Anomaly Detection in Tesla Stock Price Time Series: Approach and Findings

Approach:

- **Data Collection:** In this project, historical stock price data of Tesla (TSLA) was collected from Yahoo Finance. The dataset spans from January 1, 2020, to January 1, 2023. The data contains timestamped closing price readings, which are commonly used in time series analysis.
- **Exploratory Data Analysis (EDA):** The first step in any data analysis task is to perform EDA. We started by calculating basic statistics of the closing price data, such as mean, standard deviation, minimum, maximum, and quartiles. This helped us gain insights into the central tendency, variability, and distribution of the data.
- **Visualizing Time Series Data:** We then visualized the time series data using line plots. The x-axis represents time (date), and the y-axis represents the closing price. The line plot revealed the general trend of Tesla's stock price over the specified period, providing insights into whether the stock price increased or decreased over time and if there were any significant fluctuations.
- **Histograms to Show Data Classes:** To further understand the distribution of closing prices, we created histograms. Instead of analyzing the entire dataset as a single entity, we divided the closing prices into three classes for simplicity: Class 1 (lower prices), Class 2 (moderate prices), and Class 3 (higher prices). The histograms displayed the frequency distribution of closing prices within each class. This allowed us to observe how many data points fall into each price range.
- **Model Selection and Training (ARIMA):** For anomaly detection, we chose the ARIMA (AutoRegressive Integrated Moving Average) model. ARIMA is a popular time series model that captures temporal dependencies and trends in the data. We fitted the ARIMA model to the closing price data with an order of (5, 1, 0). The (5, 1, 0) order indicates the number of autoregressive (p), differences (d), and moving average (q) terms in the model.

Findings:

- **Visual Inspection:** The line plot of Tesla's stock price time series showed an overall upward trend, indicating that the stock price generally increased over time. However, there were fluctuations and volatility, indicating periods of high and low prices.
- **Histograms:** The histograms provided a clear view of how the closing prices were distributed across the three classes. We could observe whether the stock price mostly belonged to the lower, moderate, or higher price ranges.
- **Anomaly Detection:** To detect anomalies, we used the residuals obtained from the ARIMA model. Residuals are the differences between the actual closing prices and

the predicted prices by the model. We set a threshold of 2 standard deviations from the mean of the residuals. Data points beyond this threshold were considered anomalies.

- **Evaluation Metrics:** To assess the performance of the anomaly detection model, we used precision, recall, and F1-score. Precision measures the accuracy of the model's predicted anomalies, recall assesses the model's ability to detect actual anomalies, and the F1-score provides a balanced measure of both precision and recall.

Conclusion:

The approach employed in this project involved exploratory data analysis, visualization, and anomaly detection using the ARIMA model. The model showed promising results in detecting anomalies in Tesla's stock price time series. However, achieving high recall remains a challenge. Further improvements could be made by exploring different threshold values, experimenting with other time series models, or incorporating additional features and domain knowledge to capture more complex patterns in the data. Continuous refinement and fine-tuning of the approach are necessary to achieve a robust and effective anomaly detection system for real-world applications.