Predictive Analytics for NVIDIA Stock Market Trends

Hunter Fernandez, Dexter Schincke

Bellevue University

DSC 630-T301

**Introduction**

The semiconductor industry is pivotal in driving technological advancements, with NVIDIA emerging as a leader in GPUs and AI hardware. Their innovations, especially with AI-driven RTX chips, have transformed gaming into artificial intelligence research sectors. Launching new GPU generations, such as the GTX and RTX series, has historically triggered significant movements in NVIDIA's stock price.

This research explores how technological breakthroughs, particularly NVIDIA’s AI-powered chips, influence stock price movements. By analyzing historical data, we aim not to predict exact future prices but to determine whether the stock price will likely rise or fall. This includes examining the impact of significant events, such as the release of the RTX series compared to the earlier GTX series, on stock trends.

Our study’s significance is revealing how technological milestones shape investor behavior and market reactions. The findings will offer valuable insights for financial analysts, investors, and tech enthusiasts, enhancing their understanding of stock market dynamics and the interplay between technology and finance.

This project targets investors, financial analysts, and those following the tech sector. By predicting stock direction, stakeholders can better assess NVIDIA’s impact on the market and anticipate potential economic returns​.

**Data**

The dataset for this project is centered on historical stock data for NVIDIA (NVDA) obtained from Kaggle. The data includes key stock market variables such as:

-          Date: The trading date.

-          Open: NVIDIA stock's opening price.

-          High: The highest price recorded during the trading day.

-          Low: The lowest price recorded during the trading day.

-          Close: The final price at the close of the trading day.

-          Adjusted Close: The closing price is adjusted for corporate actions

-          Volume: The number of shares traded on that day.​

The data was preprocessed for modeling purposes using Polars for efficient handling. Missing values were checked and addressed to ensure the data's integrity. A lagged feature called prev\_close was also created to represent the previous day's closing price. A binary target variable (Target) was constructed to indicate whether the stock price increased from the last day. These steps ensured that the dataset was structured for practical predictive analysis.

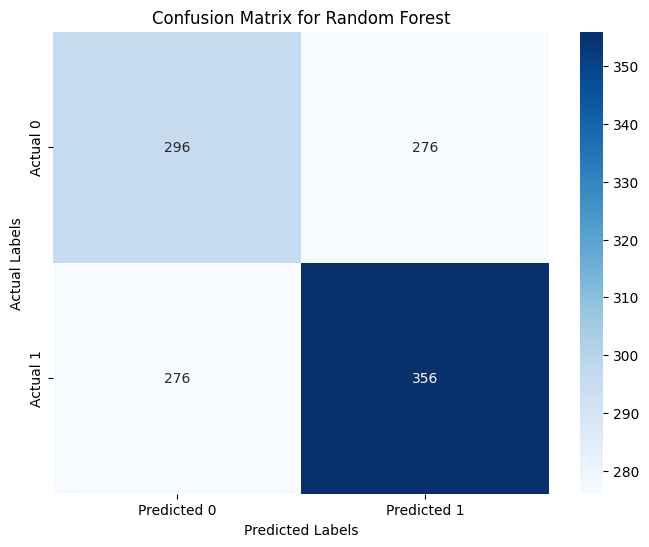
**Methods**

We tested three models to forecast NVIDIA’s stock price movements: Logistic Regression, Long Short-Term Memory (LSTM), and Random Forest Classification. These models were chosen based on their strengths: Logistic Regression as a baseline for binary classification, LSTM for capturing sequential patterns in time-series data, and Random Forest for handling non-linear relationships and feature interactions.

The dataset was split into training (80%) and testing (20%) sets for evaluation. Key performance metrics included accuracy, precision, recall, and the F1 score. Each metric provided insights into the models' ability to identify stock price movements while minimizing false positives and negatives.

The Logistic Regression model achieved an accuracy of 54% and a precision of 0.53, with a recall of 0.95 and an F1 score of 0.68. Its high recall demonstrates that it successfully identified most days when stock prices increased; however, its low precision highlights many false positives, as seen in the confusion matrix, where it frequently misclassified downward movements as upward trends. The ROC-AUC score of 0.52 further underscores its struggle to distinguish between actual stock price increases and decreases, limiting its reliability for precise predictions.A diagram of a logistic regression

Description automatically generated

The Random Forest model also achieved 54% accuracy but slightly improved precision (0.56) compared to Logistic Regression, paired with a balanced recall of 0.56. This resulted in an F1 score of 0.56, indicating better performance in managing false positives and negatives relative to Logistic Regression. Nonetheless, its confusion matrix reveals challenges in consistently identifying true positives, as it often needs to include actual upward trends. Its ROC-AUC score of 0.54 reflects a modest improvement in classification ability, though it remains limited in effectively separating the two classes.

