' performance in a dynamically changing environment.

**Keywords:** Machine learning; reinforcement learning; stock market; COVID-19.

## 1. Introduction

The COVID-19 pandemic has brought unprecedented turmoil to financial markets worldwide, thereby creating extraordinary volatility and uncertainty. Because of their grounding on historical data, traditional financial models have fallen short in capturing such rapid changes in a timely manner. It therefore follows that there is an emerging need for more enhanced methods which can deal with this complexity

ISSN 2959-6157

and uncertainty.

With regard to handling huge amounts of data and modeling nonlinear relationships, the machine-learning algorithms for stock price predictions have turned out to be one of the most efficient tools during the COVID-19 period. For instance, Saxena et al. demonstrated the efficiency of Long Short-Term Memory (LSTM) networks for stock price prediction based merely on historical data and other factors, such as sentiment from social media, which became highly relevant during the pandemic [1]. There are DRL methods applied to automatic trading strategies, such as DQN and PPO, which have proved robust in extremely volatile markets during the pandemic [2].

Moreover, Khattak et al. used Least Absolute Shrinkage and Selection Operator (LASSO) regression to predict the crisis of COVID-19 on European stock. According to the system, it was quite effective in choosing the most important predictors: S&P500, indices from Singapore, Switzerland, Spain, and most importantly France and Germany. Other evidence of internal economic shocks was given by the LASSO model, and which macroeconomic indicators determine the performance of the European stock market throughout the pandemic [3].

This paper reviews various machine learning models, namely Linear Regression, SVM, Decision Trees, Random Forests, LSTM, CNNs, and Reinforcement Learning, with regard to stock price predictability engendered by the COVID-19 pandemic. Further discussion is related to some important issues set by such a task, including interpretability challenges and how to integrate other information-like news and changes in policy. Finally, the possibility of enhancements in the future with the use of expert systems, explainable AI techniques such as SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME), Automated Machine Learning (AutoML), and transfer learning will be considered.

# 2. Machine Learning Techniques for Stock Price Prediction

## 2.1 Traditional Machine Learning Methods

### 2.1.1 Introduction and feature selection

Traditional machine learning techniques have been referred to very frequently in stock price prediction, especially during the COVID-19 pandemic. Common techniques in the domain include linear regression analysis and support vector machines. Linear regression models the relationship between stock prices and independent variables, such as historical prices or COVID-19 cases. While this is easy to apply, the conventional barrier to its application is that inherently nonlinear patterns usually come into stock-market dynamics when crises show up. Support Vector Machines can handle such issues by transforming data into higher-dimensional spaces in which complex nonlinear relationships can be modeled. These Support Vector Machines would analyze the trends of stock prices and hence it becomes relevant during market uncertainties caused by COVID-19 [4, 5]. Decision trees give relatively interpretable models by splitting data on selected features such as trading volumes and pandemic data. However, they tend to be susceptible to overfitting, especially in highly volatile markets [3, 5]. In general, ensemble methods, like Random Forests, avoid overfitting by averaging the predictions of a number of trees. Feature selection is fundamental in these models. The common features used include historical stock prices, trading volumes, and even COVID-19 case data-reliant on social media sentiment, which emerged as an important predictor during the pandemic itself [4-6].

### 2.1.2 Applications and performance

Various models were proposed, such as random forests and support vector machines, which were quite good at handling nonlinear relationships and high dimensions of data. For example, Zhu et al. presented an SVM-based model that integrated the data of the stock market and COVID-19 case numbers in order to forecast the stock price fluctuations during the pandemic period. This model achieved notable advancement compared to the traditional ones because of their higher accuracy and adaptability to the very turbulent evolution of market conditions.

In contrast, only the main preprocessing of stock market data allowed the production of valid models. The stock price normalization, filling in missing values, and scaling features are forced to be done to maintain feature consistency. Other major tasks are the aggregation from third-party sources that represent COVID-19 case counts, governmental policy indices, and social media sentiment originating from sources like Twitter [2, 4, 7]. During preprocessing, the volatile nature of the stock price data was smoothed, and outliers were removed to model the data for extreme market fluctuations during the pandemic.

In this machine learning modelling for stock price predictions, architectures varied according to the underlying methodology. Random forests and decision trees were tree-like, with data branched into different parts based on feature importance. These models also performed in chaotic nonlinear stock prices during the COVID-19 crisis 9, 4.

Another important innovation in the application of these models during the pandemic was their integration with external, non-financial data. Data included but were not limited to Google Trends data, Twitter sentiment analyses, and COVID-19 case counts that gave holistic evidence of the drivers of stock prices. These allowed the models to take into account general social and economic repercussions of the pandemic that, until then, were not part of stock market forecasting.

