5-15-2023

Customised Investment Optimization Using Genetic Algorithms

Maythaa Al Ali
mka9146@rit.edu

Shaikha Al Dossari
sra7107@rit.edu

RIT

Customised Investment Optimization Using Genetic Algorithms

By

Maythaa Al Ali & Shaikha Al Dossari

A Capstone Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Science in Professional Studies:

Department of Data Analytics

Rochester Institute of Technology

RIT Dubai

May 15, 2023

RIT

Master of Science in

Professional Studies:Department of Data Analytics

Capstone Approval

Student Name: Shaikha Al Dosari & Maythaa AL Ali

Capstone Title: Customised Investment Optimization Using Genetic Algorithms

Graduate Capstone Committee:

Name	**Prof. Sanjay Modak**	**Designation**		**Date**	**15.May.2023**
Name	**Prof.Khalil Al Huassaeni**	**Designation**		**Date**	**15.May.2023**

Acknowledgments

In light of the successful completion of this capstone project, we would like to extend our sincere gratitude to all those who contributed to its success.

As a team, we would like to sincerely thank Prof. Khalil Al Hussaeni for his valuable guidance, support, and mentorship throughout the entire process. Our research has been guided by his expertise and insights, and we've been able to overcome a variety of challenges because of his insights and expertise.

Additionally, we would like to thank the faculty members of Data Science at Rochester Institute of Technology for their constant support and encouragement. We have been able to refine our ideas and improve our work as a result of their constructive criticism and feedback. We also appreciate our project dedication and hard work. Without this collaboration, this project would not have been possible.

Finally, we are grateful for the unwavering support and encouragement we have received from our families and friends throughout our academic journey. We have received constant inspiration from their love, motivation, and belief in us. Your support and contributions are greatly appreciated! Thank you all for your contribution and support.

Abstract

Portfolio selection is an important part of fund management as it contributes to investors' economic growth. Investing is investing money to obtain an additional or specific advantage over money. Investment involves not only profit (return), but also risk that the investor bears. The higher the return an investor expects, the higher the risk the investor takes. Proper portfolio construction can minimize the level of risk to the expected value of an individual stock portfolio. Equity portfolio optimization, therefore, plays an important role in setting an investor's investment portfolio strategy. In recent decades, there have been great advances in financial mathematics.

By harnessing the increasingly powerful computational power, these problems have been addressed in new ways, developing new algorithms to trade, model, and forecast mostly automatedly. With the recent increasing adoption of machine learning approaches, genetic algorithms have emerged as one of the most widely used stochastic optimization approaches for solving complex optimization problems. The proposed approach aims to develop a machine-learning solution to simplify investment decision-making by rapidly generating optimal investment portfolios. The suggested study will train a genetic algorithm to help investors select assets with the greatest return. The proposed solution is a Genetic Algorithm-based model that: selects K assets from the universally available assets, includes them in the portfolio, and the capital weights to invest in to minimize risk and maximize portfolio returns / Allocate shares.

The study exploited 391,108 assets to expound on the utilization of genetic algorithms in constructing optimal investment portfolio and experimented two optimization approaches, that is single objective case and multi objective approach. The maximization of Sharpe ratio resulted to a high-risk portfolio with risk rising to 49.24% during asset selection step and 787.31% capital allocation while the multi objective approach yield was dependent on the objective weights. For risk seeking investors who would rather miss their risk goal than missing their return goal, the strategy allocates high amounts to risky assets. As the investor becomes more risk averse, the strategy allocates high capital proportion to less risk assets.

Keywords: *Portfolio, Investment, Genetic Algorithm, Machine Learning.*

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Background

In portfolio optimization, the optimal portfolio consists of a weight corresponding to each asset on the portfolio; these weights represent the proportion of investment capital an investor should allocate to the asset in order to optimize their object, which frequently is to minimize risk or maximize profit. Investor preferences and rules of nature constrain the optimal choice.

Research by Markowitz [1] introduced a portfolio optimization approach named portfolio selecting theory which generally focuses on allocating these weights based on the trend off between mean return and variance, later the theory was expanded by Sharpe [2] who introduced valuation of assets as a function of systematic risk. The two theories build the basis of modern portfolio theory (MPT). Many research works proposed different optimization methods, but most of them were limited to a single objective.

In the 1960s and 1970s, Genetic algorithms were introduced by John Holand and his team which imitates Darwin's theory of "evolution by way of natural solutions" [3]. The approach works by first creating a population of a feasible solution to an optimization problem, which acts as parents to the next step. At each step, population members are assigned a fitness score that determines their chances of surviving to the next step. The fittest solutions survive to the next population, crossovers are created by combining vectors from the parent population, and mutation is done by modifying some members of the previous answer. At each iteration, children's solutions are created by the three methods, i.e., elite survivors, crossover, and mutation. This natural solution is performed iteratively; most fit solutions survive to the final step.

With the increased adoption of Machine learning approaches over the recent past, genetic algorithm has become one of the most used stochastic optimization approaches to solve complex optimization problems.

1.2 Statement of The Problem

Asset allocation has been one of the complex steps in the investment decision process. It involves the selection of investments to make from diverse asset classes and the proportion of capital to allocate to each asset such that the overall investment output is optimized. This is done in an environment guided by the investor's goals and investment purpose, among other factors. A combination of factors, such as whether the investor is an individual or organization, job security, cash flow, employment type, and length of investment, among others, makes investment goals even more diverse such that a single allocation/portfolio will not work for all. Common investment outputs anticipated by investors include minimized risks and maximized returns.

Researchers in [4] and [5] found that factors such as income level, risk tolerance length of the investment period, Age of investors, asset familiarity, Financial knowledge, market sentiments, and expected return, among others, significantly affect asset allocation, as these factors shift over time, the optimal portfolio to change. With volatile financial assets, such as stocks, the optimal portfolio is also volatile. There arises a need for an algorithm that can make asset allocation less complex by analysing given assets in real time and constructing an optimal portfolio at any point in time.

1.3 Project Definition and Goals

The proposed approach aims at developing a machine learning solution that can simplify the investment decision process through the fast generation of optimal investment portfolios. At any point of time. The proposed solution is a GA-based model which;

i. Selects K assets from universally existing assets and include them in a portfolio. The investor can set K and denotes the number of assets they wish to include in their investment portfolio.
ii. Allocate weights/proportions of capital to invest in these assets to minimize risk and maximize portfolio returns.

Chapter 2: Literature Review

2.1 Literature Review

Today's businesses are increasingly looking for tools to support their negotiation process. Getting timely answers to complex problems is complex, so artificial intelligence (AI) approaches are used in decision-making to assist in this process. In the field of metaheuristic optimization, genetic algorithms (GA) use the concepts of evolution and genetics to search for optimal solutions [6]. GA is an essential tool not only for optimization solutions but also for satisfying requests involving large amounts of data, especially due to its evolutionary property, that it can perform combinations of responses to develop powerful solutions.

The scholars in [7] developed a model with real-world limitations (floor-ceil and cardinality constraints) that the Markowitz unconstrained Mean-Variance technique does not take into account when choosing the best portfolio. Used heuristic crossover to optimize the risk-return trade-off and reach an ideal solution for the choice of the portfolios and the distribution of weights to each portfolio. The outcomes demonstrate that the Heuristic crossover outperforms the other two crossovers, with a maximum return of 4.214555e-04 and a minimum risk of 1.331713e-06. The results demonstrate that Heuristic crossover is particularly helpful when an investor wishes to place his entire capital in an investment to provide the maximum return and the lowest risk.

By utilizing stable distributions as the marginal distributions and the dependence structure based on the copula function. The researchers in [8] attempted to find a solution for portfolio optimization. They formulated the portfolio optimization problem as a multi-objective mixed integer programming problem. Cardinality and quantity limitations were taken into consideration in the Model to broaden its applicability. Variants of multi-objective particle swarm optimization methods are used by the researchers in [8] to address the problem. The empirical findings show that one of the suggested MOPSOs outperforms the other important algorithms in terms of performance metrics.

In order to address the issue of optimum involvement in various power markets, the work in [9] provide a solution based on genetic algorithms (GA). Metaheuristic optimization tools based on artificial intelligence enable quick problem-solving with outcomes that are extremely comparable to those produced by deterministic techniques but at the expense of a lengthy execution time. The results obtained using a simulation scenario based on actual data from the Iberian electricity market demonstrate that the proposed method is capable of outperforming earlier implementations of the Particle Swarm Optimization (PSO) and Simulated Annealing (SA) methods while achieving very similar objective function results to those of a deterministic approach and doing so in a much shorter amount of time.

In [10] Paiva et al. developed a model for investment decisions. First, he uses the SVM algorithm to classify the assets and then the mean-variance Model to create the portfolio. Gu et al. demonstrated the utility of using machine learning to conduct empirical asset valuation [11]. This is by attributing the prediction gains to the explanation for the nonlinear predictor interactions. Trees and neural networks are most successful at predicting returns [12]. We used the Black-Scholes pricing model and machine learning to present stock price predictions. Various algorithms were used, such as decision trees, machine learning, and neural networks. Jiang et al. proposed an inventory index prediction version using deep mastering algorithms [13]. This entered version of technical and macroeconomic signs was used to forecast inventory indices on a month-to-month basis. Basak et al. used random forests to present stock market forecasts [14]. An empirical stock prediction classification framework that accounts for the previous day's price increase or decrease uses random forests and decision trees to drive this relationship. Wang et al. proposed a reinforcement learning method for optimizing investment policies [15]. To address the issue of balancing risk and return, they propose a model that uses macro market conditions as a dynamic indicator and adjusts the ratio of short and long funds to reduce the risk of market volatility. A study examined the optimization of stock portfolios using the causal relationship between the return of behavioural stocks and the bias according to the return of the holding period of behavioural stocks. We found that by mimicking real-world investment constraints, such as test costs and statistical evidence, by incorporating short selling into stock portfolio selection, available investment flexibility and stock portfolios outperform benchmarks and markets [16]. A study in [17] proposed a hybrid model with a combination of LSTM and Markowitz's mean-variance Model for streamlined modeling and asset price prediction. The work in [18] presented suitable tools and techniques for strategic project selection and alignment and portfolio balancing. In [17] the scholars proposed a portfolio construction method consisting of LSTM networks and MV models. LSTM is applied to obtain stock price patterns using various technical indicators as input variables, such as B. Relative Strength Index (RSI), Momentum Index (MOM), and True Range (TR). The researchers in [19] introduced an LSTM neural network to predict the directional movement of S&P 500 stocks from 1992 to 2015. In this study, we found that the LSTM-based portfolio outperforms other non-memory machine learning models (i.e., RF, DNN, and LR).

