Stock Price Prediction Project Documentation

Problem Statement

Problem: We aim to develop a stock price prediction model for Microsoft (MSFT) based on historical stock data. Our primary objective is to predict the closing price of MSFT stock using a machine learning model.

Objective: We are working to develop a robust stock price prediction model that provides accurate forecasts, facilitating informed investment decisions.

Design Thinking Process

1. Data Collection

- Data source: We obtained the dataset from [Microsoft Lifetime Stocks Dataset](#) on Kaggle.

2. Data Preprocessing

- Handling Missing Values: We removed rows with missing data to ensure data quality.

- Date Transformation: We converted the 'Date' column to datetime format and set it as the index.

- Feature Selection: We retained relevant columns, including 'Open,' 'Close,' and 'Volume.'

3. Feature Engineering

- Moving Averages: We calculated 50-day and 200-day moving averages to capture short-term and long-term trends.

- Relative Strength Index (RSI): We computed it as an oscillator measuring price momentum.

- Moving Average Convergence Divergence (MACD): We calculated it for trend-following momentum analysis.

4. Model Selection

- Chosen Model: We selected the Random Forest Regressor as the prediction model.

5. Model Training

- Data Split: We divided the dataset into training and testing sets (80% training, 20% testing).

- Input Data: We separated the input features from the target variable.

- Model Fitting: We trained the Random Forest Regressor model using the training data.

Dataset Description

Dataset Source

- **Dataset Link**: [Microsoft Lifetime Stocks Dataset](#)

Brief Description

The dataset used for this project contains historical stock data for Microsoft (MSFT). It spans a significant timeframe and includes the following key features:

- **Date**: Temporal index representing each data point's date.

- **Open**: Opening price of MSFT stock on each date.

- **Close**: Closing price of MSFT stock, used as the target variable.

- **Volume**: The trading volume, indicating the number of shares traded.

This dataset is a valuable resource for predicting MSFT stock prices and analyzing financial trends.

Data Preprocessing

- Handling Missing Values: Rows with missing data were removed to ensure data integrity.

- Date Transformation: The 'Date' column was converted to datetime format and set as the index.

- Feature Selection: Only relevant columns ('Open,' 'Close,' and 'Volume') were retained.

Feature Engineering

Several technical indicators were calculated, including:

- **50-Day and 200-Day Moving Averages**: To capture short-term and long-term trends.

- **Relative Strength Index (RSI)**: A momentum oscillator for price momentum analysis.

- **Moving Average Convergence Divergence (MACD)**: A trend-following momentum indicator.

Model Training

- Data Split: The dataset was divided into an 80% training set and a 20% testing set.

- Input Data: Input features were separated from the target variable ('Close').

- Model Selection: The Random Forest Regressor was chosen for its predictive capabilities.

- Model Fitting: The model was trained using the training data.

Key Findings, Insights, and Recommendations

Based on the analysis, key findings and recommendations include:

- **Model Accuracy**: The Random Forest Regressor demonstrates strong predictive power, as indicated by low Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

- **R-squared Score**: The R-squared (R^2) score suggests that a significant portion of the variance in stock prices is explained by the selected features and indicators.

Recommendation: To further enhance the model's performance, we recommend experimenting with alternative regression algorithms, hyperparameter tuning, and the inclusion of additional relevant features.