NATIONAL UNIVERSITY OF SINGAPORE

Master's of Computing (General-Track)

Alpha Tree Search and Machine Learning Approaches to Optimising Real Estate Portfolios

Leong Wei Ming

Supervisor: Professor Liu Li Li

Examiner: Professor Chin Wei Ngan

Department of Computer Science

Internal Capstone Project for AY2023/2024

ABSTRACT

An internal project about applying genetic algorithm to search for optimal alphas. State the major contribution:

DECLARATION

I hereby declare that this project report is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in this report.

This report has also not been submitted for any degree in any university previously.

ACKNOWLEDGEMENT

CONTENTS

INTRODUCTION

The saying goes that once a profitable trading formula has been discovered and traded on by enough people, its profits will be eroded away and it will cease to be profitable. Traders and investors appear to be playing a never-ending game of "Hide-and-Seek" in search of profitable trading strategies. Due to the evolving nature of the financial markets, traditional financial time-series forecasting models which are static in nature are becoming less effective than Machine Learning models and dynamic algorithms in identifying the best investments (Sheth & Shah, 2023). The success of Machine Learning has led to numerous research papers applying a myriad of Machine Learning techniques to predict stock prices (Obthong, Tantisantiwong, Jeamwatthanachai, & Wills, 2020). However, fewer papers apply these approaches in the Real Estate sector and these papers largely focus on future stock price prediction rather than portfolio allocation as a whole (Habbab & Kampouridis, 2024). Many of these papers also use price-volume data without fundamental financial data as inputs. In 2022, the global real estate sector was worth more than \$380 trillion and worth more than the global equity and bond markets combined (Tostevin & Rushton, 2023), with approximately 893 listed Real Estate Investment Trusts (REITs) (Nareit, 2024). Hence, addressing the research gaps in this sector is particularly valuable.

1.1 Problem Definition

The aim of this report is to develop an effective real estate portfolio optimisation model that maximises profits, minimises risks and outperforms the market indexes with a sharpe ratio of more than 2. A novel approach of applying Genetic Algorithms to search for outperforming Alpha formulas, as well as, three Machine Learning models, namely, Multiple Linear Regression (MLR), Neural Networks (NN) and Long-Short Term Memory (LSTM) are

used to construct models. A trading agent is also constructed to allocate funds to the portfolio based on the results from the models. The input data to the Machine Learning models will be extended beyond price-volume data to include fundamental stock data. The models will be trained on 10 years worth of real market data. Then, the performances of these approaches will be compared with traditional financial models and the market indexes.

1.2 Motivations

Interviews of traders and financial journalists have revealed that both technical and fundamental analysis are used for forecasting investments, for shorter and medium to long term investment horizons respectively (Oberlechner, 2001). Given that most investments in real estate are meant for the medium to long term, the lack of fundamental data inputs for REITs stock price predictions reveals a pressing need for fundamental data to be incorporated.

In the field of quantitative trading, quant firms seek to develop predictive algorithms for quantitative trading called "Alphas" that determine how capital is allocated to portfolios profitably (Tulchinsky, 2020). What began with investment experts manually constructing Alpha formulas has gradually been replaced with automated alpha mining techniques that employ Machine Learning algorithms (Wang et al., 2023). There appears to be an opportunity to apply the concept of quantitative finance Alphas for Real Estate portfolio allocation. In addition, researchers have been inspired by Charles Darwin's theory of evolution (Ruse, 1975) to develop a Genetic Algorithm and have applied it in finance (Aguilar-Rivera, Valenzuela-Rendón, & Rodríguez-Ortiz, 2015). Synthesising these two ideas, presents a new and innovative approach to use Genetic Algorithms to search for Alphas in the Real Estate sector.

Lastly, as most of past research is centered on price prediction, there is a

need to develop trade execution logic to allocate portfolios based on these predictions.

