Predictive Analytics for NVIDIA Stock Market Trends

Hunter Fernandez, Dexter Schincke

Bellevue University

**Introduction**

The semiconductor industry is crucial for pushing technology forward, with NVIDIA standing out as a major company in GPUs and AI hardware. Their innovations, especially with AI-focused chips, have made a big impact on everything from gaming to AI research. Launching new GPU generations, like the GTX and RTX series, has caused noticeable shifts in NVIDIA's stock price.

In this project, we will investigate how technological breakthroughs, particularly NVIDIA’s AI-powered RTX chips, have influenced the movement of their stock price. Rather than predicting exact future prices, we will focus on whether the stock price is likely to go up or down based on past data. We will also look at how big events, such as the release of the RTX series compared to the earlier GTX series, affect stock trends. By exploring these connections, our study will shed light on how tech advancements impact investor behavior and market reactions, offering useful insights for financial analysts and those following the tech sector.

**Data**

The dataset for this project is centered on historical stock data for NVIDIA (NVDA). 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 adjusted for corporate actions, such as stock splits and dividends.
* Volume: The number of shares traded on that day.

**Modeling Approach**

**Model Selection**

To predict NVIDIA's stock price movement, we will evaluate the following models:

* Logistic Regression: This will classify price movements as increases or decreases based on historical data, serving as a baseline for binary classification.
* LSTM (Long Short-Term Memory): This recurrent neural network will analyze historical trends to forecast future price direction, considering both short-term and long-term patterns.
* Random Forest Classification: This algorithm will handle complex interactions and non-linear relationships to predict stock price movements using features like historical prices, trading volume, and price volatility.

These models will be tested to assess their effectiveness in predicting NVIDIA's stock price direction, offering a range of traditional and advanced methods.

**Why these Models?**

We chose these models because of their unique strength in predicting stock price movements. Logistic Regression gives us a simple baseline for classifying price changes. LSTM is great at capturing long-term patterns and time-based trends through deep learning. Random Forest handles complex, non-linear relationships and interactions, like price volatility and trading volume. By testing these models, we will be able to compare their performance and figure out which approach works best for forecasting stock prices.

**Evaluation of Results**

To ensure accurate predictions of NVIDIA's stock price movements, we will use several metrics for evaluation. Accuracy will measure the proportion of correct predictions for classification models like Logistic Regression and Random Forest. Precision and recall will assess how well the models predict upward or downward movements, with precision focusing on the correctness of predicted movements and recall on identifying actual movements. The F1 score will balance precision and recall, particularly for imbalanced classes. AUC-ROC will evaluate how well the models distinguish between price increases and decreases. Lastly, cross-validation will help prevent overfitting by assessing model performance across different data splits. These metrics will guide us in selecting the most effective model for predicting stock price directions.

**Learning Objectives**

This project aims to deepen our understanding of the factors influencing NVIDIA's stock price movements by analyzing key variables such as trading volume, volatility, and major product releases. Through this analysis, we seek to identify the elements that significantly drive stock price increases or decreases. Additionally, by comparing models like Logistic Regression, LSTM, and Random Forest, we aim to determine which algorithms are most effective for predicting stock price movements. The project will also help refine our skills in classification-based modeling and machine learning techniques, enabling us to enhance our predictive capabilities within financial markets. Ultimately, we expect to develop a model that accurately predicts stock price movements while providing insights into the behavior of NVIDIA's stock about technological advancements.

**Risks and Ethical Concerns**

**Risks**

This project faces several risks that must be carefully managed. One key risk is overfitting, where the model may perform well on historical data but fail to generalize to new, unseen data. To address this, we will employ cross-validation techniques and avoid using overly complex models. Another concern is data quality; inaccuracies or missing data in the historical stock dataset could lead to flawed predictions. We will use thorough data preprocessing and validation steps to mitigate this risk and improve the model’s reliability.

**Ethical Concerns**

There are also ethical considerations to keep in mind for this project. One significant concern is market manipulation. The predictive models developed could potentially be used for unethical purposes, such as manipulating stock prices. While this project is for educational purposes, it is essential to consider the broader implications of sharing predictive models. Additionally, if we decide to integrate sentiment analysis from social media or financial reports, data privacy becomes another concern. We must ensure that any external data sources comply with privacy regulations to avoid misuse of personal information.

**Contingency Plan**

If the initial project fails to deliver the expected results, we will pivot to a predictive analytics project using an NCAA dataset, which contains information on sports participation, revenues, expenses, and enrollment at various institutions. The alternative project will focus on predicting the financial performance of sports programs based on participation numbers and other relevant factors. Similar preprocessing steps will be applied, such as handling missing data and feature engineering, followed by testing models like linear regression and random forest to forecast revenue and expenses. This pivot will allow us to apply the same machine-learning techniques in a different domain.

**Conclusion**

This project focuses on predicting NVIDIA's stock price movements using predictive analytics models, such as Logistic Regression, LSTM, and Random Forest Classification. These models will help capture both short- and long-term stock trends, providing valuable insights for investors and analysts. We will evaluate model performance using metrics like MAE, RMSE, and R-squared to determine the most effective approach. The project also considers risks like overfitting and data limitations, as well as ethical concerns regarding the use of predictive models in stock trading. Additionally, a contingency plan is in place to pivot to a predictive analytics project using NCAA financial data if the stock price predictions fail to yield satisfactory results. Ultimately, this project aims to refine forecasting techniques and develop accurate, actionable models for predicting NVIDIA's stock movements.