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

I would like to thank Professor Liu Lili for her guidance, support and encouragement throughout the course of this project. Working with Professor Lili has been a great learning experience, getting to learn much more about machine learning and its applications to finance in solving some challenges faced by industry practitioners.

Also many thanks to Professor Chin Wei Ngan for taking the time out to assess this project.

CONTENTS

CHAPTER 1

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 [18]. The success of Machine Learning has led to numerous research papers applying a myriad of Machine Learning techniques to predict stock prices [15]. 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 [9]. 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 [19], with approximately 893 listed Real Estate Investment Trusts (REITs) [13]. 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 [14]. 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 [20]. What began with investment experts manually constructing Alpha formulas has gradually been replaced with automated alpha mining techniques that employ Machine Learning algorithms [21]. 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 [16] to develop a Genetic Algorithm and have applied it in finance [1]. 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 macro-economic analysis is also briefly discussed.

2.1 Key Terms

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

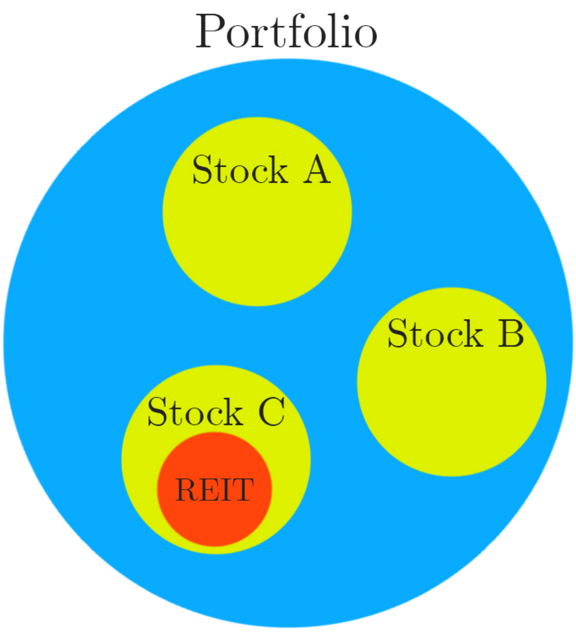


Figure 2.1: Stock vs REIT vs Portfolio

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

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

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.

$$\overrightarrow{W}_{\mathrm{i,1:T}}\overrightarrow{P}_{\mathrm{i,1:T}} = \left[\overrightarrow{W}_{\mathrm{1,1:T}}\overrightarrow{P}_{\mathrm{1,1:T}},\ \overrightarrow{W}_{\mathrm{2,1:T}}\overrightarrow{P}_{\mathrm{2,1:T}},\ \dots,\ \overrightarrow{W}_{\mathrm{i,1:T}}\overrightarrow{P}_{\mathrm{2,1:T}}\right]$$

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

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

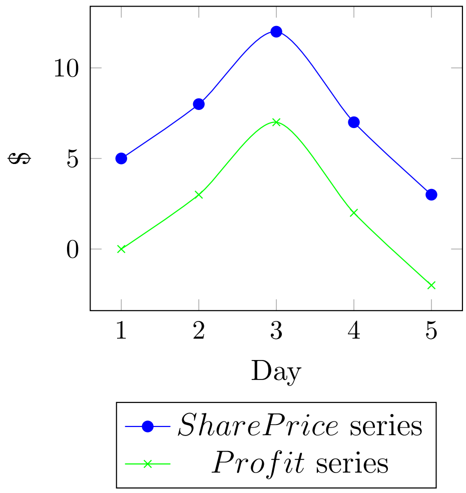
The features Open, High, Low and Close are derived from the Price of a stock at different points in the day and trends of these features can indicate future movements, while volume signals the strength of a price trend.

2.2.1.1 Profit and Loss (PnL)

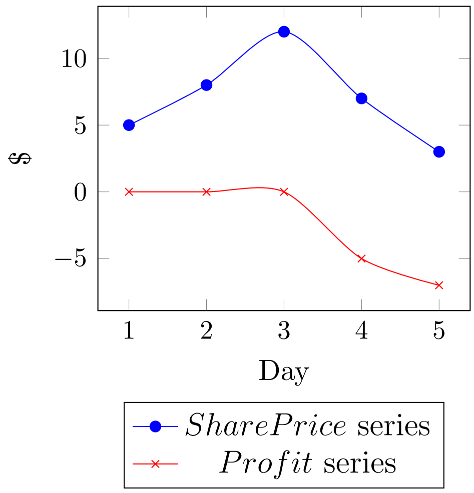
Profits from the purchase of a single stock arises when the selling price of the stock is greater than the original purchase price. However, if the share price falls below the original purchase price then a loss is incurred. Profit or Loss is given by the formula:

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

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.

