Team Project 4 (Team11)

CS 53744 Machine Learning Project – Instructor Professor: Jongmin Lee
Department of Computer Science & Engineering
20222446 Hyounsoo Kim, 20201876 Sanghyun Na, 20203009 Jaehyun Park

「Time-series prediction problem: Hull Tactical Market Prediction」

GitHub Repository: [github.com/MachineLearning-Project-Team11/Project04-Hull-Tactical-Market-Prediction]

Abstract

This project focuses on forecasting the daily forward excess returns of the S&P 500 Index using decades of factor-based financial data. The primary goal is to maximize the Modified Sharpe Ratio. We addressed the data's low signal-to-noise ratio by applying Mutual Information(MI) regression and PCA to select a robust set of 20 features. The resulting model demonstrates high performance and adaptability in risk-adjusted returns, confirming the effectiveness of our feature engineering strategy.

1. Introduction

Forecasting financial markets is challenging due to market noise, non-stationarity, and limitations imposed by the Efficient Market Hypothesis (EMH). In this project, the objective is to predict the daily forward returns of the S&P 500 and convert these predictions into position weights between 0 and 2. The strategy must outperform the S&P 500 benchmark while keeping volatility within 120% of the benchmark, as measured by the Modified Sharpe Ratio, which penalizes excessive volatility or poor performance.

The dataset contains 94 factor-based features across Macro-Economic, Volatility, Price, Momentum, and Sentiment families. Our workflow begins with exploratory data analysis to understand feature behavior and temporal structure. Based on these findings, we select a compact set of 20 high-signal features using correlation, mutual information, and PCA to reduce redundancy and noise. The overall goal is to design a machine learning model that captures market directionality while maintaining stable, risk-adjusted returns under realistic constraints. Subsequent sections describe feature preparation, model development, and evaluation.

2. Datasets

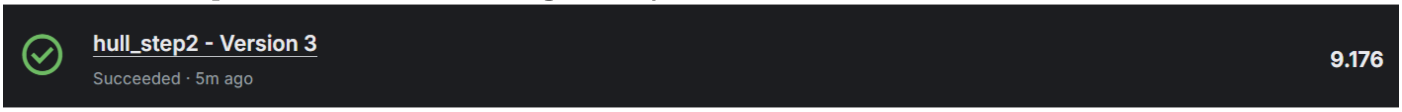
The dataset's predictor variables are organized into 8 features. The first is **Market Features (M*)**, capturing market microstructure such as price momentum, volatility dynamics, and order flow characteristics. The second, **Macro Economic Features (E*)**, consists of macroeconomic indicators reflecting the business cycle, inflation, growth expectations, and labor market dynamics. The third set, **Interest Rate Features (I*)**, includes term structure indicators, real interest rates, credit spreads, and monetary policy signals. The fourth set, **Price and Valuation Features (P*)**, comprises absolute and relative valuation metrics such as P/E ratio, dividend yield, and market capitalization-weighted valuation indicators. Fifth, **Volatility Features (V*)** measure realized volatility, implied volatility, volatility risk premiums, and tail risk indicators. Sixth, **Sentiment Features (S*)** represent market sentiment proxies, positioning data, and behavioral finance indicators. Seventh, **Momentum Features (MOM*)** capture cross-sectional and time-series momentum indicators, reflecting price trends across various rolling windows. Finally, **Dummy/Binary Features (D*)** represent calendar-based characteristics such as day-of-the-week effects, monthly effects, quarterly effects, and market regime indicators. The target variable is forward_returns, denoting the return from investing in the S&P 500 index on day t and selling at the closing price on day t+1. It is defined as follows:

$$forward\_returns\_t = \frac{P_{S\&P500,\,t+1} - P_{S\&P500,\,t}}{P_{S\&P500,t}}$$

3. Baseline Model: Adaptive Momentum Strategy

In this project, an adaptive momentum strategy utilizing traditional technical analysis indicators was constructed as a baseline for comparing and validating the performance of complex machine learning models. This model captures the strength and direction of stock price trends to dynamically determine the optimal investment allocation between 0 (cash holding) and 2 (2x leverage). The baseline model generates two key derived variables based on restructured price data to reflect short-term and medium-term market trends. The first derived variable is the SMA Ratio (Trend Signal), which uses the deviation ratio between the 21-day short-term simple moving average (SMA) and the 63-day medium-term SMA to capture trend reversal signals similar to golden crosses/death crosses. The second derivative variable is the Short-term Momentum (Speed Signal), which measures the velocity of price fluctuations through the recent 5-day return rate. The final investment weight is determined through a linear combination of the two indicators calculated above to generate a raw signal, followed by scaling and clipping processes. During the test phase, to perform real-time predictions without referencing future data, the model employs stateful processing, maintaining the price history of the most recent 63 days in memory.

