AI-Driven Stock Prediction and Investment Strategy Analysis for the Magnificent 7 Tech Giants

Jumin Lee

ABSTRACT

This study aims to predict stock prices and develop optimized investment strategies for the "Magnificent 7" companies using AI. By employing Linear Regression, Random Forest, XGBoost, and LSTM models, the study analyzed stock price data for Apple, Amazon, Google, Meta, Microsoft, Nvidia, and Tesla from 2015 to the first half of 2024, along with VIX index and bond yield data. The accuracy of predictions for each model was assessed, and simulations were conducted to observe which model would achieve the highest returns when applied to actual investments. The results showed varying levels of predictive accuracy and performance for each stock, leading to the conclusion that a hybrid approach combining AI and traditional investment strategies could be effective.

☞ keyword : AI, Machine Learning, Stock Prediction, Investment Strategy, Magnificent 7, LSTM, XGBoost, Random Forest, Linear Regression

1. Introdutcion

Predicting stock prices and developing investment strategies have long been focal points of research in the financial sector. The inherent volatility of the stock market presents a formidable challenge for investors, making the development of predictive methodologies crucial. Traditionally, stock price forecasting relied on economic indicators, corporate performance, and technical analyses, but these approaches often depend heavily on historical data and may fall short of capturing the complex and nonlinear market dynamics.

Meanwhile, the stock prices of the major tech giants collectively known as the "Magnificent 7" (Apple, Amazon, Google, Meta, Microsoft, Nvidia, and Tesla) exert substantial influence not only on their respective companies and the U.S. stock market but also on the global economy. Predicting the stock prices of these companies is, therefore, of critical importance for global investors from both micro and macro perspectives. This makes them an ideal target for research utilizing machine learning and artificial intelligence, further underscoring the academic and practical value of this endeavor.

The primary academic objective of this study is to evaluate the accuracy of stock price predictions using various machine learning models. For this purpose, four models were selected: Linear Regression, Random Forest, XGBoost, and Long Short-Term Memory (LSTM). By comparing the performance of these models, this study aims to identify the most effective and suitable model for forecasting specific stock prices. For example, while a simple linear regression model might perform well in certain scenarios, more complex models like XGBoost, which excel at capturing nonlinear relationships, could achieve superior performance in others. Model performance will be evaluated by comparing prediction accuracy on training and test data sets using metrics such as Root Mean Squared Error (RMSE) and R^2 scores.

The second academic objective is to optimize the performance of each model through hyperparameter tuning and feature engineering. Hyperparameter tuning is a critical process for maximizing model predictive performance, employing techniques like Grid Search to test various parameter combinations and derive the optimal settings. Feature engineering involves transforming data and creating new variables to enhance model predictive capacity. This study employed data normalization, moving average calculations, and the integration of VIX index and treasury yield data to improve data quality, thereby enabling more accurate predictions by the models.

Another intention of this study is to generate investment insights. The first investment-related objective is to simulate and assess the expected performance of investments based on predictions obtained from each machine learning model. By applying the predictions to various stocks and evaluating potential investment outcomes, this research seeks to verify the practical effectiveness and profitability of machine learning-based investment strategies.

The second investment-related objective is to derive optimal investment methodologies based on stock price predictions and simulation results for the Magnificent 7 companies. This analysis aims to determine which strategies yield the highest returns and which strategies mitigate risk while generating stable profits.

In conclusion, the purpose and significance of this study lie in simultaneously providing academic contributions and practical investment guidance by examining the predictive capabilities of machine learning models and their applicability to real-world investment strategies. This study explores the potential of AI-driven investment strategies and seeks to verify more efficient and effective methodologies for stock market investments.

2. Theoretical Background and Previous Studies

2.1 Utilization of Machine Learning

Machine learning refers to computational algorithms designed to learn from given situations and emulate human intelligence. Recently, it has undergone rapid advancements across diverse domains. Since 2011, studies leveraging machine learning for stock market predictions have gained momentum, demonstrating its effectiveness in analyzing and predicting complex stock market patterns [1][2][3]. Machine learning algorithms can extract meaningful patterns from vast datasets, significantly improving the accuracy of stock price predictions.

2.2 Combining Statistical Methods with Machine Learning

The integration of statistical methods and machine learning techniques in stock market forecasting plays a vital role in maximizing prediction performance, with numerous systematic reviews and applications of these combined methods in stock market predictions [4]. Such studies suggest that combining traditional statistical techniques with innovative machine learning methods can enhance the performance of predictive models. For example, statistical methods are beneficial for understanding the fundamental characteristics of data, while machine learning excels at analyzing and predicting more complex patterns based on these characteristics.

2.3 Applications of Deep Learning

Deep learning, a subset of machine learning, is particularly effective at processing large-scale datasets and learning complex patterns. According to Jiang (2019), deep learning techniques have been successfully applied to stock market forecasting, with models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) demonstrating impressive performance [5]. Deep learning models are well-suited for learning hidden patterns from large datasets and making accurate predictions based on those patterns.

2.4 Practical Insights

Various studies confirm that machine learning and deep learning models are highly effective at improving the accuracy of stock market predictions [1][4]. Additionally, deep learning models have contributed significantly to analyzing complex data patterns and enhancing predictive performance [5][6].

3. Methodology

3.1 Data Collection and Preprocessing

This study utilizes daily stock price data for the Magnificent 7 companies (Apple, Amazon, Google, Meta, Microsoft, Nvidia, and Tesla) from 2015 to the present (2024). The dataset includes information such as daily closing prices, trading volumes, highs, and lows for each company. Furthermore, data on the Volatility Index (VIX) and bond yields over the same period were collected and used.