1.3   Major Contributions and Creativity

The first major contribution of this project is the analysis on REITs stock datasets with more than 150 features of technical and fundamental indicators. This expands the dimensionality of the input training data significantly beyond price-volume features to incorporate REIT company data from financial statements. The comprehensive dataset covers 150 out of the total worldwide population of 893 REITs and includes up to 20 years of historical data.

Creativity is demonstrated in the writing of a Genetic Search Algorithm that searches for outperforming REIT Alphas. It resulted in the second major contribution of a trading model with superior returns that significantly outperforms the market indexes.

The third contribution is the implementation of a trading agent or trade execution logic that acts on individual REIT price predictions to allocate funds to selected REITS in a portfolio.

Overall, this research paper has expanded the body of knowledge on the application of Alphas and Machine Learning models in Optimising Real Estate Portfolios.

I.   Background

CHAPTER 2

FINANCIAL TERMINOLOGY AND CONCEPTS

In order to better communicate ideas, the key financial terminologies and formulas will be explained. Particularly the difference between a stock, REIT and a portfolio, the difference between technical and fundamental analysis and the concept of Alphas. This builds the foundation for further discussions on the measures of portfolio performance, namely, profits/losses and the sharpe ratio. Other methods of analysis such as sentiment and macroeconomic analysis is also briefly discussed.

2.1 Key Terms

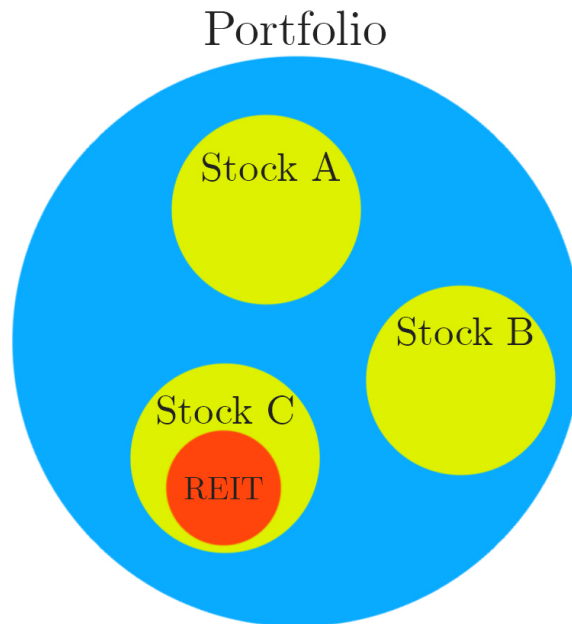The relationship between stocks, REITs and Portfolios can be illustrated in the figure below.

Figure 2.1: Stock vs REIT vs Portfolio

2.1.1 Stock

A stock also called a share, represents ownership over a company's assets. Each stock is represented in the market by a unique identifier called a ticker. For instance, Apple's ticker is "AAPL". The proportion ownership depends on how many shares are owned out of the total number of shares in the market and stocks of publicily listed companies can be purchased on the stock exchange. Over time, share prices can rise and fall, and owners can receive a share of the company's profits through dividends. The daily movement of the share prices over time will be represented in this paper by the formula:

- Prices: P is the price of a specific stock at a particular time

- Date: T in this paper is the daily date

- Ticker: i is the stock ticker for a specific stock

$$\overrightarrow{P}_{i,1:T} = \begin{bmatrix} p_{i,1} \\ p_{i,2} \\ \vdots \\ p_{i,T} \end{bmatrix} \in \mathbb{R}_{+}^{T}$$

2.1.2 Real Estate Investment Trusts (REITs)

A REIT company invests in properties. Typically these companies invest or own a mix of housing, commercial or industrial properties. Buying shares of REITs allow investors to own a share of the real estate that these REITS are managing and indirectly invest their monies in the real estate sector. Of course, each of the 893 REITs are different companies which own a wide range of different properties across the world. In this paper, a REIT is used to refer to the stock of a REIT company. A REIT is a type of stock.