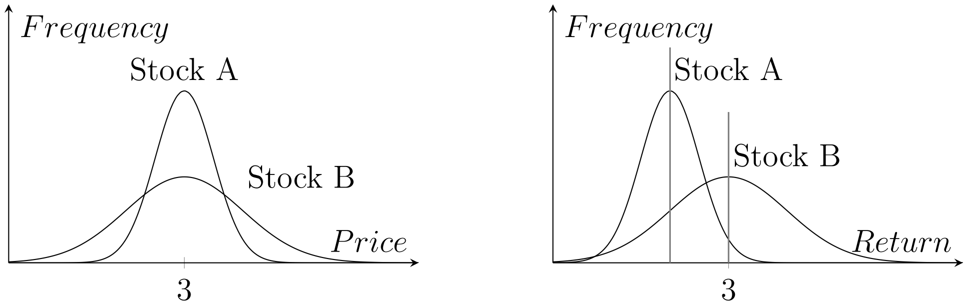
When a stock is shorted, a rise in price will result in a loss because it becomes more expensive for the investor to repurchase the share. Hence, for shorting the Profit or Loss formula is negated:

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

2.2.1.2 Risk and Volatility

Volatility is a statistical measure of how varied share price and returns on an investment are over a period of time. It is a proxy of the risk of an investment, the likelihood of an investment performing worse than expected.

Prices and Profits from Stock A, Purchasing on Different Days



(a) Price Risk (b) Return Risk

Figure 2.3

Figure 2.3a shows the distribution of prices of Stocks A and B over a period of time. Stock A has a lower standard deviation σ_A than Stock B σ_B, lower volatility and lower price risk. The returns on both stocks can also be plotted on a similiar distribution as shown in Figure 2.3b. Because this paper seeks to minimise risk, the variance of prices and returns need to be considered.

2.2.1.3 Sharpe Ratio

Deciding between stocks or portfolios with different expected returns and risk like in Figure 2.3b where Stock B has higher returns but also higher risk than Stock A, requires a formula that evaluates these two measures together. The sharpe ratio combines both measures of profit/loss and risk into a single formula as follows:

$$SharpeRatio = \frac{R_{\mathrm{p}} - R_{\mathrm{f}}}{\sigma_{\mathrm{p}}}$$

- R_p: The annualised return of portfolio

- R_f: The risk free rate
- σ_p: The standard deviation of portfolio return

The risk free rate R_f is the return from a zero risk investment such as government bonds. Hence $R_p - R_f$ gives us the excess returns of the portfolio investment compared to a zero risk investment. σ_p measures volatility as covered in section 2.2.1.2.

In this paper, the sharpe ratio is used as the overall measure of performance of the portfolio optimisation models and is incorporated in the various objective functions.

2.2.2 Fundamental Analysis with Financial Statements

Fundamental analysis is a second method of investment analysis that is widely used by finance practitioners but less often applied by Machine Learning researchers. It involves examining the assets and incomes of the companies behind the different stock tickers by looking at their financial statements to evaluate the intrinsic value of a stock.

Balance Sheet	31-Dec
Assets	5000
Liabilities	3000
Equity	2000

Income Statement	31-Dec
Revenue	8000
Expenses	3000
Profit	5000

Cash Flow Statement	31-Dec
Operating	5000
Investing	1000
Financing	1000

Figure 2.4: Highly Abbreviated 3 Main Statements

Every quarter, every company with a stock ticker will publish its financial statements which consists of the Balance Sheet, Income Statement and Cash Flow Statement. Figure 2.4 is a highly abbreviated example, the real financial statements breaks down each category into hundreds of subcategories which can be used as features for Machine Learning.

The predictive value of each statement is as follows:

- Income Statement: Profitability of the company's business and the rate of growth of its value
- Balance Sheet: The asset value and financial stability of the company
- Cash Flow Statement: How cash is generated by the company

Incorporating both technical and fundamental data into machine learning models can provide a more holistic picture to the analysis and better inform these models with an extended set of short, medium and long term features.

2.2.3 Alpha Formulas

According to quant researcher [20], an Alpha is a formula that determines how much capital to allocate to each stock every day. A good alpha allocates capital in a way that results in a high sharpe ratio for the entire portfolio.