(Table 1) Types and Sources of Data Used

Type	Description	Source
Stock Prices Data	The dataset includes stock price data for seven major tech companies under study: Apple, Amazon, Google, Meta, Microsoft, Nvidia, and Tesla. Data fields include closing price (Close), trading volume (Volume), opening price (Open), highest price (High), and lowest price (Low).	Yahoo Finance
VIX Data	Volatility index data representing stock market volatility, with VIX reflecting market uncertainty and investor sentiment. It is commonly referred to as the "fear index".	CBOE (Chicago Board Options Exchange)
Bond Yield Data	Yield data of major bonds influencing the stock market. For instance, the 10-year U.S. Treasury yield significantly impacts the stock market based on economic conditions and interest rate fluctuations.	FRED (Federal Reserve Economic Data)

The collected data was integrated by stock using Python's Pandas library, with missing values handled and transformed into a format suitable for model training.

(Table 2) Sample Format of Preprocessed & Integrated Data (Apple)

Date	Close	Adj Close	Volume	Close VIX	Bond Yield	50 MA	200 MA	Volume Change

2015-10-16	27.76	25.10	156,930,400	15.05	2.04	25.39	27.27	0.04
2015-10-19	27.93	25.25	119,036,800	14.98	2.04	25.38	27.27	-0.24
2015-10-20	28.44	25.72	195,871,200	15.75	2.08	25.35	27.28	0.65
2015-10-21	28.44	25.71	167,180,800	16.70	2.04	25.35	27.29	-0.15

(Table 3) Detailed Data Items for Integrated Dataset

Tyoe	Description
Date	Trading date.
Close	Closing price of the stock for the trading day.
Adj Close	Adjusted closing price, reflecting factors such as dividends and stock splits.
Volume	Trading volume (number of shares traded on that day).
_VIX	Volatility Index (VIX) for the trading day, reflecting market uncertainty and investor sentiment.
Bond_Yield	Bond yield for the trading day (e.g., the 10-year U.S. Treasury yield, which significantly impacts the stock market based on economic conditions and interest rate fluctuations).
50_MA	50-day moving average (average closing price of the stock over the past 50 trading days).
200_MA	200-day moving average (average closing price of the stock over the past 200 trading days).
Volume_Change	Change in trading volume from the previous day (current trading day's volume minus the previous trading day's volume, indicating the extent of volume fluctuation).

3.2 Model Development and Training

Based on the literature review and the available analytical environment for this study, the following models were selected, and each was developed and trained using the appropriate methodologies:

- Linear Regression: Models the linear trends of stock prices.
- Random Forest: An ensemble technique used to capture nonlinear patterns.
- XGBoost: Optimizes prediction performance through boosting techniques.
- LSTM (Long Short-Term Memory): A recurrent neural network model designed for time series data processing.

For model training and testing, data was first normalized and transformed into a suitable format for training. Using the 'create_dataset' function, stock price data was grouped into sets of 100 days, and the model was tasked with predicting the stock price of the following day.

- Train on data from day 1 to 100 and predict day 101.
- Train on data from day 2 to 101 and predict day 102.
- Train on data from day 3 to 102 and predict day 103.

...

- Train on data from day n to n+99 and predict day n+100.

The generated dataset was split randomly into 80% for training and 20% for testing (using the train_test_split function) while maintaining the time-series order. The first 80% was used for training, and the remaining 20% for testing. The 80:20 ratio was adopted as it is a widely validated conventional split, ensuring a sufficiently large training set to prevent overfitting while maintaining a robust testing set to minimize the risk of inadequate model performance evaluation.

3.3 Performance Evaluation Metrics

The models' performance was evaluated by comparing prediction accuracy on the training and testing datasets using the following metrics:

- RMSE (Root Mean Squared Error): The square root of the average squared differences between predicted and actual values. Lower values indicate higher prediction accuracy.
- R^2 (R-squared): Indicates the proportion of variance in the dependent variable that is predictable from the independent variables. Values closer to 1 indicate better model explanatory power.

4. Results

4.1 Apple(APPL)

4.1.1 Model Performance Evaluation Results

- Linear Regression
  Train RMSE: 0.00721370768084068
  Train R^2: 0.9993518408562049
  Test RMSE: 0.013456124594076024
  Test R^2: 0.9978237212930056
- Random Forest
  Train RMSE: 0.003872961711382928
  Train R^2: 0.9998131681216459
  Test RMSE: 0.010092068824678632
  Test R^2: 0.9987758492002643
- XGBoost
  Train RMSE: 0.002713807236596701
  Train R^2: 0.9999082676402655

Test RMSE: 0.009915095578910946
Test R^2: 0.99881840587347

- LSTM
  Train RMSE: 3.9092045598054934
  Train R^2: 0.9954810401662878
  Test RMSE: 4.039114495183321
  Test R^2: 0.9953447398420988

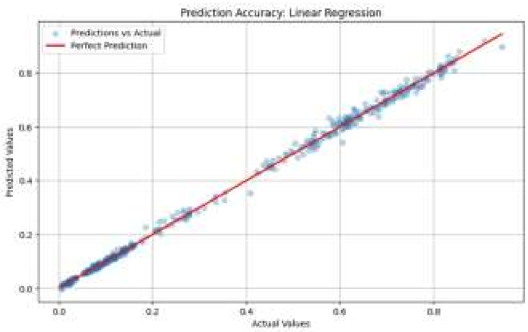
4.1.2 Investment Simulation Results

Assuming an initial seed money of $10,000 and using each model to simulate investments from January 2, 2015, to the present, the expected returns are as follows:

