Journal of Research High School
Vol. 2025 (3) 207 – 217
ISSN 2688-3651 [online]

Predicting Stock Prices with Machine Learning: A Comparative Analysis of Models and Simulations

Shoken Tsurumaru[1 *]

[1]United World College of South East Asia (Dover Campus), Singapore
*Corresponding Author: tsurucircle@gmail.com

Advisor: Odysseas Drosis, od84@cornell.edu

Received November 16, 2024; Revised January 13, 2025; Accepted January 27, 2025

Abstract

Accurate stock price prediction is a crucial aspect of financial markets, offering significant benefits for investors and policymakers. This study explores the application of machine learning techniques to forecast the opening stock prices of multiple companies across various sectors, including technology, energy, finance, and healthcare. Four regression models—Linear Regression, Decision Tree Regressor, Random Forest Regressor, and Multilayer Perceptron (MLP) Regressor—were employed, trained on five years of historical 'Open' prices obtained from Yahoo Finance. Model performance was evaluated using Mean Squared Error (MSE), with hyperparameter tuning conducted for optimization. Trading simulations with varying initial capital amounts were implemented to assess the practical applicability of the predictions. Additionally, the impact of significant events and political sentiment on the stock market was discussed, along with the potential of using social media data, particularly tweets, for stock price prediction. A trading strategy guided by the model's forecasts yielded returns more than 2.5 times higher than those from simply holding the stock, demonstrating the potential increase in profitability of model-informed investment decisions. These findings highlight the significance of machine learning in enhancing stock market predictions and suggest avenues for future research to further optimize and apply these techniques in diverse financial contexts.

Keywords: Machine Learning, Stock Market, Prediction, Stock Price Prediction

1. Introduction

The stock market has long been a cornerstone of the global financial system, serving as a critical platform for companies to raise capital and for investors to participate in corporate growth (Mishkin et al., 2012). Originating in the 17th century with the establishment of the Amsterdam Stock Exchange, stock markets have evolved into complex networks that facilitate the buying and selling of securities worldwide (Jones, 2013). Over the centuries, they have become integral to economic development, influencing everything from corporate governance to individual wealth management.

As the stock market matured, it mirrored the technological advancements of each era. The transition from open outcry trading floors to electronic trading systems in the late 20th century marked a significant shift in how securities were exchanged (Hendershott et al., 2011). This digital transformation increased the speed and efficiency of transactions and democratized access to market data, enabling a broader range of participants to engage in trading activities. The proliferation of online trading platforms and real-time data analytics tools has empowered individual investors and led to the rise of algorithmic trading strategies (Hendershott et al., 2013).

Despite these advancements, the stock market remains a volatile environment influenced by a multitude of factors, including economic indicators, geopolitical events, and investor sentiment (Fama, 1970). The 2008 global financial crisis and the 2020 COVID-19 pandemic are stark reminders of how quickly market conditions can change, resulting

Vol. 2025 (3) 207 – 217
ISSN 2688-3651 [online]

in significant financial losses or gains (Brunnermeier, 2009; Baker et al., 2020). Such events underscore the importance of developing robust predictive models that can navigate the complexities of the market.

In recent years, the advent of Artificial Intelligence (AI) has introduced new possibilities for navigating the complexities of the stock market (Samuel, 1959; LeCun et al., 2015). AI technologies, particularly machine learning algorithms, have the potential to analyze vast amounts of data at unprecedented speeds, uncovering patterns and insights beyond human capability (Silver, 2016; Schmidhuber, 2015). Deep learning models, such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks, have been employed to capture temporal dependencies in financial data (Hochreiter, 1997; Fischer and Krauss, 2018; Selvin et al., 2017). Investment firms and hedge funds are increasingly leveraging AI to enhance trading strategies, optimize portfolios, and manage risks more effectively (Kroll, et al 2017).