Here are two examples of alpha formulas used in real-life trading from a paper by [11]:

Alpha 1: $(((high * low)^{0.5}) - vwap)$

Alpha 2: $((0.25 < (((delay(close, 20) - delay(close, 10))/10) - ((delay(close, 10) - close)/10)))?(-1*1) : (((((delay(close, 20) - delay(close, 10))/10) - ((delay(close, 10) - close)/10)) < 0)?1 : ((-1 * 1) * (close - delay(close, 1)))))$

The first formula Alpha 1 is relatively simple while Alpha 2 is complex.

Figure 2.5 demonstrates how Alpha 1 is applied to portfolio allocation in a stock portfolio of 6 tickers. Every day, each stock ticker will have its own Alpha value in the first green column of Figure, derived by passing the required features as inputs to the alpha formula (e.g. in this case, Alpha 1 requires, High, Low and Volume Weighted Average Price (Vwap)).

		Features			Alpha	Ranking Alpha Values to Determine Allocation						Allocation	DoD Price	Profit
Date	Ticker	High	Low	Vwap	Alpha Value	Rank	Rank out of 1	Centered 0	Abs	Sum Rank	Normalised	(20Mil)	Change %	/Loss
10/09/2024	SPG	5	4	4.5	-0.028	3	0.4	-0.1	0.1	1.8	-0.06	-1.11	-0.01	0.01
10/09/2024	PSA	8	3	5	-0.101	2	0.2	-0.3	0.3	1.8	-0.17	-3.33	-0.02	0.07
10/09/2024	PLD	10	2	3	1.472	6	1.0	0.5	0.5	1.8	0.28	5.56	0.04	0.22
10/09/2024	DLR	4	3	3	0.464	5	0.8	0.3	0.3	1.8	0.06	1.11	-0.02	-0.02
10/09/2024	AMT	6	6	6	0.000	4	0.6	0.1	0.1	1.8	0.17	3.33	0.01	0.03
10/09/2024	EQR	5	1	3	-0.764	1	0.0	-0.5	0.5	1.8	-0.28	-5.56	0.02	-0.11
													Total PnL for the Day	**0.20**

Figure 2.5: Application of Alpha 1 to Porfolio Allocation

In the blue section of 2.5, the Alpha values are then ranked, centered and normalised to determine the relative proportion of the portfolio to allocate to each stock ticker. In a portfolio of 6 tickers, the largest 3 ranks will be bought and the lowest 3 ranks will be shorted to differing degrees.

In the pink column, 20 million in capital is allocated by multiplying the total capital of 20 million by the normalised rank. The allocation amount mulitplied by the day-on-day percentage price change (Dod Price Change %) for each ticker will then give us the Profit/Loss for each stock ticker for the day. The (Dod Price Change %) formula is as follows:

$$(Dod\ Price\ Change\ \%) = \frac{Pave_{\mathrm{t+1}} - Pave_{\mathrm{t}}}{|Pave_{\mathrm{t}}|}$$

- $Pave_{t+1}$: Average of Tommorrow's Open and Close Price
- $Pave_t$: Average of Today's Open and Close Price

The total sum of the Profit/Loss for every ticker is the PnL from the day's investment as all positions are closed and the capital is reallocated the following day.

This paper seeks to apply Alpha formulas to allocate REIT portfolios, and go further by developing a Genetic Algorithm inspired by Charles Darwin's theory of evolution to search for strong Alphas to Optimise Real Estate Portfolios.

2.2.4 Other Methods of Analyses

Sentiment Analysis: Given that share prices in the short run are largely driven by investor sentiment, some researchers have employed Natural Language Processing (NLP) techniques to process trade forum discussions and financial news for trade decision making resulting in good sharpe ratios [2].