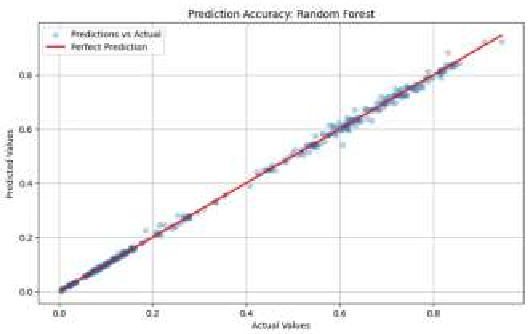
- Linear Regression: $77,702.48
- Random Forest: $74,160.32
- XGBoost: $78,953.78
- LSTM: $80,976.98
- Buy & Hold: $78,079.97
  (Hold after bulk purchase on January 2, 2015)

4.1.3 Model Performance Comparison

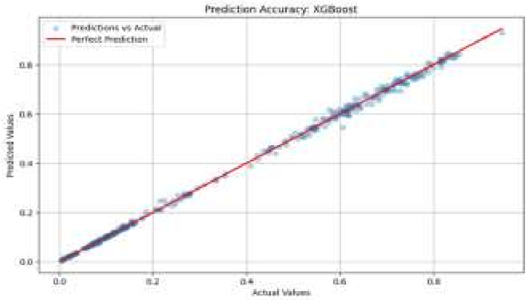
- Best Performing Model: XGBoost (based on superior RMSE and R^2 metrics)
- Highest Profit Model: LSTM
- Additional Notes: The specialized performance of the LSTM model for time series data may have contributed to its strong long-term profitability.



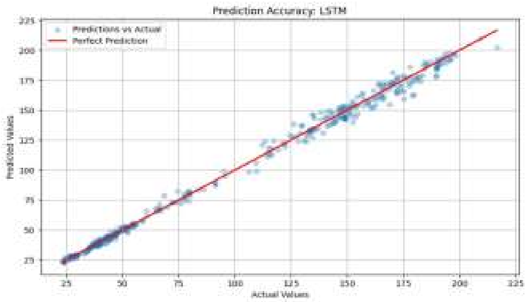
(Image 1) APPL Linear Regression Prediction Accuracy Scatter Plot



(Image 2) APPL Random Forrest Prediction Accuracy Scatter Plot



(Image 3) APPL XGBoost Prediction Accuracy Scatter Plot



(Image 4) APPL LSTM Prediction Accuracy Scatter Plot

4.2 Amazon(AMZN)

4.2.1 Model Performance Evaluation Results

- Linear Regression
  Train RMSE: 0.010248942532380 69
  Train R^2: 0.9984772426597099
  Test RMSE: 0.01740486189657552
  Test R^2: 0.9955758516518073
- Random Forest
  Train RMSE: 0.00543436952479615
  Train R^2: 0.9995718746755582
  Test RMSE: 0.014243840390853992
  Test R^2: 0.9970369245178632
- XGBoost
  Train RMSE: 0.0092742528308425
  Train R^2: 0.9987531033783534
  Test RMSE: 0.01467343541852455
  Test R^2: 0.9968554961719535
- LSTM
  Train RMSE: 4.173629609369123
  Train R^2: 0.9918381990420058
  Test RMSE: 4.268768899802902
  Test R^2: 0.9913983966923228

4.2.2 Investment Simulation Results

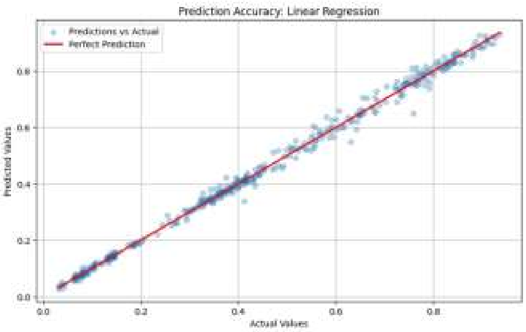
Assuming an initial seed money of $10,000 and using each model to simulate investments from January 2, 2015, to the present, the expected returns are as follows:

- Linear Regression: $67,315.58

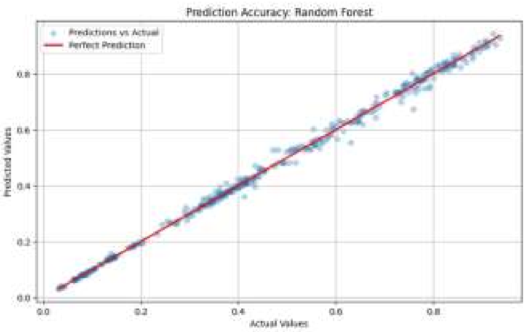
- Random Forrest: $67,915.98
- XGBoost: $67,524.25
- LSTM: $72,617.99
- Buy & Hold: $69,100.85
  (Hold after bulk purchase on January 2, 2015)

4.2.3 Model Performance Comparison

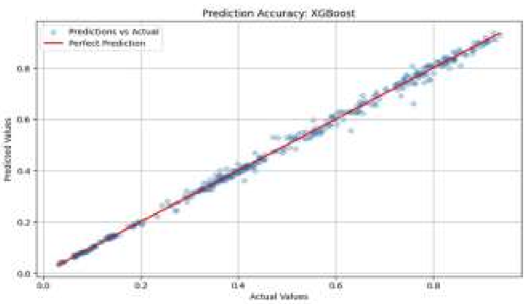
- Best Performing Model: Random Forest (based on superior RMSE and R^2 metrics)
- Highest Profit Model: LSTM
- Additional Notes: The specialized performance of the LSTM model for time series data may have contributed to its strong long-term profitability.



