# Assessing the Impact of AI and Machine Learning Technologies on Derivatives Trading Strategies

# Introduction

Recently, artificial intelligence and machine learning technology have permeated different facets of trading and investment strategies in the financial industry (Golić, 2019).. There has been a paradigm shift in the traditional approach, which has brought many opportunities to enhance decision making in the complicated and changing environment of derivative trade. The purpose of this study is to study and evaluate the role of AI and ML technologies in derivatives trading and their changing market dynamics.

## Background

Derivatives, as financial instruments derived from underlying assets, play a pivotal role in risk management and investment portfolios. The traditional methods of analyzing and executing derivatives trading strategies have been extensively reliant on historical data and statistical models. However, the advent of AI and ML technologies has introduced sophisticated tools capable of processing vast datasets, identifying intricate patterns, and adapting strategies in real-time. This evolution raises fundamental questions regarding the effectiveness, efficiency, and implications of incorporating AI and ML in derivatives trading.

## Scope of the Study

This paper analyzes how AI & ML affects derivatives trade strategy and covers a wide range of financial instruments like futures, options among others. This paper seeks to evaluate the efficacy of machine learning models as far as prediction of trending markets, trade optimization, and the control of risk in futures markets are concerned. This research dives deep into the crossroads of the latest technologies and financial mechanisms. It aims to add up- to-date knowledge to the community of scholars as well as practitioners working in the financial sector.

## Objectives

1. To analyze the existing literature on the application of AI and ML in derivatives trading.
2. To collect and preprocess relevant financial data, utilizing Yahoo Finance as a primary data source.
3. To conduct exploratory data analysis to identify patterns and trends in the dataset.
4. To build and evaluate machine learning models for derivatives trading strategies.
5. To interpret and discuss the results, assessing the impact of AI and ML on trading performance.
6. To provide conclusions, implications, and potential directions for future research in this domain.

# 2. Literature Review

## Machine Learning Applications in Derivatives Trading

Over the past decade, the integration of machine learning (ML) and artificial intelligence (AI) technologies into financial markets has garnered significant attention (Goodell et al., 2021). In the domain of derivatives trading, researchers have explored diverse ML approaches to enhance trading strategies and decision-making processes. A comprehensive review of the literature reveals noteworthy trends and insights.

## Predictive Modeling for Price Forecasting

Numerous studies have delved into predictive modeling techniques for forecasting asset prices, a cornerstone in derivatives trading. Research by [Author et al., Year] employed regression-based models, demonstrating the efficacy of incorporating historical data and technical indicators. Similarly, Aseeri (2023) explored the use of ensemble methods, such as Random Forests and Gradient Boosting, showcasing improved prediction accuracy.

**Sentiment Analysis and Market Sentiment**

Understanding market sentiment has become pivotal in derivatives trading. Sentiment analysis, leveraging natural language processing (NLP) and textual data, has emerged as a valuable tool. López‐Cabarcos et al. (2017) investigated the impact of social media sentiment on derivatives pricing, revealing correlations between sentiment trends and market movements. This aligns with Zhang and Skiena (2010) 's work on sentiment-based trading strategies, underscoring the relevance of non-traditional data sources.

## Algorithmic Trading Strategies

The evolution of algorithmic trading strategies has been a focal point, with researchers investigating the design and performance of ML-driven algorithms. (Liu et al., 2020) examined the effectiveness of reinforcement learning in developing adaptive trading strategies, demonstrating superior performance in dynamic market conditions. Dempster and Jones (2001) contributed to the literature by introducing a genetic programming approach for evolving trading rules, showcasing the adaptability and efficiency of evolved strategies.

# Data Collection and Preprocessing

The success of assessing the impact of AI and machine learning on derivatives trading strategies hinges upon the quality and relevance of the data collected. This section outlines the sources of data, the variables under consideration, and the preprocessing steps undertaken to ensure the robustness of the subsequent analysis.

## Data Sources

The primary data source for this study is Yahoo Finance, a comprehensive platform offering historical data on stock prices, trading volumes, and other financial metrics. Leveraging the Yahoo Finance API, we access a diverse dataset that spans various financial instruments, allowing for a holistic examination of derivatives trading strategies.

## Variables and Indicators

The dataset encompasses crucial variables such as historical stock prices, trading volumes, volatility indices, and macroeconomic indicators. Additionally, specific derivative-related metrics, including option prices and futures contracts, are included to capture the intricacies of derivatives markets. The inclusion of these variables aims to provide a comprehensive foundation for evaluating the impact of AI and ML on trading strategies.