Macro-economic Analysis:

CHAPTER 3

LITERATURE REVIEW OF PORTFOLIO OPTIMISATION TECHNIQUES

3.1 Optimal Portfolio Theory

3.2 Comparison of Traditional and Machine Learning Techniques

Methods	Data Inputs	Objectives	Scope for REITs	Strengths	Weaknesses	References
Autoregressive Integrated Moving Average (ARIMA)	Daily Price-Volume Data	Stock Price Prediction and Clustering	Stock Price Prediction	+Suitable for short-term time-series analysis + Widely used in the field of finance for prediction + More explanable than complex ML models	-Unsatisfactory for long term prediction -Higher RMSE than ML models	[3], [8], [15]
Generalised AutoRegressive Conditional Heteroskedasticity (GARCH)	Daily Commodities Price-Volume data	Stock Volatility Prediction	Stock Volatility Prediction	+Outperforms ARIMA for price prediction and volatility forcasting	- Model assumes volatility can be predicted based on past returns, but volatility in reality is very unpredicatable	[7], [12], [22]
Multiple Linear Regression (MLR)	Daily Price-Volume Data	Stock Price Prediction	Stock Price Prediction	+Relatively low RMSE error for stock price prediction + Relatively easy to implement	- Only considers input features and not historical price data	[15], [17]
Recurrent Neural Networks (RNN)	Daily Price-Volume Data	Stock Price Prediction		+Relatively low RMSE error for stock price prediction + Remembers historical stock prices	- Susceptible to vanishing and exploding gradient problem	[6]

3.3 Comparison of Traditional and Machine Learning Techniques (Continued)

Methods	Data Inputs	Objectives	Scope for REITs	Strengths	Weaknesses	References
Long-Short Term Memory (LSTM)	Daily Price-Volume Data	Stock Price Prediction	Stock Price Prediction	+ One of the lowest RMSE error for financial time series prediction + Remembers historical stock prices which are good indicators of future prices + Solves the vanishing and exploding gradient problem	- Mixed results when predicting REIT stock prices	[4], [8], [15]
Random Forest (RF)	Daily Price-Volume Data, Fundamental Data	Stock Price Prediction		+ One of the few papers to incorporate fundamental data into machines learning models	- Relatively low sharpe ratio compared to market indexes	[5], [10]
Genetic Algorithms	Daily Price-Volume Data	Stock Price Prediction and Clustering		+ Suitable for problems with a large search space + Works well with large time-series data		[15]

3.3.1 Predictive Accuracy

A recent paper aimed at predicting REIT share prices through 5 machine learning algorithms and 3 traditional benchmarks, which includes many of the methods in Section 3.2 above, found that machine learning models (MLR, LSTM) appear more accurate than traditional benchmarks (ARIMA). The algorithms were trained solely on daily price data, and Figure 3.1 shows the RMSE results for out-of sample predictions over different time periods (30, 60, 90, 120 and 150 days).

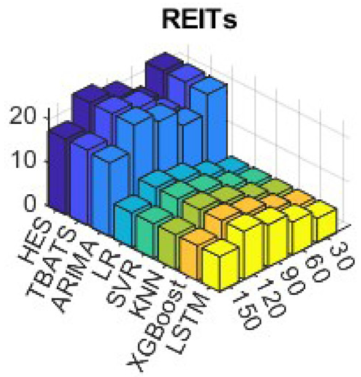


Figure 3.1: RMSE results for REITs using various ML Models (Source: [9])

3.3.2 Gaps in Past Research to be Addressed

From the literature review in Section 3.3, it appears that there are the following gaps that can be further explored in this paper:

1. From the "Data Inputs" column, it appears most papers use price-volume (technical) data without fundamental data to make predictions.

2. From the "Objectives" column, most papers are focussed on predicting stock prices and rarely address portfolio allocation or optimisation (i.e. to use these predictions to inform buy/sell decision of each stock in a portfolio or to find the optimal weights of each stock in a portfolio to maximise the sharpe ratio).

3. There are no papers about searching for Alphas for REIT portfolios
4. Even though genetic algorithms have been used to predict stock prices, there appears to be an opportunity to apply it to the search of outperforming REIT Alphas.

II. Innovation

CHAPTER 4

DATASETS

4.1 Extended Fundamental Data Features

While many papers have obtained price-volume data from Yahoo Finance's API (yfinance), this data only contains 10 or so price-volume indicators listed in Appendix A. This paper goes further by obtaining a comprehensive list of 161 fundamental features from FinancialModelingPrep's (FMP) API[1], listed in Appendix B.

