

Amazon Stock Price Prediction

Introduction

In this project, we set out to predict Amazon's stock prices using a variety of market indicators and commodity prices. Stock price prediction is a critical area in financial analytics, offering valuable insights for investors and stakeholders. By leveraging historical data and machine learning techniques, we aimed to build a predictive model that could capture the complex relationships between Amazon's stock price and other influential financial variables.

Challenges

Our journey began with a dataset that contained numerous inconsistencies and missing values. The primary challenges we faced included:

- **Data Quality Issues:** The dataset had missing entries (NaN values) in several columns, particularly in volume data for commodities like Natural Gas and Crude Oil.
- **Data Type Mismatches:** Some numerical values were stored as objects (strings) due to formatting issues like commas in numbers (e.g., "43,194.70" instead of 43194.70).
- **Feature Selection:** With 39 columns, determining which features would be most predictive for Amazon's stock price was non-trivial.
- **Scaling and Normalization:** The features had varying scales, which could adversely affect the performance of our machine learning model.

Solutions

To address these challenges, we implemented the following steps:

- **Data Loading and Initial Exploration:** We used Pandas to read the CSV file and explored the first few rows to understand the data structure.
- **Data Cleaning:** We handled missing values and data type mismatches. For columns with numerical data stored as strings (due to commas), we removed the commas and converted them to numeric types.
- **Feature Selection:** We selected relevant features like commodity prices (Natural Gas, Crude Oil, Gold) and stock prices of companies like Meta and Netflix.
- **Data Scaling:** Applied StandardScaler to standardize the features.
- **Model Building and Evaluation:** Split the data into training and testing sets, trained a Linear Regression model, and evaluated using MAE and RMSE. Performed cross-validation for robustness.

Roadblocks

During the project, we encountered several roadblocks:

- **Data Imbalance:** Some features had a high number of missing values that couldn't be addressed by simple imputation.
- **Outliers:** The presence of outliers (e.g., negative Crude Oil prices) required careful consideration.
- **Multicollinearity:** High correlation between certain features could affect the model's reliability.

Conclusion

Despite the challenges, we successfully developed a predictive model for Amazon's stock price. The Linear Regression model achieved an MAE of approximately 8.52 and an RMSE of about 10.62, indicating a reasonable level of accuracy. Feature importance analysis revealed that Netflix's and Meta's stock prices were significant predictors of Amazon's stock price.