The integration of AI into stock market operations promises to revolutionize traditional investment paradigms. By automating analytical processes and providing predictive analytics, AI can assist investors in making more informed decisions (Agrawal, et al., 2018). Machine learning models can process diverse data sources, including historical prices, economic indicators, and even textual data from news articles and social media (Zhang and Zhao, 2021; Gu et al., 2020). However, although AI-driven stock prediction models offer notable advantages, they raise several ethical and regulatory concerns that warrant deeper examination. For instance, large-scale data collection—particularly from social media—can infringe on user privacy if consent and data protection measures are insufficient. Furthermore, algorithmic bias may systematically disadvantage certain market participants, while 'black-box' AI models can reduce transparency and accountability in trading decisions. In extreme cases, the rapid, automated nature of AI-driven trades could exacerbate market volatility (e.g., flash crashes), affecting overall market stability. Addressing these ethical concerns through proper data governance, model interpretability, and regulatory frameworks is crucial for responsible AI adoption in finance. Significant events, such as economic crises, political upheavals, and pandemics, have profound impacts on the stock market. The market is highly sensitive to changes in political sentiment, with elections, policy shifts, and geopolitical tensions influencing investor confidence and market volatility (Pastor and Veronesi, 2013). Moreover, the rise of social media platforms like Twitter(X) has introduced new avenues for gauging public sentiment. Analyzing tweets and other social media data can provide real-time insights into market perceptions, potentially enhancing predictive models (Bollen et al, 2011; Mittal and Goel, 2012). This study aims to explore these facets by applying machine learning techniques to stock price prediction and assessing the influence of external factors on market dynamics.

1.1 Literature Review

Historical Foundations and Market Volatility

Early stock exchanges date back to 17th-century Amsterdam, evolving into modern electronic and algorithmic trading systems that dramatically increase accessibility (Jones, 2013; Hendershott, Jones, & Menkveld, 2011). Despite the Efficient Market Hypothesis (Fama, 1970), events like the 2008 financial crisis and the COVID-19 pandemic underscore persistent volatility and the importance of adaptive forecasting models (Brunnermeier, 2009; Baker et al., 2020).

Machine Learning in Stock Prediction

AI and deep learning methods—such as LSTMs and CNNs—allow for sophisticated time-series analysis in financial contexts (Hochreiter & Schmidhuber, 1997; LeCun, Bengio, & Hinton, 2015; Fischer & Krauss, 2018). These methods can reveal non-linear relationships in market data, prompting widespread adoption in professional trading and risk management (Agrawal, Gans, & Goldfarb, 2018; Gu, Kelly, & Xiu, 2020).

Sentiment Analysis and Alternative Data

Beyond price-based indicators, researchers have integrated social media sentiment and macroeconomic signals to enhance predictive power (Bollen, Mao, & Zeng, 2011; Mittal & Goel, 2012). While these alternative data sources can yield more robust forecasts, issues of data noise, bias, and overfitting remain (Li et al., 2017; Hu, Zhao, & Khushi, 2021).

Vol. 2025 (3) 207 – 217
ISSN 2688-3651 [online]

Ensemble Approaches and Practical Challenges

Ensemble methods (e.g., Random Forests) and hybrid statistical-ML models (e.g., ARIMA-NN) often outperform individual models by reducing variance and capturing diverse patterns (Breiman, 2001; Zhang, 2003; Zhang & Qi, 2005). However, inconsistent hyperparameter tuning across studies limits generalizability, and most approaches do not fully account for transaction costs and market liquidity (Hasbrouck, 1993).

Political and Geopolitical Factors

Policy shifts, elections, and geopolitical tensions can drastically affect stock prices (Pastor & Veronesi, 2013; Baker, Bloom, & Davis, 2016). While some models incorporate these elements, many still rely solely on historical price trends, underscoring a need for more comprehensive frameworks (Caldara & Iacoviello, 2018).

Gaps and Study Rationale

Existing work often fails to integrate multiple data streams and real-world trading constraints in a single predictive framework. By addressing these gaps and leveraging both price and sentiment data, this study aims to develop a more robust, adaptable model for stock price forecasting.

