McGill - FIAM Asset Management Hachakton

EXECUTIVE SUMMARY

At **LYTA Strategy Analytics**, we leverage advanced machine learning and data-driven insights to design optimized investment strategies that consistently outperform traditional market approaches

STRATEGY

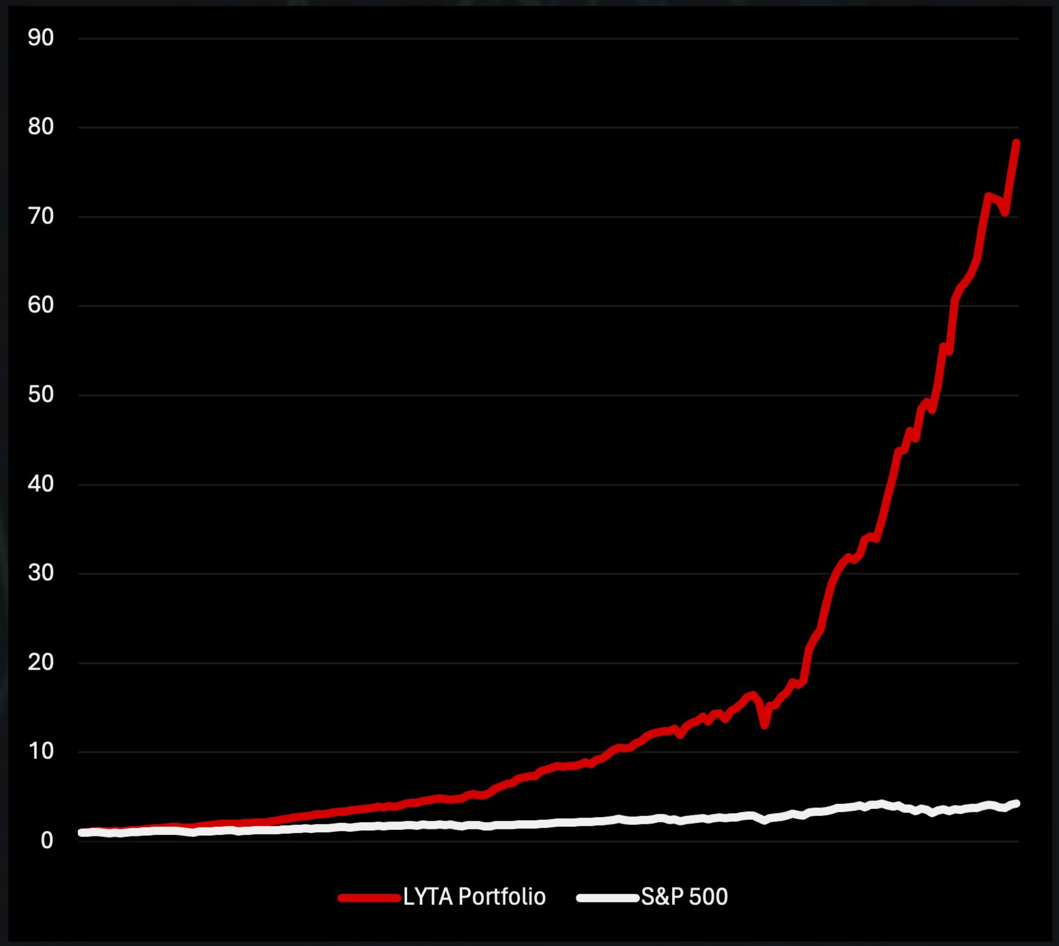
Through extensive analysis, we developed an optimal portfolio strategy consisting of **long** and **short** positions. delivering superior returns while maintaining a high Sharpe ratio