2.1.3 Portfolio

A portfolio is a collection of stocks i.e a portfolio is a combination of more than one stock. For example, the portfolio in Figure 2.1 consists of 10% of stock A, 11% of stock B and 9% of stock C. The relative proportions of each stock in a portfolio can be represented by weights. This paper will represent a portfolio of stocks with the vector and matrix below:

- Weight: W is the percentage value of a ticker in the portfolio.

$$\vec{W}_{i,1:T}\vec{P}_{i,1:T} = \left[\vec{W}_{1,1:T}\vec{P}_{1,1:T},\ \vec{W}_{2,1:T}\vec{P}_{2,1:T},\ \ldots,\ \vec{W}_{i,1:T}\vec{P}_{2,1:T}\right]$$

$$= \begin{bmatrix} w_{1,1}\,p_{1,1}, & w_{2,1}\,p_{2,1}, & \ldots, & w_{i,1}\,p_{i,1} \\ w_{1,2}\,p_{1,2}, & w_{2,2}\,p_{2,2}, & \ldots, & w_{i,2}\,p_{i,2} \\ \vdots & \vdots & \ddots & \vdots \\ w_{1,T}\,p_{1,T}, & w_{2,T}\,p_{2,T}, & \ldots, & w_{i,T}\,p_{i,T} \end{bmatrix} \in \mathbb{R}_+^{MxT}$$

As with prices, the weights of each stock ticker can change daily if the value of the portfolio is redistributed. Depending on the weights assigned and what stocks are in the portfolio, the total portfolio value can change rise or fall. The decision of which stocks to select as part of a portfolio and how much of each stock to buy or hold at every point in time is the central idea of portfolio optimisation. This paper seeks to develop an effective model to select stocks and determine these weights.

2.1.4 Shorting

Conventionally, we need to own a certain stock before we can sell it. However, market makers also allow investors to borrow stocks to sell them first, with the contractual obligation for the investor to repurchase the stocks from the market at a later date to return to the market maker. This mechanism of selling first and buying back later is called shorting and is represented by a negative weight. This paper's analysis allows shorting.

2.2 Evaluating Investments with Data

In this subsection, various measures of investment performance and how they are applied in the paper are discussed. It covers the two main schools of thought in investment analysis, technical and fundamental analysis before discussing Alpha formulas.

2.2.1 Technical Analysis with Price Volume Data

Technical analysis is one method of evaluating investments in stocks and identifying prospective stocks to invest in. It focusses primarily on identifying short term trends in price and volume data. Table 2.1 provides the key features required for technical analysis.

Table 2.1: Daily Price Volume Data for ticker PLD

Date	Open	High	Low	Close	Volume
31/08/2023	125.36	125.91	123.89	124.2	3277600
01/09/2023	125.4	125.57	124.05	124.59	1591000
05/09/2023	125.36	125.91	123.89	122.05	2854100

- Open: The price of the stock at the start of a trading day

- High: The highest price of the stock in a trading day

- Low: The lowest price of the stock in a trading day

- Close: The price of the stock when the market closes for the day

- Volume: The total number of shares traded throughout the day

The features Open, High, Low and Close are derived from the Price of a stock at different points in the day and trends of these features can indicate future movements, while volume signals the strength of a price trend.

2.2.1.1 Profit and Loss (PnL)

Profits from the purchase of a single stock arises when the selling price of the stock is greater than the original purchase price. However, if the share price falls below the original purchase price then a loss is incurred. Profit or Loss is given by the formula:

