

Stock Market Prediction Project Documentation

1. Introduction

The primary objective of this project is to analyze and predict stock market prices using historical stock data. By leveraging various data analysis and machine learning techniques, the project aims to provide insights into stock price movements and develop a predictive model for forecasting future prices.

2. Data Loading and Initial Exploration

Data Source: The dataset is loaded from Google Drive, ensuring easy access and persistence.

Initial Inspection: Initial data inspection is conducted using methods like `head()`, `tail()`, `info()`, and `describe()` to understand the dataset's structure, summary statistics, and potential data quality issues.

Missing Values: Identified columns with missing values and their respective counts to strategize appropriate handling techniques.

3. Handling Missing Values

Mean Imputation: Specific columns with missing values were filled using the mean value of the respective columns to retain data integrity.

Forward and Backward Fill: Remaining missing values were addressed using forward fill (`ffill`) and backward fill (`bfill`) methods, ensuring no gaps in the dataset.

4. Exploratory Data Analysis (EDA)

Trend Analysis: Plotted stock prices over time to visualize the historical trend and identify any significant patterns or anomalies.

Distribution Analysis: Analyzed the distribution of stock prices using histograms to understand the spread and central tendency of the data.

5. Feature Engineering

Technical Indicators: Calculated several technical indicators such as Simple Moving Averages (SMA), Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), and Bollinger Bands. These indicators provide additional insights into stock price movements and are used as features for modeling.

Handling Missing Values in Indicators: Ensured all calculated technical indicators were free of missing values by filling any gaps with their mean values.

6. Model Building

Feature Selection: Selected relevant features for the predictive model, including stock prices and technical indicators.

Train-Test Split: Split the data into training and testing sets to evaluate the model's performance.

Model Training: Built a Linear Regression model to predict stock prices based on the selected features.

Model Evaluation: Evaluated the model's performance using the Mean Squared Error (MSE) metric to measure the accuracy of predictions.

7. Results and Visualization

Prediction Visualization: Visualized the actual vs. predicted stock prices to assess the model's performance. This comparison helps in understanding how well the model captures the trend and volatility of stock prices.

8. Insights and Conclusions

Methodology Effectiveness: Documented the effectiveness of different methodologies used for handling missing values, feature engineering, and model building.

Predictive Model Performance: Provided insights into the predictive model's performance, highlighting its strengths and areas for improvement.

9.Challenges faced

It took a long time to load dataset when ever I tried to run jupyter notebook

So I have uploaded it to my google drive and mounted drive in Google Colab and it made my work easier in importing csv file