The effectiveness of emergency response systems is pertinent in saving lives during crisis situations; hence innovative approaches such as using genetic algorithms have been crucially studied by researchers in [25]. This method results in reduced response times and improved resource efficiency. Additionally combining genetic algorithms with simulated luminescence methods has reported successful outcomes when used to improve parameter settings of machine learning systems according to [26]. The hybrid approach proposes a fast strategy that identifies optimal options while maximizing classification accuracy and minimizing computation time. Researchers in [27] developed a genetic algorithm to resolve the routing problem in a logistics network. The study proved the algorithm's capability to find the optimal alternatives for lowering total transportation costs while improving service. In [30] the researchers developed a hybrid strategy that combines a genetic algorithm with a differential evolution algorithm in order to handle the machine learning feature selection problem more efficiently. The research demonstrated the effectiveness of the recommended technique in locating optimal solutions that enhance classification accuracy and minimize feature subset size. In [27] researchers used a genetic algorithm to optimize the production planning issue in a flexible manufacturing system. The results proved the algorithm's efficiency in identifying the best potential selections that minimize total production costs and maximize resource efficiency. By merging a genetic algorithm with a particle swarm optimization method, according to [25] proposed a hybrid technique to address the inventory routing problem in a supply chain. The analysis demonstrated that the recommended algorithm is capable of swiftly discovering the optimal solutions that simultaneously lower overall expenses and increase customer satisfaction.

Genetic algorithms have shown promise in tackling manufacturing layout design problems according to [25]. These smart methods provide speedy solutions in choosing the optimal approach towards lowering material handling expenses whilst boosting production output. Meanwhile in [30] the research offers a hybrid technique that combines genetic and tabu search algorithms in solving university scheduling issues. By maximizing resource compatibility while keeping waiting times to a minimum this approach provides the most efficient solution to the problem at hand. The results showed that the algorithm is capable of quickly locating the best options that shorten total processing time and optimize resource utilization. In order to solve the wind power generation scheduling problem more effectively, the researchers in [28] created a hybrid algorithm that combines a genetic algorithm with a harmony search method. The study showed that the suggested algorithm can efficiently locate the best options that reduce the overall cost of generation and increase the stability of the power system.

2.2 Key Take Aways

Taking into account a comprehensive evaluation of the existing literature, an examination has yielded the subsequent outcomes that can be considered as the key findings:

- Many scholars have conducted different studies aimed at developing a suitable model for simplifying the investment decision process through the fast generation of optimal investment portfolios.
- Most of the scholars relied on prejudice assumptions and hence lacked methodologies for selecting the most suitable models in their work.
- The existing literature presented unreliable models for simplifying the investment decision process and thus creates a need for future research to address the identified research gaps.
- GA is an essential tool not only for optimization solutions but also for satisfying requests involving large amounts of data, especially due to its evolutionary property, that it can perform combinations of responses to develop powerful solutions.

Chapter 3: Methodology

3.1Dataset

3.1.1 Dataset Collection

Yahoo Finance provides real-time and historical data for various types of securities, including Stocks, ETFs, Futures, Indices, and currencies. Each security has a unique "symbol" code that can be used to retrieve its information, such as historical or real-time prices. However, Yahoo Finance does not provide a complete list of all the listed assets and their symbols, making it difficult to experiment with different symbols. To address this issue, a GitHub project was created by Mlapenna in 2021 [20]. The project includes a Python script that retrieves the asset symbol and name, and has collected data on 391,108 assets, including Stocks, ETFs, Futures, Indices, and currencies. The retrieved data is saved in CSV format and can be used as the asset universe. Figure 3.1 shows a snapshot of the retrieved data.

Symbols	Name
<chr>	<chr>
0P0001FDND.TO	Manulife Global Stra Bal Yld Fd
^NQMY65LMGBPT	Nasdaq Malaysia Utilities Large
CRBL.L	MULTI UNITS LUXEMBOURG LY COMMO
0MNL.IL	ISHARES STOXX EUROPE 600 BASIC
OBGRF	ORBIT GARANT DRILLING INC
VNX.BE	NXP SEMICONDUCTORS EO-,20

Figure 3.1: A snapshot of asset symbol and name data

In this study, out of the 391,108 assets collected from the aforementioned GitHub project, 100,000 assets were randomly selected for analysis due to the computational resources needed to process all of the assets. The symbols of the selected assets were used to retrieve their historical prices. Figure 3.2 shows a snapshot of the historical price data for a particular asset.

Date	Open	High	Low	Close	Volume	Adjusted
<date>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
2018-12-11	9.74680	9.74680	9.74680	9.74680	0	9.74680
2018-12-12	9.76770	9.76770	9.76770	9.76770	0	9.76770
2018-12-13	9.76310	9.76310	9.76310	9.76310	0	9.76310
2018-12-14	9.67370	9.67370	9.67370	9.67370	0	9.67370
2018-12-17	9.59750	9.59750	9.59750	9.59750	0	9.59750

Figure 3.2: Snapshot of historical data for sample assets

3.1.2 Inclusion Criteria

In order to capture price trends and covariance structure between patterns, assets were selected for inclusion in the dataset based on specific criteria. Only assets that reported prices for at least 60% of the days in the recent past were included. Assets that did not report any data in the last 6 months were also excluded to prevent the GA model from recommending investments in delisted assets. Additionally, assets for which data could not be retrieved were excluded. As a result of these inclusion criteria, only 355 assets met the requirements and were included in the final dataset. It's worth noting that the number of included assets is dependent on the sampled assets and may fluctuate with each new sample.

3.2 Data Cleaning and Pre-processing

The daily closing prices and corresponding dates for each asset were extracted and organized into a final table to ensure that the prices are in a format suitable for analysis, as shown in Figure 3.3. Daily returns were calculated using the formula:

$$R_t = \frac{p_t - p_{t-1}}{p_t}$$

Where R_t represents the asset return at date t, p_t represents the asset price at date t, and p_{t-1} represents the asset price at date t-1.

A data.frame: 1826 × 343

Date	CRBL.L	OBGRF	VNX.BE	NUIIX	EMNT.V	0P00014CKP.HK	JPEH.L	BRIS.JK	ALVAL.PA
<date>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
2018-04-05	14.6642	1.92	94.0	27.09	0.27	NA	36.540	NA	7.888928
2018-04-06	14.5077	1.92	92.0	26.95	0.27	NA	36.540	NA	7.928571
2018-04-07	NA	NA	NA	NA	NA	NA	NA	NA	NA
2018-04-08	NA	NA	NA	NA	NA	NA	NA	NA	NA
2018-04-09	14.6549	1.90	92.0	27.17	0.27	8.7428	36.540	NA	7.988035
2018-04-10	14.7698	1.90	93.0	27.47	0.27	8.7617	36.540	NA	8.186250
2018-04-11	14.8573	1.84	93.0	27.38	0.27	8.7202	36.540	NA	8.424107
2018-04-12	14.9741	1.84	93.5	27.46	0.27	8.7272	36.540	NA	8.325000
2018-04-13	14.9926	1.84	92.5	27.44	0.24	8.7514	36.540	NA	8.225892

Figure 3.3: A snapshot of asset prices on the final data

The asset returns were divided into training and testing sets, with the evaluation case covering the final year of the sample period.

3.3Analysis Method

3.3.1 Portfolio Optimization Theory

Portfolio optimization theory aims to develop investment strategies that select a mix of assets to invest in at a given level of risk, in order to maximize the expected return on investment. The basic idea is that spreading investment capital across several assets can help minimize investment risk [21], Mean variance optimization is one of the commonly used approaches in modern portfolio theory (MPT). It involves choosing an asset mix that maximizes the portfolio Sharpe ratio, which can be thought of as the excess return per unit risk. In other words, the goal is to maximize the amount of return for each unit of risk taken. The formula for the Sharpe ratio is shown below.

$$\text{Sharpe} = \frac{ER - R_f}{\sigma}$$

Where:

$$\text{ER} = \sum_{i=1}^{n} w_i r_i \ , \sigma^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j (r_i r_j) = W^T (S) W$$

ER is the portfolio expected return, R_f is the risk free rate (i.e., the expected return if we choose to invest in a risk-free investment), σ is the portfolio's standard deviation, and $r_1, r_2 \dots r_n$ are the expected portfolio returns for each individual asset, while $w_1, w_2 \dots w_n$ are the asset weights. High values of the Sharpe ratio indicate that the chosen mix yields high returns with very little risk according to [31]. Conversely, negative values of the Sharpe ratio indicate that taking risks on the investment is not worthwhile because the risk-free investment yields higher returns.

3.3.2 Formulating the Optimization Problem

3.3.2.1 Single Objective

The research problem at hand involves an optimization challenge, where the aim is to select a portfolio of assets that maximizes the Sharpe ratio [22]. However, managing a portfolio with a large number of assets can be unwieldy, so a decision must be made about which assets to invest in. To address this issue, the experimentation is broken down into two optimization steps:

3.3.2.1.1 Step 1: Asset Selection:

The first step involves choosing which assets to include in the portfolio to achieve an optimal Sharpe ratio. The objective function is:

$$\text{Maximize Sharpe } \frac{ER - R_f}{\sigma}$$

The search for weights $w_1, w_2 \dots w_n$ is subject to the following constraints: w_i = 0 or 1 , $\sum_{i=1}^{n} w_i \leq$ K, and $\sum_{i=1}^{n} w_i r_i \geq 0$. These constraints mean that the weights are either 0 (for dropped assets) or 1 (for selected assets), the sum of the weights should be equal to or less than the user-specified value K (which represents the number of assets to select), and the expected return should be positive (to ensure profitability). For simplicity, we assume that the investment capital is unlimited, and that the investor invests an equal amount of capital in each asset.