2. Materials and Methods

2.1 Data Collection

Five years of historical stock data (from November 6, 2019 to November 5, 2024) for several companies across different sectors were collected using the yfinance Python library. In assembling the dataset, a higher number of technology firms was included due to their market prominence, high liquidity, and historically rapid growth patterns, which make them ideal test cases for machine learning algorithms. By contrast, sectors such as energy, finance, and healthcare are represented by fewer companies, reflecting both practical data availability and the desire to showcase how the models perform in industries with different risk profiles and regulatory landscapes. Preliminary results indicate that predictive accuracy can vary notably by sector—technology stocks often exhibit more volatile price movements, potentially amplifying both gains and errors, whereas healthcare stocks may present smoother trends that favor simpler models. These variations underscore the importance of sector-specific considerations when evaluating machine learning approaches for stock price forecasting. The selected companies and their ticker symbols are:

- Technology Sector:
  - Apple Inc. (**AAPL**)
  - Alphabet Inc. (**GOOG**)
  - Nvidia Corporation (**NVDA**)
  - Microsoft Corporation (**MSFT**)
  - Amazon.com Inc. (**AMZN**)
  - Meta Platforms Inc. (**META**)
  - Tesla Inc. (**TSLA**)
- Energy Sector:
  - ExxonMobil Corporation (**XOM**)
  - Saudi Aramco (**2222.SR**)
- Financial Sector:
  - JPMorgan Chase & Co. (**JPM**)
- Healthcare Sector:
  - Johnson & Johnson (**JNJ**)

The dataset provides daily records of the 'Open' prices, which were used exclusively for this study. Focusing on a five-year period allowed us to capture recent market trends and ensured a substantial dataset for model training and evaluation.

Vol. 2025 (3) 207 – 217
ISSN 2688-3651 [online]

Rationale for Sector and Company Selection

The selection of technology, energy, finance, and healthcare sectors was made to ensure a diverse range of market conditions and risk profiles. Technology stocks (e.g., Apple, Alphabet, Nvidia, Microsoft, Amazon, Meta, Tesla) exhibit rapid growth and are often sensitive to innovation cycles and investor sentiment. By contrast, energy companies (e.g., ExxonMobil, Saudi Aramco) are typically influenced by commodity prices and geopolitical factors, offering a distinct investment dynamic. In the financial sector, JPMorgan Chase & Co. represents a key player affected by interest rates, credit cycles, and regulatory changes, whereas Johnson & Johnson in healthcare provides a comparatively stable growth trajectory and is subject to different regulatory and market forces. Additionally, all of the stocks chosen are big-market-capitalization stocks, which mean that they are less prone to flukes or anomalies in terms of unusually large purchases of small-cap stocks.

Collectively, these selections capture a broad cross-section of market behaviors and industry-specific variables, enhancing the study's ability to assess the adaptability and robustness of machine learning models. By spanning multiple sectors, the research evaluates how well predictive techniques generalize across diverse economic conditions, thereby strengthening conclusions about their practical utility in real-world trading scenarios.

2.2 Data Preprocessing and Feature Engineering

Feature Selection

To simplify the models and reduce computational complexity, only the "Open" price was utilised for each stock. This decision aligns with our objective of predicting the next day's opening price based on recent opening prices.

Input and Output Construction

Input features (X) and the target variable (Y) were constructed using a sliding window approach:

- Inputs (X): For each day t, the input vector consisted of the opening prices from the previous three days:

$$X_t = \left[Open_{t-3}, Open_{t-2}, Open_{t-1}\right]$$

- Target Variable (Y): The opening price on day t, denoted as $Y_t = Open_t$

This method captures short-term temporal dependencies, assuming that recent opening prices have a significant influence on the immediate future.

2.3 Dataset Splitting

Each company's dataset was split into training and testing sets:

- Training Set: Approximately 67% of the data, used to train the models.
- Testing Set: The remaining 33%, reserved for evaluating model performance on unseen data.

The performance that is attained on the training set is a good indicator of how the algorithm will perform on the testing set.