Other trends involved hybrid models that were ensembled with both deep learning and traditional machine learning methods. One is that many researchers combined random forests and LSTM models to take advantage of the interpretability of a random forest and the time-series capabilities of LSTM for better improvements in performance. This hybrid approach preceded others due to the highly volatile market conditions resulting from this pandemic, in accuracy and adaptability.

## 2.2 Neural Networks

### 2.2.1 Introduction and feature selection

These are powerful tools that can model complex nonlinear associations-a highly relevant capability in the case of using neural networks like MLPs, CNNs, and LSTMs to predict the stock prices during the COVID-19 pandemic. Being the simplest variant, neural network models learn underlying nonlinear associations between features and stock prices. However, it performs very poorly with sequential data and therefore cannot be applied to a time series forecasting problem. Typically, the architecture of CNNs is utilized in image recognition; however, it has been modified to adapt to stock prediction by analyzing historical trends in prices. It works well in the identification of short-term patterns and fluctuations in stock prices [1, 4].

Meanwhile, LSTM, as a variant of RNN, captures long-run dependency, which becomes suitable for time series prediction. According to [1, 4], during the pandemic time, LSTM was in high demand as it can incorporate other factors such as COVID-19 case count variables, political policies, and sentiment variables of the general public to which stock prices will be affected.

Most of these neural network models rely heavily on effective feature selection [8], which includes common features, namely, historical prices and trading volumes, along with technical indicators such as Moving Averages and RSI. Besides external variables, the inclusion of government interventions, COVID-19 statistics, and social sentiment became highly important for capturing the general market trend in the pandemic scenario [1, 4]. Feature reduction techniques have been used, therefore, like PCA to eliminate data that is unnecessary for the models to focus on features that were impactful.

### 2.2.2 Applications and performance

The different neural networks-MLP, CNN, and LSTM-performed exceptionally well in stock price prediction during the COVID-19 pandemic. The volatility and uncertainty in financial markets ahead created a need for models with capabilities to capture intricate details nonlinearly that traditional models such as ARIMA were not aware of. Thus, LSTM played a very important role in modeling during the pandemic period, as it has strengths in time-series data analysis and in catching up with long dependencies. LSTM was also applied in the sphere of stock price prediction, which uses historical price data and external data, such as social media sentiments and government policy. The results indicated that LSTM outperformed the classical time-series models due to its ability to process data sequentially and consider external factors [3, 4].

There are several applications of CNNs in financial forecasting. The analysis of stock price charts helped identify local patterns in market movements. Thus, the CNN model will be very appropriate for short-term prediction of stock prices at least during turbulent periods such as during pandemic periods [1, 4]. Hybrid models combining the use of CNN with LSTM further improved this prediction accuracy by leveraging the strength of CNN in extracting a local trend and LSTM in modelling a long-term sequence [4].

Preprocessing of neural network models was indeed one of the main drivers for these improved performances. Normalization and cleaning of data, especially feature engineering, was important for the preparation of data before training the models. Traditional features have included historical prices, trading volumes, and technical indicators such as Moving Averages and RSI, while new features introduced during the pandemic included external factors such as COVID-19 case counts, government lockdowns, and social media sentiment for rapid market responsiveness modelling [1, 3, 4].

Wavelet transforms were performed to reduce noise and enhance model accuracy by reducing irrelevant data fluctuations during high volatility periods of the pandemic period. Since LSTM has a recurrent architecture, it could keep information for a very long time. Therefore, LSTM is very effective in capturing prevailing trends and stock price forecasts that are influenced by past data. For instance, CNN can detect short-term trends with higher accuracy because of their convolutional layers using candlestick charts. Further hybridization of approaches was performed in order to arrive at more accurate and robust stock predictions [4].

The new feature of neural networks during the pandemic was that they are able indeed to integrate external data

ISSN 2959-6157

such as social sentiment and COVID-19 statistics, which considerably enhances the forecast's accuracy. Hybrid models combined LSTM also with reinforcement learning in order to optimize real-time trading strategies when turbulent market conditions occurred [1, 4].

In a nutshell, neural networks and, in particular, their variant-LSTM-performed exceedingly well in stock price prediction during the pandemic, besides being an inseparable tool for investors to make decisions in turbulent markets by effectively integrating external data with advanced pre-processing techniques.

