**FINAL REPORT**

1. **Summary**

The objective of the project was to test whether the market gains of major companies within the Nasdaq 500 index could be outperformed using machine learning-based return predictions. Weekly return predictions were made, and profits were calculated based on the buy-sell transactions of the top 5 stocks with the strongest predictions for each week. The stocks that were no longer among the top 5 strongest predictions in the following week were sold, while the position weights of the remaining stocks in the top 5 were increased. Additionally, a 5% take-profit condition was implemented. Optimization was performed based on the ROI and Sharpe ratios calculated from the transactions at the end of the year. Additionally, metrics such as RMSE and MAE were calculated for the weekly predictions.

1. **ETL Process (PySpark with PostgreSQL)**

The dataset was obtained using the yfinance library by downloading the following stock data for the period 2012–2024: ['AAPL', 'NVDA', 'MSFT', 'AMZN', 'META', 'ADBE', 'TSLA', 'FFIE', 'ASTI', 'ALLR']. The dataset included the fields: 'Date', 'High', 'Low', 'Open', 'Close', 'Adj Close', and 'Volume'.

The data was transformed into a PySpark DataFrame, and the multi-index structure was removed. Stock names were added as a column. Missing values in the Date column were identified and removed, and the Date column was converted to a datetime format. As a result, a multi-stock DataFrame was created.