3.3.2.1.2 Step 2: Capital allocation

In this step, the focus is on selecting an allocation strategy that maximizes the Sharpe ratio. The objective function is formulated as follows:

$$\text{Maximize Sharpe} = \frac{ER - R_f}{\sigma}$$

Subject to the constraints: $0 \leq w_i \leq 1$, $\sum_{i=1}^{n} w_i$= 1, and $\sum_{i=1}^{n} w_i r_i \geq 0$ [23], the allocated weights should range between 0 and 1 and sum up to 1. Additionally, the expected return should be positive (profit). The weights represent the amount of investment capital to allocate to each of the K chosen assets.

3.3.2.2 Multi Objective Approach

In some cases, users may have multiple target objectives. A common strategy in such situations is to use the weighted sum of deviation approach, which seeks to minimize the sum of deviations from the target objectives [24], In this research, we experimented with a two-

objective case, with the aim of maximizing expected returns while minimizing risk. Our approach aims to achieve zero risk while achieving a 100% growth goal for returns.
The objective is therefore:

$$\text{Minimize, } x|(ER-1)|+(1-x)|(\sigma-1)|$$

Where x is the weight assigned to the objective for return and can be interpreted as the importance a user places on return, while 1-x is the weight assigned to the objective for risk. The objective function is subject to the same constraints as in the single objective case, plus an additional constraint for x; $0 \leq x \leq 1$. We experimented with different values of x to understand how users' risk tolerance affects the optimal portfolio. There are several search algorithms available for finding the optimal solution. For this research, we focused on the Genetic algorithm due to its robustness and ability to converge to the optimal solution even without the initial population exploiting the whole set of feasible solutions.

3.3.3 Genetic Algorithm

The genetic algorithm is a heuristic search algorithm that is based on Darwin's theory of evolution. The theory explains how populations evolve over time, with natural selection favouring the most fit species. The algorithm is guided by the following principles (Hendry et al., 2011):

1) Variation: That the populations contain different species with different traits according to the respective genetic make-up which is influence by the surrounding environment.
2) Inheritance: these traits are passed to the next generations through genes
3) Overproduction: Populations are able to grow exponentially but are constrained by limited resources.
4) Competition: there is a completion for the limited resources on the populations, and not all individuals will survive.
5) Differential survival and reproduction: only individuals with traits have a higher chance to survive and give offspring's which will thrive on the next generation.

When these principles are in effect, natural selection acts on the variation in the population, leading to different numbers of individuals from each species in the population. Species with favourable traits thrive and are referred to as fit.

The genetic algorithm optimization model imitates this process to come up with the best solution. Initially, it generates an initial population of feasible solutions that satisfy the variation principle. Subsequent generations are then generated from the population by applying genetic operators including elitism, crossover, and mutation. Elitism allows a proportion of fit solutions to pass into the next generation, crossover involves swapping traits of two fit solutions to produce modified child solutions, and mutation alters parent solutions to produce better solutions in the next population. This process continues until convergence is reached or a specified final iteration is reached. The most fit solution seen over the generations is kept.

Figure 3.4 below visualizes how genetic operators work on the initial solution to generate more fit solutions.

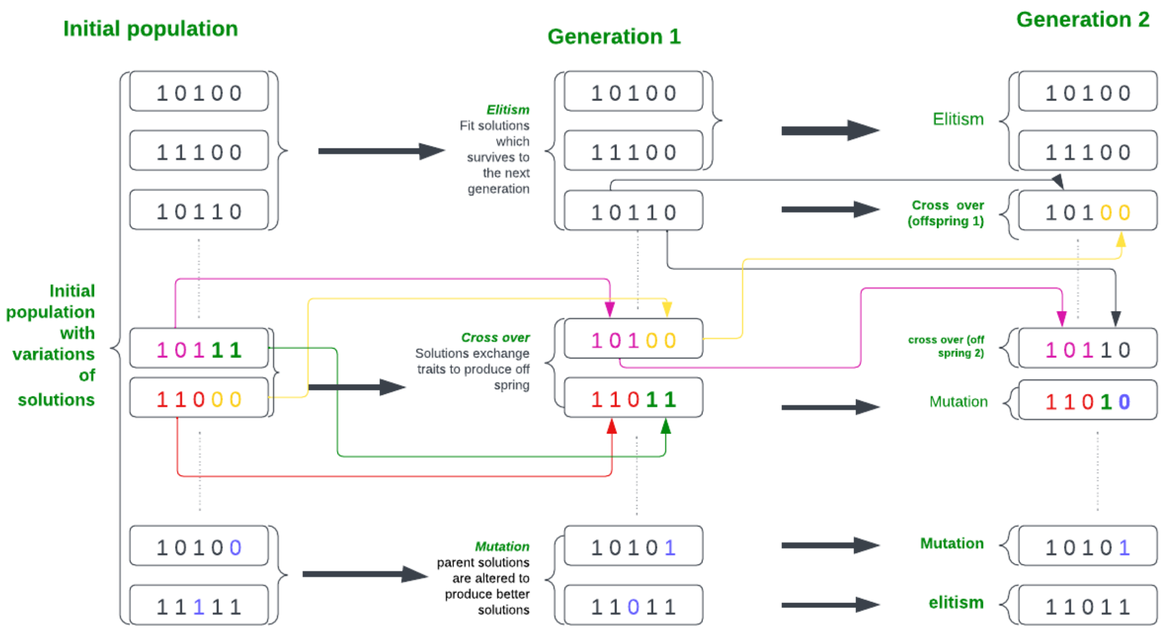


Figure 3.4: This figure visualizes how GA, applies genetic operators on initial population.

To use a genetic algorithm for constrained optimization problems, it is often necessary to convert the problem into an unconstrained form. This can be achieved by penalizing the objective function for solutions that violate the constraints. In the case of a single objective, the constraints can be rewritten as follows:

- The range $0 \le w_i \le 1$ be broken into $0 \le w_i$ and $w_i \le 1$ respectively, then the two rewritten as - $w_i \le 0$ and $w_i - 1 \ge 0$.
- Functions $\sum_{i=1}^{n} w_i \le$ K and $\sum_{i=1}^{n} w_i r_i \ge 0$ be rewritten to $\sum_{i=1}^{n} w_i$-K≥0 and $-\sum_{i=1}^{n} w_i r_i \le 0$ respectively.

The following penalty functions were used at the first step:

- For - $w_i \le 0 \Rightarrow \max\,(0, -w_i)^2$ penalizes objective function for values of w_i less than 0, while
- For $w_i - 1 \ge 0 \Rightarrow \max\,(0, w_i - 1)^2$ penalizes objective function for values of w_i greater than 1.
- For $\sum_{i=1}^{n} w_i - \mathrm{K} \ge 0 \Rightarrow \max\,(0, \sum_{i=1}^{n} w_i - \mathrm{K})^2$ penalizes objective function for solutions with assets greater than the user specified K.
- For $-\sum_{i=1}^{n} w_i r_i \le 0 \Rightarrow \max\,(0, -\sum_{i=1}^{n} w_i r_i)^2$ penalizes objective function for solutions with expected return less 0(solution leading to losses in long run).

The constraint formalism used in this research is related to the minimization of the Sharpe ratio, which increases (making it worse) whenever a constraint is not met. To transform this into a minimization problem, the Sharpe ratio was negated and the GA algorithm was set to search for binary weights only. The fitness function is shown below.
Maximize:

$$\text{-Sharpe} + 10[\max(0, -w_i)^2 + \max(0, w_i - 1)^2 + \max(0, \textstyle\sum_{i=1}^{n} w_i - \text{K})^2 + \max(0, -\textstyle\sum_{i=1}^{n} w_i r_i)^2]$$

The capital allocation step differs from the asset choice step in that the weights should sum up to 1. Hence, the penalty function $\max(0, \sum_{i=1}^{n} w_i - \text{K})^2$ was slightly changed to $(\sum_{i=1}^{n} w_i - 1)^2$ to make the fitness function. The algorithm was restricted to search real values.

Maximize:

$$\text{-Sharpe} + 10[\max(0, -w_i)^2 + \max(0, w_i - 1)^2 + (\textstyle\sum_{i=1}^{n} w_i - 1)^2 + \max(0, -\textstyle\sum_{i=1}^{n} w_i r_i)^2]$$

The multi objective case differs from the single objective case on the objective function, instead of maximizing Sharpe we minimize the function $x|(ER - 1)| + (1 - x)|(\sigma - 1)|$. The fitness function on the asset choice stage is therefore:

Minimize:

$$-x|(ER - 1)| + (1 - x)|(\sigma - 1)| + 10[\max(0, -w_i)^2 + \max(0, w_i - 1)^2 + \max(0, \textstyle\sum_{i=1}^{n} w_i - \text{K})^2 + \max(0, -\textstyle\sum_{i=1}^{n} w_i r_i)^2]$$

While the fitness function on the capital allocation step is:

Minimize:

$$-x|(ER - 1)| + (1 - x)|(\sigma - 1)| + 10[\max(0, -w_i)^2 + \max(0, w_i - 1)^2 + (\textstyle\sum_{i=1}^{n} w_i - 1)^2 + \max(0, -\textstyle\sum_{i=1}^{n} w_i r_i)^2]$$

Chapter 4: Data Analysis Results

4.1 Data Analysis Results

Visualizing returns for a sample of 12 assets shows that most of the assets had a close to stationary returns for the period between 2018-04-06 and 2023-04-07. A common trend is a highly volatile returns for the period around 2020, which indicates high risk to investments made around that time. Figure 4.1 below visualizes this information.