### 2.3 Reinforcement Learning

#### 2.3.1 Introduction and feature selection

Recently, reinforcement learning has turned out to be one of the main means of predicting stock prices, especially under extremely high volatility conditions that arise due to instances such as the COVID-19 pandemic outbreak. Unlike the supervised learning paradigm, RL aims at sequential decision-making while learning an interaction with the environment. It thus fits for optimizing a stock trading strategy in conditions of uncertainty [9, 10]. Reinforcement learning models, including Deep Q-Network(DQN) and Proximal Policy Optimization(PPO), have achieved phenomenal performance in stock price prediction, which learn from interaction of market data. The RL models approach stock trading as one Markov Decision Process(MDP): an agent interacts with the market; depending on the present state, the agent takes an action-buy, sell, or hold-and receives a reward depending upon the future movement of the price of the stock. These include several algorithms like Deep Q-Network and Proximal Policy Optimization, which have turned in very impressive results in dynamic market conditions [9].

Feature selection is an important factor in influencing the performance of RL-based models. Standard features include raw market data represented by open-to-close prices, volume, and returns, all reflecting the real status of the market [2]. Moving Average, RSI, and Bollinger Bands are among technical indicators to capture market trends and price momentum [2, 10]. Features representing volatility and risk portrayed by VIX and the Sharpe ratio became very important during the COVID-19 period for the management of risks [10].

#### 2.3.2 Applications and performance

Reinforcement learning has become a powerful tool for the prediction of stock prices in high-volatility periods, such as during the COVID-19 pandemic. Unlike typical supervised models, which cannot adapt so easily and quickly to rapid and unpredictable changes in the market, RL continuously improves trading strategies over time by interacting through learning with dynamic market environments. It was during this period of high volatility brought into being by the pandemic that one of the most valued aspects of RL proved to be of immense value: handling real-time feedback while maximizing long-term rewards.

Alameer et al. explored how methods from reinforcement learning, such as DQN and PPO, can perform well in highly volatile conditions. These are models concerned not with the actual prediction of stock prices but, through learning, with making the best possible trading decisions by optimizing the cumulative rewards [9]. Their findings indicated that the adaptiveness of RL, combined with its event data processing capability, enabled it to adapt to sudden market fluctuations and to make profitable trades in real time. Alameer et al. then forthcoming sophisticated mechanisms of memory replay and prior experiences allowed reinforcement learning models to further improve their learning efficiency by drawing on past experiences while reacting to present market dynamics.

An RL-based approach was proposed by Lee et al. in formulating the problem of stock price prediction as an MDP [10]. Preprocessing involves the conversion of raw data at each stock into state vectors that the RL model will use in making decisions: opening prices, closing prices, highs, lows, and trading volume. This architecture allows the agent to switch between states according to market conditions, in which the reward definition is based on stock price changes. Employing the Temporal Difference learning algorithm, Jae Won et al. showed that RL can predict quite accurately both short- and long-run price trends. The main novelty of their work was using neural networks for function approximation, which allows for generalization across a continuous state space, even unseen.

Besides their ability for adaptive learning, advanced feature selection methods further enhance the performance of RL models. This can improve the model's performance, narrowing it to the main broad variables driving the movements in the stock price during these volatile times. Sagiraju et al. emphasized the power of explainable AI techniques complementary to SHapley Additive exPlanations in offering interpretability to RL models [2]. This helps investors determine, under normal conditions, which features are influencing model predictions with a view of enhancing confidence and reliability in using RL-based stock trading systems.

## 3. Disscussion

At the same time, it was advantageous and challenging how different machine learning models helped with stock price predictions in the COVID-19 pandemic. While the linear regression model is more interpretable and simpler, it misses nonlinear patterns that are very common in violent periods of this pandemic [3, 5]. Thus, Support Vector Machines could be useful for modeling nonlinear relationships in such uncertain markets. However, their computational complexity and limited interpretability reduce their practical utility [6]. Decision trees are naturally interpretable, since there is transparency in the decision-making structure. Still, they also tend to overfit in volatile markets. Random Forests reduce variance by averaging across trees, but this comes at the cost of interpretability [3, 6].

Regarding deep learning, the LSTMs are ideal in grasping the long-term dependencies from time series data and work pretty well, but their black-box nature and high demands on computational resources are serious limiting factors to wider applicability [5]. In turn, reinforcement learning models, including Deep Q-Networks, have shown impressive performance in dynamic market environments when trying to optimize trading strategy. However, they require extensive training and easily cannot be deployed in fast-changing markets [5, 6].