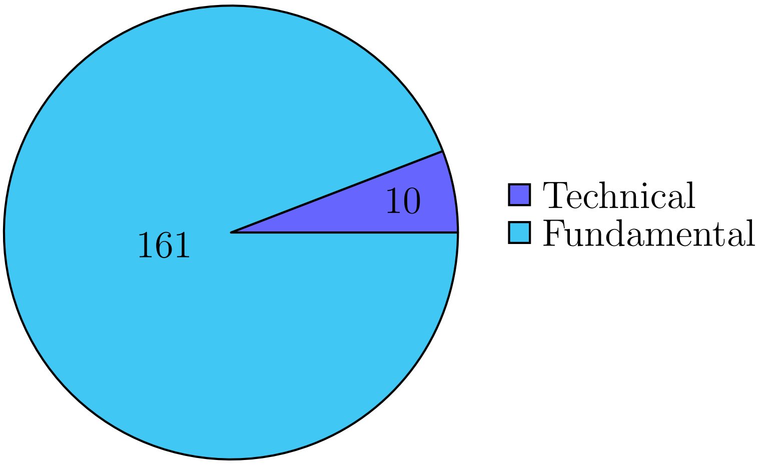


Figure 4.1: Count of Fundamental vs Technical Features

A script was written to combine the data from various api endpoints and backfill any missing values. Because price-volume data is available daily while fundamental data is generally available quarterly, a function is written to backfill the fundamental data such that daily numbers are available.

[1] Although the datasource is proprietory and the author does not have permission to distribute the raw data, it can be easily purchased for a small fee at https://site.financialmodelingprep.com/.

4.2 Feature Exploration with Decision Trees

Running a 25 layer Decision Tree on all 171 features reveals the strongest predictors for the next day's closing price. The top 4 layers are in Figure 4.2:

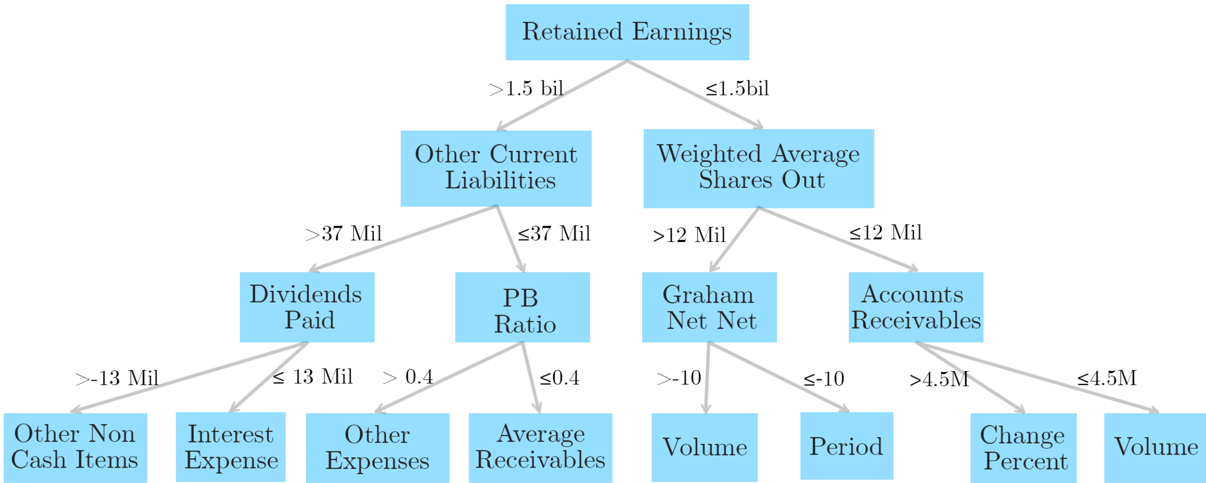


Figure 4.2: Top 4 Layers of Decision Tree To Predict T+1 Close

Figure 4.3 also shows the weights of the top 10 features in the Decision Tree. From both figures, it appears many of the balance sheet, income and cash flow statement features are indeed strong predictors of share price, coupled with technical features like Volume and Percentage Change in Share Price.