The train_test_split function from scikit-learn with a random state for reproducibility was employed.

2.4 Model Selection and Training

Four regression models using the scikit-learn library were implemented:

Linear Regression Model

- *Purpose*: To model the linear relationship between the input features and the target variable.
- *Training*: Used the Ordinary Least Squares method to minimize the residual sum of squares between the observed and predicted values.
- *Equation*: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$

  Where β are coefficients, and ϵ is the error term.

Vol. 2025 (3) 207 – 217
ISSN 2688-3651 [online]

Decision Tree Regressor

- *Purpose*: To capture non-linear relationships by partitioning the data based on feature thresholds.
- *Hyperparameter Tuning*: Tested various maximum depths (6, 7, 10, 20) to prevent overfitting.
- *Training*: Implemented using the DecisionTreeRegressor class with the max_depth parameter adjusted accordingly.
- *Algorithm*: Recursive binary splitting based on minimizing the mean squared error in child nodes

Random Forest Regressor

- *Purpose*: An ensemble method that builds multiple decision trees and merges their predictions.
- *Hyperparameter Tuning*:
  - Maximum Depth: Tested depths of 6, 7, 8, 10, and 12.
  - Number of Estimators: Tested with 50, 100, and 200 trees.
- *Training*: Used the RandomForestRegressor class, adjusting max_depth and n_estimators parameters.
- *Algorithm*: Combines bootstrapping and feature randomness to create an uncorrelated forest of trees whose prediction is more accurate than any individual tree.

Multilayer Perceptron (MLP) Regressor

- *Purpose*: A feedforward artificial neural network capable of modeling and capturing complex non-linear relationships.
- *Hyperparameter Tuning*:
  - Hidden Layers and Neurons: Experimented with different configurations.
  - Maximum Iterations: Set to 120,000 to ensure convergence.
- *Training*: Employed the MLPRegressor class with the max_iter parameter set.
- *Algorithm*: Uses backpropagation to adjust weights, minimizing the loss function.

2.5 Model Evaluation

Model performance was assessed using the Mean Squared Error (MSE):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

where:

n is the number of observations.

Y_i is the actual opening price

$\hat{Y}_i$ is the predicted opening price.

Both of the following were calculated:

- Training MSE: To evaluate how well the model learned from the training data.
- Testing MSE: To assess the model's ability to generalize to new, unseen data.

2.6 Model Combination Strategies

To potentially enhance predictive accuracy, ensemble methods were explored:

Simple Averaging

- *Approach*: Calculated the mean of the predictions from the Linear Regression, Random Forest, and MLP Regressor models.
- *Rationale*: Averages can smooth out individual model biases and reduce variance.

Vol. 2025 (3) 207 – 217
ISSN 2688-3651 [online]

Weighted Averaging Based on Inverse MSE

- *Approach*: Assigned weights inversely proportional to each model's testing MSE:

$$\omega_i = \frac{1/MSE_i}{\sum_j \left(1/MSE_j\right)}$$

- *Rationale*: Models with lower MSE (higher accuracy) have more influence on the final prediction.

Dynamic Weight Adjustment

- *Approach*:
  - Initialization: Set equal weights for all models.
  - Iteration: After each prediction, adjusted weights based on the absolute error:

$$\omega_i \leftarrow \frac{\omega_i}{1.01^{\left|Y_{true} - Y_{pred_i}\right|}}$$

  - Normalization: Ensured that the sum of weights equals one.
- *Rationale*: Allows the ensemble to adapt by giving more weight to models performing better on specific instances.

2.7 Trading Simulation

Trading simulations were conducted to assess the practical applicability of our predictions.

Simulation Parameters

- Initial Capital: Varied amounts of $1,000, $2,500, $5,000, and $10,000.
- Stock Holdings: Initially zero shares.
- Trading Period: Corresponded to the testing dataset period for each company.