4. Model Development: Gradient Boosting with Dynamic Feature Generation



To overcome the simple linearity of baseline models and capture the nonlinear patterns in financial time series, this study proposes a machine learning pipeline based on **LightGBM (LGBMRegressor)**. This model predicts future forward returns using historical data and converts the predicted values into final portfolio weights through an optimization process. To incorporate auto-correlation and trend information from the time series data into the model, derived variables were created by applying various laggings and rolling windows to the original numerical data. Past data points at 1, 3, 5, 10, 21 and 63 days before were used as variables to directly incorporate short-term and medium-to-long-term historical information. Additionally, rolling statistics were applied to calculate rolling mean and rolling std for 5-day (weekly), 21-day (monthly), and 63-day (quarterly) windows. This design enabled the model to learn market trend, volatility, and regime information. LightGBM was adopted as the prediction model. Mean squared error (MSE) was used as the loss function. Furthermore, the output of a simple regression model is merely the expected return value and cannot directly serve as an investment weight between 0 and 2. Therefore, this study applied a post-processing technique combining Z-Score normalization and scope optimization to convert the predicted values into investment weights.

5. Exploratory Data Analysis (EDA)

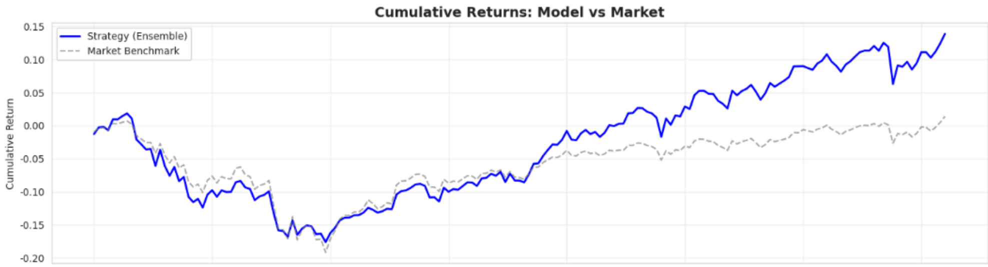
The exploratory analysis revealed several structural characteristics of the dataset that directly shaped the modeling strategy. The target variable, forward_returns, displayed clear non-normality, heavy tails, and near-zero mean, confirming that short-horizon equity returns are dominated by noise and require nonlinear models to capture weak predictive structure. Rolling windows of mean and volatility further highlighted strong non-stationarity, with volatility clustering and regime-like shifts that are typical in financial time series. Across the original 94 factor features, correlation analysis showed extensive multicollinearity, which many Economic (E*) and Interest-rate (I*) features exhibited $|r| > 0.7$ with one another, reducing the effective dimensionality of the dataset. This redundancy was quantified through PCA, where the first ~25 components explained nearly 80% of the variance, indicating that a large portion of features carried overlapping information.

Dummy variables (D1–D9) were also evaluated for potential regime-related structure, but Pearson correlations with the selected factors were all below 0.13 and mutual information with the target was near zero, confirming that they did not capture meaningful market states. These results collectively motivated a relevance-driven and redundancy-aware feature selection strategy rather than relying on the full raw factor set. Consequently, they were excluded from the final feature set, allowing the modeling process to focus on higher-signal and more stable factors.

