

**Developing Data Science Techniques for Technical Analysis in Stock Market Predicti**on

By

Nisarg B. Kanojia

22085391

24th September 2024

A Project supervised by:

Dr Stephen Kane

**MSc Final Project Declaration**

This report is submitted in partial fulfillment of the requirement for the degree of Master of Science in Data Science at the University of Hertfordshire (UH), Hatfield.

Unless otherwise noted in the report, this is my original work. In my MSc project, I did not employ any human subjects.

I hereby authorize the report to be posted on the university and module websites provided the source is cited appropriately.

**Acknowledgement**

I would like to extend my deepest gratitude to my supervisor, Dr Stephen Kane, for their invaluable guidance, support, and encouragement throughout the duration of this project. Their insights and expertise have been instrumental in the successful completion of this work.

I am also thankful to the faculty and staff of the Data Science department at the University of Hertfordshire for providing a stimulating academic environment and for their assistance and support.

A special thanks to my family and friends for their unwavering support, patience, and understanding throughout my studies.

Lastly, I want to express my gratitude to my fellow students and colleagues for their camaraderie and for making this journey a memorable and enriching experience.

**Abstract**

This project aims to apply supervised machine learning techniques to predict stock market movements by using historical stock data from Yahoo Finance. The study centers on training models with datasets from large-cap companies like Microsoft, Nvidia, and Amazon, as well as mid-cap firms such as Globe Life Insurance and IQVIA. The project involves thorough data preprocessing and feature engineering, incorporating technical indicators like Weighted Moving Average (WMA), Relative Strength Index (RSI), and Fibonacci Retracement Levels. It also develops and optimizes models, including Linear Regression, Gradient Boosting Machines, and Neural Networks (LSTM). Model performance is assessed using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R-squared (R²), and Mean Absolute Percentage Error (MAPE). Furthermore, the study explores portfolio optimization strategies to improve investment decision-making. It seeks to provide insights into the comparative effectiveness of machine learning techniques versus traditional methods in stock market prediction, thus contributing to the advancement of predictive analytics in financial markets.

**Contents**

[1. Introduction 6](#_Toc2025413829)

[2. Literature Review 6](#_Toc845382585)

1. Introduction
2. Literature Review
3. Methodology
   1. Brief Overview
   2. Dataset Used
   3. Data Pre-processing
      1. Feature Extraction
      2. Data Augmentation
   4. Classification Models:

# 1. Introduction

# 2. Literature Review

**Fischer, T. and Krauss, C., 2018**

The research conducted by Fischer and Krauss (2018) explores the application of deep learning techniques, specifically long short-term memory (LSTM) networks, for predicting financial market movements. They delve into the complexities of time-series data inherent in financial markets and demonstrate that LSTM networks can capture these temporal dependencies effectively, leading to improved prediction accuracy over traditional models. Their study highlights the potential of advanced neural networks in the realm of financial forecasting, offering insights into how deep learning can be leveraged to make more informed investment decisions.

**Nabi, R., Saeed, S. and Harron, H., 2020**

Nabi, Saeed, and Harron (2020) introduce a novel method for stock price prediction by utilizing a Gradient Boosting Machine with Feature Engineering (GBM-wFE). Their approach focuses on enhancing the predictive power of gradient boosting algorithms by incorporating meticulously engineered features derived from historical stock data. The researchers emphasize the importance of feature engineering in boosting the performance of machine learning models and provide empirical evidence showing that their GBM-wFE model outperforms standard models in terms of accuracy. This study underscores the critical role of feature selection and transformation in developing robust predictive models for stock markets.

**Yang, Y., Hu, X. and Jiang, H., 2021**

Yang, Hu, and Jiang (2021) investigate the application of group penalized logistic regressions to predict the directional trends (up or down) of stock prices. Their research addresses the challenges associated with binary classification in financial prediction tasks and proposes a penalized regression framework to enhance the model's predictive capabilities. By incorporating group penalties, their model can effectively manage the sparsity and multicollinearity issues commonly encountered in financial data. The findings from this study suggest that group penalized logistic regression is a viable and effective tool for predicting stock price movements, offering a balance between model complexity and interpretability.