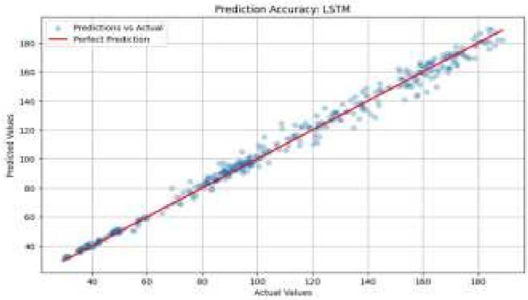
(Image 5) AMZN Linear Regression Prediction Accuracy Scatter Plot



(Image 6) AMZN Random Forrest Prediction Accuracy Scatter Plot



(Image 7) AMZN XGBoost Prediction Accuracy Scatter Plot



(Image 8) AMZN LSTM Prediction Accuracy Scatter Plot

4.3 Google(GOOG)

4.3.1 Model Performance Evaluation Results

- Linear Regression
  Train RMSE: 0.007584069124673875
  Train R^2: 0.9990256817711023
  Test RMSE: 1.306391154418243
  Test R^2: -27.65752534522947
- Random Forest
  Train RMSE: 0.004066965400911 82
  Train R^2: 0.9997198199438097
  Test RMSE: 0.012584873126705991
  Test R^2: 0.9973405617786385
- XGBoost
  Train RMSE: 0.006815078229777993
  Train R^2: 0.9992132478227236
  Test RMSE: 0.012310704901908068
  Test R^2: 0.9974551741655481
- LSTM
  Train RMSE: 3.406891096143106
  Train R^2: 0.9923034011633383
  Test RMSE: 3.605051117097021
  Test R^2: 0.9914571818277564

4.3.2 Investment Simulation Results

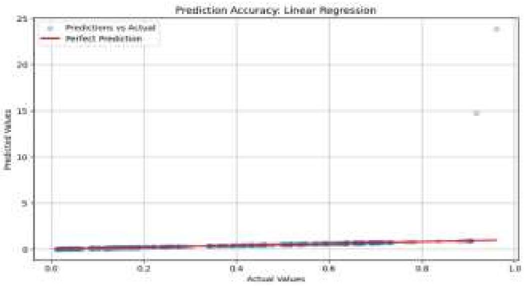
Assuming an initial seed money of $10,000 and using each model to simulate investments from January 2, 2015, to the present, the expected returns are as follows:

- Linear Regression: $57,874.78
- Random Forrest: $53,630.16
- XGBoost: $53,844.67
- LSTM: $54,501.17
- Buy & Hold: $55,720.33
  (Hold after bulk purchase on January 2, 2015)

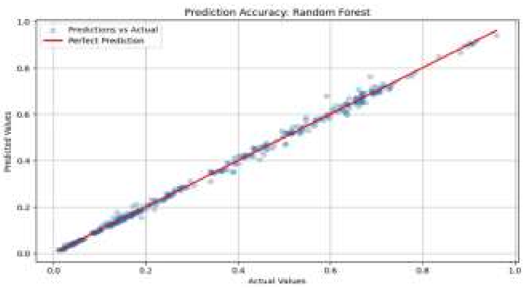
4.3.3 Model Performance Comparison

- Best Performing Model: XGBoost, Random Forest (based on superior RMSE and R^2 metrics)
- Highest Profit Model: Linear Regression
- Additional Notes: The Linear Regression model

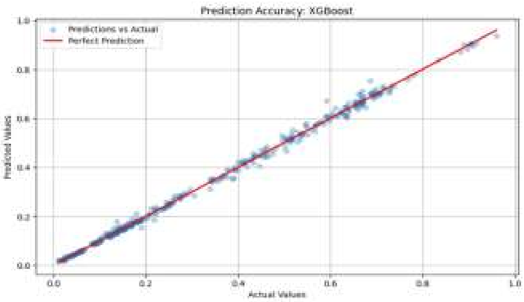
has good Train RMSE and Train R^2, but poor Test RMSE and Test R^2 indicating overfitting and particularly the low Test R^2 suggests that the model does not explain the test data well. Despite this, it shows strong performance in terms of returns.



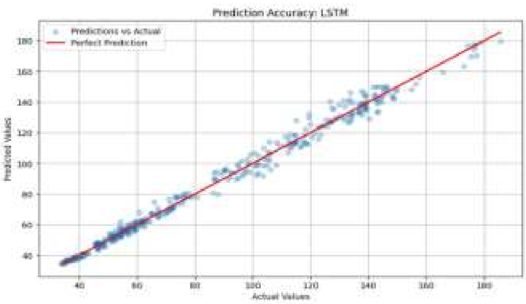
(Image 9) GOOG Linear Regression Prediction Accuracy Scatter Plot



(Image 10) GOOG Random Forrest Prediction Accuracy Scatter Plot



(Image 11) GOOG XGBoost Prediction Accuracy Scatter Plot



(Image 12) GOOG LSTM Prediction Accuracy Scatter Plot

4.4 Meta(META)

4.4.1 Model Performance Evaluation Results

- Linear Regression
  Train RMSE: 0.009598516669102537
  Train R^2: 0.997895457708489
  Test RMSE: 0.017520560791958403
  Test R^2: 0.9923682025495548
- Random Forest
  Train RMSE: 0.005150499879096363
  Train R^2: 0.999394034117066
  Test RMSE: 0.014090458652083021
  Test R^2: 0.9950639325031585
- XGBoost
  Train RMSE: 0.007878474457522076
  Train R^2: 0.9985821389464132
  Test RMSE: 0.013624728776867986
  Test R^2: 0.9953848421314374
- LSTM
  Train RMSE: 9.973018666277179
  Train R^2: 0.9888300906047836
  Test RMSE: 10.663743413230975
  Test R^2: 0.9861005639621295

