**Predicting Stock Prices with ARIMA and Gradient Boosting**

**Project Summary**

This project leverages ARIMA and Gradient Boosting, two powerful machine learning models, to predict stock market prices. Utilizing a year’s worth of daily stock prices and volumes from select companies, we undertake a comprehensive process encompassing data preparation, feature engineering, model training, and performance evaluation.

**Data Source**

Our analysis is based on historical daily OHLC (Open, High, Low, Close) and volume data for 5-10 stocks from January 1 to December 31, 2023, obtained from Yahoo Finance.

**Workflow**

**Data Acquisition**

We gather daily stock price data for major companies like Apple (AAPL), Microsoft (MSFT), Google (GOOG), Amazon (AMZN), and Tesla (TSLA).

**Data Preprocessing**

We address missing values through forward filling and index the dataset by date to ensure consistency.

**Exploratory Analysis**

Historical prices and volumes are visualized to discern trends and inform feature selection.

**Feature Creation**

New predictive features are engineered:

* **Lagged Variables**: Prior closing prices.
* **Moving Averages**: 5-day and 10-day rolling means.
* **Price Changes**: Daily close price percentage variations.

**ARIMA Implementation**

**Model Optimization**

Optimal ARIMA parameters (p, d, q) are determined via ACF and PACF analysis.

**Price Forecasting**

Future stock prices are forecasted with the ARIMA model and visualized against historical data.

**Gradient Boosting Approach**

**Data Setup**

Feature selection is refined, and data is split into training and testing subsets.

**Model Training**

A Gradient Boosting Regressor is trained on the training set.

**Performance Assessment**

Predictions are made on the test set and evaluated using RMSE and MAE.

**Comparative Analysis**

Both models’ performances are assessed using RMSE, MAE, and MAPE to identify the superior predictor.

**Insights**

* A critical analysis of ARIMA and Gradient Boosting models reveals distinct advantages.
* Performance metrics are scrutinized to gauge accuracy and reliability.

**Conclusions**

The project showcases the efficacy of ARIMA and Gradient Boosting in forecasting stock prices. The comparative analysis aids in selecting the optimal model for informed trading decisions.

**Visual Insights**

Included are visual representations of:

* Stock price movements and volume trends.
* ARIMA’s ACF/PACF insights.
* Predictive versus actual pricing for both models.
* Comparative performance metrics.

This README offers a concise yet detailed account of the project’s methodology and insights gleaned from the predictive modeling of stock prices.