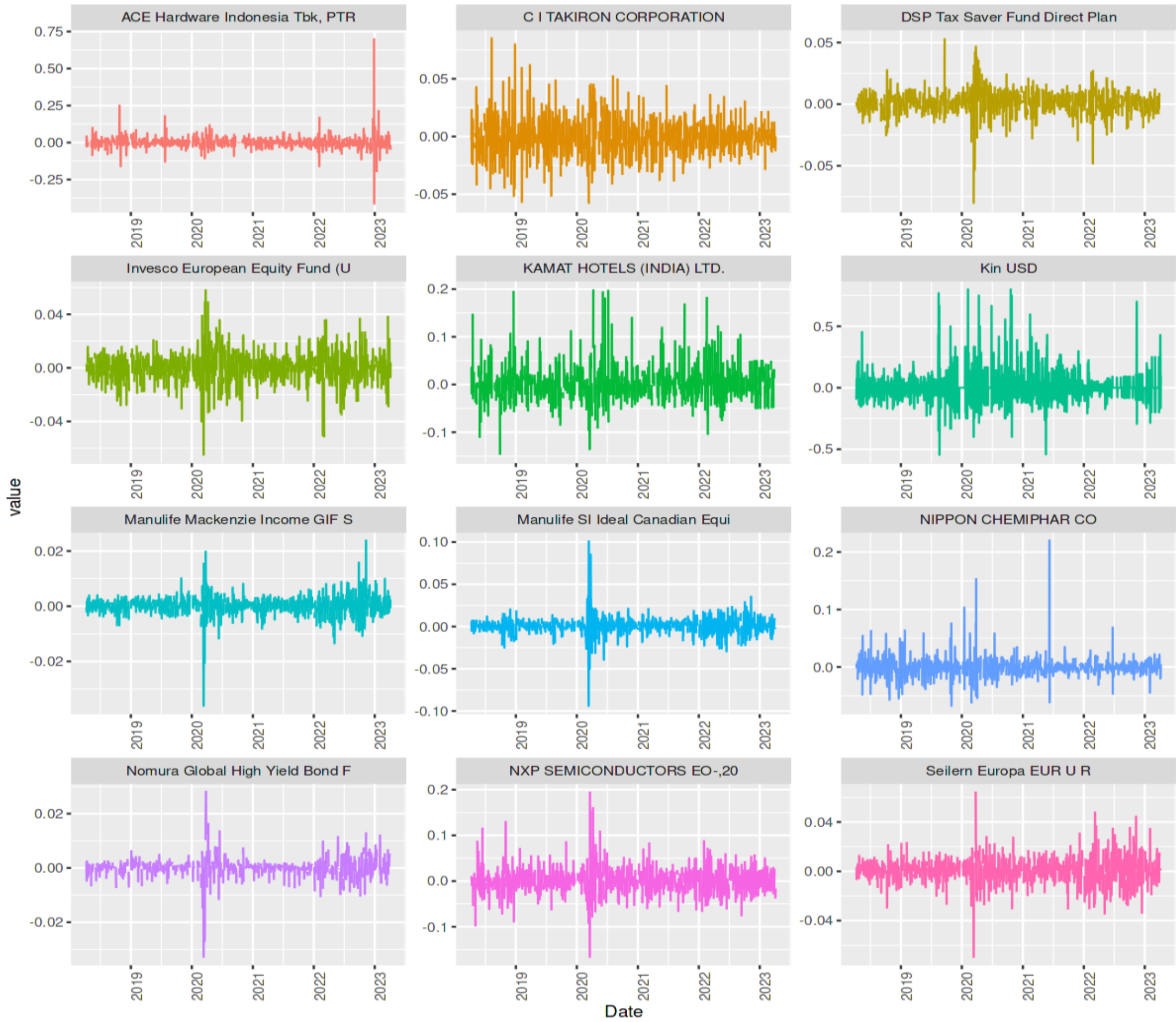


Figure 4.1: Visualizing the trend in returns for a sample of 12 assets.

For all the scenarios Genetic algorithm was run with initial population of 50 solutions, 5% elitism, with 80% probability of cross over and 1% mutation probability, maximum iteration was set to 1000 generations, an early stopping was set to declare the model as converged after 50 consecutive iterations without best solution changing. We assumed a risk-free rate equal to 1.1%.

4.2Single Objective Case

4.2.1 Step 1: Asset selection:

For the asset selection step, we configured the asset selection step to select by default a portfolio of 10 assets at most, users can change this number according to their preference. For 1000 iterations, the algorithm selected 10 assets. They are listed in the table 4.1 below:

Asset	Symbol
VERICEL CORP.	ATQP.F
Aditya Birla Sun Life Asset All	0P00008THR.BO
China Merchants Port Hldgs Co.R	CPM.SG
Legg Mason ClearBridge Value Fu	0P0000MOLV.L
Match Group, Inc.	MTCH
Quilter Investors Cirilium Cons	0P0000VW6D.L
TALON INTERNATIONAL INC	TALN
iSun, Inc.	ISUN
SEI Global Assets Fund plc – Th	0P0001EXXS.L
AUTECO FPO	AUT.AX

Table 4.1: selected portfolio for the single objective case.

Figure 4.2 below visualized mean, median and the best fitness value in each iteration, the algorithm. To start converging at around 650 iterations where mean, median and best fitness values are almost the same.

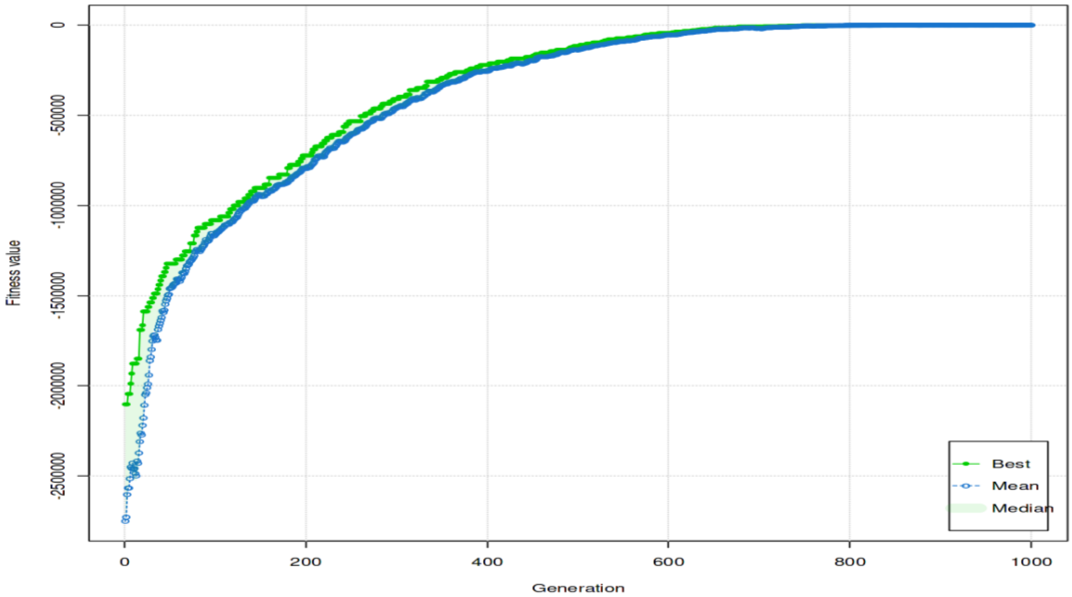


Figure 4.2 Model Training History

Based on the assumption that, Investors plan to invest equal proportion on all the assets, and capital allocated to those assets not selected was not spend. The expected return equals to 5.24% and Sharpe ratio 5.89%, the expected risk is 49.24%. table 4.2 below reports the expected performance of the optimal portfolio. Figure 4.3 shows the trend on the 3 metrics over the iterations. The dotted line are the optimal values also reported in table 4.2. It is evident from figure 4.3 that the objective function was only greed on the Sharpe ratio, and admitted any solution with high Sharpe, regardless of the risk associated with it. On this case, the risk levels are not controlled which resulted to the model recommending a portfolio with the optimal risk.