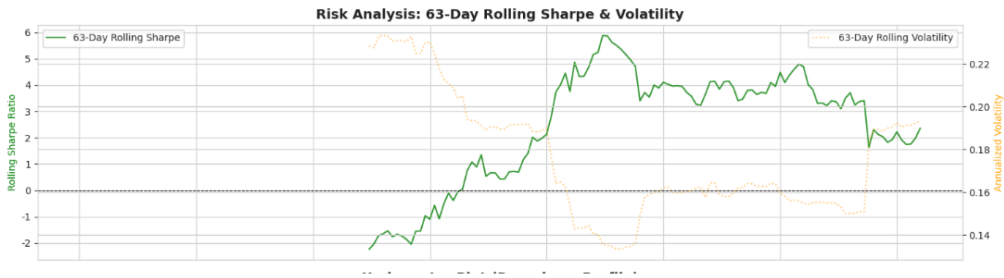
6. Feature engineering and validation strategy

Guided by the EDA findings, feature engineering focused on constructing a compact set of high-signal and low-redundancy predictors. Feature relevance was quantified using three complementary metrics: correlation with the target to capture linear effects, mutual information to detect nonlinear dependencies, and random forest feature importance to assess model-based contribution. These signals were combined into a composite ranking, and low-relevance or highly collinear features were systematically removed. Dummy variables and handcrafted regime indicators were excluded, as both correlation and MI analyses showed negligible predictive value. This process produced a final set of 20 features: **V13, V7, E19, M4, P5, S5, P8, V9, S8, V10, S2, M1, P10, M12, I2, M2, M3, P7, E1, M10**. The result subset covers all major factors, Market, Volatility, Price, Sentiment, Economic, and Interest. The resulting subset captures complementary short-term and macro-level signals while minimizing overfitting risk, and serves as the foundation for all subsequent modeling and validation.

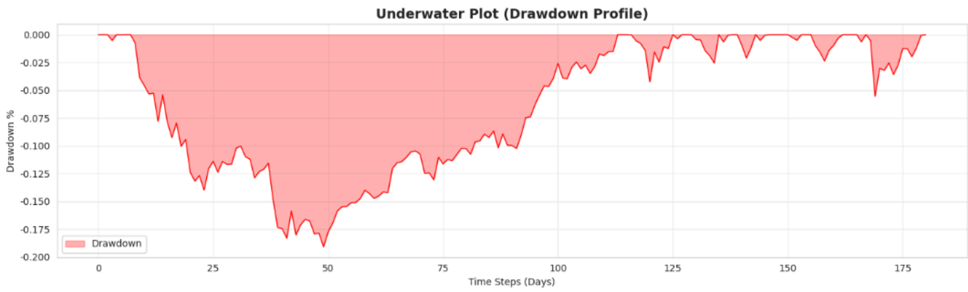
7. Evaluation and Backtesting



Comparing the performance of this model against the market benchmark represented by the S&P 500, the proposed Dynamic Ensemble strategy achieved a distinct excess return (Alpha) that outperformed the market return. Although it initially showed a somewhat sluggish trend, moving in tandem with the market's downturn during the early validation period, decoupling from the market was observed as data accumulated and online learning progressed. Notably, even during periods when the benchmark moved sideways or saw only marginal gains, this strategy maintained an upward trend, ultimately recording a significantly superior cumulative return compared to the market. This suggests the model is not merely a market-tracking (Beta) strategy but captures subtle market patterns to generate independent sources of return.



Analysis of the trends in the 63-day rolling Sharpe ratio and annualized volatility, key indicators of risk-adjusted returns, revealed a clear inverse correlation between the two metrics. During periods of volatility exceeding 20% at the start of the validation, the Sharpe Ratio remained negative or at low levels. However, once volatility stabilized below 15%, the Sharpe Ratio rose sharply, reaching values exceeding 5.0 at its peak. This demonstrates that the OnlineOptimizer and Volatility Targeting mechanism applied to the model effectively reduced risk exposure during periods of heightened market uncertainty and maximized return opportunities when the market stabilized. In other words, the high Sharpe ratio is not a product of chance but rather the result of rigorous volatility control.



Drawdown analysis (Underwater Plot), which shows the maximum decline in assets, revealed the model's risk defense capabilities and resilience. The maximum drawdown (MDD) occurring in the early training phase was approximately -19%, but the model subsequently demonstrated a swift recovery to its previous peak. Notably, the depth of subsequent declines after the initial adaptation period became significantly shallower. This indicates the model strengthened its asset defense logic by learning from past loss experiences: it now detects market crash signals earlier and reduces portfolio exposure (de-leveraging).