MACHINE LEARNING ALGO

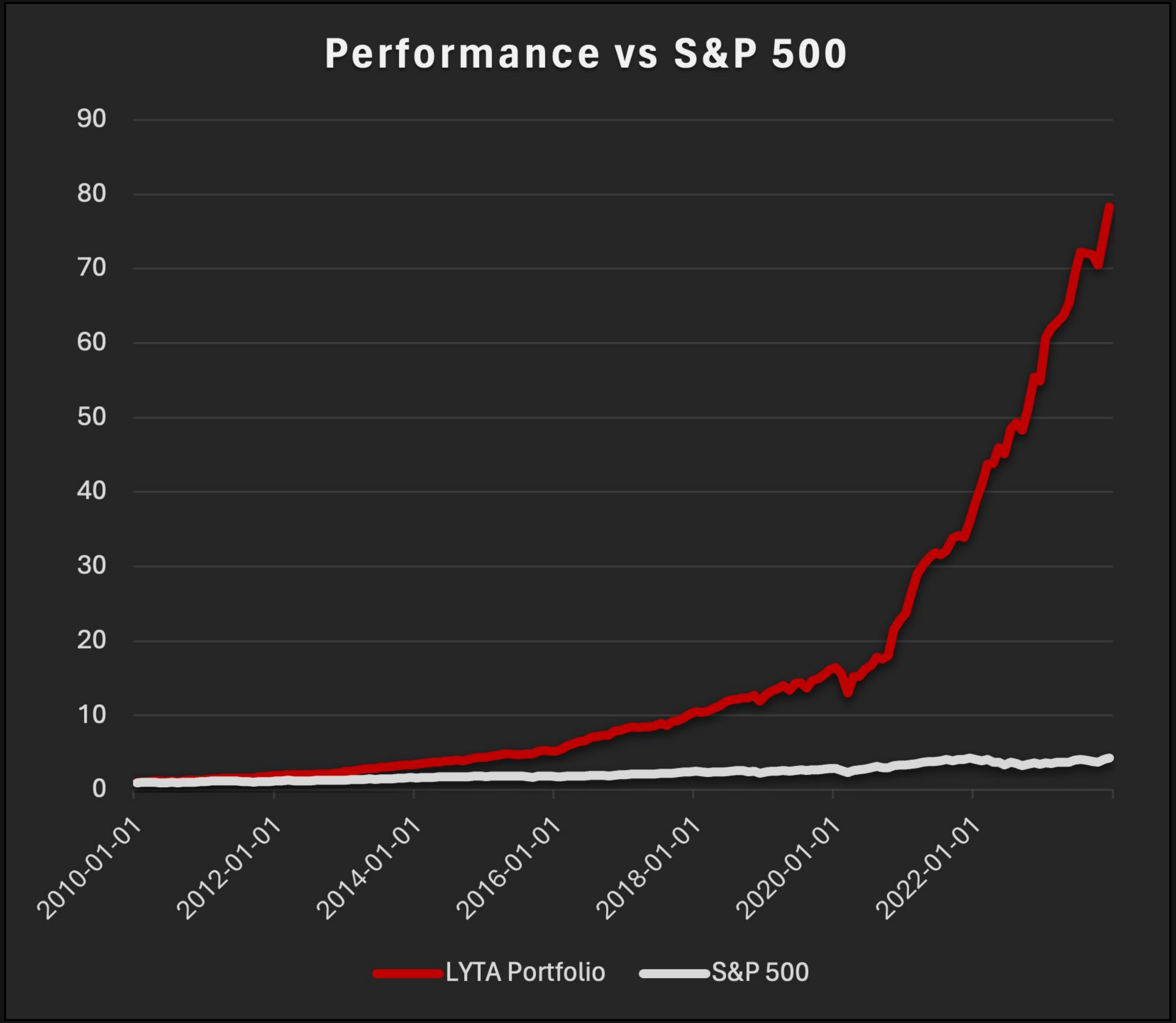
Leveraging **XGBoost** for return prediction, our strategy accurately identifies market trends and asset behaviors. The **Recursive Feature Elimination** for feature selection ensures that only the most relevant variables inform the decision-making process.

