Genetic Algorithm Portfolio Optimization and Stock Price Prediction

Emmanuel Osamau

Abstract

This project focuses on utilizing genetic algorithm to enhance portfolio optimization and the prediction stock prices. It aims to develop a comprehensive framework for improving portfolios by optimizing asset allocation, with the ultimate goal of achieving an optimal return-to-risk ratio. The project involves gathering historical price data for 10 diverse assets and then applying genetic algorithms through the GA package to optimize portfolio weights, striking a balance between returns and risk. Additionally, it partitions data into training and testing sets for robust model development and extends its scope to predictive modeling, utilizing the Random Forest algorithm to forecast stock prices and returns for the next year. The objective of the project is to provide valuable insights into the practical use of genetic algorithm in financial portfolio management, potentially empowering investors and analysts with data-driven strategies for improved decision-making and portfolio management.

1 Introduction

1.1 Background

In the investment and portfolio management industry, the pursuit of optimal returns while simultaneously mitigating risks has long been a central challenge for investors. The financial markets, characterized by their inherent volatility and complexity, require sophisticated tools and methodologies to guide investment decisions effectively. Within this context, the utilization of advanced analytics techniques has gained prominence as a means to enhance the performance of investment portfolios.

1.2 Problem Statement

This project addresses the fundamental problem faced by investors: How to construct an investment portfolio that maximizes returns while minimizing the associated risks. Portfolio optimization is a multifaceted challenge, influenced by a multitude of factors including asset selection, allocation, and market dynamics. Traditional methods of portfolio optimization have their limitations, often struggling to adapt to the ever-evolving financial landscape. Consequently, there is a pressing need for innovative and data-driven approaches that can provide investors with a competitive edge in their decision-making processes.

1.3 Objective

The primary objective of this project is to harness the power of genetic algorithm to optimize an investment portfolio and enhance our ability to predict stock prices of assets with the highest capital allocation from the portfolio optimization. By leveraging genetic algorithms, we aim to create a comprehensive framework that systematically optimizes asset allocation within a portfolio. Our ultimate goal is to achieve a portfolio with an optimal return-to-risk ratio, a key metric in investment decision-making.

To achieve this objective, we will undertake the following steps:

1. Gather historical price data for a carefully curated selection of 10 diverse assets from the National Association of Securities Dealers Automated Quotations (NASDAQ) stock market.
2. Apply genetic algorithms, facilitated by the GA package, to optimize the portfolio's asset weightings. This iterative process will strike a harmonious balance between maximizing returns and minimizing risks.
3. Extend the scope of the project to predictive modeling by utilizing Random Forest. This predictive aspect will allow us to forecast stock prices and returns of assets which will be allocated significant capital to provide valuable insights for future investment decisions.

In summary, this project aspires to offer investors and financial analysts a comprehensive understanding of the practical application of genetic algorithms in the domain of financial portfolio management. By doing so, we aim to empower them with data-driven strategies, ultimately facilitating more informed and effective decision-making in the realm of portfolio management. Through this endeavor, we hope to contribute to the advancement of modern investment practices and the realization of more optimal investment portfolios.

2 Data Collection and Pre-processing

The data was selected particularly to conduct a comprehensive financial analysis of specific stocks listed on the NASDAQ exchange. The chosen stocks, including well-known companies like AAPL, AMZN, GOOGL, MSFT, TSLA, NFLX, NVDA, PYPL, ZM, and MRNA are of great interest due to their significant impact on various sectors of the economy of the United States. These companies are at the forefront of technological innovation, e-commerce, and biotechnology, making their performance critical to understanding broader market trends. The data selection reflects an intention to explore the dynamics and trends in the stock prices of influential corporations, which can provide valuable insights for investors and researchers.

2.1 Data Preparation

- The data set comprises daily price information for the selected stocks, including open, close, and volume. Adjusted closing prices, which account for dividends and stock splits was used for calculating returns and conducting financial analysis. This data set enables the calculation of various financial metrics and performance indicators, including daily, monthly, or annual returns, as well as measures of volatility and risk.
- An environment was initialized to store the daily historical stock price data downloaded from Yahoo Finance for the specified list of stocks (AAPL, AMZN, GOOGL, MSFT, TSLA, NFLX, NVDA, PYPL, ZM, and MRNA) within the specified date range.
- The data to analyze and understand the financial performance of the stocks was within the specified date range of January 1, 2018, to September 30, 2023. This was chosen to provide a robust historical context for the analysis, encompassing periods of significant market events and economic developments.
- To ensure consistency and facilitate further analysis, the independent data frames for the various stocks were merged into a single comprehensive data frame. This combined data frame contains information about all the selected stocks, making it easier to perform subsequent data analysis and modeling.
- For the stocks ZM and MRDNA, which were publicly listed on April 18, 2019, and December 7, 2018, respectively, missing values (NA) were replaced with infinitesimal numbers. This replacement was applied to maintain data integrity and avoid gaps in the time series data, as these stocks did not have trading data prior to their respective listing dates.