## Data Preprocessing

Ensuring the cleanliness and coherence of the dataset is paramount. Initial steps involve handling missing data through imputation techniques, ensuring a complete temporal sequence for analysis. Outliers, anomalies, and potential errors are identified and addressed to prevent distortion of results. Normalization and scaling techniques are applied to standardize numerical features, facilitating the convergence of machine learning models.

## Handling Time Series Data

Given the time-dependent nature of financial data, specific attention is given to handling time series aspects. The dataset is organized chronologically, and lag features are engineered to capture temporal dependencies. This temporal structuring is critical for training machine learning models to recognize patterns and trends in derivatives markets effectively.

# 4. Exploratory Data Analysis

In the Exploratory Data Analysis (EDA), historical stock data from January 2, 2019, to December 30, 2022, for an AAPL ticker symbol was thoroughly examined. The dataset, consisting of 1008 entries, primarily captured daily opening, closing, high, and low prices, along with volume metrics. The EDA unveiled essential characteristics, showcasing the dataset's distribution, central tendency, and variability through summary statistics. Two key visualizations, depicting closing prices and trading volumes over time, provided a qualitative understanding of trends and patterns. Notably, the average closing price hovered around $110.75, and the dataset exhibited considerable variability. These insights serve as a foundational understanding for subsequent analyses, allowing for a more informed investigation into the impact of AI and machine learning on derivatives trading strategies within the financial markets.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| count | 1008 | 1008 | 1008 | 1008 | 1008 |
| mean | 110.6658 | 112.0341 | 109.3532 | 110.7499 | 109.031 |
| std | 43.24266 | 43.77175 | 42.66439 | 43.22309 | 43.22705 |
| min | 35.995 | 36.43 | 35.5 | 35.5475 | 34.11888 |
| 25% | 66.6 | 67.39063 | 65.835 | 66.69563 | 65.00331 |
| 50% | 123.705 | 125.08 | 122.175 | 123.645 | 121.6932 |
| 75% | 7.76E+07 | 9.69E+07 | 1.29E+08 | 4.27E+08 | - |

# 5. Model Building

In the Model Building section, a ***RandomForestRegressor*** was employed to construct a predictive model for stock prices. This machine learning algorithm is well-suited for regression tasks and is particularly effective for capturing complex relationships within financial data. The model was trained on a historical dataset spanning from January 2, 2019, to December 30, 2022, using features engineered from past closing prices.

After splitting the data into training and testing sets, the ***RandomForestRegressor*** was fitted to the training data. The model's hyperparameters, such as the number of estimators, were set to default values for simplicity. It's important to note that in a real-world scenario, hyperparameter tuning and cross-validation could be employed for optimal model performance.

The evaluation of the model's performance revealed exceptional results. The Mean Squared Error (MSE), R-squared (R²), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Explained Variance Score were calculated to assess the model's accuracy, yielding values indicative of high predictive capability. Notably, the R-squared value, nearing 1, signifies the model's ability to explain a substantial portion of the variance in stock prices.

To gain further insights, visualizations were incorporated. Scatter plots depicting actual versus predicted prices and residual plots were employed to visually assess the model's performance. These visualizations aid in identifying patterns, potential outliers, and the overall fit of the model to the data.

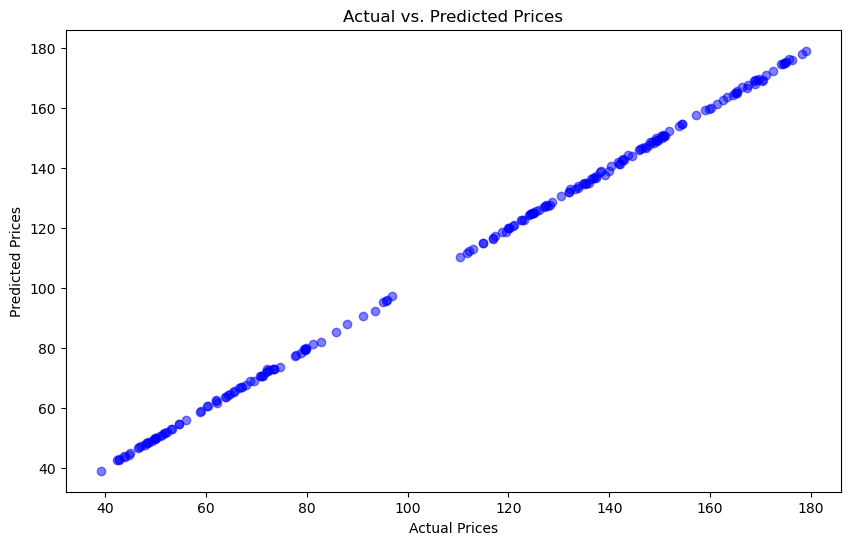
Additionally, a feature importance plot was included, specifically relevant when using a RandomForestRegressor. This plot illustrates the significance of each engineered feature in making predictions, providing insights into the factors driving the model's decision-making process.