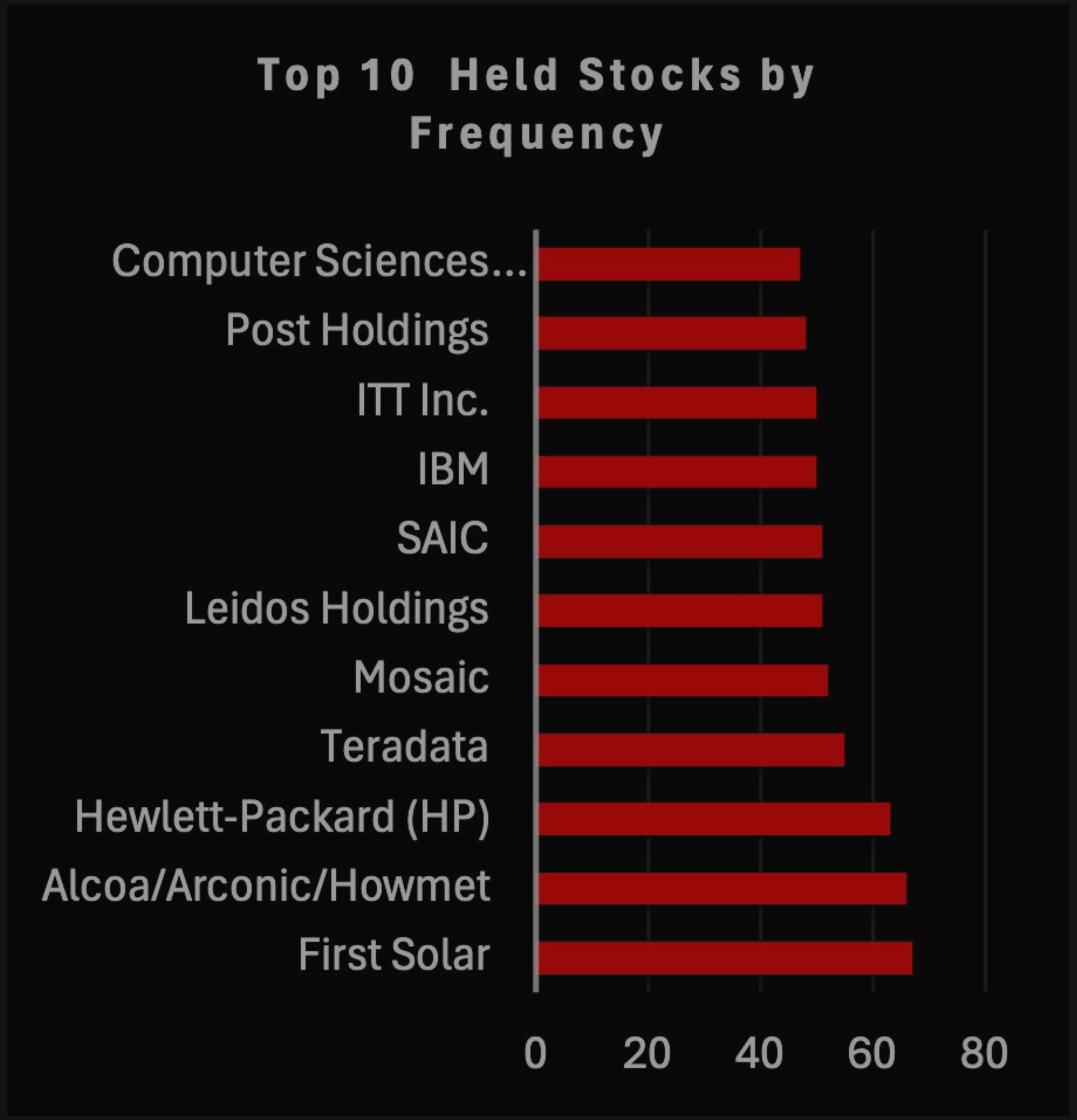
PERFORMANCE

Utilizing advanced machine learning, our strategy delivered over **7500% return** from 2010 to 2024, far surpassing the S&P 500.



INVESTMENT STRATEGY





Predicted returns from XGBoost guide a mixed long-short strategy, with 70% long and 30% short positions of 51 stocks. This mixed portfolio strategy reiterated every month insure a good balance between return and risk delivering a sharp ratio of 2.466

Performance is based on real returns, showing consistent outperformance over the S&P 500 during the out of sample period. This data-driven approach ensures consistent outperformance in changing market conditions.