Metric	**Value**
Returns	0.0524
sharpe	0.0589
variance	0.4924

Table 4.2: Expected Performance

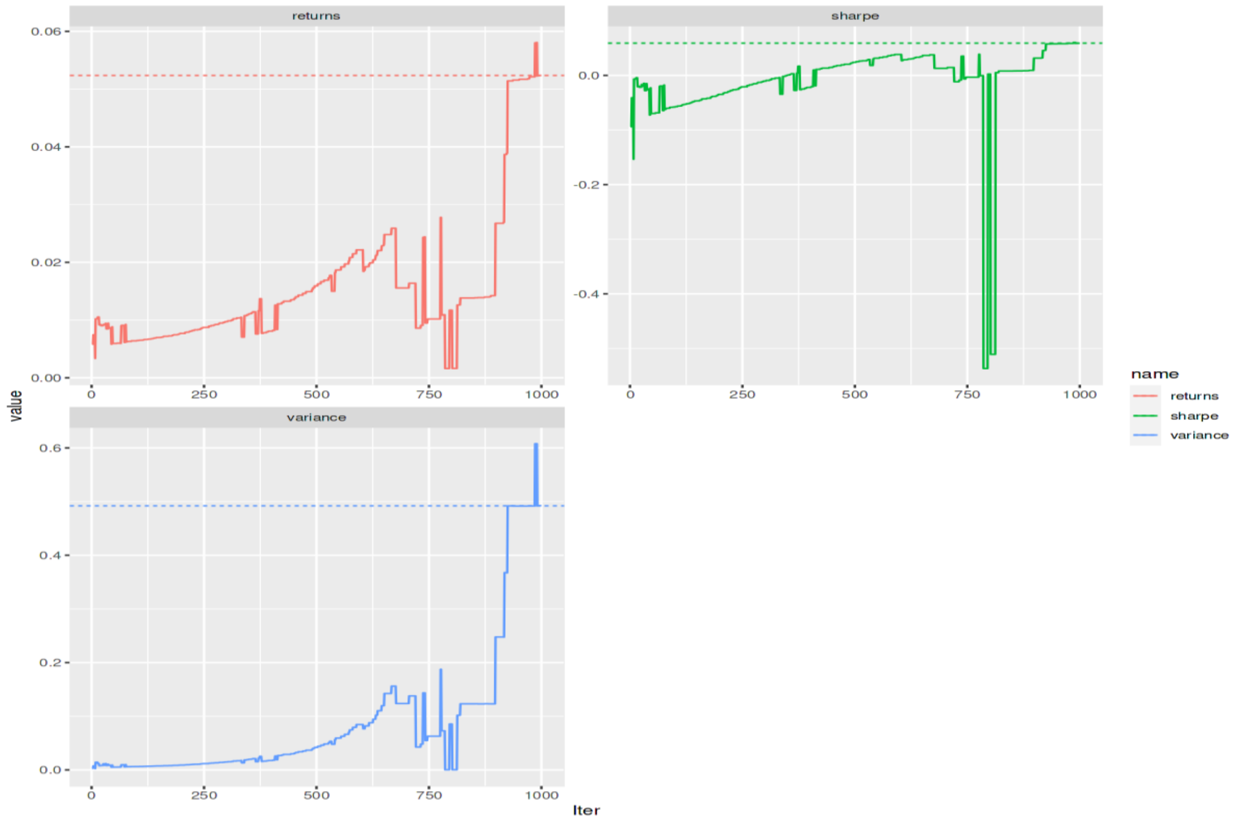


Figure 4.3: Expected returns, variance, and Sharpe ratio for the best solution in each iteration vs the global best solution.

4.2.2 Step 2: Capital Allocation

The capital allocation step was run with 1000 iterations but the model converged too quickly(with 414 iterations) the optimal allocation strategy involved allocating 53.66% of the capital to SEI Global Assets Fund plc – Th, then 18.51% of the capital to Quilter Investors Cirilium Cons and 17.91% to Legg Mason ClearBridge Value Fu, the remaining 9.02% of the capital should be allocated to the other 7 assets with iSun, Inc. and China Merchants Port Hldgs Co.R receiving the least share. Table 4.3 below reports this information. Figure 4.4 visualizes the trends in mean, median and best fitness value during each training iteration.

Name	Weight	Weight (%)
SEI Global Assets Fund plc - Th	0.5366	53.66%
Quilter Investors Cirilium Cons	0.1851	18.51%
Legg Mason ClearBridge Value Fu	0.1791	17.91%
TALON INTERNATIONAL INC	0.0260	2.60%
Aditya Birla Sun Life Asset All	0.0248	2.48%
AUTECO FPO	0.0162	1.62%
Match Group, Inc.	0.0128	1.28%
VERICEL CORP.	0.0112	1.12%
iSun, Inc.	0.0068	0.68%
China Merchants Port Hldgs Co.R	0.0013	0.13%

Table 4.3: optimal capital allocation strategy

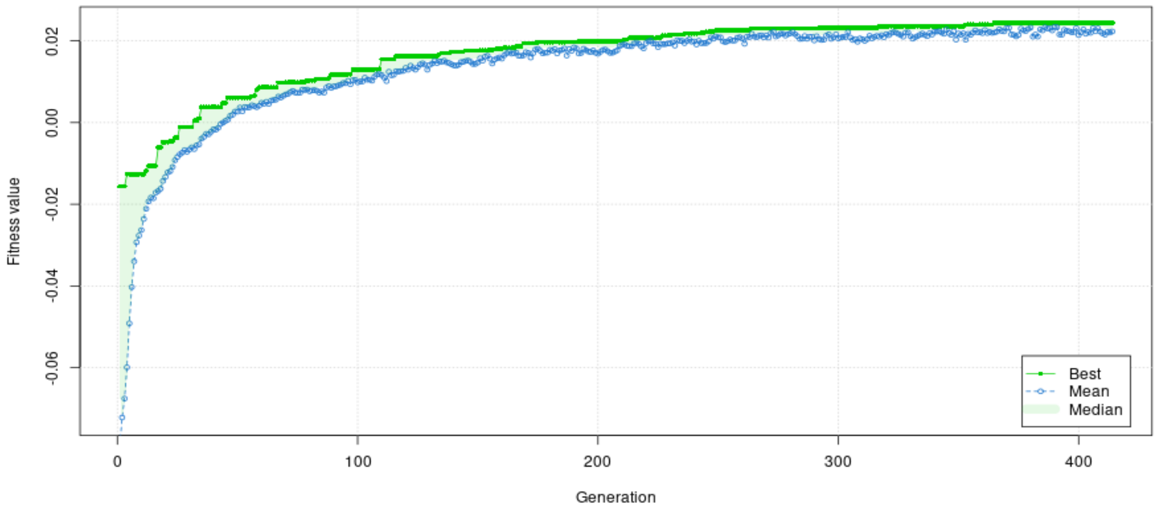


Figure 4.4:the trend in performance during model training.

Based on the above allocation strategy, the expected return is 17.85%. with Sharpe ratio 5.97% and risk 787.31%. Figure 4.5 below visualizes the trend in Expected returns, Sharpe ratio and risk, over the training iterations. The dotted line shows the corresponding values with the optimal strategy. Again, we see that the optimal solution attracts high risks.

Metric	Value
returns	0.1785
sharpe	0.0597
variance	7.8731

Table 4.4: Expected performance with the optimal strategy

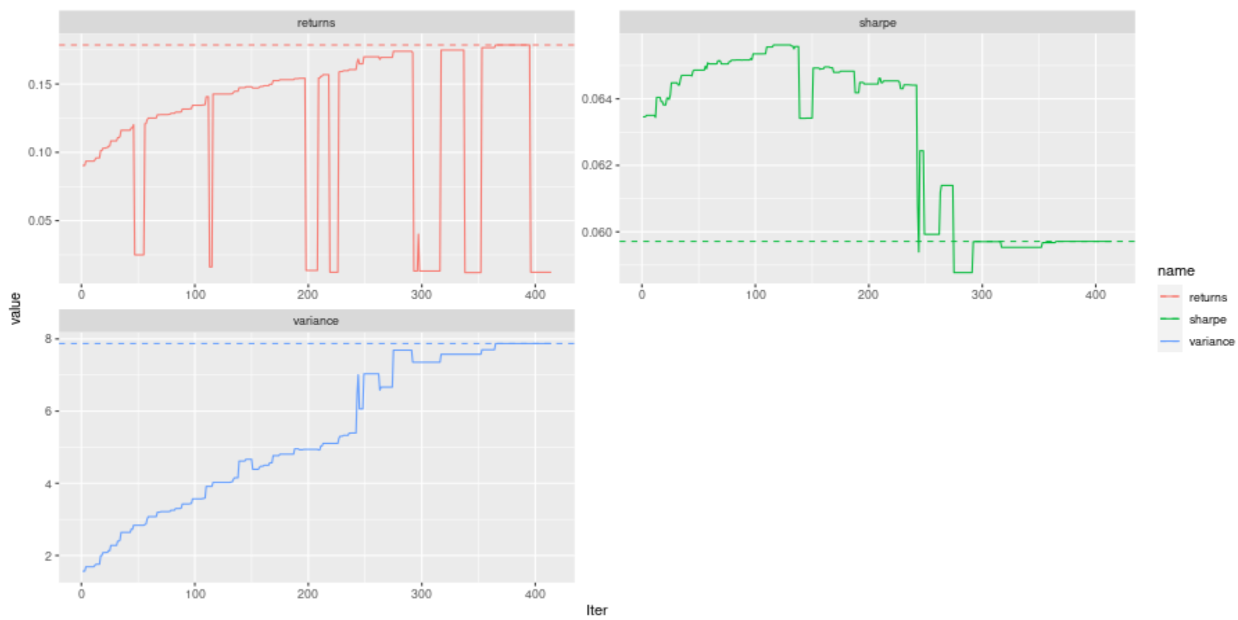


Figure 4.5: Trends in expected return, risk, and Sharpe ration during the training period.

To evaluate the worthiness of this investment strategy we compared it with Investing equal amounts on all the selected assets also randomly allocating the investment capital to the selected assets. GA approach yielded the highest expected return (27.21%) but attracts the highest risk 1448.48%, random allocation of weights came the second with 8.57% expected return and 142.46% risk. Equal allocation of capital had the least expected return (5.12%) but attracts the least risk (50.14%). Figure 4.6 below visualized the trend in daily expected return during the evaluation period.