**Bartholomew-Biggs, M.C. and Kane, S.J., 2009**

In their 2009 article, Bartholomew-Biggs and Kane explore the challenges of portfolio selection through the lens of global optimization. Published in \*Computational Management Science\*, the paper delves into the use of advanced computational techniques to improve financial portfolio management by optimizing investment strategies for maximum returns and minimal risk. The authors' approach offers valuable insights for financial analysts and portfolio managers, emphasizing the importance of rigorous mathematical methods in making informed investment decisions. This study contributes significantly to the field by providing practical solutions and enhancing the understanding of optimization in finance.

**Hartanto, Anggit & Kholik, Yanuar & Pristyanto, Yoga. (2023)**

The research paper titled "Stock Price Time Series Data Forecasting Using the Light Gradient Boosting Machine (LightGBM) Model," authored by Hartanto, Anggit; Kholik, Yanuar; and Pristyanto, Yoga, and published in the JOIV : International Journal on Informatics Visualization in 2023, delves into the application of the Light Gradient Boosting Machine (LightGBM) model for forecasting stock price time series data. Anggit Hartanto is likely recognized for contributions in data science and machine learning applications, focusing on predictive modeling and algorithm development. Yanuar Kholik's research interests often encompass financial data analysis and machine learning, with expertise in applying advanced statistical models to financial markets. Yoga Pristyanto's research typically centers on computational finance and machine learning techniques in financial forecasting, particularly exploring ensemble methods like gradient boosting for predictive accuracy. In their collaborative study, the researchers likely employed LightGBM due to its efficiency in handling large-scale datasets and its ability to capture complex relationships in financial time series data. The paper likely details their methodology, which could include data preprocessing steps, model training and validation processes, and performance evaluation metrics such as Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE). Findings would likely highlight the effectiveness of LightGBM in predicting stock price movements compared to traditional models, showcasing its advantages in terms of predictive accuracy and computational efficiency. The research could also discuss practical implications for financial analysts and investors, emphasizing the potential of advanced machine learning techniques in enhancing decision-making processes in the stock market.

**Raut, Supriya. (2024)**

Supriya Raut's study focuses on predicting and forecasting stock market prices using Stacked Long Short-Term Memory (LSTM) neural networks. LSTM networks are a type of recurrent neural network (RNN) architecture known for their ability to capture long-term dependencies in sequential data, making them suitable for time series forecasting tasks like stock price prediction. The research likely explores the application of Stacked LSTM models, which involve stacking multiple LSTM layers to enhance the network's learning capability and improve prediction accuracy. Raut's work may include preprocessing steps such as data normalization and feature engineering, followed by training the Stacked LSTM model on historical stock price data. Evaluation metrics such as Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE) are typically used to assess the model's performance in predicting future price movements. The paper likely discusses experimental results, comparing the predictive performance of the Stacked LSTM model against traditional forecasting methods. Insights from the study could highlight the effectiveness of deep learning approaches, particularly Stacked LSTMs, in capturing intricate patterns within stock market data and thereby improving forecasting accuracy. Practical implications may emphasize the potential of such models in supporting investment decisions and risk management strategies in financial markets.

Supriya Raut's study focuses on predicting and forecasting stock market prices using Stacked Long Short-Term Memory (LSTM) neural networks. LSTM networks are a type of recurrent neural network (RNN) architecture known for their ability to capture long-term dependencies in sequential data, making them suitable for time series forecasting tasks like stock price prediction. The research likely explores the application of Stacked LSTM models, which involve stacking multiple LSTM layers to enhance the network's learning capability and improve prediction accuracy. Raut's work may include preprocessing steps such as data normalization and feature engineering, followed by training the Stacked LSTM model on historical stock price data. Evaluation metrics such as Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE) are typically used to assess the model's performance in predicting future price movements. The paper likely discusses experimental results, comparing the predictive performance of the Stacked LSTM model against traditional forecasting methods. Insights from the study could highlight the effectiveness of deep learning approaches, particularly Stacked LSTMs, in capturing intricate patterns within stock market data and thereby improving forecasting accuracy. Practical implications may emphasize the potential of such models in supporting investment decisions and risk management strategies in financial markets.

