Machine Learning Project

Project #4

20211109 배현우
20213948 최원용
20203901 지승후

1. Description of Dataset and Task

The Hull Tactical Market Prediction competition aims to forecast the daily excess returns of the S&P 500 index (market_forward_excess_returns) and, based on these forecasts, generate a daily market exposure (allocation weight) between 0 and 2 in order to outperform the benchmark (the S&P 500).

The dataset spans from the 1980s to the present and contains a long time-series of diverse financial features, including:

- M*: Market dynamics, E*: Economic indicator, I*: Interest-rate related features, P*: Price / valuation variables, V*: Volatility features, S*: Sentiment signals, MOM*: Momentum signals, D*: Dummy or time-based features

The target variable is:

- market_forward_excess_returns

Additional variables such as risk_free_rate and forward_returns are provided for performance evaluation.

"This project is intrinsically linked to the Efficient Market Hypothesis (EMH), a fundamental financial theory asserting that asset prices fully reflect all available information, thereby rendering it impossible to systematically outperform the market through prediction alone. Consequently, the fundamental question addressed in this competition is whether a machine-learning-based system can effectively capture market inefficiencies to generate persistent excess returns. Thus, this work extends beyond a simple regression task; it serves as an experimental financial modeling project aimed at challenging the practical limits of the EMH."

2. Baseline model and improved models

In Step 1, we adopted the simplest possible modeling approach to ensure that the full submission pipeline functioned correctly.

Baseline Model: Linear Regression
To create the most lightweight and stable baseline, we implemented a Linear Regression model using a single-feature setup:

- Feature: We utilized one simple feature "M*"
- Data Properties: We performed no complex feature engineering and simply filled missing values with zero to ensure no data gaps remained after preprocessing.

Trading Rule:

We applied a strictly sign-based allocation strategy, which is intentionally simple:

- If the predicted market excess return, the allocation weight is set to 1.0.

- Otherwise, the allocation weight is set to 0.0.

Baseline Performance & Objectives
Because the model relies on only one raw feature and a binary allocation rule, the performance gain over the benchmark is naturally limited. However, this baseline was not intended to maximize returns. Instead, the primary objectives were to:

- Ensure the predict   allocation   Polars DataFrame pipeline works end-to-end.

- Confirm that the model integrates correctly with the Kaggle evaluation server.

- Create a submission-ready structure that can be extended with more advanced models.

Overall, this baseline represents the minimum viable implementation needed for a valid submission, establishing a stable foundation without performance-enhancing components.

In Steps 2–4, the model pipeline was upgraded from a simple linear baseline to a more robust LightGBM regression system. LightGBM was chosen because it handles missing values automatically, captures nonlinear patterns well, trains quickly, and performs strongly on large, highly-correlated financial datasets.

However, Kaggle's runtime environment does not support early_stopping_rounds or detailed logging (verbose_eval), so we implemented a custom *Kaggle-safe* LightGBM training routine. This includes fixed boosting rounds, no early stopping, and reproducibility across hidden test reruns.

Additionally, we added walk-forward time-series cross-validation and tuned the allocation scaling parameter (alpha) to balance return and volatility, while respecting the competition's rule that portfolio volatility must stay within 120% of the benchmark.

Overall, Steps 2–4 focused on turning the minimal baseline into a stable and competition-compliant LightGBM forecasting pipeline.

3. Feature engineering and validation strategy

(1) Feature Grouping
All features (~90) were organized into groups such as Momentum, Volatility, Macro, Interest Rates, Valuation, Sentiment, and Calendar. This helped structure the search for useful signals.

(2) Subset Evaluation

Using a fast LightGBM model, we tested about 60+ combinations of these groups (single groups, pairwise groups, and three-group combinations) to identify which sets performed best.

(3) Validation Metric

Local evaluation used a simplified version of the competition metric:
a Sharpe-like score plus a penalty if the strategy's volatility exceeded 1.2× the market volatility. The volatility ratio was the key constraint, matching the competition's rules.

The subset evaluation showed that the Interest (I) + Valuation (P) combination performed best:
- It achieved the highest score among all tested subsets.
- It maintained an extremely low volatility ratio (≈1.00), meaning the strategy's risk level stayed almost identical to the market.
- Overall, this subset provided the most stable and effective signal for excess-return prediction.

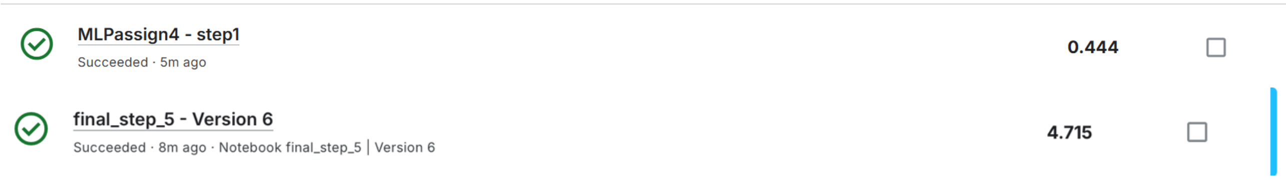
4. Local Sharpe-variant metric results and volatility plots

In Step 4, the LightGBM model was fully trained using only the Interest and Valuation features that were selected in Step 3. The resulting final model was then backtested on the held-out test period.