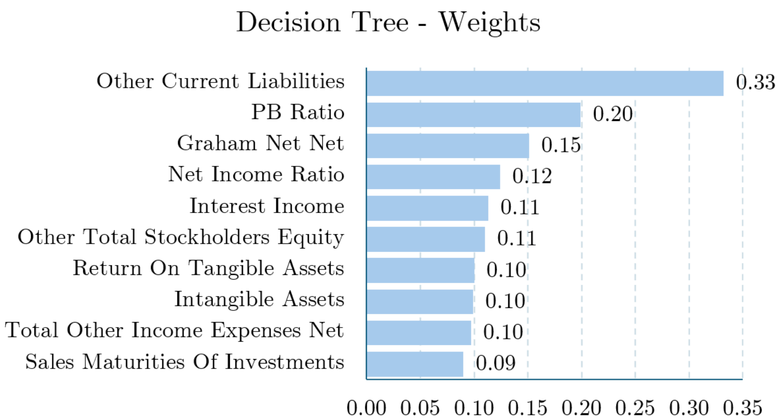


Figure 4.3: Top 10 Feature Weights from Decision Tree

4.3 Feature Selection

From the decision tree, it also appears that many of the top features can be encapsulated by other features. For example, Other Current Liabilities is a component of Total Current Liabilities, Expenses are part of Net Income and Cash Items are a part of Cash From Activities. Given that the tree was generated from only one stock ticker, to make the feature selection generalisable to all 150 tickers, the results are contextualised with financial knowledge to derive the final 22 Selected Features presented in Figure 4.4. The financial significance of each feature is explained in the column "Domain Importance".

Feature	**Source**	**Domain Importance**	**Feature**	**Source**	**Domain Importance**
open	Price Volume Data	Initial market sentiment and potential direction	Total Equity	Balance Sheet	Shareholders' stake in the company
high	Price Volume Data	Maximum bullish sentiment and potential resistance levels	Net Cash Provided By Operating Activities	Cash Flow Statement	Cash generated from core business activities
low	Price Volume Data	Extent of bearish sentiment and potential support levels	Net Cash Used For Investing Activites	Cash Flow Statement	Company's Investment in long-term growth
close	Price Volume Data	Final consensus value of the stock for the day	Net Cash Used Provided By Financing Activities	Cash Flow Statement	Financing activities' impact on cash flow
volume	Price Volume Data	Level of activity in a stock. Signals the strength of a price movement	Net Change In Cash	Financial Analysis Ratios	Overall cash flow health and liquidity
Volume Weighted Average Price (VWAP)	Price Volume Data	Average price paid per share within a trading period	Capital Expenditure	Financial Analysis Ratios	Investment in the company's future growth
Net Income	Income Statement	Company's Profitability, affecting investor perception	Cash Per Share	Financial Analysis Ratios	Company's liquidity
Total Assets	Balance Sheet	Size of the company and its resource base	Current Ratio	Financial Analysis Ratios	Ability to pay off its short-term liabilities with short-term assets
Total Current Assets	Balance Sheet	Short-term financial health and liquidity.	Return On Tangible Assets	Financial Analysis Ratios	Effectiveness of using physical assets to generate profits
Total Debt	Balance Sheet	Signal financial risk	Earnings Yield	Financial Analysis Ratios	Earnings per share over the share price
Total Current Liabilities	Balance Sheet	Short-term obligations and ability to meet them			

Figure 4.4: 22 Input Features Selected for Models

In the preprocessing step, up to 20 years of historical data with these 22 features are retrieved for 150 stock tickers, and appended to a dictionary which stores all the required data for each ticker.

CHAPTER 5

METHODOLOGY

This paper takes two approaches to portfolio optimisation. 1) Optimisation with Machine Learning and 2) Alpha Tree Search using Genetic Algorithms.