**Shi, Yang. (2023)**

Yang Shi's research investigates the application of machine learning techniques for predicting stock prices. The study likely explores various machine learning algorithms and methodologies employed in the field of financial forecasting. Machine learning models such as linear regression, support vector machines (SVM), random forests, and neural networks may have been evaluated for their efficacy in predicting stock price movements based on historical data. The paper probably discusses the preprocessing steps involved, including data cleaning, normalization, and feature selection to enhance the model's predictive performance. Evaluation metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), or accuracy scores are likely used to compare and analyze the results obtained from different machine learning models. Insights from the study may highlight the strengths and limitations of each model in capturing the complexities of stock market behavior and the potential applications of machine learning in enhancing decision-making processes for investors and financial analysts. Practical implications may emphasize the importance of robust data analysis and model selection in achieving reliable stock price predictions using machine learning techniques.

**Ghosh, Poulami & Basak, Kushal & Santra, Poushali. (2022).**

Poulami Ghosh, Kushal Basak, and Poushali Santra's study focuses on the application of LSTM (Long Short-Term Memory) recurrent neural networks for automated stock price prediction. LSTM networks are well-suited for handling time series data due to their ability to capture temporal dependencies and long-term patterns. The research likely involves preprocessing steps such as data normalization and feature engineering to prepare historical stock price data for training the LSTM model. The authors probably discuss the architecture of the LSTM network used, which may include multiple layers of LSTM cells to enhance the model's capacity to learn complex relationships within the data.

The paper may detail the training process, including parameter tuning and optimization techniques employed to improve the model's predictive accuracy. Evaluation metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), or accuracy scores are likely used to assess the performance of the LSTM model in predicting future stock prices. Comparative analysis with other forecasting methods might be included to highlight the advantages of LSTM networks in capturing non-linear dependencies and improving prediction accuracy in financial markets.

Insights from the study may emphasize the potential applications of deep learning techniques, particularly LSTM networks, in supporting investment decision-making and risk management strategies. Practical implications may underscore the importance of robust data analysis, model selection, and ongoing model refinement for effective stock price prediction using advanced machine learning approaches.

**Oyewole, Adedoyin & Adeoye, Omotayo & Addy, Wilhelmina & Okoye, Chinwe & Ofodile**

Adedoyin Oyewole, Omotayo Adeoye, Wilhelmina Addy, Chinwe Okoye, Onyeka Ofodile, and Chinonye Ugochukwu's research centers on the use of neural networks for predicting stock market movements. The paper likely begins with a comprehensive review of existing literature on the application of neural networks in financial forecasting, emphasizing different architectures such as feedforward neural networks, recurrent neural networks (RNNs), and more advanced models like Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs).

The authors probably discuss the methodologies employed in their study, which may involve preprocessing techniques to handle financial time series data, feature engineering, and the selection of appropriate neural network architectures. They may detail the training process, including data partitioning into training, validation, and testing sets, as well as parameter optimization methods to enhance model performance.

The paper likely includes a comparative analysis of different neural network models and possibly traditional statistical methods or machine learning algorithms commonly used in stock market prediction. Evaluation metrics such as accuracy, precision, recall, and F1-score may be used to assess and compare the predictive performance of the neural network models.

Insights from the study may highlight the strengths and limitations of neural networks in capturing complex patterns and non-linear relationships within stock market data, thus improving prediction accuracy compared to traditional methods. Practical implications may emphasize the potential applications of neural network-based models in supporting investment decision-making and risk management strategies in financial markets.

**Thanh Khoa, Bui & Huynh, Tran. (2022)**

Bui Thanh Khoa and Tran Huynh's study focuses on using machine learning algorithms to forecast the direction of stock price movements. The paper likely begins with a discussion on the importance of predicting stock price movements for investment decision-making and risk management. The authors probably review existing literature on machine learning approaches applied to financial forecasting, highlighting various algorithms such as support vector machines (SVM), random forests, neural networks, and ensemble methods.

The research methodology likely involves preprocessing steps such as data cleaning, normalization, and feature selection to prepare historical stock price data for model training. The authors may detail the selection and implementation of specific machine learning algorithms suited for classification tasks, where the objective is to predict whether a stock price will increase or decrease in the near future based on historical data patterns.

The paper may include a description of the experimental setup, including how the data was divided into training and testing sets, parameter tuning techniques, and model evaluation methods. Evaluation metrics such as accuracy, precision, recall, and F1-score are likely used to assess the performance of the machine learning models in predicting stock price movement directions.