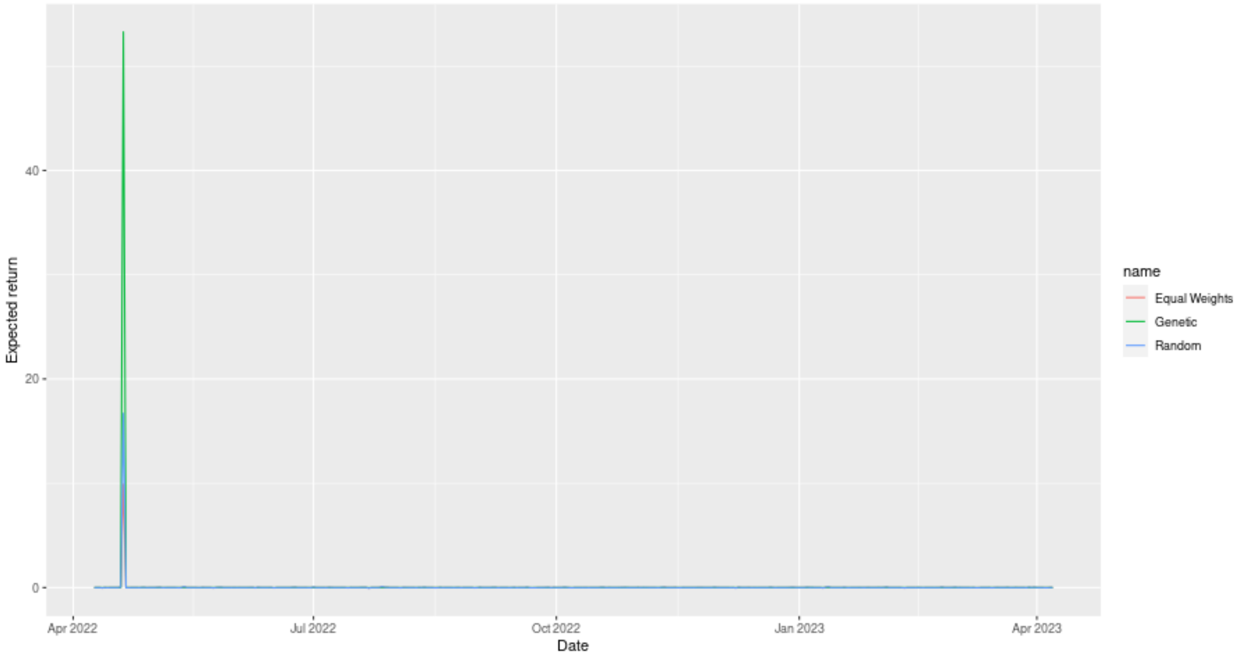


Figure 4.6: trend in Daily expected returns with the GA optimal strategy, equally allocated capital, and randomly allocated capital.

4.3Multi Objective Approach

4.3.1 Step 1: Asset selection

Similar to the single objective approach. The algorithm was configured to construct a portfolio of 10 assets at most, which minimize the sum of weighted deviations from the objective goals. The objective weights were set at 0.5. The algorithm converged at 589 iterations. The following 10 assets were selected.

Asset	Symbol
XTRACKERS PHYSICAL GOLD EUR HED	XAD1.MI
PRINCIPAL FONDOS DE INVERSION P	PRINLS0FA.MX
Bankinter Ahorro Activos Euro C	0P0001DEY5.F
PIMCO Funds:Real Return Fund,	PARRX
Aditya Birla Sun Life Asset All	0P00008THR.BO
Invesco BulletShares 2026 Corpo	BSCQ
The Hartford Strategic Income F	HSNRX
ADX ENERGY	GHU.MU
E-Commodities Holdings Ltd. Reg	WWY1.SG
Pioneer Global High Yield Fund	GHYYX

Table 4.5: selected portfolio for the single objective case.

Figure 4.7 below visualizes mean, median and the best fitness value in each iteration. The algorithm starts converging at around 650 iterations where mean, median and best fitness values are almost the same.

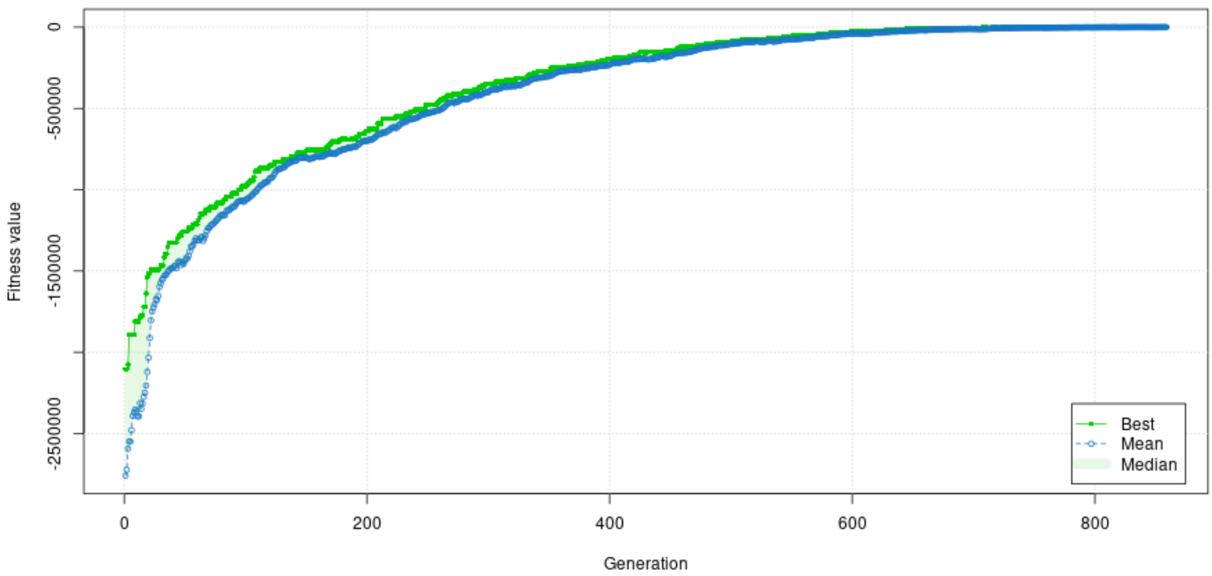


Figure 4.7: the trend in performance during model training.

For a scenario where Investors plan to invest equal proportion on all the assets, and capital allocated to those assets not selected is not spend. The selected portfolio leads to 0.24% expected return, -69.31% Sharpe ratio and 0.01% risk. Figure 4.8 shows the trend in returns, Sharpe, and risk. It is seen that both returns and risk continue to fall over the iterations, until the risk is almost the same as or goal of 0. Returns drop and rise to the maximal possible for the minimal variance. The negative Sharpe ratio means investment on the risk-free rate is more profitable than this risk, additionally the value of expected return here means any risk-free investment with the rate above 0.01% renders this scenario less feasible.

metric	Value
returns	0.0024
sharpe	-0.6931
variance	0.0001

Table 4.6: Expected performance with the optimal strategy

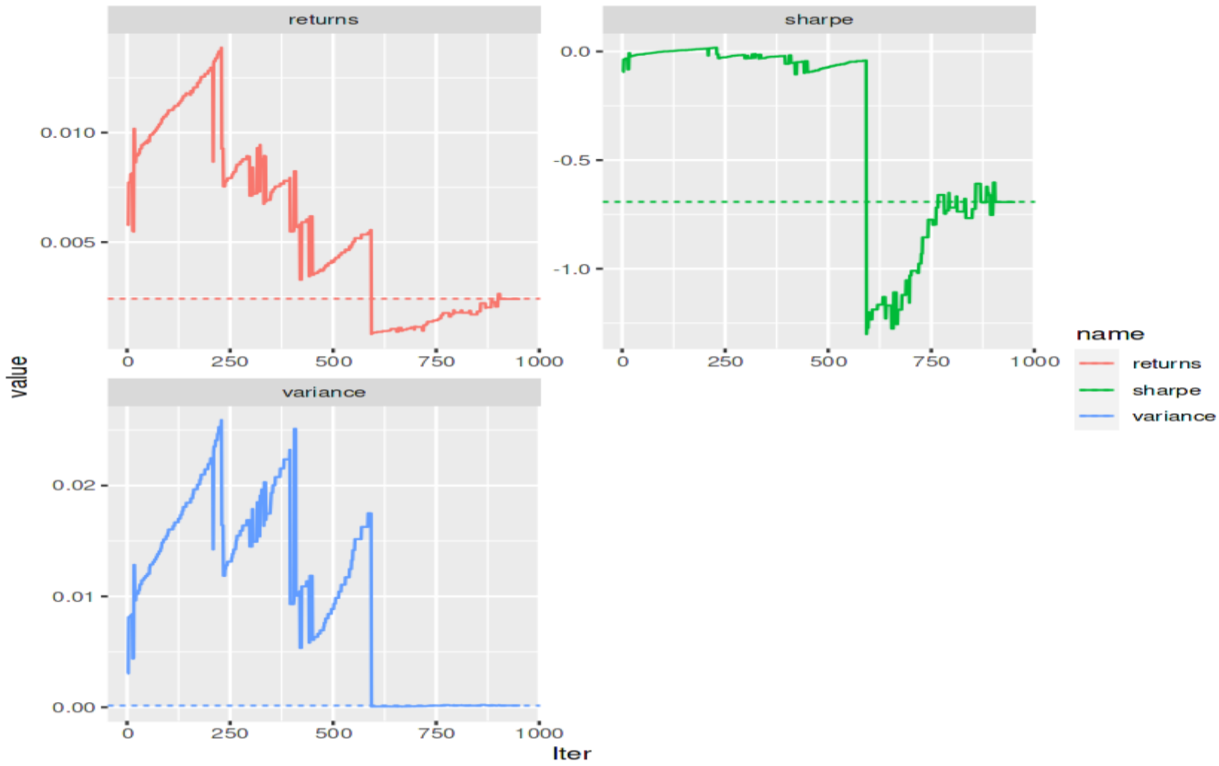


Figure 4.8: Trends in expected return, risk, and Sharpe ratio during the training period.

4.3.2 Step 2: Capital Allocation

For the capital allocation step, the algorithm was run with 1000 iterations, the objective weights for return were changed from 0 to 1 to observe the change in performance for investors with different balance between risk and return. For an investor placing all (x=0) their weight on risk. The model recommends investing 86.19% of their capital on AUTECO FPO, 3.11% on PROCTER & GAMBLE, 2.32% on Federated Hermes European Alpha and the remaining 8.40% on the tother 7 assets. Table 4.7 below reports the optimal strategy for each level of objective weights. The second column shows capital allocation for an investor putting 20% weight on expected returns and 80% on risk column 5 shows the optimal allocation for investors putting 80% on expected return and 20% weight on risk. Moving from left to right of this table represents an investor becoming less risk averse.