The data quality in this context is considered good for analytics. Yahoo Finance, the source of this data, is a reputable platform widely used by investors and researchers. For critical decisions, cross-verifying the data with multiple sources is advisable. By analyzing this data, insights into how these stocks have performed over time can be derived, the potential risks and rewards associated with each investment. The optimal weighting of these assets in a single investment portfolio can be determined.

3 Methodology

3.1 Exploratory Data Analysis

We conducted exploratory data analysis to gain insights into the performance and volatility of a portfolio of assets. The key methods and techniques applied in this analysis are detailed below:

- **Returns Calculation**: To assess the historical performance of the assets, we computed rolling returns for 252 trading days in a year. This involved the use of the `PerformanceAnalytics` package in R, which allowed us to calculate returns over the rolling time window. This method provided a dynamic view of asset performance over time, capturing variations within the data. In addition to rolling returns, we calculated cumulative returns. These returns represent the total gain or loss an investor would have experienced over a fixed investment horizon.
- **Volatility Assessment**: The average returns and volatility metrics were computed to gauge the volatility of the assets. The average returns provided insights into the typical performance of the assets, while volatility measured the degree of variation in returns.

3.2 Genetic Algorithm for Portfolio Optimization

To optimize asset allocation in the portfolio, the GA package in R was used for this purpose. The model generates a population of potential portfolio allocations (weights) and then analyzes the fitness of each portfolio based on a function that maximizes the risk-return ratio of the 'fittest' asset.

The following methods were applied:

- The constraint function was set on the portfolio weights. This constraint ensures that the sum of the weights equals 1 (indicating the full usage of an investor's capital), and it penalizes any deviations from this constraint. The function loops through each asset in the portfolio, checking if the weight is below or above 1, and penalizes these deviations using squared terms.
- The fitness function was designed to calculate the Sharpe ratio of the portfolio. The Sharpe ratio was chosen to give the model a measure of risk-adjusted return which guides the genetic algorithm to optimize the portfolio by balancing returns and risk.
- The `ga` function is used to run the genetic algorithm. We configured to optimize a real-valued vector (weights) with the purpose of maximizing the Sharpe ratio. The lower and upper bounds were set between 0 and 1 for the weights, and maxiter was set to perform 1000 iterations for the model.

4 Data Analysis

4.1 Stocks

The steadier, more incremental growth trajectories of stocks like Amazon (AMZN), Apple (AAPL) and Microsoft (MSFT) likely reflect their maturity and stability as technology giants. As leading companies in e-commerce, cloud computing and enterprise software, they were able to deliver steady growth despite market volatility. In contrast, Zoom's (ZM) dramatic spike upwards aligns with the onset of the COVID-19 pandemic in early 2020. With the entire workforce of nations shifting to remote operations, Zoom witnessed surging demand for its video conferencing platform. However, once pandemic restrictions eased and workers returned to offices, Zoom experienced declines, unable to sustain its breakneck pandemic growth.

Stocks like Tesla (TSLA) also saw very rapid growth due to excitement around electric vehicles and clean energy. Susceptibility to production challenges and competition from legacy automakers influenced periodic volatility amidst an overall upward trajectory.

