**Stock Data Analysis and Prediction Report**

**Introduction**

**Background**

In this analysis, we explore real-time daily stock data for the symbol "META" obtained from the Alpha Vantage API. The primary objectives include understanding stock trends, conducting exploratory data analysis (EDA), and building a machine learning model to predict stock prices.

**Objectives**

* Fetch real-time daily stock data from Alpha Vantage API.
* Transform data into a structured Pandas Data Frame.
* Set up a Spark environment for streaming analysis.
* Conduct EDA to gain insights into stock trends.
* Build and evaluate a machine learning model for stock price prediction.

**Architecture Design**

**Data Retrieval**

The Alpha Vantage API is utilized to fetch daily stock data. The function ‘fetch\_daily\_stock\_data’ accepts an API key and stock symbol, retrieves data, and transforms it into a Pandas Data Frame.

**Data Transformation**

Data is transformed into a Pandas Data Frame with proper indexing, data type conversion, and column renaming. This Data Frame is then converted to a Spark Data Frame for further analysis.

**Spark Streaming Setup**

Spark is initiated with a SparkSession, and SparkContext is obtained. A StreamingContext with a batch duration of 1 second is created to facilitate streaming analysis.

**Challenges and Solutions**

**API Communication**

**Challenges:**

* Potential HTTP errors, connection errors, or timeouts during API requests.
* Solutions:
* Implemented error handling to address different types of exceptions raised during API communication.

**Data Transformation**

**Challenges:**

* Ensuring data consistency, handling missing values, and converting datatypes.

Solutions:

* Implemented proper datatype conversion and handling missing values in the Pandas DataFrame.

**Spark Streaming**

**Challenges**:

* Setting up Spark Streaming and handling potential issues with data processing.

Solutions:

* Created a streaming context (ssc) and utilized foreachRDD to process each RDD, enabling EDA and window operations.

**Data Analysis**

**Descriptive Statistics**

Descriptive statistics, including mean, standard deviation, and percentiles, were computed for the stock data.

**Exploratory Data Analysis (EDA)**

EDA was performed on the data, showcasing key statistics and implementing window operations, such as calculating the moving average.

**Machine Learning Model**

**Feature Engineering**

Features include 'Open', 'High', 'Low', and 'Volume'. A linear regression model is trained to predict the 'Close' price.

**Model Training**

The model is trained using Spark's machine learning library on the training data.

**Model Evaluation**

The model is evaluated on test data, and metrics such as RMSE and R-squared are calculated.

We got

Root Mean Square Error (RMSE) on test data = 0.6780771772029048

R-squared on test data = 0.9783647633691587

**Findings and Insights**

**Investment Opportunities**

Insights from the analysis can be used to identify potential investment opportunities, considering stock trends and machine learning predictions.

**Data Patterns**

EDA revealed patterns and anomalies that may influence decision-making in the context of stock trading.

**Conclusion**

The analysis provides valuable insights into stock trends, and the machine learning model shows promising predictive performance. The integration of Pandas for initial data manipulation and Spark for scalable streaming analysis proves effective.

**Future Work**

Future work may include refining the machine learning model, exploring additional features, and incorporating more advanced algorithms for improved predictions.