Quantitative Analyst Assessment

Introduction

Complete the following assignment in the stipulated time.
Time limit: 72 hours

Please do not hold back on your creativity while implementing the below data science assignment.
You can use any package you deem necessary or any open source code available on the internet to achieve the following results.

For any doubts contact:
debraj@somish.com

CC: suraj@hillroute.com

Primary Objective - Summary

1. Select appropriate machine learning or deep learning algorithms for predicting the percentage change in close price of the ETH/USDT data provided.
2. Predict the close price change for ETH/USDT based on the following time horizon in the future.
    a. 15 minutes
    b. 30 minutes
    c. 1 hour
    d. 1 day
    e. 30 days
3. This assignment is designed in a way for you to not hold back on your creativity and knowledge to solve a data science quantitative trading problem. Please choose one of the approaches below.The idea is to know your thought process while solving this assignment.
    a. Solve this assignment using a classification approach.
    b. Solve this assignment using a regression / forecasting approach.
4. You will be evaluated on the following metrics based on the type of approach you will take to solve the problem.
    a. Assessing the accuracy and performance of the developed models using metrics such as loss function graphs, accuracy graphs, precision, recall, F1-score, and confusion matrix for classification based problems.
    b. MAE, MSE, RMSE, MAPE, percentage of correct directional prediction if you find regression / forecasting as an appropriate way to solve this assignment.

Details

Data Analysis and Preprocessing

- Understand what OHLCV stock price data and understand how to run predictive models on OHLCV data.
    - OHLCV stands for Open, High, Low, Close, and Volume. It is a common format used to represent financial market data, particularly in the context of trading. Each of these attributes provides valuable information about the price movement and trading activity of a financial instrument, such as a stock, cryptocurrency, or commodity, over a specific time period.
    - Here's a brief explanation of each attribute:
        - Open: The opening price is the price of the financial instrument at the beginning of the time period being observed (e.g., a day, an hour, a minute). It's the first traded price after the market opens for that period.
        - High: The high price represents the highest price reached by the financial instrument during the specified time period. It reflects the peak value the asset attained within that timeframe.
        - Low: The low price is the lowest price the financial instrument reached during the given time period. It indicates the bottommost value the asset touched within that timeframe.
        - Close: The closing price is the price of the financial instrument at the end of the observed time period. It's the last traded price before the market closes for that period.
        - Volume: Volume refers to the total number of units (shares, contracts, coins, etc.) of the financial instrument that were traded during the specified time period. It indicates the level of trading activity and market interest. Higher volume can suggest greater market participation and potential price significance.
    - By analyzing OHLCV data, traders and analysts can gain insights into price trends, volatility, and market sentiment. Technical analysis tools, such as candlestick charts, moving averages, and indicators, are often applied to OHLCV data to make informed trading decisions and predictions about future price movements.
- Perform exploratory data analysis on the provided ETH/USDT price data.
- Calculate percentage change in the data for every time period. This will be the data on which prediction will be done.
- Analyze the distribution, trends, and patterns in the data.
- Handle missing values, outliers, and data inconsistencies appropriately.
- Add further statistical features or technical indicators if deemed necessary.
- Preprocess and transform the data to ensure it is suitable for input into machine learning models.
- Apply techniques such as scaling, normalization, or feature engineering as needed.

Important Note: We have been observing that some candidates are predicting the prices and keeping the prices as the features and labels instead of percentage returns.
This is incorrect.

In order to know why kindly refer to the link below if you are trying to predict prices instead of percentage change in price. (Returns)

https://stats.stackexchange.com/questions/497877/forecasting-prices-vs-returns-by-deep-learning

Some of the marketing content on the internet predicts prices instead of returns using LSTM's.

Thus, it is one of the common mistakes made when creating prediction models for non stationary data in financial analysis and quantitative trading.
Please ensure the fact that we are predicting **percentage change (returns) and not prices**.

Algorithm Development

- Select appropriate machine learning or deep learning algorithms for predicting the percentage change in price.
- Design the architecture and model structure considering the nature of the data and the trading task.
- Implement the chosen algorithms using Python and relevant frameworks (e.g., XGBoost, TensorFlow, PyTorch).
- Optimize the models by tuning hyperparameters and exploring different configurations.
- Consider techniques such as cross-validation, grid search, or Bayesian optimization for model selection and hyperparameter tuning.

Model Implementation

- Split the dataset into training and testing sets, considering the 2-year timeframe.
- Train the developed models using the training data.
- Select appropriate machine learning or deep learning algorithms for predicting the percentage change in close price for the following time horizon in the future.
    - **15 minutes**
    - **30 minutes**
    - **1 hour**
    - **1 day**
    - **30 days**

Evaluation and Performance Metrics

- Assess the accuracy and performance of the developed models using metrics such as loss function graphs, accuracy graphs, precision, recall, F1-score, and confusion matrix.
- MAE, MSE, RMSE, MAPE, percentage of correct directional prediction if you find regression / forecasting as an appropriate way to solve this assignment.
- Compare and contrast the performance of different models and variations.
- Interpret the results and provide insights into the strengths and limitations of the models.

Conclusion and Recommendations

- Summarize the findings and conclusions from the analysis and backtesting.
- Provide recommendations for further improvements or modifications to enhance the strategies.
- Highlight the significance and potential applications of machine learning-based trading strategies.

Share the code and Technical Report for assessment

- Share the code repository and the models for our engineers to test the feasibility of the mode.
- Write a detailed technical report limited to 1000 words.
- Ensure clear and concise communication of the entire process, including data analysis, algorithm development, model implementation, backtesting, and evaluation.
- Include visualizations, charts, and tables to support the findings.
- Provide a well-structured and organized report with appropriate sections and headings.
- Pay attention to grammar, spelling, and overall readability of the report.

Note: The technical report should provide a comprehensive overview of the assignment, highlighting the candidate's proficiency in machine learning experimentation, implementation, and their ability to develop profitable trading strategies based on the provided data.