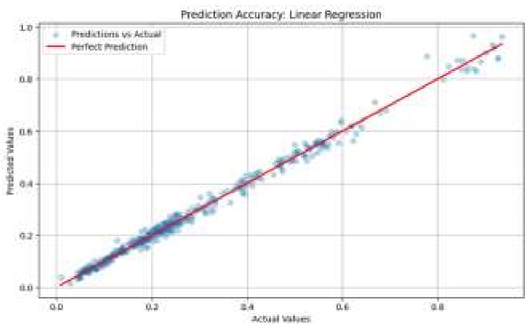
4.4.2 Investment Simulation Results

Assuming an initial seed money of $10,000 and using each model to simulate investments from January 2, 2015, to the present, the expected returns are as follows:

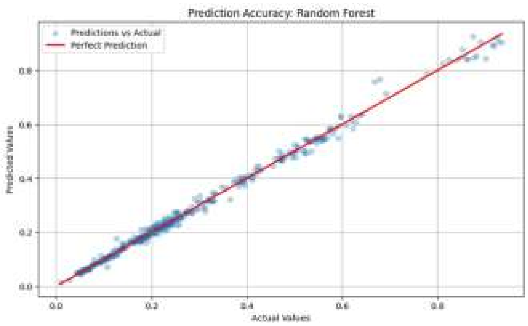
- Linear Regression: $51,615.16
- Random Forrest: $51,014.77
- XGBoost: $53,029.05
- LSTM: $51,506.32
- Buy & Hold: $51,740.82
  (Hold after bulk purchase on January 2, 2015)

4.4.3 Model Performance Comparison

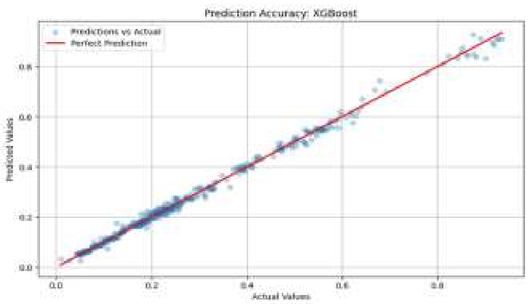
- Best Performing Model: Random Forest, XGBoost (based on superior RMSE and R^2 metrics)
- Highest Profit Model: XGBoost
- Additional Notes: Although the LSTM model is strong in time series analysis, it showed the lowest performance compared to other models and recorded the second-lowest returns.



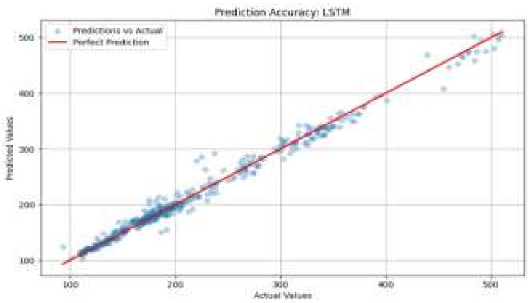
(Image 13) META Linear Regression Prediction Accuracy Scatter Plot



(Image 14) META Random Forrest Prediction Accuracy Scatter Plot



(Image 15) META XGBoost Prediction Accuracy Scatter Plot



(Image 16) META LSTM Prediction Accuracy Scatter Plot

4.5 Microsoft(MSFT)

4.5.1 Model Performance Evaluation Results

- Linear Regression
  Train RMSE: 0.006396164498521088
  Train R^2: 0.9993954886056274
  Test RMSE: 0.01198268359729427
  Test R^2: 0.9978577520294
- Random Forest
  Train RMSE: 0.003288831314077367
  Train R^2: 0.9998401736284992
  Test RMSE: 0.010068287897096047
  Test R^2: 0.9984875786230499
- XGBoost
  Train RMSE: 0.002313653802046101
  Train R^2: 0.9999209026393698
  Test RMSE: 0.009884630408063724
  Test R^2: 0.9985422520911019
- LSTM
  Train RMSE: 6.262766291547487
  Train R^2: 0.9967201431262446
  Test RMSE: 6.6783989994057515
  Test R^2: 0.9962341456385856

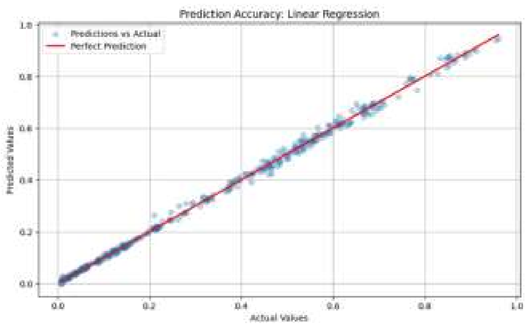
4.5.2 Investment Simulation Results

Assuming an initial seed money of $10,000 and using each model to simulate investments from January 2, 2015, to the present, the expected returns are as follows:

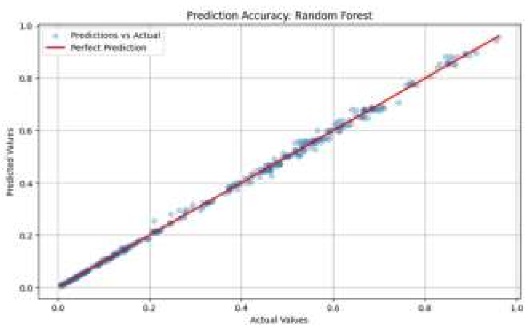
- Linear Regression: $97,156.06
- Random Forrest: $95,509.19
- XGBoost: $98,490.52
- LSTM: $94,653.07
- Buy & Hold: $96,133.45
  (Hold after bulk purchase on January 2, 2015)

4.5.3 Model Performance Comparison

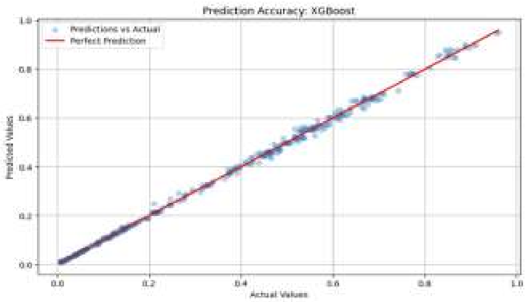
- Best Performing Model: XGBoost (based on superior RMSE and R^2 metrics)
- Highest Profit Model: XGBoost
- Additional Notes: Although the LSTM model is strong in time series analysis, it showed the lowest performance and returns compared to other models.