$$Profit/Loss = CurrentPrice \times Quantity - PurchasePrice \times Quantity$$

Prices and Profits from Stock A, Purchasing on Different Days

(a) Day 1 Purchase

(b) Day 3 Purchase
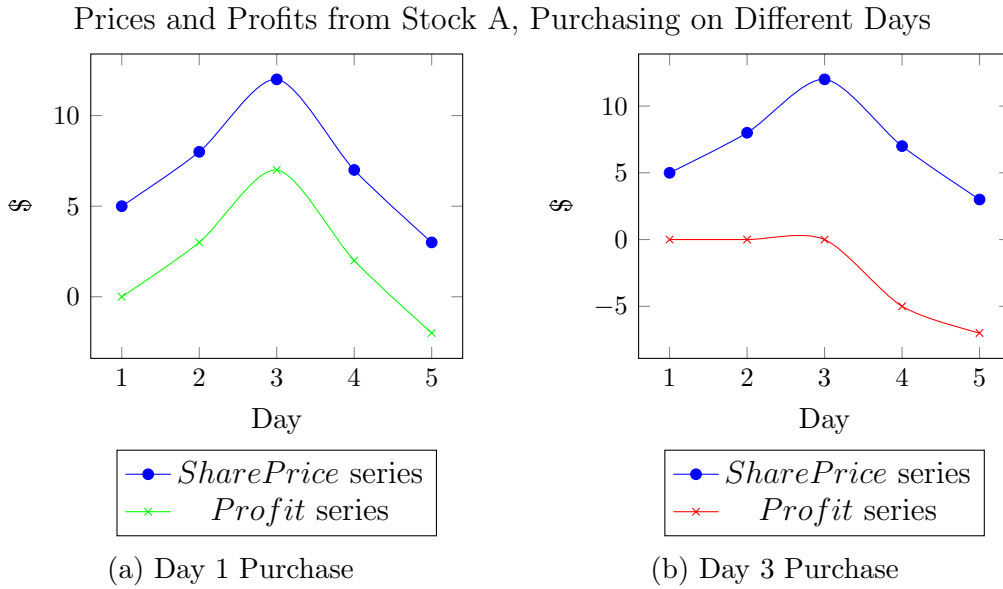
Figure 2.2

It's important to note that profits or losses are only realised when the closing transaction is made i.e. when the stock is sold. Figure 2.2 illustrates that buying the same stock on day 1 versus day 3 can have vastly different results.

When a stock is shorted, a rise in price will result in a loss because it becomes more expensive for the investor to repurchase the share. Hence, for shorting the Profit or Loss formula is negated:

$$Profit/Loss = -(CurrentPrice \times Quantity - PurchasePrice \times Quantity)$$

