**Description:** As a consultant specializing in financial analytics, you've been tasked with developing a stock price forecasting system for Netflix using Apache Spark. The firm aims to leverage advanced machine learning techniques to predict future stock prices accurately, enabling informed investment decisions. Your objective is to analyze historical stock market data posted [here](https://www.kaggle.com/datasets/jainilcoder/netflix-stock-price-prediction), build predictive models, and deploy a real-time forecasting pipeline using Apache Spark.

**Tasks:**

1. **Data Investigation and Preprocessing:**
   * Investigate and understand the data.
   * Preprocess the data by handling missing values, adjusting for stock splits and dividends, and aligning timestamps.
2. **Feature Engineering:**
   * Extract relevant features from the historical stock data that may influence future price movements.
   * Engineer technical indicators such as moving averages, relative strength index (RSI), stochastic oscillator, and Bollinger Bands.
3. **Model Development:**
   * Split the historical data into training and testing sets, considering the temporal nature of the data.
   * Build stock price forecasting models using machine learning algorithms supported by Apache Spark, such as linear regression, time series models (ARIMA, SARIMA), or deep learning architectures (LSTM, GRU).
   * Experiment with ensemble methods and hybrid models to improve forecast accuracy.
4. **Model Evaluation:**
   * Evaluate the performance of each forecasting model using appropriate metrics (e.g., mean absolute error, mean squared error, directional accuracy) on the test dataset.
   * Conduct backtesting to assess the models' performance on historical data and validate their effectiveness in real-world scenarios.
5. **Real-Time Forecasting Pipeline:**
   * Develop a Spark streaming application to ingest real-time stock market data and generate forecasts on the fly.
   * Integrate the trained forecasting models into the streaming pipeline to make predictions for incoming data.
   * Implement alerting mechanisms to notify traders or portfolio managers when significant price movements or anomalies are detected.
6. **Performance Optimization:**
   * Optimize Spark job configurations and resource allocation to enhance the scalability and efficiency of the real-time forecasting pipeline.
   * Explore techniques for distributed model training and inference to handle large volumes of data and improve processing speed.
7. **Documentation and Deployment:**
   * Document the entire stock price forecasting system, including data preprocessing steps, feature engineering techniques, model selection criteria, and deployment architecture.
   * Provide guidelines for maintaining and updating the forecasting pipeline over time, including retraining models and adapting to changing market conditions.

**Deliverables:**

1. Jupyter notebook or Python script containing the Spark code for data preprocessing, modeling, and real-time forecasting.
2. Report documenting the stock price forecasting process, including data exploration, feature engineering techniques, model evaluation results, and performance optimizations.
3. Visualization of forecasted stock prices, model performance metrics, and real-time monitoring dashboards.
4. Deployment package for the forecasting system, including configuration files and setup instructions.

**Additional Notes:**

* Understand the factors influencing stock price movements and refine the forecasting models accordingly.
* Emphasize the importance of continuous monitoring and model recalibration to adapt to changing market dynamics and evolving trading strategies.
* Consider the regulatory requirements and compliance standards applicable to financial forecasting systems when designing and deploying the solution.
* Provide training and support to the Netflix's technical team for maintaining and operating the stock price forecasting system.