8. Final Strategy performance and Analysis

Our final submission model is a multi-scale regression framework that uses lagged factors, rolling statistics, and volatility-related signals to predict daily forward excess returns. The model output is transformed into portfolio weights between 0 and 2, and evaluated under the Modified Sharpe metric used in the competition. The model is trained using walk-forward validation to avoid look-ahead bias and ensure that only information available at each time step is used.

First, during the data preprocessing stage, we generate not only simple price information but also technical indicators like RSI, MACD, and Bollinger Bands, along with higher-order statistics such as skewness and kurtosis. This allows us to capture market momentum and extreme tail risk in a multidimensional way. Furthermore, RobustScaler is applied to ensure robustness against outliers in financial data, and a stateful structure that preserves historical data during inference guarantees the continuity of indicator calculations.

The model architecture features a unique structure that processes input features by separating them into short-term, medium-term, and long-term time scales. Each time-scale branch focuses on important features via the SE-Block (Squeeze-and-Excitation) and ensures learning stability in deep neural networks through the Residual Block. The distinct time-series information extracted in this manner is dynamically combined by a learnable Gating Network at the optimal ratio for the current market situation. To maximize the model's generalization performance, the SAM (Sharpness-Aware Minimization) optimizer is used to find the flat minimum of the loss function. Ensemble effects are achieved by applying EMA (Exponential Moving Average) to the weights of multiple seed models.

Operationally, we go beyond static models by introducing an Online Learning mechanism. As new data arrives, the model continuously fine-tunes itself via a SlidingWindowBuffer, ensuring historical data is not forgotten. A Hybrid Loss combining prediction error (MSE) and the financial performance metric Sharpe Ratio is used to optimize risk relative to return. Finally, during the inference phase, the Online Optimizer analyzes recent prediction performance to dynamically adjust investment weights. It also applies Volatility Scaling, which reduces

On the validation set, the strategy achieved a cumulative return of 13.89%, outperforming the S&P 500 benchmark (1.41%) with a reduced maximum drawdown of -19.11%. The Modified Sharpe score was 0.761, and the volatility ratio remained below the 1.2 constraint throughout the period. The investment weights were stable, with an average allocation close to 1.0 and no extreme leverage. On the Kaggle public leaderboard, the final submission reached a score of 0.761, confirming that the model generalizes to unseen test data, though with reduced strength compared to the validation period.

9. Discussion

Although the final strategy achieved strong performance and outperformed the S&P 500 benchmark, several limitations remain. Short-horizon equity returns are inherently difficult to predict, and even well-engineered features provide only weak and regime-dependent signals. Structural breaks—such as macroeconomic shocks or abrupt volatility shifts—are not explicitly modeled, making the system vulnerable to rapid market transitions. Moreover, the backtests assume frictionless trading with daily rebalancing, excluding transaction costs and slippage that would materially reduce real-world performance. The feature-selection and modeling pipeline is also largely static; more adaptive methods such as online learning or dynamic factor selection may yield greater robustness.

From the perspective of the Efficient Market Hypothesis (EMH), our results offer a nuanced interpretation. The model's ability to extract modest predictive structure suggests that markets are not perfectly efficient, particularly at very short horizons where temporary behavioral or macro-driven anomalies appear. However, the fragility of these signals—together with their rapid deterioration on unseen data—supports the view that the S&P 500 is *nearly efficient*. Any exploitable patterns are small, unstable, and easily competed away once translated into actionable trading rules. Thus, while machine learning can marginally improve risk-adjusted performance under controlled assumptions, persistent excess returns remain difficult to achieve after accounting for market frictions and real-world constraints.

Overall, this project demonstrates that careful feature engineering, nonlinear modeling, and risk-aware evaluation can produce a strategy with improved historical performance and stable volatility characteristics. Yet the broader findings reinforce a central implication of EMH: predictive signals in highly liquid equity markets are limited, short-lived, and sensitive to structural changes—highlighting the continued challenge of generating reliable alpha in practice.