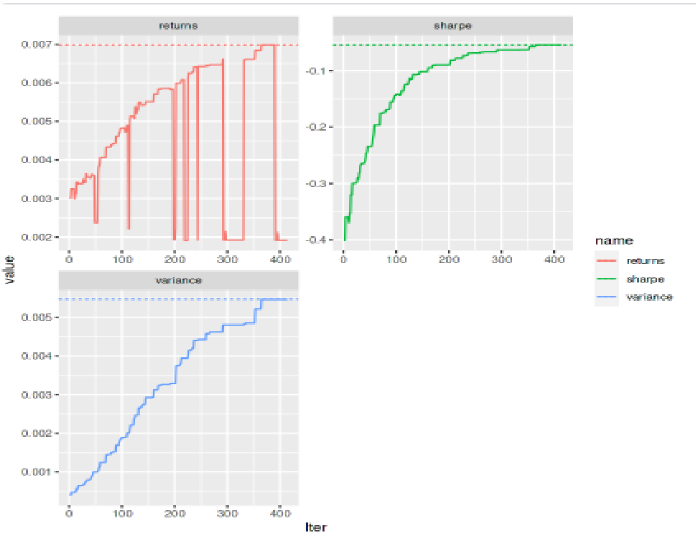
Objective weight for expected return(x)						
Asset	**0**	**0.2**	**0.4**	**0.6**	**0.8**	**1**
Fidelity MSCI Real Estate Index	1.78%	0.30%	22.83%	19.57%	11.44%	9.02%
PROCTER & GAMBLE	3.11%	1.48%	0.44%	1.35%	1.28%	0.55%
AQR Risk-Balanced Commodities F	0.21%	1.86%	1.69%	2.81%	0.45%	8.88%
HONDA MOTOR CO	1.39%	1.38%	0.42%	0.37%	0.21%	0.67%
XTRACKERS II USD OVNI RATE SWA	1.32%	1.40%	**33.24%**	**39.59%**	**36.25%**	**40.23%**
USHA MARTIN LTD.	1.48%	0.13%	1.12%	0.81%	0.80%	0.52%
DLF LIMITED	1.36%	1.53%	0.44%	1.21%	1.52%	0.55%
UPAMC Asian Bric Fund	0.86%	1.89%	10.72%	2.88%	17.46%	8.49%
AUTECO FPO	**86.19%**	**88.30%**	0.46%	0.48%	0.50%	0.57%
Federated Hermes European Alpha	2.32%	1.73%	28.64%	30.94%	30.10%	30.54%

Table 4.7: optimal strategy for each level of objective weights.

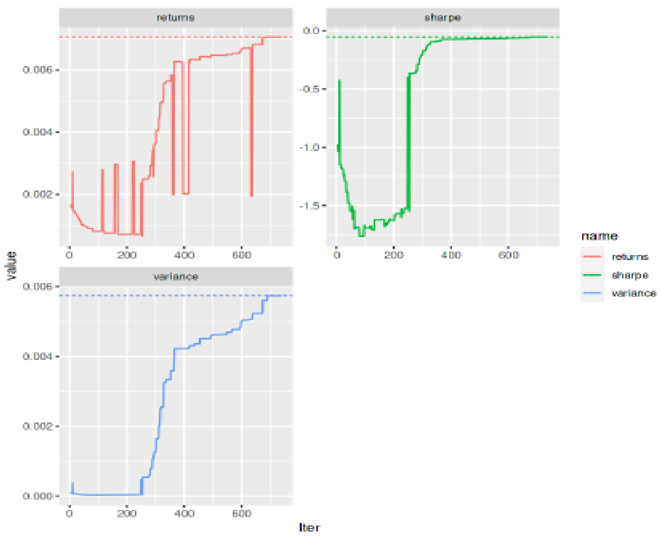
Figure 4.9 reports the trends in expected return, Sharpe ratio and risk. The graph in panel (a) visualizes a less risk averse case where investors can allow deviation from risk goal but 0 deviation on expected return. Panel(b) visualizes less risk averse case which can allow 80% of the total deviation from objective goals to come from risk. Moving through panels (a) to (f) represent investors turning more risk averse, panel (f) represents an investor who is not ready to miss the risk goal at the expense of returns. Table 4.8 below reports the expected return, Sharpe ratio and risk corresponding to the optimal allocation strategy for each balance in objective weight.

Metric	**0**	**0.2**	**0.4**	**0.6**	**0.8**	**1**
returns	0.70%	0.71%	0.06%	0.06%	0.06%	0.06%
sharpe	-5.42%	-5.20%	-178.94%	-191.82%	-194.78%	-215.95%
variance	0.55%	0.57%	0.00%	0.00%	0.00%	0.00%

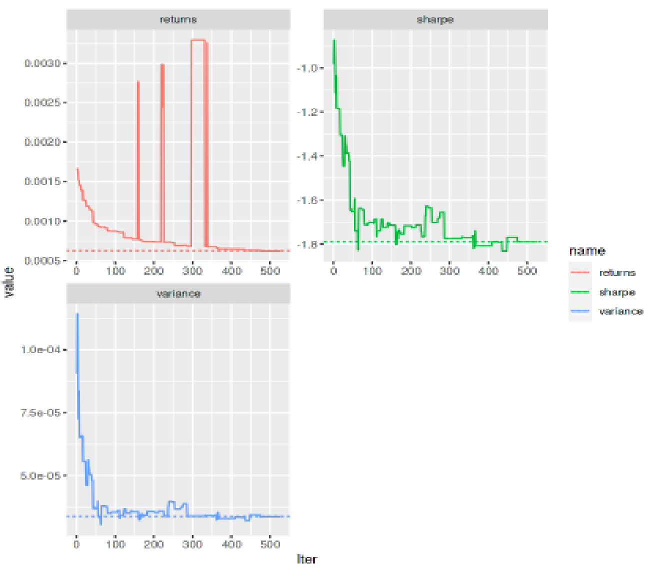
Table 4.8: returns, Sharpe ratio and risk for the optimal strategies.



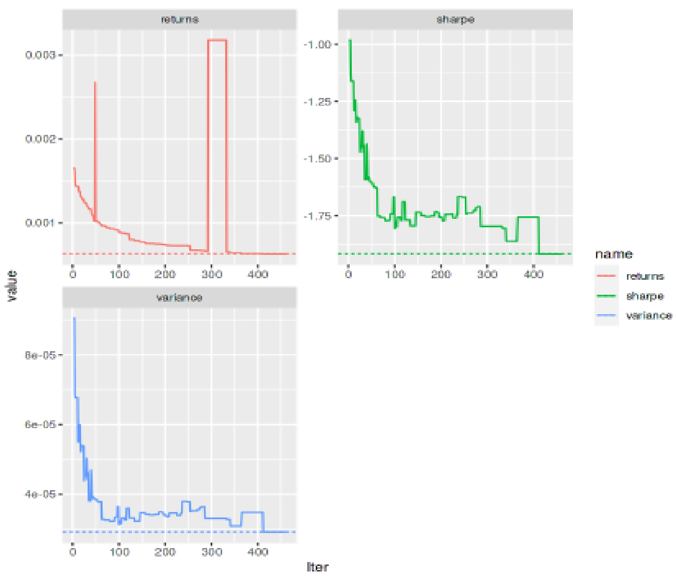
(a)



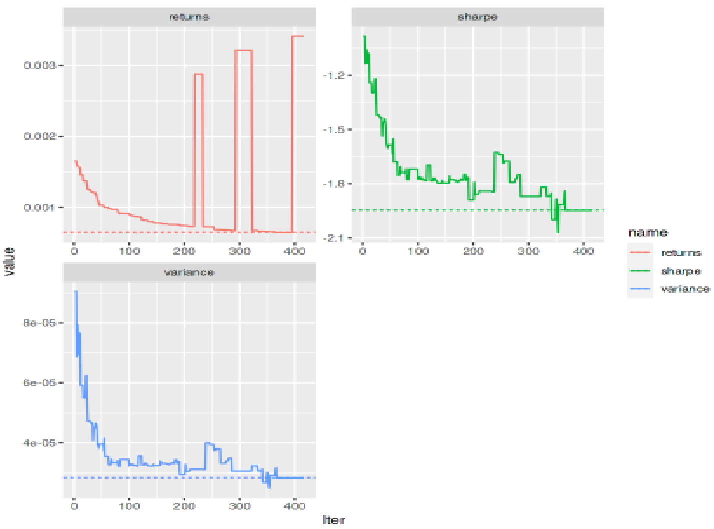
(b)



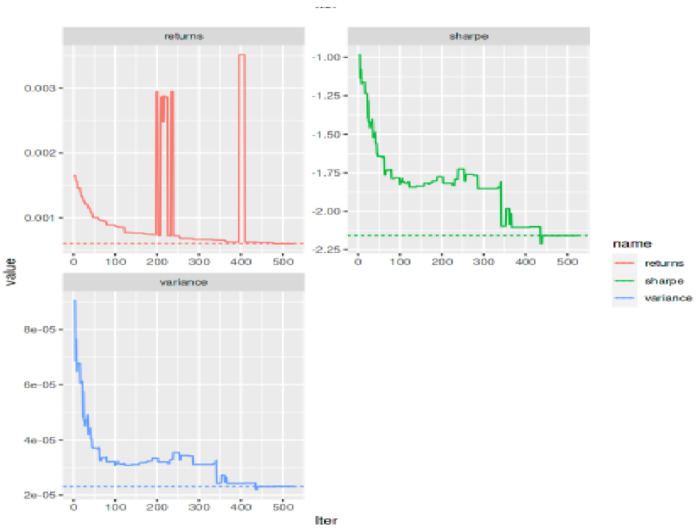
(c)



(d)



(e)



(f)

Figure 4.9: expected returns, variance, and Sharpe ratio for the best solution in each iteration vs the global best solution.

