Industrial project description "Forecasting price dynamics in financial market for investment purposes"

The chosen role is an **analyst**.

1 Planning the industrial research project

1. **Goal of the project.** The goal is to develop a machine learning model that predicts future prices of financial assets.

2. **Applied problem solved in the project.** The result will be used to enhance investment strategies and risk management for investors and financial analysts. To illustrate the result, will be created time-series plots to compare forecasted prices with actual observed prices over time. These graphs will highlight how closely the model's predictions align with real market data, making it easier to assess the model's accuracy visually.

3. **Description of historical measured data.** Historical data will include five years of daily asset prices along with trading volumes, financial indicators and a variety of technical and macroeconomic factors. This data will be organized with timestamps to capture the sequential nature of the financial market. The dataset will be stored in tabular formats such as CSV files or databases, with columns representing timestamp and various financial indicators.

4. **Quality criteria.** The model's performance will be evaluated using regression metrics such as Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). The error function to optimize could be Huber Loss as it is differentiable and combines the strengths of both MSE and MAE. It is less sensitive to extreme price movements (outliers) compared to MSE, but still handles large errors well.

5. **Project feasibility.** The feasibility of the project can be demonstrated through exploratory data analysis (EDA) and model prototyping. EDA will help to confirm that the dataset is suitable for time series forecasting, while prototyping using baseline models will provide initial insights into the predictability of the price dynamics. Possible risks include data quality issues, market volatility, model under– and overfitting and limited hardware.

6. **Conditions necessary for successful project implementation.** The success of the project will depend on several key factors. First, it is crucial to have access to a high-quality dataset that includes comprehensive historical data. The dataset should cover a sufficient time period with a high resolution. In addition, the data should be clean, with minimal missing values and outliers, and pre-processed appropriately for time series analysis. To track changes in data and models, the project will require an organized workflow, including version control.

7. **Solution methods.** The solution will involve regression-based machine learning models, with a focus on methods that handle time series data. The models to be considered will include Random Forest, Gradient Boosting for their ability to handle complex, non-linear relationships, and the LSTM network for its capability in capturing long-term dependencies in sequential data. Additionally, probabilistic models such as ARMA, ARIMA will be explored, as these models help capture trends and seasonality. Libraries such as statsmodels will be used for ARMA/ARIMA models, while scikit-learn will be employed for traditional regression techniques. The PyTorch library will be used for LSTM implementation. Also, to ensure the robustness of the model, cross-validation techniques and hyperparameter tuning methods will be applied. The hypotheses to be tested will include the assumption that incorporating macroeconomic factors will enhance model performance and that hybrid models will outperform individual models in capturing both trends and volatility.

2 Research or development?

What impact will the research have on the field of knowledge? How useful will it be?

The research impact lies in developing more robust, accurate and interpretable models for predicting price dynamics in highly volatile financial market. The research will benefit financial analysts, traders, and institutions by offering a forecasting tool, which will improve decision-making processes and risk management strategies.