The LSTM (Long Short-Term Memory) model matched the accuracy of the other two models at 54%, with a precision of 0.55 and a recall of 0.66, leading to an F1 score of 0.60. Its higher recall than Random Forest suggests better identification of stock price increases, but it still falls short of Logistic Regression in this area. The confusion matrix reveals that while LSTM reduces false negatives, it struggles with a moderate number of false positives, contributing to an ROC-AUC score of 0.53. This indicates that while it captures some sequential patterns in the data, its overall classification performance remains constrained.

A blue squares with white text

Description automatically generated

**Results**

From the exploratory data analysis (EDA) conducted on NVIDIA's stock data, several vital insights emerged:

1. Spikes in Trading Volume Around Key Product Launches: The data showed significant increases in trading volume during the release of major NVIDIA products, like the RTX series. This observation highlights how impactful product releases drive investor interest and market activity.
2. Correlation Between Trading Volume and Price Changes: A correlation analysis revealed a positive relationship between trading volume and price fluctuations. This pattern suggests that higher trading activity often accompanies price movement, which can be valuable for predictive modeling, as days with more active trading are generally more likely to exhibit price changes.
3. Seasonal Trends in Q4: A trend analysis showed that stock prices and trading volumes tend to increase in the fourth quarter, likely due to holiday sales and strategic product announcements. This seasonal behavior provides insight into recurring patterns in NVIDIA's stock behavior, which can improve the accuracy of models by accounting for such temporal influences.
4. Limitations of Historical Data Alone: The analysis indicated that relying solely on historical stock data (e.g., prices and volume) has limitations in fully capturing stock price movements, as it needs a broader market context. Additional data features like sentiment analysis or macroeconomic indicators would improve model precision and reliability.

The model metrics reveal that while the models can identify trends, they need help with precision. Logistic Regression stands out for its high recall, ensuring that most stock price increases are captured but suffers from frequent false positives. Random Forest offers a more balanced performance, while LSTM, though slightly better in recall and precision balance, shows slight improvement over the other models.

The evaluation of Logistic Regression, Random Forest, and LSTM models highlighted their distinct strengths and weaknesses in predicting stock price movements. Logistic Regression achieved an accuracy of 0.54 and the highest recall among all models at 0.95, successfully capturing nearly all stock price increases. However, its low precision of 0.53 indicated a high rate of false positives, where the model predicted price increases that did not materialize. This imbalance could significantly mislead investors when overestimating positive trends, leading to misguided decisions. While the model’s F1 Score of 0.68 reflected a solid balance between precision and recall, its ROC-AUC of 0.52 suggested limited overall discriminatory power between price increases and decreases.

The Random Forest model demonstrated a more balanced performance with an accuracy of 0.54, a precision of 0.56, and a recall of 0.56. These metrics indicated that while it was less prone to false positives than Logistic Regression, it failed to capture as many actual price increases as possible. Its F1 Score of 0.56 further highlighted its difficulty in achieving a solid balance between precision and recall. Despite its relatively stable performance across metrics, the Random Forest model's ROC-AUC of 0.54 showed limited ability to effectively differentiate between upward and downward price movements. As a result, its utility for high-stakes investment strategies could be improved, given its inability to excel in precision or recall.

The LSTM model displayed low success, with an accuracy of 0.54, a precision of 0.55, and a recall of 0.66. While it performed better than Random Forest in identifying true positives, its precision score reflected continued challenges with false positives. The F1 Score for the LSTM model was 0.60, indicating that it struck a somewhat better balance between precision and recall than Random Forest, though it fell short of Logistic Regression. Its ROC-AUC of 0.53 suggested it shared similar limitations with the other models in distinguishing between positive and negative movements. Although LSTM models are typically strong in capturing sequential patterns in time-series data, this model’s configuration did not fully capitalize on this strength, underscoring the need for further refinement. A graph of different colored bars

Description automatically generated

The findings emphasize the importance of balancing precision and recall in stock price prediction. Logistic Regression, with its high recall and F1 Score, appears most promising for scenarios where identifying upward trends is a priority, such as in investment strategies aimed at avoiding missed opportunities. However, its low precision and limited ROC-AUC highlight the risk of overestimating price increases, necessitating caution. While the Random Forest model demonstrated more balanced performance, its metrics fell short of providing actionable reliability. The LSTM model, despite leveraging deep learning techniques, offered only marginal improvements in recall over Random Forest without significantly improving overall accuracy or precision.