To evaluate the feasibility of these investment strategies we compared Genetic algorithm approach with Investing equal amounts on all the selected assets also randomly allocating the investment capital to the selected assets. GA approach yielded the highest expected returns when larger objective weight is placed on risk i.e greater than 50%. On the other hand, with risk averse investors the approach yields risk conservative strategies.

x	Model	Return	variance
0	Genetic	0.31%	0.17%
	Random	0.17%	0.01%
	Equal weight	0.13%	0.01%
0.2	Genetic	0.31%	0.18%
	Equal weight	0.13%	0.01%
	Random	0.13%	0.01%
0.4	Random	0.14%	0.01%
	Equal weight	0.13%	0.01%
	Genetic	0.03%	0.00%
0.6	Random	0.14%	0.01%
	Equal weight	0.13%	0.01%
	Genetic	0.04%	0.00%
0.8	Random	0.14%	0.01%
	Equal weight	0.13%	0.01%
	Genetic	0.04%	0.00%
1	Random	0.17%	0.01%
	Equal weight	0.13%	0.01%
	Genetic	0.04%	0.00%

Table 4.9: Evaluation of genetic algorithms in comparison to Equal and random capital allocation

Chapter 5: Conclusion

5.1 Summary

The study has successfully examined the use of genetic algorithms in constructing optimal investment portfolio, based on a universe of 391,108 assets. A multi - step approach was employed to select assets and allocate investment capital to each. Two optimization approaches were experimented, which include a single objective case maximizing Sharpe ratio and an objective case which involve minimizing sum of absolute deviation from a risk goal of 0 and end expected return of 1. Genetic algorithm was run with a 50-solution initial population, 5% elitism, 80% probability of crossover, 1% mutation probability, and 1,000 maximum iterations with an early stopping criterion of 50 consecutive iterations without any changes in the best solution. The maximum number of assets allowed to the portfolio was limited to 10 assets.

Maximization of Sharpe ratio resulted to a high-risk portfolio with risk rising to 49.24% during asset selection step and 787.31% capital allocation. we can therefore conclude that the strategy maximizes Sharpe ratio by allocating high capital proportion to high-risk asset. The approach is not friendly to risk adverse investors as it goes for strategies yielding higher return per unit risk, regardless of how risky the investment is. The portfolio from this approach yields better returns than investing on risk free opportunities with rates below 1.1%. Additionally, the approach outperforms random capital allocation or equal capital allocation.

The multi objective approach yield, is dependent on the objective weights. For risk seeking investors who would rather miss their risk goal than missing their return goal, the strategy allocates high amounts to risky assets. As the investor becomes more risk averse, the strategy allocates high capital proportion to less risk assets.

5.2 Research Limitations

In this research, the asset selection step is based on randomly selected 100,000 assets due to computational intensity, this means the whole asset universe was not exploited, hence good assets combinations might be left out of the constructed portfolio. Additionally, the inclusion criteria allow only assets which registered prices for at least 60% of the sampling time. This excludes newly listed assets from the experimentation.

The experimentation didn't consider important asset groups, such as industry asset types and sectors, hence exposes investors to risks affecting a specific grouping (sector, industry, or country) in case the recommended strategy is to invest all the capital on same grouping. For the multi objective approach, the study didn't consider experimenting different levels of risk aversiveness, which limits investors from selecting assets based on their risk aversiveness.

5.3 Future Research

From the limitations, we propose future research with enough computation power, larger asset universe and involving risk aversiveness in asset selection and Sharpe ratio computation. Additionally, deployment of this model on webapp or other accessible repositories, where users can use filters to select their desired combinations.

References

[1] H. Markowitz, “PORTFOLIO SELECTION*,” The Journal of Finance, vol. 7, no. 1, p. 77–91, 1952.

[2] W. F. Sharpe, “CAPITAL ASSET PRICES: A THEORY OF MARKET EQUILIBRIUM UNDER CONDITIONS OF RISK*,” The Journal of Finance, vol. 19, no. 3, p. 425–442, 1964.

[3] J. Holland, “Adaptation in natural and artificial systems : an introductory analysis with application to biology,” Control and Artificial Intelligence, 1975.

[4] A. Seetharaman, I. Niranjan, N. Patwa and A. Kejriwal, “A Study of the Factors Affecting the Choice of Investment Portfolio by Individual Investors in Singapore,” Accounting and Finance Research, vol. 6, no. 3, p. 153, 2017.

[5] M. Kritzman, “The Role of Factors in Asset Allocation,” The Journal of Portfolio Management, vol. 47, no. 5, pp. 58-64, 2021.

[6] J. Luan, Z. Yao, F. Zhao and X. Song, “A novel method to solve supplier selection problem: Hybrid algorithm of genetic algorithm and ant colony optimization,” Mathematics and Computers in Simulation, vol. 156, pp. 294-309, 2019.

[7] J. Ackora-Prah, S. A. Gyamerah and P. S. Andam, “A heuristic crossover for portfolio selection,” Applied Mathematical Sciences, vol. 8, p. 3215–3227, 2014.

[8] S. Babaei, M. M. Sepehri and E. Babaei, “Multi-objective portfolio optimization considering the dependence structure of asset returns,” European Journal of Operational Research, vol. 244, no. 2, p. 525–539, 2015.

[9] R. Faia, T. Pinto and Z. Vale, “GA optimization technique for portfolio optimization of electricity market participation,” 2016 IEEE Symposium Series on Computational Intelligence (SSCI), 2016.

[10] F. D. Paiva, R. T. N. Cardoso, G. P. Hanaoka and W. M. Duarte, “Decision-making for financial trading: A fusion approach of machine learning and portfolio selection,” Expert Systems with Applications, vol. 115, pp. 635-655, 2019.

[11] S. Gu, B. Kelly and D. Xiu, “Empirical asset pricing via machine learning,” The Review of Financial Studies, vol. 33, no. 5, pp. 2223-2273, 2020.

[12] R. Chowdhury, M. R. C. Mahdy, T. N. Alam, G. D. Al Quaderi and M. A. Rahman, “Predicting the stock price of frontier markets using machine learning and modified Black–Scholes Option pricing model,” Physica A: Statistical Mechanics and its Applications, vol. 555, p. 124444, 2020.

[13] M. Jiang, J. Liu, L. Zhang and C. Liu, “An improved Stacking framework for stock index prediction by leveraging tree-based ensemble models and deep learning algorithms,” Physica A: Statistical Mechanics and its Applications, vol. 541, p. 122272, 2020.

[14] S. Basak, S. Kar, S. Saha, L. Khaidem and S. R. Dey, “Predicting the direction of stock market prices using tree-based classifiers,” The North American Journal of Economics and Finance, vol. 47, pp. 552-567, 2019.

[15] Z. Wang, B. Huang, S. Tu, K. Zhang and L. Xu, “DeepTrader: A Deep Reinforcement Learning Approach for Risk-Return Balanced Portfolio Management with Market Conditions Embedding,” in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, 2021, pp. 643-650.

[16] K. H. Chang and M. N. Young, “Behavioral stock portfolio optimization considering holding periods of B-stocks with short-selling,” Computers & Operations Research, vol. 112, p. 104773, 2019.

[17] W. Wang, W. Li, N. Zhang and K. Liu, “Portfolio formation with preselection using deep learning from long-term financial data,” Expert Systems with Applications, vol. 143, p. 113042, 2020.

[18] D. Santos, A. D. D. Souza and F. Mota, Techniques and Tools for Selection and Strategic Alignment of Projects and Projects Portfolio Balancing: A Systematic Mapping, Springer, Cham, 2021, pp. 195-201.

[19] T. Fischer and C. Krauss, “Deep learning with long short-term memory networks for financial market predictions,” European journal of operational research, vol. 270, no. 2, pp. 654-669, 2018.

[20] M. Mlapenna, “GitHub - mlapenna7/yh_symbol_universe: Retrieves all the symbols available on Yahoo Finance,” 5 June 2021. [Online]. Available: https://github.com/mlapenna7/yh_symbol_universe. [Accessed 1 April 2023].

[21] N. A. Hali and A. Yuliati, “Markowitz Model Investment Portfolio Optimization: a Review Theory,” International Journal of Research in Community Service, 4 Octomber 2014.

[22] C.-F. Huang, “A hybrid stock selection model using genetic algorithms and support vector regression,” Applied Soft Computing, vol. 12, no. 2, pp. 807-818, February 2012.

[23] M. Montgomery, T. Rupasinghe and M. Kurz, “A Hybrid Portfolio Asset Selection Strategy Using Genetic Algorithms (GA),” Industrial Engineering Research Conference, May 2009.

[24] R. T. Marler and J. S. Arora, “The weighted sum method for multi-objective optimization: new insights,” Structural and Multidisciplinary Optimization, vol. 41, no. 6, pp. 853-862, 1 June 2020.

[25] Wang, W., Li, W., Zhang, N., & Liu, K. (2020). Portfolio formation with preselection using deep learning from long-term financial data. *Expert Systems with Applications, 143*, 113042.

[26] Wang, Z., Huang, B., Tu, S., Zhang, K., & Xu, L. (2021). DeepTrader: A Deep Reinforcement Learning Approach for Risk-Return Balanced Portfolio Management with Market Conditions Embedding. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 35, pp. 643-650).

[27] Wang, C., Chen, Z., Liu, X., & Zhang, X. (2021). Maintenance scheduling optimization in manufacturing systems using genetic algorithm. The International Journal of Advanced Manufacturing Technology,113(11-12), 3687-3704.

[28] Wang, R., Cui, X., & Zhang, L. (2021). Optimization of warehouse layout design based on genetic algorithm. Computers & Industrial Engineering, 160, 107916.

[29] Wu, Y., Cai, X., & Yang, H. (2021). Optimization of laser cladding process parameters using a hybrid genetic-neural network algorithm. Metals, 11(7), 1088.

[30] Zhao, L., Zhang, H., & Lu, Y. (2021). Optimization of location-routing problem using genetic algorithm in supply chain. Journal of Industrial and Production Engineering, 38(4), 281-294.

[31] TEAM, C. (2023, May 9). Sharpe ratio. Corporate Finance Institute. https://corporatefinanceinstitute.com/resources/risk-management/sharpe-ratio-definition-formula/