DATA AND METHODOLOGY

Data Preprocessing

The data preprocessing step aims to ensure the integrity and usability of the dataset by cleaning and selecting relevant factors and stocks based on missing values and zero values. Factors with fewer than 30% missing values and less than 20% zero values are retained. Stocks are selected based on the number of available months, keeping those with the most available data and removing stocks that have all missing values for any factor. Missing values in smaller gaps are filled using mean or median imputation, though this method might not fully capture temporal dynamics. Additionally, a ranking and normalization process is applied, where each factor is ranked based on its values to enable comparison between stocks. The normalization adjusts the dataset to ensure that factors contribute equally to the ranking process, preventing scale bias

Feature Selection

By using Robust Scaler for feature scaling, it reduces the impact of outliers, which are common in financial data, ensuring that the model's performance is not skewed by extreme values. The Recursive Feature Elimination (RFE) method helps to automatically identify and retain the most important features, enhancing interpretability and reducing overfitting. The use of 500 estimators and subsampling helps generalize better on unseen data by preventing overfitting and speeding up training without sacrificing too much accuracy. Finally, XGBoost flexibility with parameters like learning rate and tree sampling provides precise control over model performance, making it highly adaptable to various financial prediction tasks. While rule of thumb may suggest to keep from 30-45 out of 145, several tests have been conducted to give out result of 50 features leading to the highest R-squared.

Predictive Model

XGBoost is a highly efficient gradient boosting algorithm that excels at handling non-linear relationships of large complex datasets, using techniques like regularization and parallel processing to reduce overfitting and reduce computational cost. Hyperparameter tuning via GridSearchCV optimizes the model's performance by testing various parameter combinations, enhancing predictive accuracy. By applying Robust Scaler, it helps the model reduce the impact of outliers, especially in complex financial data and ensuring stability. Also, the model employs time series cross-validation and a time window approach, ensuring realistic training and validation over time. Together, they strengthen XGBoost ability to handle features interaction and imbalances in data, which enhances its suitability and ability to capture the complex pattern in the big financial data. Hyperparameter values are selected based on several test conducted to give the most precise result but still keeping the moderate-to-low computational cost and duration.

PORTFOLIO PERFORMANCE

LYTA PORTFOLIO

Average Annual Return	36.54%
Standard Deviation	0.1313
Alpha (CAPM)	0.0286
Sharpe Ratio	2.47
Information Ratio	2.61
Max Drawdown	-23.02%
Max 1-Month Loss	-16.52%
Turnover (Long)	35.09%
Turnover (Short)	49.80%

2010 -2024

S&P 500

Average Annual Return	13.42%
Standard Deviation	0.157
Alpha (CAPM)	0
Sharpe Ratio	bellow 0.90
Information Ratio	N/A
Max Drawdown	-18.11%
Max 1-Month Loss	-9.18%
Turnover (Long)	Bellow 5%
Turnover (Short)	N/A

2010 -2024

The **LYTA Portfolio** achieves a higher return of **36.54%** compared to **13.42%** for the **S&P 500** but with a larger **drawdown** of **-23.02%** versus **-18.11%**

Despite greater drawdown, the **LYTA Portfolio** holds a **Sharpe ratio** of **2.47**, significantly outperforming the S&P 500's ratio below **0.90**

STRATEGY REVIEW

Our strategy performed as expected, with a **Sharpe ratio of 2.47**, indicating a strong risk-adjusted return. The **average return** of **36.54%** significantly outperformed the S&P 500's **13.42%**

The model's focus on **fundamental signals** like **market equity, price-to-high, and volatility** allowed it to capture both upside potential and downside risk effectively. The use of **alternative data and machine learning techniques** contributed to the success by identifying patterns.

Profitable Stocks

Top performing positions included companies like **First Solar** and **Hewlett-Packard(HP)** resilient in economic downturns.The **long positions** in companies with consistent **revenue growth** and **low volatility** provided stability, while **shorting overvalued stocks** capitalized on market corrections.