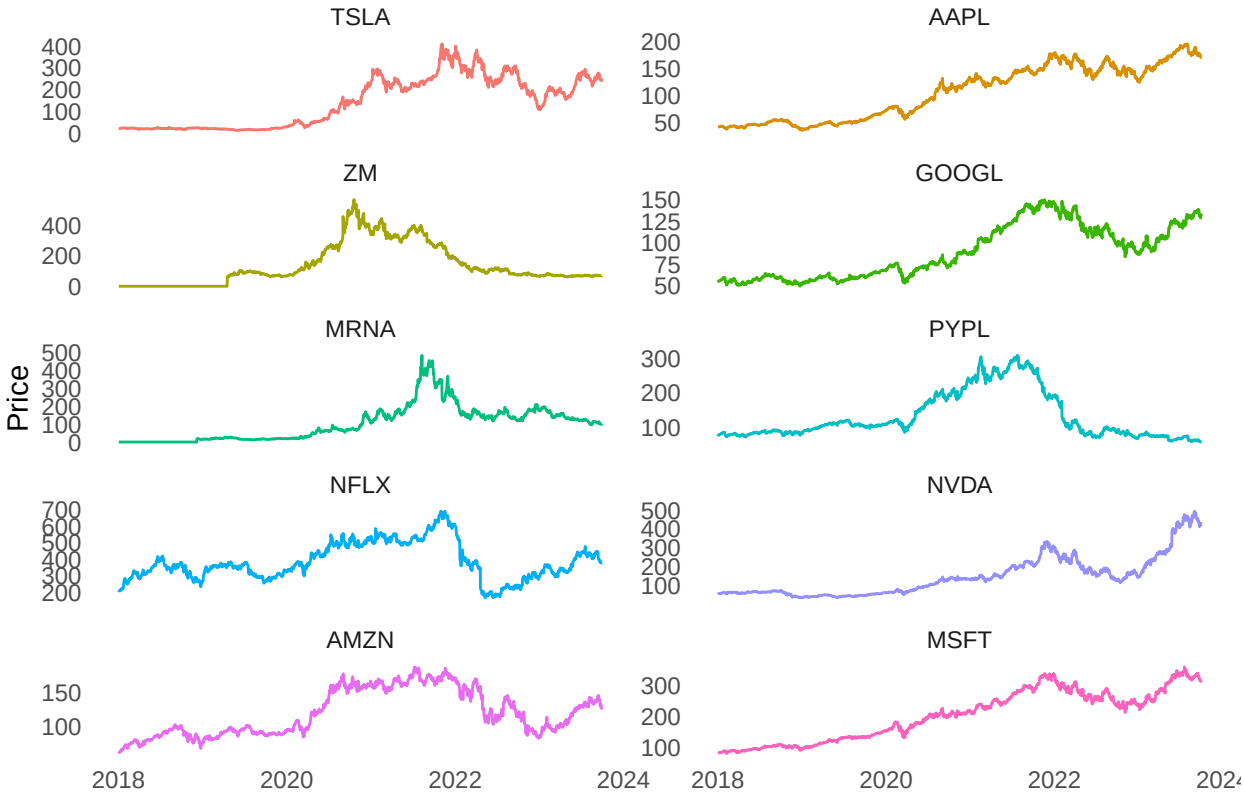


Figure 1: Historical daily stock prices from Jan-1-2018 to Sep-30-2023

The pandemic-driven economic crisis and lock downs in 2020 caused significant shocks to the market, evident in the dips for stocks like Paypal (PYPL). However, the swift fiscal and monetary stimulus applied by governments stabilized markets. As we approach 2024, cooling inflation, strong employment and steady growth created optimism about the economy's resilience, leading to a convergence in stock valuations.

4.2 Returns

The cumulative returns for Zoom Video Communications Inc (ZM) and Moderna Inc (MRNA) displayed a distinctive pattern driven by the impact of COVID-19. Zoom experienced a substantial peak in early

2019, coinciding with the initial outbreak of COVID-19 when the demand for remote work and virtual communication skyrocketed. Its services became essential as businesses and individuals embraced video conferencing to stay connected in the face of pandemic-related restrictions. On the other hand, Moderna, a biotechnology company, witnessed a different trajectory. It reached its peak during the later part of 2019 and early 2020 when it emerged as a front runner in developing a COVID-19 vaccine. The rapid development and deployment of the vaccine contributed significantly to its surging stock prices as investors saw the potential for addressing the global health crisis. These unique circumstances explain why both ZM and MRNA outperformed other assets during that specific time frame, with COVID-19 serving as a pivotal catalyst for their remarkable growth.