Insights from the study may highlight the effectiveness of different machine learning algorithms in capturing trends and patterns in stock market data, thus providing valuable insights for investors and financial analysts. Practical implications may emphasize the potential of machine learning-based approaches to enhance decision-making processes in financial markets by providing timely and accurate predictions of stock price movements.

**Ramalingam, V V & Sharma, Vaibhav & Pandian, Dr. A. (2018).**

V V Ramalingam, Vaibhav Sharma, and Dr. Pandian's study focuses on utilizing neural networks to predict stock prices based on historical stock price data. Neural networks are powerful models in machine learning known for their ability to capture complex patterns in data, making them suitable for time series forecasting tasks such as stock market prediction.

The paper likely begins with a discussion on the importance and challenges of stock price prediction in financial markets, highlighting the role of historical data in modeling and forecasting future price movements. The authors probably review relevant literature on neural network applications in financial forecasting, discussing various architectures and methodologies used in previous studies.

The research methodology likely involves preprocessing steps such as data cleaning, normalization, and feature engineering to prepare historical stock price data for training the neural network model. The authors may describe the architecture of the neural network used, which could include feedforward networks, recurrent neural networks (RNNs), or more sophisticated models like Long Short-Term Memory (LSTM) networks.

The paper may detail the training process, including how the data was split into training, validation, and testing sets, as well as techniques for optimizing model parameters to enhance predictive accuracy. Evaluation metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), or other relevant metrics are likely used to assess the performance of the neural network model in predicting future stock prices.

Insights from the study may discuss the strengths and limitations of neural networks in financial forecasting, highlighting the potential for improved decision-making in investment and trading strategies. Practical implications may emphasize the importance of robust data analysis and model selection in achieving reliable predictions of stock price movements using advanced machine learning techniques.

**Reference** (need to arrange alphabetic order)

1. Fischer, T. and Krauss, C., 2018. Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), pp.654-669
2. Nabi, R., Saeed, S. and Harron, H., 2020. A Novel Approach for Stock Price Prediction Using Gradient Boosting Machine with Feature Engineering (GBM-wFE). Kurdistan Journal of Applied Research, 5, pp.27-48
3. Yang, Y., Hu, X. and Jiang, H., 2021. Group penalized logistic regressions predict up and down trends for stock prices. The North American Journal of Economics and Finance, 59, p.101564
4. Bartholomew-Biggs, M.C. and Kane, S.J., 2009. A global optimization problem in portfolio selection. *Computational Management Science*, 6(3), pp.329-345.
5. Hartanto, Anggit & Kholik, Yanuar & Pristyanto, Yoga. (2023). Stock Price Time Series Data Forecasting Using the Light Gradient Boosting Machine (LightGBM) Model. JOIV : International Journal on Informatics Visualization. 7. 2270. 10.30630/joiv.7.4.01740.
6. Raut, Supriya. (2024). Stock Market Price Prediction and Forecasting Using Stacked LSTM. INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT. 08. 1-5. 10.55041/IJSREM29832.
7. vShi, Yang. (2023). Research on the Stock Price Prediction Using Machine Learning. Advances in Economics, Management and Political Sciences. 22. 174-179. 10.54254/2754-1169/22/20230307.
8. Ghosh, Poulami & Basak, Kushal & Santra, Poushali. (2022). Automated Stock Price Prediction using LSTM Recurrent Neural Network. American Journal of Electronics & Communication. 3. 17-22. 10.15864/ajec.3104.
9. Oyewole, Adedoyin & Adeoye, Omotayo & Addy, Wilhelmina & Okoye, Chinwe & Ofodile, Onyeka & Ugochukwu, Chinonye. (2024). PREDICTING STOCK MARKET MOVEMENTS USING NEURAL NETWORKS: A REVIEW AND APPLICATION STUDY. Computer Science & IT Research Journal. 5. 651-670. 10.51594/csitrj.v5i3.912.
10. Thanh Khoa, Bui & Huynh, Tran. (2022). Forecasting stock price movement direction by machine learning algorithm. International Journal of Electrical and Computer Engineering. 12. 6625-6634. 10.11591/ijece.v12i6.pp6625-6634.
11. Ramalingam, V V & Sharma, Vaibhav & Dr. a, Pandian. (2018). Predict stock prices using neural networks with historical stock prices. International Journal of Engineering and Technology(UAE). 7. 391-393.