Macro-Economic Events

The portfolio benefitted from **market recovery post-2008 financial crisis** and the **stimulus-driven bull market** from 2010 to 2023. **COVID-19 recovery** also provided strong opportunities for long positions in sectors like **technology** and **consumer staples**

Potential Improvements

- **Enhance Risk Management:** Implement **dynamic hedging strategies**
- **Feature Engineering:** By having more time, we could have implemented more advanced techniques like **natural language processing (NLP)** to capture sentiment analysis from financial news or earnings reports could improve predictive accuracy

APPENDIX 1

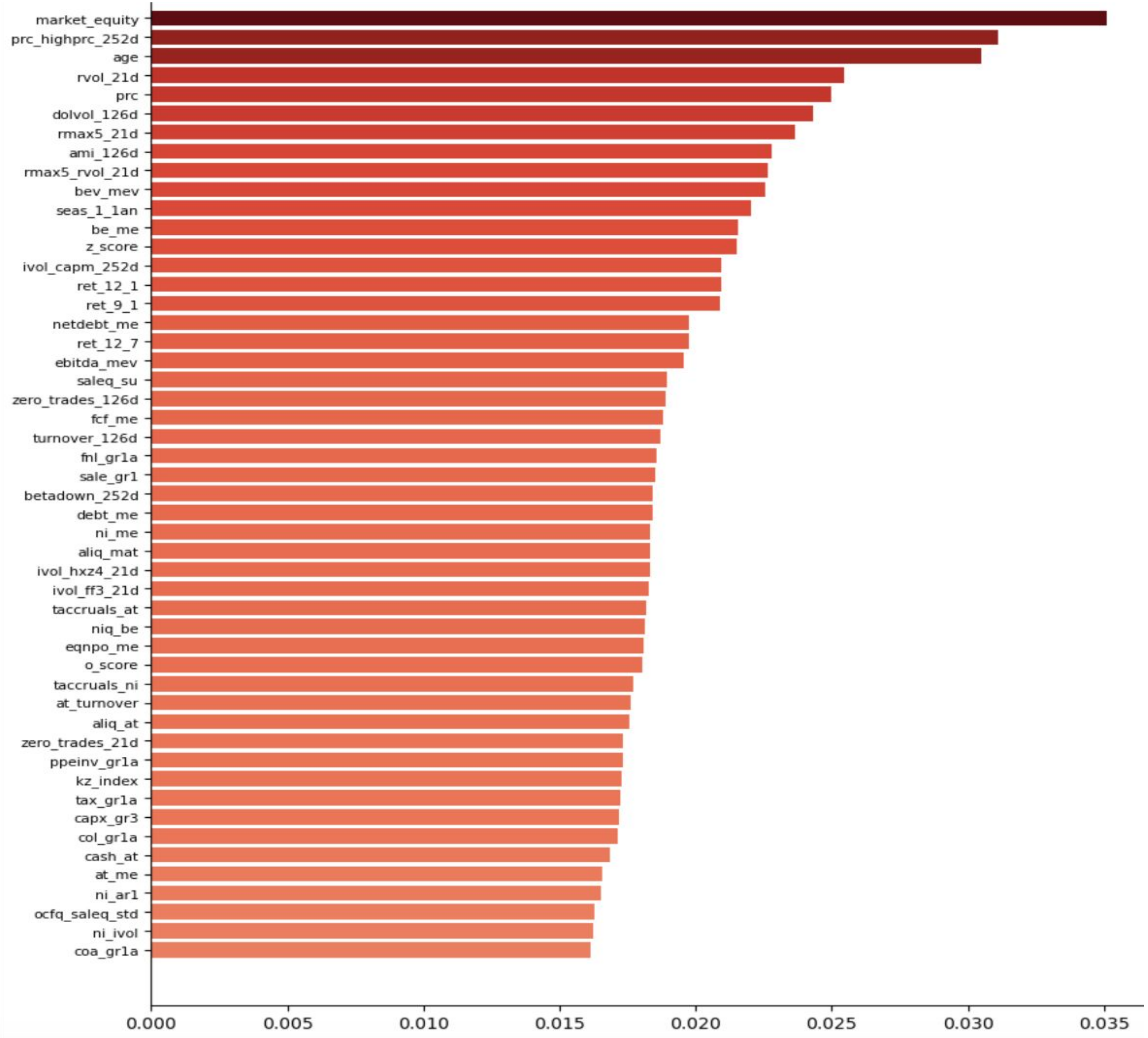
Summary of final strategy (Mixed strategy 70% Long and 30% short)

```
The best number of stocks is: 51 with a Sharpe ratio of: 2.4663284755789285
Final portfolio Sharpe ratio: 2.4663284755789285
Sharpe Ratio (Mixed Strategy Portfolio): 2.4663284755789285
                            OLS Regression Results
==============================================================================
Dep. Variable:         weighted_return   R-squared:                       0.005
Model:                             OLS   Adj. R-squared:                 -0.001
Method:                  Least Squares   F-statistic:                    0.9665
Date:                 Thu, 03 Oct 2024   Prob (F-statistic):              0.327
Time:                         02:30:35   Log-Likelihood:                 312.30
No. Observations:                  168   AIC:                            -620.6
Df Residuals:                      166   BIC:                            -614.4
Df Model:                            1
Covariance Type:                   HAC
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.0286      0.003      8.772      0.000       0.022       0.035
rf            -2.2709      2.310     -0.983      0.327      -6.831       2.290
==============================================================================
Omnibus:                       29.118   Durbin-Watson:                   2.157
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              226.325
Skew:                           0.075   Prob(JB):                     7.15e-50
Kurtosis:                       8.684   Cond. No.                         866.
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 3 lags and without small sample correction
CAPM Alpha: 0.028596770205470117
t-statistic: 8.771757243375754
Information Ratio: 2.611385342858769
Max 1-Month Loss (Mixed Strategy Portfolio): -0.16516121665490197
Maximum Drawdown (Mixed Strategy Portfolio): -0.2302101788894957
Long Portfolio Turnover: 0.3508982035928144
Short Portfolio Turnover: 0.49820359281437127
Portfolio Annualized Return: 0.3654
Portfolio Annualized Standard Deviation: 0.1313
```