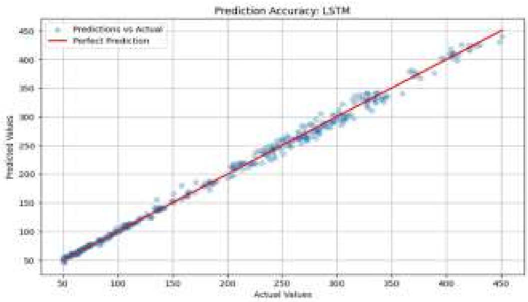
(Image 17) MSFT Linear Regression Prediction Accuracy Scatter Plot



(Image 18) MSFT Random Forrest Prediction Accuracy Scatter Plot



(Image 19) MSFT XGBoost Prediction Accuracy Scatter Plot



(Image 20) MSFT LSTM Prediction Accuracy Scatter Plot

4.6 Nvidia(NVDA)

4.6.1 Model Performance Evaluation Results

- Linear Regression
  Train RMSE: 0.003770818938547538
  Train R^2: 0.9994855066070386
  Test RMSE: 0.009463293046525049
  Test R^2: 0.9964062844613787
- Random Forest
  Train RMSE: 0.002500311935477405
  Train R^2: 0.9997737977966868
  Test RMSE: 0.005681129619470234
  Test R^2: 0.9987048248121455
- XGBoost
  Train RMSE: 0.0004672211128859668
  Train R^2: 0.9999921013405159
  Test RMSE: 0.007505210468866724
  Test R^2: 0.9977396022260528
- LSTM
  Train RMSE: 1.6066052684654817
  Train R^2: 0.9948715722217412
  Test RMSE: 1.4947424804646756
  Test R^2: 0.9950767793363464

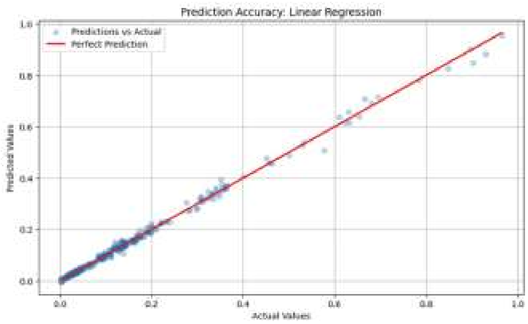
4.6.2 Investment Simulation Results

Assuming an initial seed money of $10,000 and using each model to simulate investments from January 2, 2015, to the present, the expected returns are as follows:

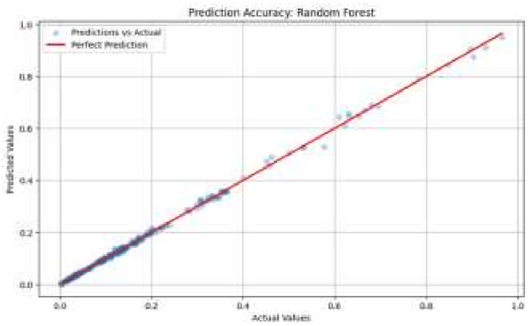
- Linear Regression: $1,995,663.85
- Random Forrest: $1,765,615.30
- XGBoost: $1,766,698.25
- LSTM: $1,553,832.96
- Buy & Hold: $1,784,637.52
  (Hold after bulk purchase on January 2, 2015)

4.6.3 Model Performance Comparison

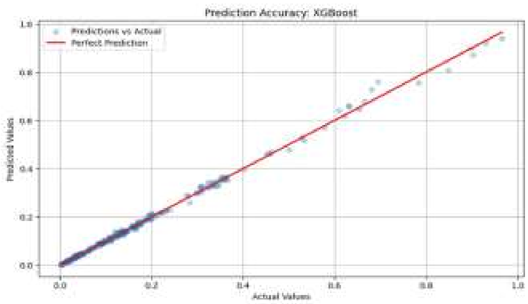
- Best Performing Model: Random Forest, XGBoost (based on superior RMSE and R^2 metrics)
- Highest Profit Model: Linear Regression
- Additional Notes: Although the LSTM model is strong in time series analysis, it showed the lowest performance and returns compared to other models.



