Stock Price Prediction for NVIDIA (NVDA)

Shengyi Qian
Zi Wang

Abstract

This project aims to forecast NVIDIA's (NVDA) stock price by comparing multiple regression models that capture different aspects of market behavior. Specifically, we will implement Polynomial Linear Regression (PLR), a Deep Neural Network Regressor (DNNR) using PyTorch, a Random Forest Regressor and potentially an LSTM-based Regressor to predict the next day's closing price using historical data. Our approach leverages both raw stock data (Open, High, Low, Close, Volume) and engineered features (such as 10-day and 50-day moving averages and daily returns) to improve predictive accuracy. We will evaluate these models using performance metrics such as Mean Squared Error (MSE) and R^2, and our findings are expected to provide insights into the effectiveness of each modeling strategy for financial time-series forecasting.

Introduction

Predicting stock prices remains a challenging problem due to the inherent volatility and noise in financial markets, as well as the complex, non-stationary patterns that characterize stock data. NVIDIA, a major player in the semiconductor industry, provides an interesting case study because its stock exhibits significant fluctuations influenced by rapid technological advancements and market dynamics. This project is motivated by the need to explore robust machine learning methods that can accurately capture both the short-term fluctuations and long-term trends in stock prices. The successful prediction of stock movements has wide-ranging applications, including risk management, algorithmic trading, and investment decision support, making it a valuable area of study for both academic research and practical financial strategies.

Proposed Project

We propose to tackle the prediction problem as a regression task where the goal is to forecast the next day's closing price for NVDA. The dataset will be sourced from Kaggle (https://www.kaggle.com/datasets/muhammaddawood42/nvidia-stock-data/data), which contains 1,697 daily records of NVIDIA stock data. The available features include Date, Open, High, Low, Close, Adjusted Close, and Volume; we will further engineer features by calculating 10-day and 50-day moving averages as well as daily returns. After cleaning the data, converting date fields, and scaling the features with StandardScaler and MinMaxScaler, we will split the dataset chronologically into 80% training and 20% testing sets to avoid lookahead bias. Our models will include a PLR approach with degree-2 polynomial features for a baseline comparison, a DNN regressor built using PyTorch to capture non-linear relationships, a Random Forest regressor that captures complex patterns while reducing overfitting and an LSTM regressor also implemented in PyTorch to explicitly model temporal dependencies in the data. Model performance will be assessed using MSE and R^2, and visualizations will be generated to compare actual versus predicted closing prices. The project will follow a four-week timeline, starting with data exploration and preprocessing, followed by model development and tuning, and concluding with performance evaluation and final documentation.