Appendix (Optional Bonus): Cross-Market Extension - NASDAQ Forecasting Study

This appendix presents an extension of our main S&P 500 forecasting framework to a second market: the NASDAQ Composite Index (IXIC). Because NASDAQ is a high-beta, growth-oriented index that reacts strongly to interest rate changes and global risk sentiment, this experiment evaluates the robustness and generalizability of our methodology across different market environments. The extension follows the same structure as the main task—feature engineering, model training, volatility-constrained backtesting, and interpretability analysis—while using an independently collected dataset.

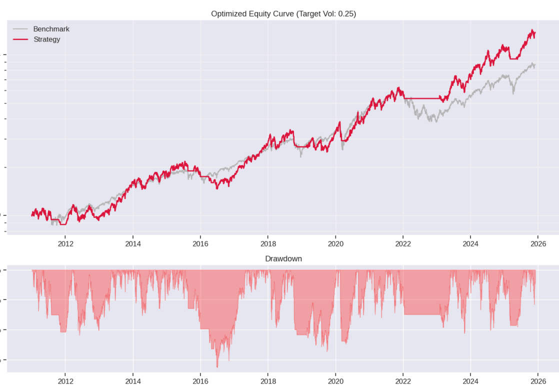
1. Data Collection and Experimental Setup

This extension evaluates whether the S&P 500 forecasting framework generalizes to a structurally different market. We selected the NASDAQ Composite, a high-beta index deeply influenced by interest-rate cycles and global liquidity conditions. To build a dataset suitable for macro-sensitive modeling, we collected daily NASDAQ prices and a broad set of external indicators using the Yahoo Finance API. The dataset spans the period from 2011 to 2025 and includes fifty-two engineered features representing global equity movements (e.g., Nikkei and DAX), credit-spread stress, Treasury yield shocks, volatility clustering, and trend-momentum signals.
All price-based inputs were converted to log-returns for stationarity, while volatility levels and yield levels were preserved to allow the model to learn threshold behaviors. Because Asian and European markets close before the U.S. opens, their same-day returns serve as legally usable pre-market sentiment indicators. The target variable was constructed in the same manner as the main task: next-day NASDAQ returns paired with a bounded investment-weight prediction in the range 0–2. The appendix figures (Equity Curve, Drawdown, Daily Weights) reflect this full dataset and testing period.

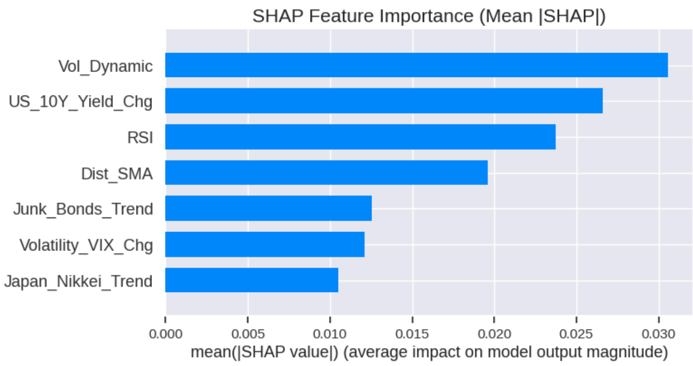
2. Modeling Framework and Volatility Control

To transfer the S&P pipeline to NASDAQ, we adopted a hybrid architecture combining a Random Forest regime classifier with an inverse-volatility targeting module. The Random Forest, deliberately shallow (max depth = 5, high minimum leaf size), was optimized to detect broad market regimes rather than fit short-term noise. Its continuous output represents the estimated "investability" of the next day. This signal does not directly dictate leverage; instead, it is processed through a volatility-scaling rule in which realized NASDAQ volatility determines the allowable position size.
Exposure is computed as a function of both model signal and dynamic volatility, capped at 2.0×. In low-volatility uptrends, exposure naturally increases, while volatility spikes force the system into reduced positions even when the predictive model signals optimism. To meet the project requirement that strategy volatility not exceed 1.2× the benchmark, a hard constraint was integrated into the optimization loop: any configuration violating the limit was immediately discarded rather than softly penalized. This ensured that the final selected model is structurally incapable of producing excessive risk.
The resulting weight trajectories, shown in the appendix images, demonstrate smooth and economically interpretable adjustments, avoiding the unstable leverage behavior typical of unconstrained machine-learning strategies.