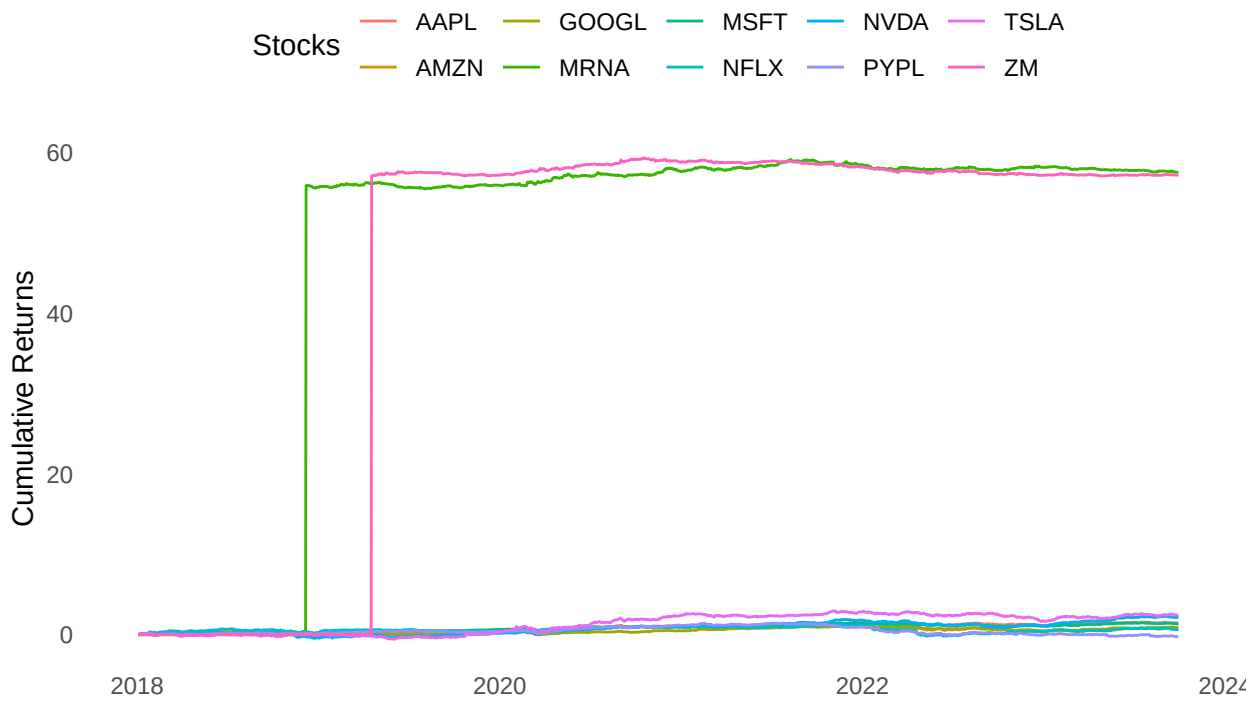
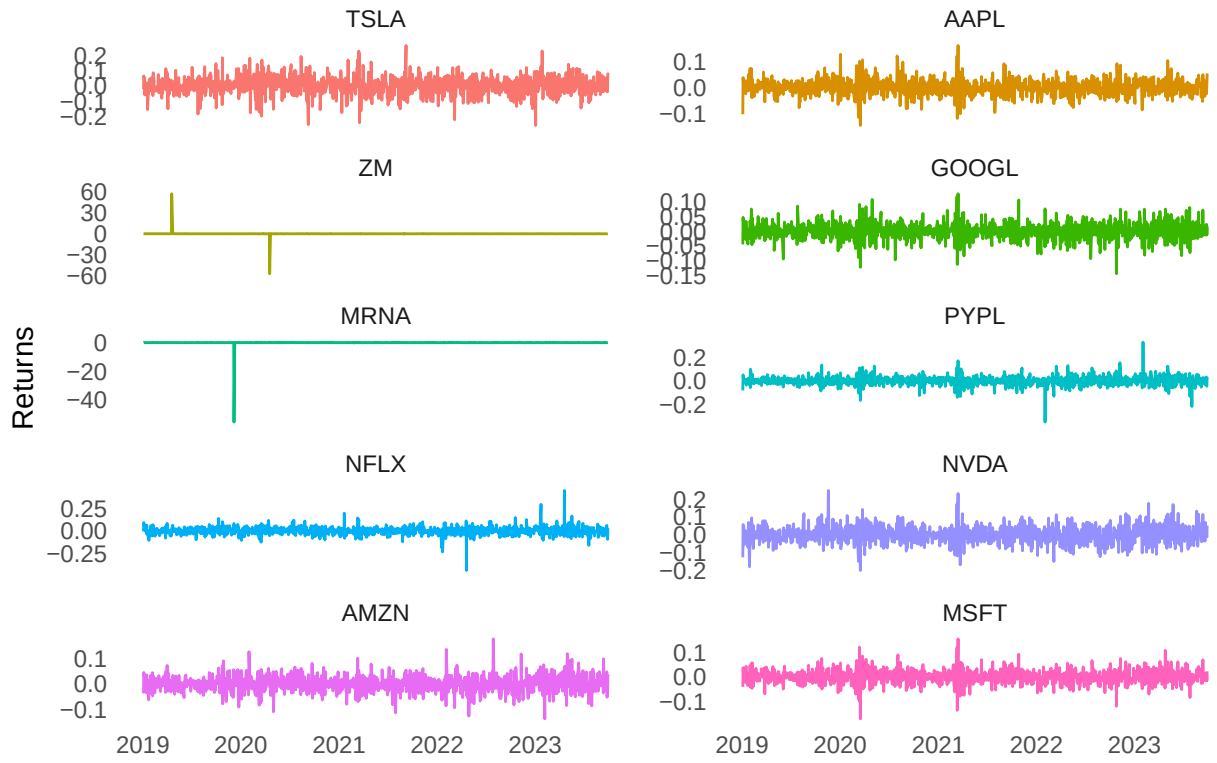


