PROJECT SUMMARY | HUBBLEMIND

Introduction

The goal of this project was to develop a machine learning model to predict Amazon stock prices using historical data. By leveraging features such as commodity prices, volume traded, and previous stock performance, the project aimed to create an accurate forecasting tool. Predicting stock prices is crucial for making informed investment decisions, which can help investors manage risk and optimize their portfolios.

Challenges

Ch 1: Data Quality Issues

Some columns contained missing values, requiring a strategy to handle them in order to avoid biasing the model results.

Ch 2: Data Type Issues

Some attributes used commas as thousands separators (e.g., "8,198.60"), which led to difficulties in data processing, as these values were read as strings instead of numerical data.

Ch 3: Variance in Data Magnitude

Some attributes had values that were significantly larger compared to others, creating a disparity in scale. This large variance required feature scaling to ensure all features contributed equally to the model's performance.

Solutions

- **Filling missed values**

  Handle missing values by filling them with the mean of the respective columns.

- **Thousands Separator**

  Use the `thousands` parameter to account for these variables when importing the data.

➢ **Feature Scaling**

The dataset was standardized using `StandardScaler` to ensure that all features had similar variances, preventing certain features from dominating the model.

Roadblocks

During the project, I encountered a smooth workflow with no significant roadblocks that hindered my progress. The planning and execution phases were well-structured, allowing me to address challenges as they arose without any major disruptions. My ability to manage tasks effectively and solve problems contributed to maintaining a steady pace throughout the project, ensuring that I met my deadlines and objectives without encountering any significant obstacles.

Conclusion

Table 1: Model evaluation.

	MSE	RMSE	MAE	R^2 Score
Linear regression	0.0449	0.212	0.165	0.954

The final model performed reasonably well, with a Mean Absolute Error (MAE) of 0.16 and a Root Mean Squared Error (RMSE) of 0.21 on the test data. Additionally, the R^2 value of 0.95 indicates that the model can explain a significant portion of the variance in the data, demonstrating its effectiveness. This project showcased that by leveraging historical data and machine learning, it is possible to create a reliable stock price forecasting tool. However, the model's limitations suggest that future work could involve incorporating more complex models or additional external data sources to improve accuracy. Overall, the project was a valuable exercise in data preprocessing, model development, and performance evaluation, offering insights into the complexities of stock price prediction.

You can find the code for this project:
here:https://github.com/HammadaMohamed/ML-Intern-HUBBELMIND