Trading Strategy

- Buy Signal: If the predicted opening price for the next day was higher than the current day's opening price.
- Sell Signal: If the predicted opening price for the next day was lower than the current day's opening price.
- Execution Rules:
  - Purchase: Bought shares if sufficient capital was available.
  - Sale: Sold shares if holdings were available.
  - Constraints: No leverage, margin trading, or short selling.

Variations in Trading

- Fixed Share Trading: Traded fixed quantities per signal (e.g. one, two, or twenty shares).

Ethical Considerations

- Data Usage: Utilized only publicly available data, complying with data privacy regulations.
- Model Limitations: Acknowledged that models are based on historical data and may not account for unforeseen events.

3. Results

3.1 Model Performance

Testing Mean Squared Error (MSE)

Table 1 presents the Testing Mean Squared Errors (MSE) for the four machine learning models—Linear

Vol. 2025 (3) 207 – 217
ISSN 2688-3651 [online]

Regression, Decision Tree, Random Forest, and MLP—across Apple, Alphabet, Nvidia, and ExxonMobil. These results quantify each model's predictive accuracy on unseen data, allowing for direct comparison between simpler linear methods and more complex ensemble or neural network approaches.

Table 1. Testing Mean Squared Error (MSE) for Different Models

Company	Linear Regression	Decision Tree (Depth=6)	Random Forest (Depth=7)	MLP Regressor Max Iter=120,000
Apple	6.40	10.66	6.93	8.10
Alphabet	4.10	8.12	5.77	4.32
Nvidia	2.77	2.81	3.56	2.67
ExxonMobil	2.03	3.63	2.46	3.21

Training Mean Squared Error (MSE)

Table 2 shows the training MSE for the same four models. Comparing these values with Table 1 helps identify potential overfitting (when training errors are markedly lower than testing errors).

Table 2. Training Mean Squared Error (MSE) for Different Models

Company	Linear Regression	Decision Tree (Depth=6)	Random Forest	MLP Regressor
Apple	6.95	7.60	4.46	7.83
Alphabet	4.63	4.85	3.03	5.08
Nvidia	2.53	2.09	1.05	2.43
ExxonMobil	2.12	2.41	1.39	3.36

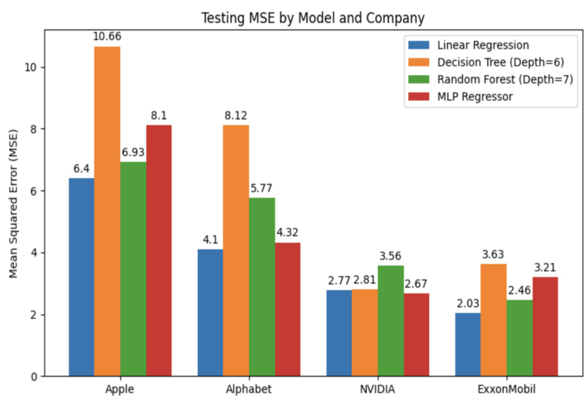

Figure 1. Testing Mean Squared Error (MSE) for Different Models

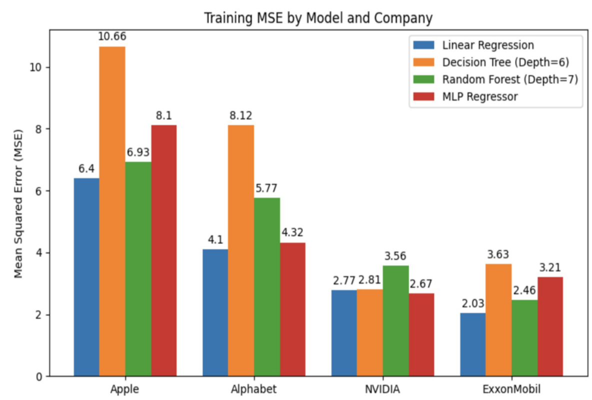

Figure 2. Training Mean Squared Error (MSE) for Different Models

Table 3 demonstrates how varying the max_depth parameter affects Decision Tree performance for Apple Inc. This experiment aims to balance complexity against overfitting.