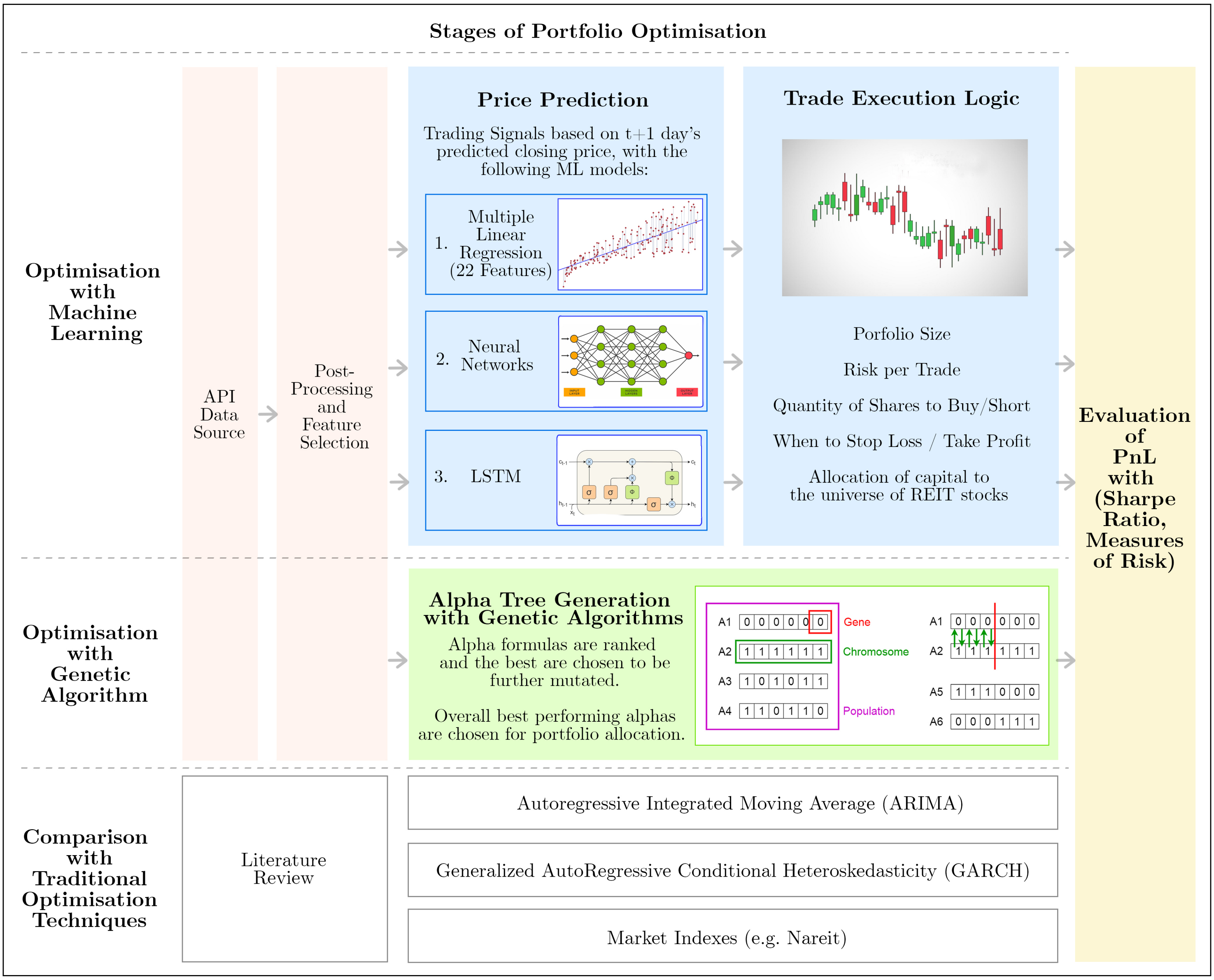


Figure 5.1: Overall Workflow For This Paper

5.1 Machine Learning for REITs Portfolio Optimisation

With reference to Figure 5.1, the first approach of machine learning for optimisation consists of two parts, A) Price Prediction and B) Trade Execution.

5.1.1 MLR / NN / LSTM Predictions with Extended Features

At the pre-processing stage of all three models, data for each ticker is split into the training and test sets with a 80:20 ratio. The close price for each row is also shifted one day forward so that all 21 features can work on predicting the next day's close.
The implementation for Multiple Linear Regression is as follows:

Listing 5.1: Multiple Linear Regression Pseudocode

```
for each ticker in dictionary:
    # Training Phase
    X_norm = normalised matrix of 21 features (training set)
    Y_norm = normalised array of T+1 Close Price (training set)

    model = nn.Linear(21,1) # Single Layer LR
    criterion = nn.MSELoss() # MSE Loss Function
    optimizer = optim.SGD(model.parameters(), lr=0.001)

    for epoch in range(10000):
        optimizer.zero_grad()
        Y_pred = model(X_norm)
        loss = criterion(Y_pred, Y_norm)
        loss.backward()
        optimizer.step()

    # Testing Phase
    x_test_norm = normalised matrix of 21 features (test set)
    y_test_norm = normalised array of T+1 Close Price (test set)
```

```
model.eval()
with torch.no_grad():
  predicted = denormalise(x_test_norm)

x_test_dates = list of dates from x_test_norm
df_predicted = pd.DataFrame({
    'Date':x_test_dates,
    'Predicted': pd.DataFrame(predicted)[0].to_list(),
    'Actual': test_set['close'].to_list()
    })
df_predicted['Actual␣(t-1)'] = df_predicted['Actual'].shift(1)
df_predicted['%␣Change␣Predicted'] = (df_predicted['Predicted'] -
        df_predicted['Actual␣(t-1)'])/df_predicted['Actual␣(t-1)']
df_predicted = df_predicted[['Date','Predicted','Actual',
                            'Actual␣(t-1)','%␣Change␣Predicted']]
df_predicted.to_csv('predictions/'+filename+'_pred.csv')
```

5.1.2 Trade Execution Logic

The latter extends past research by implementing the trading logic that decides how to act on predicted prices in order to optimise portfolios.