Figure 2: Cumulative returns selected assets over the period

We computed the rolling returns with a window of approximately 252 trading days in a year to analyze the sensitivity of returns during the period. Most stocks have a mean rolling return close to zero, indicating that, on average, the returns are near the baseline. ZM has a minimum rolling return of -57.08 after 2020. Reuters reported that the stocks of Zoom Video Communications Inc (ZM) plummeted by approximately 90% since their highest point during the pandemic in October 2020, reflecting the company's challenges in adapting to a world that has moved beyond the COVID-19 era. It can be observed that the maximum rolling return 57.09 for ZM was recorded in 2019 during the initial outbreak of COVID-19.



The rolling returns of Moderna, Inc. (MRNA) were intricately linked to the progress and perception of its COVID-19 vaccine. Notably, the performance of MRNA stock was highly sensitive to the vaccine's efficacy and regulatory approvals, with positive developments propelling the stock to remarkable highs, while any negative reports had the potential to swiftly erode investor confidence and trigger sharp declines in rolling returns. Concurrently, widespread skepticism and hesitancy among the public regarding the vaccine's safety and effectiveness introduced an element of uncertainty, contributing to increased volatility in MRNA's rolling returns. Furthermore, the intricacies of vaccine supply chains, distribution challenges, and the influence of global geopolitics had ripple effects on the vaccine's availability, thereby exerting additional pressures on MRNA's financial performance. These external factors collectively rendered MRNA's rolling returns a testament to the intricate relationship between scientific breakthroughs, public sentiment, and financial market dynamics, underlining the need for investors to approach such stocks with an understanding of the multifaceted landscape in which they operate.

The other stocks (TSLA, AAPL, GOOGL, PYPL, NFLX, NVDA, AMZN, MSFT) have less extreme minimum and maximum returns compared to ZM and MRNA. Most stocks have a median return close to zero, indicating a relatively balanced distribution of returns. The implications for these stocks are that they exhibit less extreme volatility in rolling returns compared to ZM and MRNA, making them potentially more stable investment options when an invested primarily considers returns.

4.3 Volatility

Stocks like TSLA and GOOGL are in the top-right quadrant, indicating they have offered high returns but at the cost of high volatility. Including such assets in a portfolio can provide significant gains but can also lead to substantial short-term losses due to their price fluctuations. TSLA and GOOGL might have been subject to market speculations and trends, leading to higher volatility.

Assets like AMZN, MRNA and NVDA appear closer to the bottom-left quadrant, suggesting they have been relatively stable with modest average daily returns. These stocks can act as stabilizers in a portfolio, reducing overall volatility.