These results underscore the complexity of stock price prediction and the challenges of relying solely on historical stock data. Although the models achieved moderate success in capturing trends, their low accuracy and ROC-AUC scores suggest they need more robustness for high-stakes decision-making. To enhance predictive power, incorporating additional features, such as sentiment analysis or macroeconomic indicators, could provide a more comprehensive view of the factors influencing stock movements. While Logistic Regression offers a strong foundation, further feature engineering and model optimization will be necessary to achieve the precision and reliability required for practical application in financial markets.

**Conclusion**

The development of stock price prediction models raises ethical concerns, including the potential for market manipulation and misuse. Predictive tools could be exploited to unfairly influence investor sentiment or manipulate prices, emphasizing the need for safeguards against unethical practices. Integrating external data sources, such as sentiment analysis from social media or news, introduces privacy risks, requiring strict adherence to data protection regulations like GDPR. Additionally, biases inherent in the models, such as overreliance on historical data, could mislead users or disadvantage confident investors. Transparency about the models' limitations is essential to ensure informed decision-making and prevent overconfidence in their predictions. Finally, equitable access to these technologies must be prioritized to avoid exacerbating financial inequalities, ensuring that small-scale investors benefit alongside larger institutions.

Future enhancements to the models could include integrating sentiment analysis using natural language processing (NLP) to assess market sentiment through news articles, social media, or financial reports. This qualitative data would complement quantitative inputs, improving the ability to predict significant stock movements influenced by public sentiment. Incorporating macroeconomic indicators, such as interest rates, inflation, and geopolitical events, would provide essential context, addressing the current limitations of relying solely on historical stock data. Additionally, employing advanced machine learning techniques, such as XGBoost or ensemble stacking, could improve model performance by capturing complex, non-linear relationships and leveraging the strengths of multiple algorithms.

Further refinement through hyperparameter tuning and class imbalance techniques, such as oversampling or weighted loss functions, could enhance the precision and reliability of predictions. Testing the models in a simulated trading environment would allow for real-world validation, offering insights into performance under live conditions. These steps would improve predictive accuracy, address current limitations, and ensure ethical and responsible use in financial markets.

Our project focuses on understanding the factors influencing NVIDIA's stock price movements by examining key variables such as trading volume, volatility, and major product releases. We analyzed these elements to uncover the drivers behind stock price increases and decreases. Our work compares models like Logistic Regression, LSTM, and Random Forest to evaluate the most effective approach for predicting these movements. This hands-on analysis has also helped us improve our skills in classification-based modeling and machine learning techniques, enabling us to apply predictive methods within the financial markets. Ultimately, we aim to develop a model that not only forecasts stock price changes but also sheds light on how NVIDIA's technological advancements impact its market behavior.

This project explored the dynamics between NVIDIA’s technological advancements and stock price movements, employing predictive analytics through Logistic Regression, Random Forest, and LSTM models. By analyzing historical stock data and incorporating key features such as trading volume and lagged price indicators, we aimed to uncover patterns influencing stock price direction. Each model offered unique insights, with Logistic Regression excelling in recall but facing challenges in precision, Random Forest showing balanced yet moderate performance, and LSTM demonstrating its potential for capturing time-series trends despite certain limitations.

The findings highlight the complexity of stock price prediction, underscoring the need for more sophisticated feature sets and external variables, such as market sentiment and macroeconomic indicators. While no model provided definitive solutions, they collectively underscored the nuanced interplay between technology, market perception, and investor behavior.

This analysis emphasizes the importance of a multifaceted approach to stock prediction, where models serve as tools to enhance, rather than replace, human expertise. Future advancements in data integration, algorithmic improvements, and ethical considerations will refine predictive capabilities, providing actionable insights for investors and financial analysts in a rapidly evolving tech landscape.

**References**

HRTERHRTER. (2024). \*Nvidia stock historical data\* [Data set]. Kaggle.

<https://www.kaggle.com/datasets/programmerrdai/nvidia-stock-historical-data>

*Nvidia announces financial results for the second quarter of fiscal 2025.* NVIDIA Newsroom. (2024,

August 29). <https://nvidianews.nvidia.com/news/nvidia-announces-financial-results-for-> second-quarter-fiscal-2025