STOCK PRICE PREDICTION PROJECT DOCUMENTATION

INTRODUCTION

This project aims to predict stock prices using historical data through machine learning techniques. The motivation behind this project stems from the growing interest in understanding financial markets and the potential of predictive analytics to make informed investment decisions. By utilizing historical stock price data, I aimed to develop a model that can help in forecasting future prices, providing insights into market trends.

PROJECT OVERVIEW

In this project, I utilized the **Yahoo Finance** library to gather real-time stock price data, making it easier to access and analyse the relevant financial information. The main libraries and tools used in this project include:

- **Libraries**:
    - **Pandas**: For data manipulation and analysis.
    - **NumPy**: For numerical operations.
    - **Scikit-learn**: For building the machine learning model and evaluation metrics.
    - **Matplotlib and Seaborn**: For data visualization.
- **Models**:
    - **Linear Regression**: This model was chosen for its simplicity and effectiveness in predicting continuous outcomes like stock prices.
- **Datasets Used**:
    - Historical stock price data was obtained from Yahoo Finance, providing a robust dataset for training and testing the model.

STEP-BY-STEP BREAKDOWN

1. **Data Collection and Preprocessing**:
    - Utilized the Yahoo Finance library to fetch stock price data.
    - Cleaned the dataset by checking for missing values and duplicates to ensure data quality.
2. **Exploratory Data Analysis (EDA) and Visualization**:
    - Analysed the data to understand trends and patterns.
    - Created visualizations such as time series plots to observe stock price fluctuations over time.
3. **Model Training and Evaluation**:

    - Split the dataset into training and testing sets.
    - Trained the linear regression model on the training data.
    - Evaluated the model's performance using metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

LEARNINGS AND CHALLENGES

Throughout this project, I learned valuable insights about stock price prediction and machine learning:

- **Insights Gained**: The importance of data quality and preprocessing in building a reliable predictive model.
- **Challenges Faced**:
    - Data cleaning was crucial, as missing or duplicate values could skew results. I addressed this by implementing checks to ensure a clean dataset.
    - Model tuning was necessary to improve predictions; I experimented with different parameters and evaluation metrics to achieve better performance.
- **New Skills Acquired**:
    - Gained proficiency in using the Yahoo Finance library for data retrieval.
    - Enhanced my understanding of regression models and their applications in finance.

FUTURE IMPROVEMENTS

In the future, I would like to implement the following enhancements:

- Experiment with more complex models such as Decision Trees or Neural Networks to compare performance with Linear Regression.
- Incorporate additional features, such as economic indicators or news sentiment analysis, to improve prediction accuracy.
- Develop a user-friendly interface for the model to allow users to input stock symbols and retrieve predictions easily.

REFERENCES

- Yahoo Finance API documentation for retrieving stock data.
- Scikit-learn documentation for understanding model training and evaluation metrics.
- Relevant tutorials and articles on stock price prediction and machine learning techniques.