2.2.1.2   Risk and Volatility

2.2.1.3   Sharpe Ratio

2.2.2   Fundamental Analysis with Financial Statements

2.2.2.1   Income Statement

2.2.2.2   Balance Sheet

2.2.2.3   Cash Flow Statement

2.2.3   Alpha Formulas

2.2.4   Other Methods of Analyses

CHAPTER 3

LITERATURE REVIEW OF PORTFOLIO OPTIMISATION TECHNIQUES

3.1 Optimal Portfolio Theory

3.2 Traditional Time-Series Analysis

3.3 Machine Learning Techniques

3.4 Comparison of Techniques

3.4.1 Scope

3.4.2 Profitability

3.4.3 Predictive Accuracy

II. Innovation

CHAPTER 4

DATASETS

4.1 Extended Fundamental Data Features

4.2 Feature Selection with Decision Trees
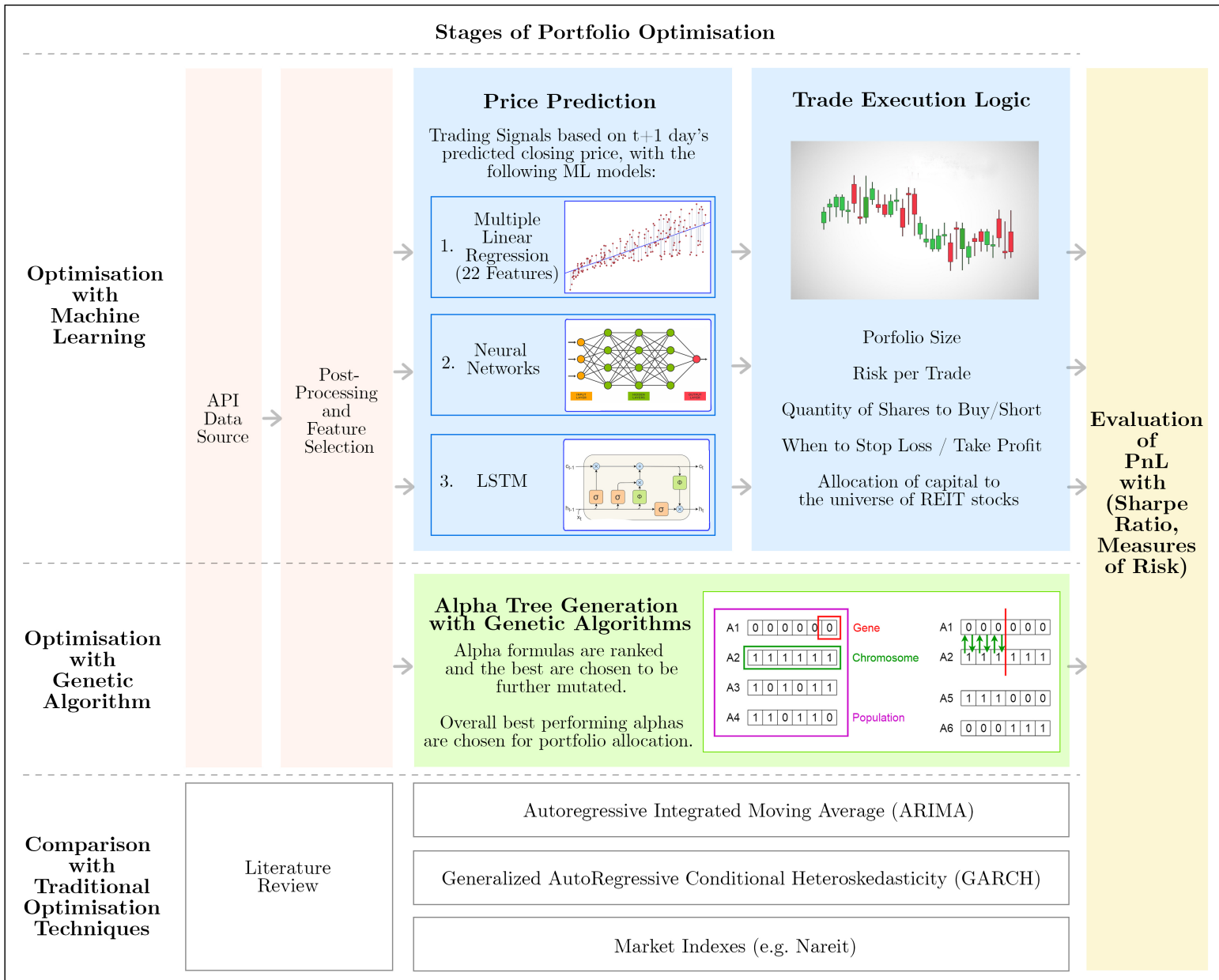
CHAPTER 5

METHODOLOGY

Figure 5.1: Write some caption here

5.1 Machine Learning for REITs Portfolio Optimisation

5.1.1 MLR / NN / LSTM Predictions with Extended Features

5.1.2 Trade Execution Logic

5.1.3 Performance Evaluation

5.2 Genetic Algorithm Search for Outperforming Alphas

5.2.1 Alpha Tree

5.2.2 Application of Genetic Algorithms to Alpha Trees

5.2.2.1 Objective Function

5.2.2.2 Selection

5.2.2.3 Crossover

5.2.2.4 Mutation

5.2.3 Portfolio Allocation Using Alpha

5.2.4 Performance Evaluation

III. Experiments

CHAPTER 6

POST-PROCESSED FINANCIAL DATASETS

CHAPTER 7

MACHINE LEARNING RESULTS

7.1 Evaluating Stock Price Predictions

7.1.1 Multiple Linear Regression (MLR)

7.1.2 Neural Networks (NN)

7.1.3 Long-Short Term Memory (LSTM)

7.2 Trade Execution Results with Different Parameters

7.3 Overall Evaluation of Performances

CHAPTER 8

ALPHA TREE SEARCH RESULTS

8.1 Alphas Generated

8.1.1 Initial Set

8.1.2 Intermediate Alphas

8.1.3 Best Performing Alphas

8.2 Portfolio Allocation Results with Best Performing Alphas

8.3 Overall Evaluation of Performance

CHAPTER 9

CONCLUSION

9.1 Benchmarking Against Index Funds

9.2 Comparing Results with Literature Review

9.3 Key Findings

9.4 Major Contribution and Creativity

9.5 Future Work

9.5.1 More Operators and Features for Alphas

Figure 9.1: The Universe

REFERENCES

Aguilar-Rivera, A., Valenzuela-Rendón, M., & Rodríguez-Ortiz, J. (2015, November). Genetic algorithms and Darwinian approaches in financial applications: A survey. *Expert Systems with Applications*, *42*, 7684–7697. doi: 10.1016/j.eswa.2015.06.001