In spite of major steps forward, many limitations still exist. The major concerns are that interpretability-advanced models like LSTM and RL are black boxes, and hence, their predictions are hard to interpret. In fact, lack of transparency is already a big concern in financial decision-making [5]. External factors like market news, government policy, and social sentiment are hard to incorporate into any machine learning model.

These have turned out to be factors of high influence during the pandemic, while most of the stock price prediction models could not capture these in real time [3, 6]. Also, the models measure future market to be exactly similar to the past. However, sudden breaks in market dynamics due to the pandemic caused shifts in data distribution, leading to pre-trained model performance degradation [4]. Others are some future potential solutions for these limitations: for example, a combination of human expertise with machine learning and ARG in order to make better decisions. Works like SHAP and LIME propose an interpretability solution to explain how the complex models obtained produce their predictions [6, 7]. Another promising research area is AutoML, which automatically selects models, tunes hyperparameters, and selects features. This reduces the broad expertise needed and speeds up model deployment, especially in stormy market conditions [5, 7]. Two ways to address such sudden changes in market behavior are transfer learning and domain adaptation. Transfer learning is the ability of models trained on a particular dataset to adapt to new data, but related; domain adaptation loophole is the generalization across different distributions. They might be very useful in crises situations, like COVID-19, where the behaviour of the market became unforeseen. While the machine learning models are doing a great job in stock price prediction, interpretation, factors involved, and adaptability to new changes are some of the major challenges which a machine learning model still faces. Future research in expert systems, explainable AI, AutoML, and transfer learning will be helpful to make these models robust, transparent, and adaptable to dynamic market conditions.

## 4. Conclusion

The paper explained how different machine learning models were able to predict the stock price during the COVID-19 pandemic and both the strengths and limitations of the different models. Some of the models include Linear Regression, SVM, Decision Trees, Random Forests, LSTM, and Reinforcement Learning, which came through with varied successful results in handling the never-seen volatility. Some of the models, such as LSTM and Reinforcement Learning, were successful in grasping nonlinear relationships and optimization of strategies but at the cost of non-interpretable models and highly expensive computational resources. Apart from this, inability to incorporate exogenous variables related to government policies and news, together with the distribution shifts, provided further complications to model accuracy.

The promising solutions involve future research in Expert Systems, explainable AI tools like SHAP and LIME, AutoML, and transfer learning. These evolvements will definitely ensure that machine learning models become more interpretable, adaptable, and efficient in handling stock price predictions under crises like the COVID-19 pandemic.

## References

[1] Saxena UR, Sharma P, Gupta G. Comprehensive study of machine learning algorithms for stock market prediction during COVID-19. Journal of Computers, Mechanical and Management. 2023 Feb 28;2(1):01-7.

[2] Sagiraju HK, Mogalla S. Application of multilayer perceptron to deep reinforcement learning for stock market trading and analysis. Indonesian Journal of Electrical Engineering and Computer Science. 2021 Dec;24(3):1759-71.

[3] Khattak MA, Ali M, Rizvi SA. Predicting the European stock market during COVID-19: A machine learning approach. MethodsX. 2021 Jan 1;8:101198.

[4] Jena PR, Majhi R. Are Twitter sentiments during COVID-19 pandemic a critical determinant to predict stock market movements? A machine learning approach. Scientific African.

ISSN 2959-6157

2023 Mar 1;19:e01480.
[5] Adekoya AF, Nti IK. The COVID-19 outbreak and effects on major stock market indices across the globe: A machine learning approach. Indian Journal of Science and Technology. 2020 Oct 5;13(35):3695-706.
[6] Napitupulu H, Sambas A, Murniati A, Kusumaningtyas VA. Artificial Neural Network-Based Machine Learning Approach to Stock Market Prediction Model on the Indonesia Stock Exchange During the COVID-19. Engineering Letters. 2022 Sep 1;30(3).
[7] Yang C, Zhang H, Weng F. Effects of COVID-19 vaccination programs on EU carbon price forecasts: Evidence from explainable machine learning. International Review of Financial Analysis. 2024 Jan 1;91:102953.
[8] Verikas A, Bacauskiene M. Feature selection with neural networks. Pattern recognition letters. 2002 Sep 1;23(11):1323-35.
[9] Alameer A, Saleh H, Alshehri K. Reinforcement learning in quantitative trading: A survey. Authorea Preprints. 2023 Nov 1.
[10] Lee JW. Stock price prediction using reinforcement learning. InISIE 2001. 2001 IEEE International Symposium on Industrial Electronics Proceedings (Cat. No. 01TH8570) 2001 Jun 12 (Vol. 1, pp. 690-695). IEEE.