3. Results, Interpretation, and Cross-Market Implication



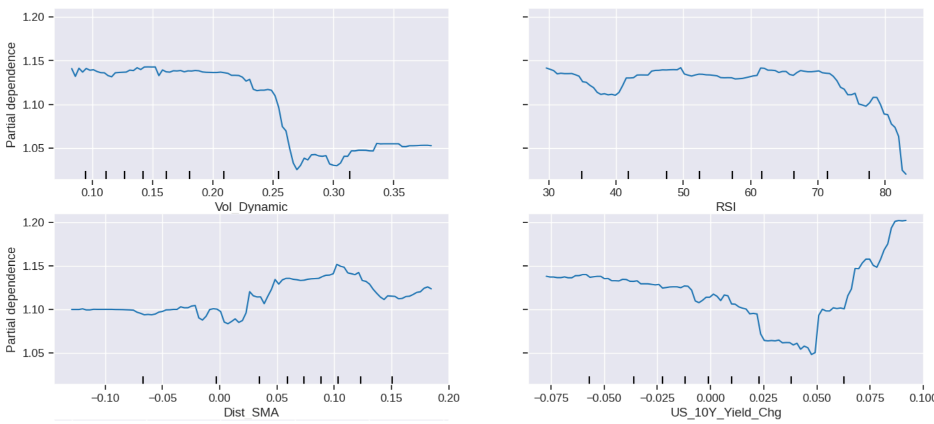
The NASDAQ extension achieved substantially higher long-term performance compared with the benchmark while satisfying the volatility constraint. Over the 2011–2025 test window, the strategy produced a cumulative return of 1,273% compared with 769% for buy-and-hold. Annualized volatility was 24.21%, corresponding to a 1.17× ratio relative to the benchmark, thus meeting the ≤1.2× requirement. Notably, the model avoided the severe 2022 rate-shock drawdown by rotating to near-zero exposure, as reflected in the equity curve where the strategy remains flat during the inflationary regime. This demonstrates genuine crisis-mitigation behavior rather than overfitting to historical momentum.



SHAP analysis confirms that the model's learned structure aligns with macroeconomic intuition: dynamic volatility was the most influential feature, followed by Treasury yield changes, RSI momentum, and distance-from-trend metrics. Partial dependence plots reveal that exposure drops sharply when realized volatility exceeds roughly 25%, despite this threshold never being explicitly encoded into the model. Similarly, rapid increases in Treasury yields depress predicted investability, reflecting the high duration sensitivity of NASDAQ components. Volatility-derivative signals, such as VIX changes, show an asymmetric effect, with exposure rising after volatility collapses and falling after fear spikes. These findings, together with the interpretable weight paths and drawdown patterns, indicate that the forecasting pipeline transports well across markets while preserving economic coherence.

Overall, the cross-market extension demonstrates that the methodology is flexible, risk-aware, and capable of adapting to structurally different regimes. Its ability to generalize from the S&P 500 to NASDAQ—while delivering superior returns, controlled volatility, and interpretable behavior—supports the broader applicability of the framework to multi-asset portfolios.

4. Partial Dependence Diagnostics



To complement the SHAP-based global feature importance, we compute Partial Dependence Plots (PDP) for the seven most influential drivers of the NASDAQ strategy. These plots verify that the model responds to macroeconomic and risk conditions in a structurally meaningful way rather than memorizing noise. Vol_Dynamic shows a sharp deterioration above ~0.25, matching the volatility level where risk-adjusted returns historically weaken. RSI increases exposure during healthy momentum (40–65) but reduces it once markets become overbought (>70). US_10Y_Yield_Chg displays an asymmetric decline under sudden rate shocks, consistent with the duration sensitivity of tech equities. **Dist_SMA** cleanly separates trend regimes, assigning higher leverage only when prices stay above their medium-term trend. **Volatility_VIX_Chg** captures fear dynamics by lowering exposure during volatility spikes and raising it when fear subsides. **Junk_Bonds_Trend** reflects tightening credit conditions by signaling risk-off behavior when spreads widen. **Japan_Nikkei_Trend** provides time-zone sentiment information, with stronger foreign momentum modestly improving expected returns. Together, these PDPs confirm that the model's behavior is economically coherent and grounded in stable cross-market signals.