Table 3. Decision Tree Hyperparameter Tuning for Apple Inc.

Max Depth	Testing MSE
6	10.66
7	10.44
10	11.11.
20	12.23

Table 4 explores max_depth and n_estimators for the Random Forest Regressor, also using Apple Inc. as the test case. Ensemble methods often benefit from multiple estimators, but increasing depth can lead to diminishing returns or overfitting.

Table 5 shows simulated portfolio outcomes for Apple Inc. under a model-driven strategy. An additional column indicates the percentage change relative to the initial capital.

Table 4. Random Forest Hyperparameter Tuning for Apple Inc.

Max Depth	n_estimator	Testing MSE
6	100	6.93
7	100	6.70
8	100	6.85
10	100	7.27
12	100	7.46

Table 6 presents outcomes for Nvidia, showing particularly large gains—an indication that the model captured significant upward trends in the stock's opening price.

Table 7 focuses on ExxonMobil, an energy-sector giant. While absolute returns are high, the percentage increase from smaller initial capitals is notable for capturing periods of growth in the fossil-fuel market segment.

Vol. 2025 (3) 207 – 217
ISSN 2688-3651 [online]

Table 5. Final Portfolio Values for Different Initial Investments (Apple Inc.)

Initial Capital	Final Portfolio Value
$1,000	$2,064.54
$2,500	$6,584.70
$5,000	$13,366.45
$10,000	$33,476.14

Table 6. Final Portfolio Values for Different Initial Investments (Nvidia Corporation)

Initial Capital	Final Portfolio Value
$1,000	$2,064.54
$2,500	$6,584.70
$5,000	$13,366.45
$10,000	$33,476.14

Table 7. Final Portfolio Values for Different Initial Investments (ExxonMobil Corporation)

Initial Capital	Final Portfolio Value
$1,000	$16,137.28
$2,500	$31,954.54
$5,000	$56,792.35
$10,000	$89,088.02

Note: The significant returns observed, especially for ExxonMobil with $10,000 initial capital, suggest that the model captured favorable trends during the testing period.

4. Discussion

The Linear Regression model consistently achieved lower testing MSE compared to the more complex models for several companies, indicating that recent opening prices have a strong linear relationship with the next day's opening price. This finding aligns with the efficient market hypothesis, suggesting that historical prices contain valuable information for prediction (Fama, 1991).

The Random Forest Regressor also performed well after hyperparameter tuning, demonstrating the effectiveness of ensemble methods in capturing non-linear patterns without overfitting. This result is consistent with prior research highlighting the robustness of ensemble techniques in financial modelling (Mittal and Goel, 2012; Breiman, 2001).

The Decision Tree Regressor showed increased testing MSE with greater depth, highlighting the risk of overfitting when the model becomes too complex. The MLP Regressor's performance varied, suggesting that neural networks require careful tuning and may benefit from additional data or features. Similar challenges have been reported in previous studies utilizing neural networks for stock prediction (Zhang, 2003; Zhang and Qi, 2005).

4.1 Trading Simulation Analysis

The trading simulations revealed that applying the model predictions could result in substantial portfolio growth. For instance, an initial investment of $10,000 in Nvidia led to a final portfolio value of over $226,000, a 2260% increase, higher than Nvidia's natural increase of 1600%. However, these simulations are idealized and do not account for:

- Transaction Costs: Brokerage fees and taxes.
- Market Impact: The effect of large trades on stock prices.
- Slippage: The difference between expected and actual transaction prices.
- Liquidity Constraints: Availability of shares to buy or sell.

Real-world trading would likely yield lower returns due to these factors. Moreover, the models do not account for market microstructure noise, which can significantly impact high-frequency trading strategies (Hasbrouck, 1993).

4.2 Impact of Significant Events on the Stock Market

Significant events, such as financial crises, pandemics, and political upheavals, can cause abrupt market shifts. For example, the COVID-19 pandemic led to unprecedented volatility, with rapid declines followed by swift recoveries in stock prices (Brunnermeier, 2009; Ashraf, 2020). These events can invalidate models trained on historical data that do not account for such anomalies.