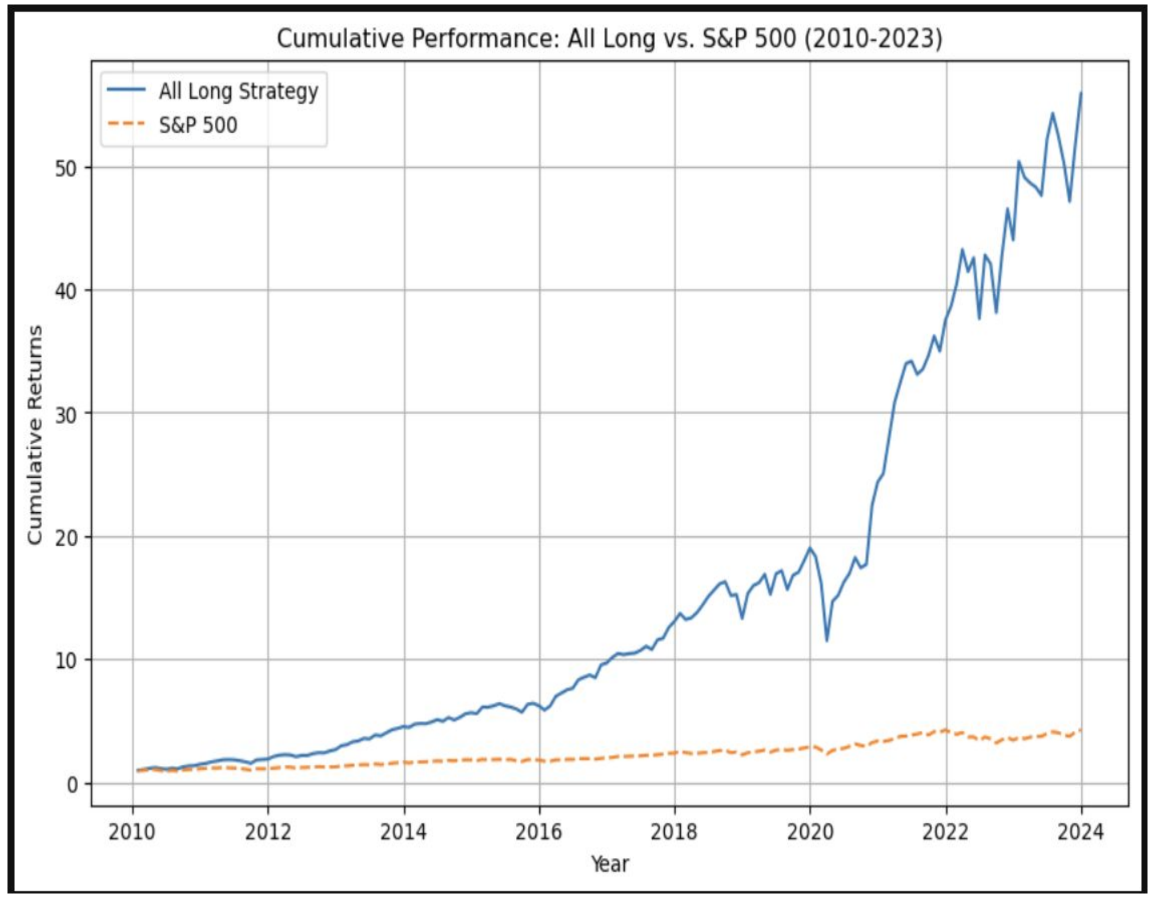
APPENDIX 2

Feature Importance Analysis

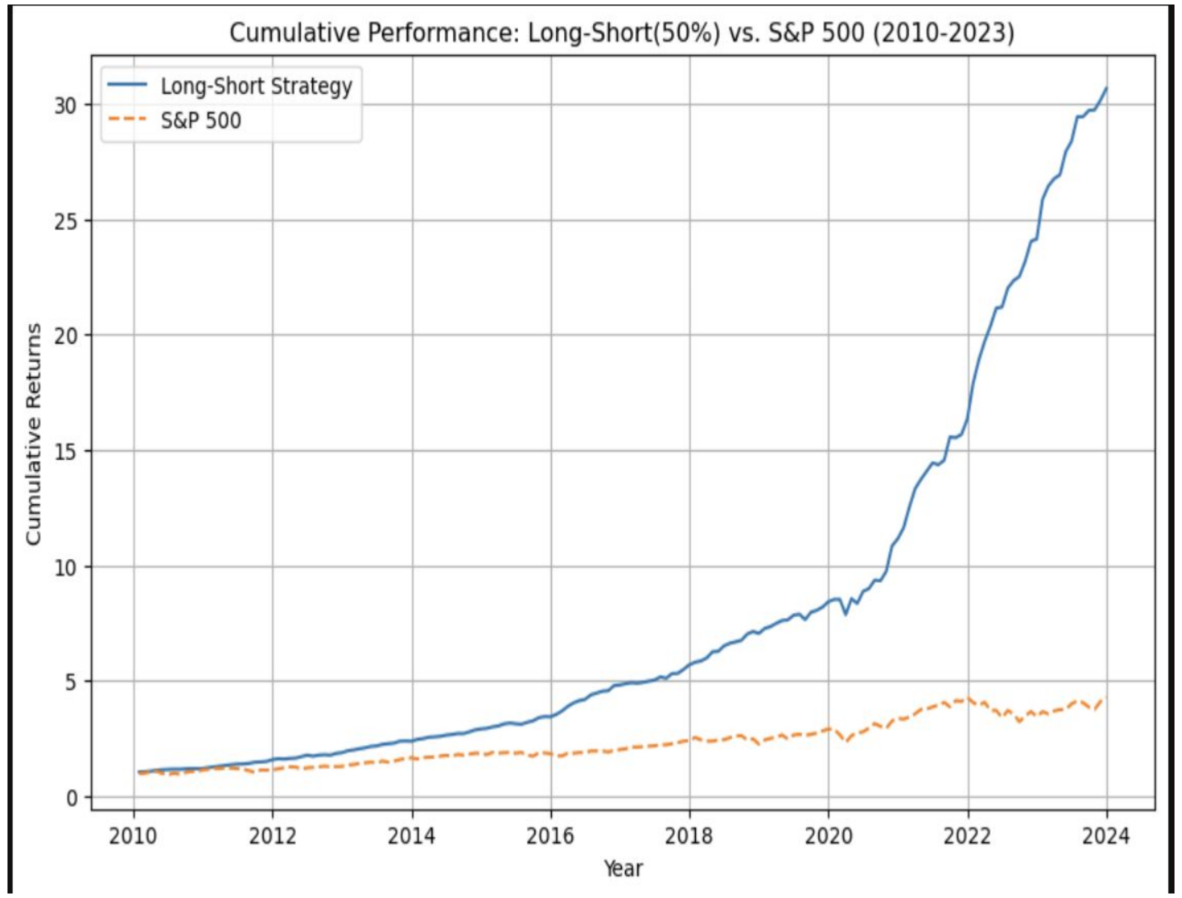