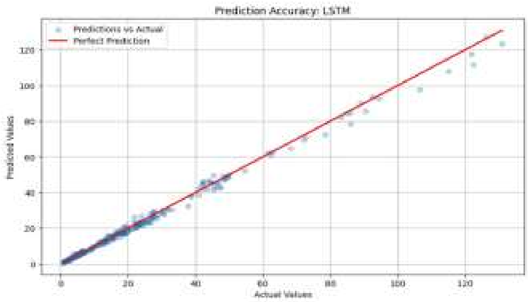
(Image 21) NVDA Linear Regression Prediction Accuracy Scatter Plot



(Image 22) NVDA Random Forrest Prediction Accuracy Scatter Plot



(Image 23) NVDA XGBoost Prediction Accuracy Scatter Plot



(Image 24) NVDA LSTM Prediction Accuracy Scatter Plot

4.7 Tesla(TSLA)

4.7.1 Model Performance Evaluation Results

- Linear Regression
  Train RMSE: 0.011075039168475045
  Train R^2: 0.998384971868137
  Test RMSE: 0.02171689158060674
  Test R^2: 0.9941871356753815
- Random Forest
  Train RMSE: 0.005993219225495276
  Train R^2: 0.9995270558281623
  Test RMSE: 0.017311966357605693
  Test R^2: 0.9963060785254125
- XGBoost
  Train RMSE: 0.0009925117622869 82
  Train R^2: 0.9999870293758624
  Test RMSE: 0.017603887595299603
  Test R^2: 0.9961804514585545
- LSTM
  Train RMSE: 10.010356379463106
  Train R^2: 0.9917696587410665
  Test RMSE: 10.218756656331266
  Test R^2: 0.9919717629150182

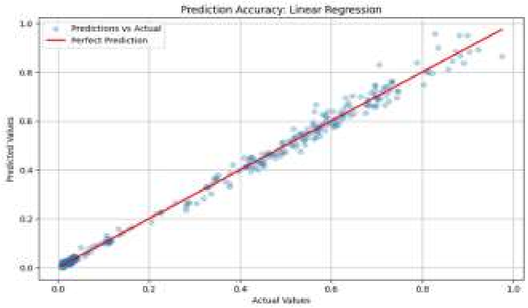
4.7.2 Investment Simulation Results

Assuming an initial seed money of $10,000 and using each model to simulate investments from January 2, 2015, to the present, the expected returns are as follows:

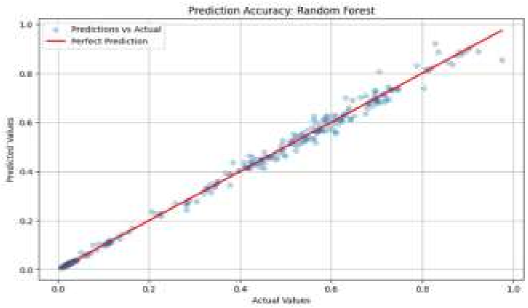
- Linear Regression: $142,466.32
- Random Forrest: $132,559.25
- XGBoost: $139,615.28
- LSTM: $141,050.34
- Buy & Hold: $138,667.90
  (Hold after bulk purchase on January 2, 2015)

4.7.3 Model Performance Comparison

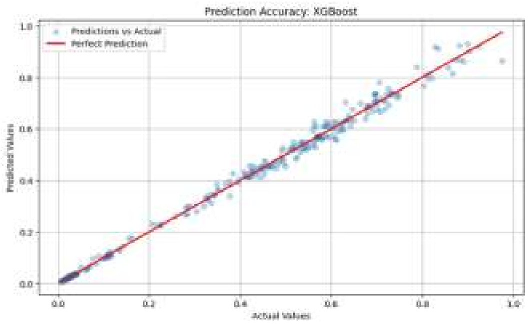
- Best Performing Model: Random Forest, XGBoost (based on superior RMSE and R^2 metrics)
- Highest Profit Model: Linear Regression
- Additional Notes: Despite its strengths in time series analysis, the LSTM model exhibited the lowest performance compared to the other models. However, it ranked second in terms of investment returns.



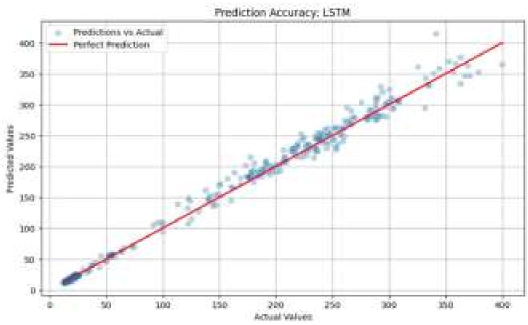
(Image 25) TSLA Linear Regression Prediction Accuracy Scatter Plot



(Image 26) TSLA Random Forrest Prediction Accuracy Scatter Plot



(Image 27) TSLA XGBoost Prediction Accuracy Scatter Plot



(Image 28) TSLA LSTM Prediction Accuracy Scatter Plot

5. Conclusion and Recommendations

5.1 Limitations of the Study

The data used in this study is limited to the period from 2015 to 2023. Although this is not a short period, it may not be sufficient to fully reflect macroeconomic cycles, market trends, and structural changes in stocks. Therefore, it may be necessary to analyze data over several decades to identify long-term patterns and cycles, which would enhance the stability and reliability of the predictive model. For instance, macroeconomic events such as economic recessions or booms can be more clearly analyzed using data spanning a longer time frame to assess their impact.

Complex machine learning models, particularly deep learning models, may face issues of overfitting, where the model becomes too tailored to the training data and loses its ability to generalize to new data. In this study, efforts were made to avoid overfitting by using methods such as cross-validation, regularization, and callbacks, and to prevent excessive complexity and hyperparameter tuning for each model. However, the results from the LSTM model suggest that further improvements may be necessary. Additionally, complex models require high computational costs and training time, which may pose challenges for real-time predictions and applications.

Moreover, the stock market is influenced by complex and diverse factors such as political events, economic policy changes, natural disasters, and technological innovations. Fully accounting for these external factors in the model is a challenging task. For example, unexpected events like a pandemic may significantly differ from the existing data patterns and affect the model's predictive performance. To integrate all these external factors, it would be necessary to incorporate real-time data feeds, news analysis, and social media data, but this would come with additional costs and technical challenges.

5.2 Recommendatinos

To enhance investment strategies using artificial intelligence, it is important to analyze the stock price patterns of each company and identify the external factors affecting them to understand the primary causes of stock price fluctuations.