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

BIBLIOGRAPHY

[1] Anton Aguilar-Rivera, Manuel Valenzuela-Rendón, and J. Rodríguez-Ortiz. Genetic algorithms and Darwinian approaches in financial applications: A survey. *Expert Systems with Applications*, 42:7684–7697, November 2015.

[2] Alexandria. Unlocking Japanese Market Sentiment with Nikkei FTRI, 2023.

[3] Adebiyi A. Ariyo, Adewumi O. Adewumi, and Charles K. Ayo. Stock Price Prediction Using the ARIMA Model. In *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, pages 106–112, March 2014.

[4] Birger Axelsson and Han-Suck Song. Univariate Forecasting for REITs with Deep Learning: A Comparative Analysis with an ARIMA Model, 2023.

[5] Kai Cao. Fundamental Analysis via Machine Learning. 2021.

[6] Polash Dey, Emam Hossain, Md Ishtiaque Hossain, Mohammed Armanuzzaman Chowdhury, Md Shariful Alam, Mohammad Shahadat Hossain, and Karl Andersson. Comparative Analysis of Recurrent Neural Networks in Stock Price Prediction for Different Frequency Domains. *Algorithms*, 14(8):251, August 2021. Number: 8 Publisher: Multidisciplinary Digital Publishing Institute.

[7] Piotr Fiszeder and Tin Fah Chung. What are the major limitations of Econometric Models like GARCH and TARCH?, 2020.

[8] Fatim Z. Habbab and Michael Kampouridis. Machine Learning for Real Estate Time Series Prediction. Sheffield, UK, July 2022.

[9] Fatim Z. Habbab and Michael Kampouridis. An in-depth investigation of five machine learning algorithms for optimizing mixed-asset portfolios including REITs. *Expert Systems with Applications*, 235:121102, 2024.

[10] Yuxuan Huang, Luiz Fernando Capretz, and Danny Ho. Machine Learning for Stock Prediction Based on Fundamental Analysis. In *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 01–10, Orlando, FL, USA, December 2021. IEEE.

[11] Zura Kakushadze. 101 Formulaic Alphas, March 2016. arXiv:1601.00991 [q-fin].

[12] Achal Lama, Girish K. Jha, Ranjit K. Paul, and Bishal Gurung. Modelling and Forecasting of Price Volatility: An Application of GARCH and EGARCH Models. *Agricultural Economics Research Review*, 28(1):73, 2015.

[13] Nareit. Global Real Estate Investment, 2024.

[14] Thomas Oberlechner. Importance of technical and fundamental analysis in the European foreign exchange market. *International Journal of Finance & Economics*, 6(1):81–93, January 2001.

[15] Mehtabhorn Obthong, Nongnuch Tantisantiwong, Watthanasak Jeamwatthanachai, and Gary Wills. A survey on machine learning for stock price prediction: algorithms and techniques. pages 63–71, May 2020. Num Pages: 9.

[16] Michael Ruse. Charles Darwin's Theory of Evolution: An Analysis. *Journal of the History of Biology*, 8(2):219–241, 1975. Publisher: Springer.

[17] Shruti Shakhla, Bhavya Shah, Niket Shah, Vyom Unadkat, and Pratik Kanani. Stock Price Trend Prediction Using Multiple Linear Regression. May 2020.

[18] Dhruhi Sheth and Manan Shah. 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, February 2023.

[19] Paul Tostevin and Charlotte Rushton. Total Value of Global Real Estate: Property remains the world's biggest store of wealth, September 2023. Section: Market trends.

[20] Igor Tulchinsky, editor. *Finding Alphas: A Quantitative Approach to Building Trading Strategies*. Wiley, 1 edition, September 2019.

[21] Saizhuo Wang, Hang Yuan, Leon Zhou, Lionel M. Ni, Heung-Yeung Shum, and Jian Guo. Alpha-GPT: Human-AI Interactive Alpha Mining for Quantitative Investment, July 2023. arXiv:2308.00016 [cs, q-fin].

[22] Ya-Ping Yuan, Jiong Sun, and Hong-Kun Zhang. Garch Models in Value-At-Risk Estimation for REIT. 2017.