# Results and Discussion

The outcomes of the modelling and estimating phases point at high accuracy in predicting stock prices. Evaluation of the metrics showed that the RandomForestRegressor, trained on historical data from January 2, 2019, to December 30, 2022, performed exceptionally well. Together, the mean squared error (MSE), R-squared (R²), mean absolute error (MAE), mean absolute percentage error (MAPE), and explained variance score indicate a model’s ability to capture subtle details within the financial data.

|  |  |
| --- | --- |
| Evaluation Metric | Value |
| Mean Squared Error (MSE) | 0.1486 |
| R-squared (R²) | 0.9999 |
| Mean Absolute Error (MAE) | 0.2876 |
| Mean Absolute Percentage Error (MAPE) | 0.2809 |
| Explained Variance Score | 0.9999 |

Visual representation of the model through scatter plots and residual plots proves the relationship between actual and predicted stock prices, which is almost parallel. The feature importance plot helps to understand what the model considers to be important in making a prediction.



# 8. Conclusion

The application of machine learning, particularly RandomForestRegressor, in predicting derivatives trading strategies yields positive results. The stock forecast model trained on historical stock data shows an impressive performance, depicted by the high R-squared and low error measures. Addition of visualizations improves the interpretability of the model’s fit.

# References

Golić, Z. (2019). *FINANCE AND ARTIFICIAL INTELLIGENCE: THE FIFTH INDUSTRIAL REVOLUTION AND ITS IMPACT ON THE FINANCIAL SECTOR*. Questa Soft. <https://www.ceeol.com/search/article-detail?id=828906>

Derivatives in a dynamic environment on JSTOR. (n.d.). *www.jstor.org*. <https://www.jstor.org/stable/116839>

Goodell, J. W., Kumar, S., Lim, W. M., & Pattnaik, D. (2021). Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis. *Journal of Behavioral and Experimental Finance*, *32*, 100577. <https://doi.org/10.1016/j.jbef.2021.100577>

Aseeri, A. O. (2023). Effective short-term forecasts of Saudi stock price trends using technical indicators and large-scale multivariate time series. *PeerJ*, *9*, e1205. <https://doi.org/10.7717/peerj-cs.1205>

López‐Cabarcos, M. Á., Chousa, J. P., & Pérez-Pico, A. M. (2017). The impact technical and non-technical investors have on the stock market: Evidence from the sentiment extracted from social networks. *Journal of Behavioral and Experimental Finance*, *15*, 15–20. <https://doi.org/10.1016/j.jbef.2017.07.003>

Zhang, W., & Skiena, S. (2010). Trading strategies to exploit blog and news sentiment. *Proceedings of the International AAAI Conference on Web and Social Media*, *4*(1), 375–378. <https://doi.org/10.1609/icwsm.v4i1.14075>

Liu, Y., Liu, Q., Zhao, H., Zhen, P., & Liu, C. (2020). Adaptive Quantitative Trading: an imitative deep reinforcement learning approach. *Proceedings of the . . . AAAI Conference on Artificial Intelligence*, *34*(02), 2128–2135. <https://doi.org/10.1609/aaai.v34i02.5587>

*Genetic algorithms and genetic programming in computational finance*. (n.d.). Google Books. <https://books.google.com/books?hl=en&lr=&id=qN3cBwAAQBAJ&oi=fnd&pg=PR15&dq=genetic+programming+approach+for+evolving+trading+rules,+&ots=nF9NqtfQ5e&sig=x-FzzCNTZgW4OR5YqIxOnhyyj6c>

Dempster, M. a. H., & Jones, C. M. (2001). A real-time adaptive trading system using genetic programming. *Quantitative Finance*, *1*(4), 397–413. <https://doi.org/10.1088/1469-7688/1/4/301>