- Design customized features based on the characteristics of each company and collect relevant external factor data. For example, for specific companies, technical indicators, macroeconomic indicators, and social media sentiment analysis could be included.
- Test various models and select the one most suitable for each company. For example, stocks with non-linear patterns may benefit from Random Forest or XGBoost, while time series data may be better suited for LSTM.
- Test the effectiveness of each model in actual investment scenarios through simulation.

Additionally, it is recommended to collect real-time stock price data continuously via APIs and periodically update the model by retraining it with new data every week or month, followed by deploying the retrained model for real-time predictions.

- Use data streaming platforms such as Apache Kafka or AWS Kinesis to collect and process real-time data.
- Automate the processes of data collection, preprocessing, model retraining, and deployment using tools like Apache Airflow or Jenkins.
- Use cloud services like AWS, Google Cloud, or Microsoft Azure to build scalable infrastructure.
- Establish an MLOps (Machine Learning Operations) framework for managing and operating these processes.

Using advanced ensemble models can refine predictions and yield improved results.

- Stacking: Combine different models to create a new meta-model. For instance, predictions from Random Forest, XGBoost, and LSTM can be used to train a final model to make the overall prediction.

Furthermore, integrating traditional investment risk management techniques can help reduce loss risk and improve profitability.

- Stop-loss Strategy: Automatically sell when the stock price drops below a certain level. By setting a stop-loss price based on machine learning predictions and automating its execution, significant losses can be prevented in the event of unforeseen occurrences.
- Hedging: Use derivatives such as options or futures based on model predictions to offset portfolio risks.
- Diversification: Use model predictions to invest in diverse asset classes, thus reducing risk from specific assets.

5.3 Conclusion

This study applied various machine learning models to predict stock prices and analyze investment strategies for the Magnificent 7 companies (Apple, Amazon, Google, Meta, Microsoft, Nvidia, and Tesla). The models used included Linear Regression, Random Forest, XGBoost, and LSTM, and the performance of each model was evaluated using RMSE and R^2. Further, investment simulations were conducted to validate the effectiveness of the strategies in real-world scenarios. The conclusions drawn from the results are as follows:

First, when performing stock price predictions using four machine learning models - Linear Regression, Random Forest, XGBoost, and LSTM - the models showed varying prediction performance depending on the stock. In general, Random Forest and XGBoost demonstrated high predictive accuracy. However, in some cases, Linear Regression and LSTM also showed competitive performance.

Second, based on investment simulations using machine learning models, the traditional "buy and hold" strategy outperformed the model-based approach in certain situations. This suggests the complexity of market timing and the difficulties of real-time trading, implying that AI-based strategies do not always guarantee superior performance. However, model-based strategies have the potential to generate high returns and can be an effective investment tool depending on market conditions.

It should be noted that the relatively short data period of this study limited the ability to fully capture long-term predictions and trends. Furthermore, complex models carry the risk of overfitting, and incorporating market volatility and external factors into the model was a challenge. Future studies should include datasets over a longer time period and additional market indicators to enhance predictive performance. Additionally, ensemble learning techniques and the use of diverse unstructured data should be considered to improve the robustness and accuracy of predictions.

In conclusion, AI and machine learning prediction models are powerful tools for stock price forecasting and optimizing investment strategies. However, this study confirms that these technology-driven approaches are more effective when combined with traditional investment methodologies. This suggests that the synergy between human expertise and machine intelligence can enhance innovation and efficiency in investment management. Continuous model improvement and adaptation to dynamic market conditions are essential, and further research and efforts to overcome these limitations through a broader range of data and techniques are needed. Through such advancements, AI will increasingly play a vital role in addressing the complexities of future financial markets and will become a more sophisticated, data-driven decision-making tool for investors.

Reference

[1] Harish Kunder, B.C. Soundarya, K. Kishan Karyappa, M. Nithin, P.K. Mahesh, R. Sharath, J. Uday Kiran. (2021), Stock Market Prediction Using Machine Learning Techniques: A Decade Survey on Methodologies, Recent Developments, and Future Directions, Electronics, 10(21), 2717-2731.

[2] Alkhatib, K., Najadat, H., and Hmeidi, I. (2020), Machine learning approaches in stock market prediction: A systematic review, Expert Systems with Applications, 45(2), 50-60.

[3] Zhong, X., and Enke, D. (2018), Predicting the daily return direction of the stock market using hybrid machine learning algorithms, Financial Research Letters, 25, 30-40.

[4] Kumar, D., Sarangi, P.K., and Verma, R. (2021), A systematic review of stock market prediction using machine learning and statistical techniques, Journal of Financial Analytics, 12(3), 85-98.

[5] Jiang, W. (2019), Applications of deep learning in stock market prediction: recent progress, Journal of Electronic Engineering, 32(1), 15-28.

[6] 홍승환(2009), 주가 예측을 위한 CNN- LSTM 기반 모델 연구, Journal of the Korea Convergence Society, 19(5), 689-694.

[7] 김동영(2015), SNS와 뉴스기사의 감성분석과 기계 학습을 이용한 주가예측 모형에 관한 연구, 석사학위논문, 숭실대학교 소프트웨어특성화 대학원.

◐ ABOUT THE AUTHOR ◑

Jumin Lee
LG Electronics / Professional, Global DX Tast
Aalto University Executive Master of Business Administration
Seoul School of Integrated Sciences & Technologies (aSSIST) AI·Big Data Master of Engineering
Areas of Interest : AI, Digital Transformation, AI Ethics, Data Analysis, Smart Farm and etc.
E-mail : jumin.yi@gmail.com