Political sentiment also plays a critical role. Elections, policy changes, and international relations can influence investor confidence and market dynamics (Pastor and Veronesi, 2013; Baker et al., 2013). Incorporating political indicators or sentiment analysis into predictive models could enhance their responsiveness to such events. Recent studies have begun to integrate geopolitical risk indices into financial models, showing improved predictive capabilities (Caldara and Iacoviello, 2018).

Vol. 2025 (3) 207 – 217
ISSN 2688-3651 [online]

4.3 Using Tweets to Predict the Stock Market

Social media platforms like Twitter offer real-time data reflecting public sentiment. Studies have shown that analyzing tweets using natural language processing (NLP) techniques can improve stock market predictions (Bollen et al., 2011; Mittal and Goel, 2012; Li et al., 2017). Sentiment analysis can gauge public reactions to news, earnings reports, or corporate actions, providing an additional layer of information beyond historical prices.
Challenges in using tweets for prediction include:

- Noise: Distinguishing meaningful signals from irrelevant content.
- Data Volume: Handling the vast amount of unstructured data.
- Biases: Accounting for demographic and regional biases in social media usage.

Integrating social media data requires sophisticated algorithms capable of real-time analysis. Advanced NLP techniques, such as transformer-based models like BERT, have shown promise in extracting sentiment with higher accuracy (Devlin et al., 2018).

4.4 Limitations and Future Work

Limitations:

- Data Scope: The study focused on 'Open' prices and did not include other potentially informative features like trading volume, closing prices, or technical indicators.
- Model Complexity: More advanced models like LSTM networks and hybrid models were not explored.
- Overfitting Risks: Models may overfit to past data and fail to generalize during unprecedented events.

Future Work:

- Feature Expansion: Incorporate additional features such as technical indicators (e.g., Moving Averages, RSI), macroeconomic indicators, and corporate financial statements.
- Advanced Models: Explore deep learning architectures like LSTM and Convolutional Neural Networks (CNNs) for time-series analysis (Selvin et al., 2017; Kim and Won, 2018).
- Sentiment Analysis: Implement NLP techniques to analyze news articles and social media data for enhanced predictions (Bollen and Huina, 2011; Delvin et al., 2018).
- Robust Testing: Conduct backtesting over different time periods, including during significant market events, to evaluate model resilience.

4.5 Risk Management:

Incorporate risk-adjusted performance metrics like Sharpe Ratio and Sortino Ratio in evaluating trading strategies.

5. Conclusion

This study demonstrates the potential of machine learning models, particularly Linear Regression and Random Forest Regressor with optimized hyperparameters, in predicting stock opening prices based on historical data. While the models showed promising results, especially in trading simulations, turning $10,000 into almost $90,000 with ExxonMobil and turning $10,000 into $226,000 with Nvidia, practical application in real-world trading requires accounting for additional factors like transaction costs and market impact.

Incorporating alternative data sources, such as social media sentiment and political indicators, holds potential for improving prediction accuracy and model robustness. Future research should focus on integrating these data sources and employing advanced machine learning techniques to navigate the complexities of the stock market.

Acknowledgments

I would like to express my gratitude to the researchers and developers whose work has contributed to the field of financial machine learning. Their pioneering efforts laid the groundwork for this study. I would also like to thank the open-source community for providing the tools and libraries that made this research possible.

References

Agrawal, A., et al. (2018). Prediction Machines: The Simple Economics of Artificial Intelligence. *Harvard Business Press.*

Ashraf, B. N. (2020). Stock Markets' Reaction to COVID-19: Cases or Fatalities? *Research in International Business and Finance,* 54, 101249.

Baker, S. R., et al. (2016). Measuring Economic Policy Uncertainty. *The Quarterly Journal of Economics,* 131(4), 1593-1636.