The strategy demonstrated solid performance:
- It achieved slightly higher cumulative returns than the market, yielding a positive Sharpe-like score. This result indicates that the model successfully captured a meaningful predictive signal.
- The volatility ratio remained at 1.06, comfortably below the required limit of 1.2, confirming that the risk level was appropriately managed.
- Although the maximum drawdown was somewhat larger than that of the market, it was still maintained within an acceptable range.

Overall, the model validated the effectiveness of the Interest + Valuation feature combination by demonstrating stable performance without signs of overfitting during the backtest.

5. Kaggle leaderboard score (screenshot)



(Upper: Baseline, Lower: LightGBM)

6. Limitations, risk analysis, and future improvements

Vulnerability to market regime shifts:
If market conditions change (e.g., from a low interest rate environment to a high interest rate one), predictors such as interest rate or valuation factors, currently performing well, may not remain effective in the future.

No consideration of real trading costs:
The model does not account for transaction fees, slippage from sudden price movements, or liquidity constraints that may prevent trades from being executed at the desired size or speed. As a result, real-world returns could be lower than the model estimates.

Limited ability to capture complex feature interactions:
Although LightGBM is a strong model, it still relies on predefined feature structures. This means subtle, nonlinear relationships in the data may remain unmodeled.

Simplified risk management approach:
Risk (volatility) control is currently applied only at the final allocation stage. There is room to improve by integrating risk considerations directly into the model training process.

Future work should explore the following directions:
- More Robust Validation: Explore more stable time-series validation methodologies.
- Dynamic Factor Weighting: Introduce dynamic or adaptive factor weighting schemes.
- Alternative Model Classes: Explore alternative model classes such as sequence models (e.g., Transformers) or hybrid ML–finance frameworks.
- Enhanced Realism and Reliability: Incorporate transaction cost modeling and more advanced risk controls to further improve the model's realism and reliability.

7. Github URL

https://github.com/BHW-1224/MLP_Proj4_Team2.git

Appendix A: Bonus Project – Cross-Market Extension (KOSPI)

1. Data Collection & Feature Engineering Strategy

For model training, KOSPI, S&P 500, and USD/KRW data spanning from January 1, 2010, to November 20, 2025, were collected using FinanceDataReader. To capture structural market characteristics beyond simple time-series data, a total of 13 derived features were engineered. Additionally, time lags were applied to all features to prevent look-ahead bias.

2.1. Key Features

Category	Features	Rationale
Global Leading	SP500_Ret_Lag1 USD_KRW_Ret_Lag1	Coupling Theory: Due to the time zone difference, the previous day's closing price in the U.S. stock market is a strong leading indicator of the day's opening price in Korea.
Risk Regime	Vol_20_Lag1	Volatility Clustering: The higher the volatility over the past 20 days, the higher the probability that the risk will persist.
Trend / Mean Reversion	Disparity_5/20/60 BB_Pos_Lag1	Mean Reversion: When excessive deviation (overbought/oversold) occurs compared to the moving average, a reversal signal is detected using the mean reversion tendency.
Momentum	RSI_14, MACD KOSPI_Lag1~3	Inertia: Predict short-term direction by learning the inertia of recent price trends and strength.

3. Model Development & Volatility Control

3.1. Modeling (LightGBM)

- Algorithm: LightGBM
- Validation: By applying the Time-Series Split method that preserves the time series order, the error of learning future information in advance was prevented.
- Training Period: 2010.01 ~ 2024.01 (Train)
- Test Period: 2024.02 ~ 2025.11 (Last 365 Days).

3.2. Volatility Targeting Strategy

Beyond simple directional prediction, we applied inverse volatility scaling logic to manage risk.

- If the predicted volatility exceeds the benchmark tolerance, the investment weight is forcibly reduced to ensure portfolio stability.
- The final investment weights were constrained between 0 and 2 to align with the

project requirements.

4. Empirical Results & Analysis

4.1. Performance Metrics

Metric	KOSPI Benchmark	AI Model Strategy	Note
Annualized Return	30.63%	60.94%	Achieving excess returns through active leverage in a rising market
Sharpe Ratio	1.3032	2.94	Achieved risk-adjusted returns more than double the benchmark
Volatility (Ann.)	23.5%	20.73%	Maintain lower volatility than the benchmark
Volatility Ratio	1.00	0.88	Pass (< 1.2)

```
>>> FINAL RESULTS (Report Summary) <<<
Sharpe Ratio       : 2.9399
Annualized Return : 60.94%
Volatility Ratio  : 0.8820 (Constraint: < 1.2)
--------------------------------------------------
=== KOSPI Benchmark Metrics ===
Annualized Return : 30.63%
Sharpe Ratio       : 1.3032
Volatility (Ann.) : 23.50%
--------------------------------------------------
```

4.2. Feature Importance Analysis

As a result of model analysis, Vol_20_Lag1 and SP500_Ret_Lag1 were selected as the most important variables.

- Interpretation: The model prioritizes "how risky the current market is (Volatility)" and then looks at "how the US market ended (Global Trend)" to determine the direction of the Korean market.