Habbab, F. Z., & Kampouridis, M. (2024). An in-depth investigation of five machine learning algorithms for optimizing mixed-asset portfolios including REITs. *Expert Systems with Applications*, *235*, 121102. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0957417423016044` doi: https://doi.org/10.1016/j.eswa.2023.121102

Nareit. (2024). *Global Real Estate Investment.* Retrieved 2024-04-08, from `https://www.reit.com/investing/global-real-estate-investment`

Oberlechner, T. (2001, January). Importance of technical and fundamental analysis in the European foreign exchange market. *International Journal of Finance & Economics*, *6*(1), 81–93. Retrieved 2024-04-08, from `https://onlinelibrary.wiley.com/doi/10.1002/ijfe.145` doi: 10.1002/ijfe.145

Obthong, M., Tantisantiwong, N., Jeamwatthanachai, W., & Wills, G. (2020, May). A survey on machine learning for stock price prediction: algorithms and techniques. In (pp. 63–71). Retrieved 2024-04-08, from `https://eprints.soton.ac.uk/437785/` (Num Pages: 9) doi: 10.5220/0009340700630071

Ruse, M. (1975). Charles Darwin's Theory of Evolution: An Analysis. *Journal of the History of Biology*, *8*(2), 219–241. Retrieved 2024-04-08, from `https://www.jstor.org/stable/4330635` (Publisher: Springer)

Sheth, D., & Shah, M. (2023, February). Predicting stock market using machine learning: best and accurate way to know future

stock prices. *International Journal of System Assurance Engineering and Management*, *14*(1), 1–18. Retrieved 2024-04-07, from `https://link.springer.com/10.1007/s13198-022-01811-1` doi: 10.1007/s13198-022-01811-1

Tostevin, P., & Rushton, C. (2023, September). *Total Value of Global Real Estate: Property remains the world's biggest store of wealth.* Retrieved 2024-04-08, from `https://www.savills.com/impacts/market-trends/the-total-value-of-global-rea` (Section: Market trends)

Tulchinsky, I. (Ed.). (2020). *Finding alphas: a quantitative approach to building trading strategies* (Second edition ed.). Chichester, West Sussex: Wiley.

Wang, S., Yuan, H., Zhou, L., Ni, L. M., Shum, H.-Y., & Guo, J. (2023, July). *Alpha-GPT: Human-AI Interactive Alpha Mining for Quantitative Investment.* arXiv. Retrieved 2024-04-07, from `http://arxiv.org/abs/2308.00016` (arXiv:2308.00016 [cs, q-fin])