APPENDIX 3

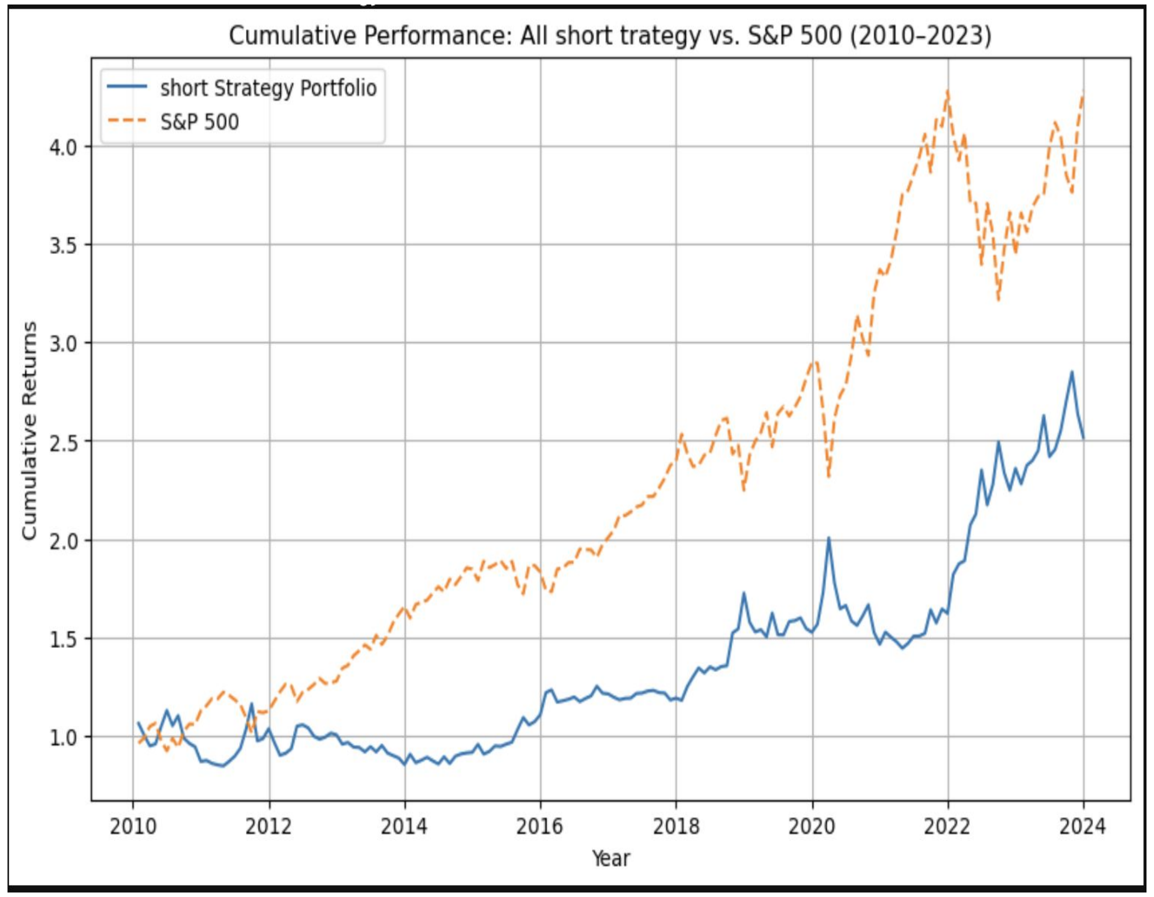
Alternative strategies tested



Sharpe Ratio : 1.34948



Sharpe Ratio : 3.30520



Sharpe Ratio : 0.46555

APPENDIX 4

Comparison between Linear model on raw dataset and LYTA model on clean dataset