Baker, S. R., et al. (2020). COVID-Induced Economic Uncertainty. *National Bureau of Economic Research, No. w26983.*

Bartram, S. M., & Grinblatt, M. (2018). A Survey of Tail Risk Premia in Financial Markets. *Journal of Financial Markets, 42, 1-44.*

Bollen, J., et al. (2011). Twitter Mood Predicts the Stock Market. *Journal of Computational Science, 2(1), 1-8.*

Bollen, J., & Huina, M. (2011). Twitter Mood as a Stock Market Predictor. *IEEE Computer, 44(10), 91-94.*

Breiman, L. (2001). Random Forests. *Machine Learning, 45(1), 5-32.*

Brunnermeier, M. K. (2009). Deciphering the Liquidity and Credit Crunch 2007–2008. *Journal of Economic Perspectives, 23(1), 77-100.*

Caldara, D., & Iacoviello, M. (2018). Measuring Geopolitical Risk. *American Economic Review, 109(4), 1194-1222.*

Devlin, J., et al. (2018). *BERT:* Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv prep.*

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance, 25(2), 383-417.*

Fama, E. F. (1991). Efficient Capital Markets: II. *The Journal of Finance, 46(5), 1575-1617.*

Fischer, T., & Krauss, C. (2018). Deep Learning with Long Short-Term Memory Networks for Financial Market Predictions. *European Journal of Operational Research, 270(2), 654-669.*

Gu, S., et al. (2020). Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies, 33(5), 2223-2273.*

Hendershott, T., et al. (2011). Does Algorithmic Trading Improve Liquidity? *The Journal of Finance, 66(1), 1-33.*

Hendershott, T., & Riordan, R. (2013). Algorithmic Trading and the Market for Liquidity. *Journal of Financial and Quantitative Analysis, 48(4), 1001-1024.*

Hasbrouck, J. (1993). Assessing the Quality of a Security Market: A New Approach to Transaction-Cost Measurement. *The Review of Financial Studies, 6(1), 191-212.*

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation, 9(8), 1735-1780.*

Hu, Z., et al. (2021). A Survey of Forex and Stock Price Prediction Using Deep Learning. *Applied System Innovation, 4(1), 9.*

Jones, C. M. (2013). What Do We Know About High-Frequency Trading? *Columbia Business School Research Paper, (13-11).*

Kim, K. J., & Won, H. K. (2018). Forecasting the Volatility of Stock Price Index: A Hybrid Model Integrating LSTM with Multiple GARCH-Type Models. *Expert Systems with Applications, 103, 25-37.*

Kroll, J. A., et al. (2017). Accountable Algorithms. *University of Pennsylvania Law Review, 165(3), 633-705. LeCun, Y., et al. (2015). Deep Learning. Nature, 521(7553), 436-444.*

Li, X., et al. (2014). News Impact on Stock Price Return via Sentiment Analysis. *Knowledge-Based Systems, 69, 14-23.*

Li, X., et al. (2017). Improving Stock Market Prediction by Integrating Both Market News and Stock Prices. *2017 IEEE 14th International Conference on E-Business Engineering, 197-204.*

Mishkin, F. S., & Eakins, S. G. (2012). Financial Markets and Institutions. *Pearson Education.*

Mittal, A., & Goel, A. (2012). Stock Prediction Using Twitter Sentiment Analysis. *Stanford University CS229.*

Pastor, Ľ., & Veronesi, P. (2013). Political Uncertainty and Risk Premia. *Journal of Financial Economics, 110(3), 520-545.*

Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development, 3(3), 210-229.*

Schmidhuber, J. (2015). Deep Learning in Neural Networks: An Overview. *Neural Networks, 61, 85-117.*

Selvin, S., et al. (2017). Stock Price Prediction Using LSTM, RNN and CNN-sliding Window Model. 2017 *International Conference on Advances in Computing, Communications and Informatics, 1643-1647.*

Silver, D., et al. (2016). Mastering the Game of Go with Deep Neural Networks and Tree Search. *Nature, 529(7587), 484-489.*

Zhang, G. P. (2003). Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model.