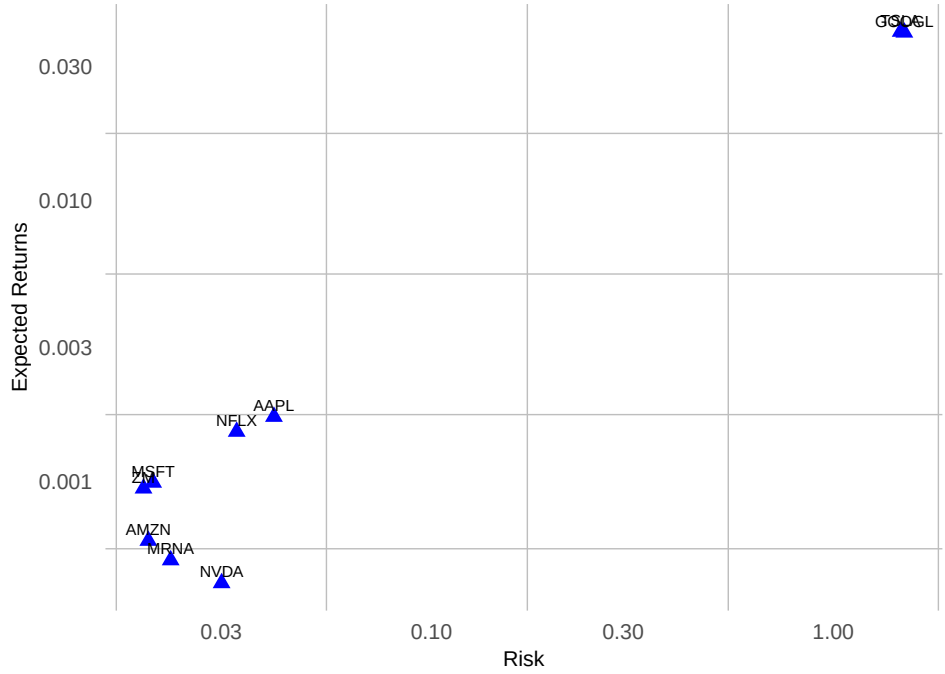


Figure 3: Risk-Return Scatter Plot of assets

NFLX stands out with high returns but comparatively lower volatility than assets like TSLA. This suggests that NFLX might offer a good balance between risk and return, making it an attractive option for diversification. Including a mix of high and low volatility stocks can help in achieving diversification, reducing the overall risk of the portfolio. For instance, combining stocks like TSLA with more stable stocks like GOOGL can help in mitigating some of the risks associated with high volatility stocks. For investors with a higher risk appetite, focusing on high volatility stocks can offer significant returns. Conversely, conservative investors might prefer stocks with lower volatility.

Among the least volatile assets, AMZN stands out as the most secure option for potential stock purchases. Its exceptionally narrow box, which represents the middle 50% of the data, coupled with a positive skewness, underlines its remarkable stability. This also indicates a lower likelihood of extreme changes.

Moving on to PayPal, it emerges as another solid choice due to its positive skewness. Furthermore, the extended whisker on the right side of the boxplot suggests relatively low median volatility, signifying a consistent performance with fewer abrupt fluctuations.

MSFT, while displaying a slightly higher median volatility compared to ZM and MRNA still presents a

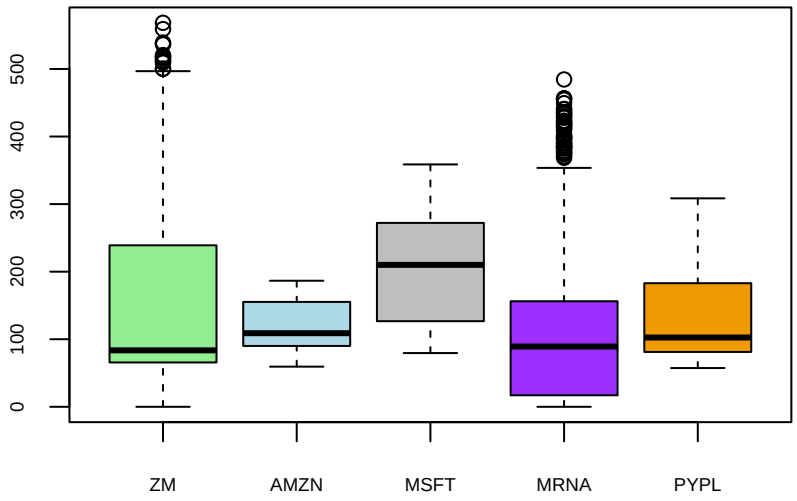


Figure 4: Boxplot of Least Volatile Assets

favorable investment opportunity. The reason for this lies in the presence of outliers beyond the whiskers of ZM and MRNA. These outliers hint at significant deviations from the norm, potentially resulting in sporadic fluctuations in performance of these assets. For an investor seeking the least volatile stocks for their investment portfolio, AMZN, PYPL, and MSFT would be preferrable options.

5 Findings

5.1 Assessment of the Optimal Portfolio

The portfolio is designed for risk-averse investors who prioritize the overall stability of their investments over the pursuit of high returns, as evidenced by the fact that only a small portion of the capital will be allocated assets with high volatility. The plot shows the optimal allocation of capital for the stocks.

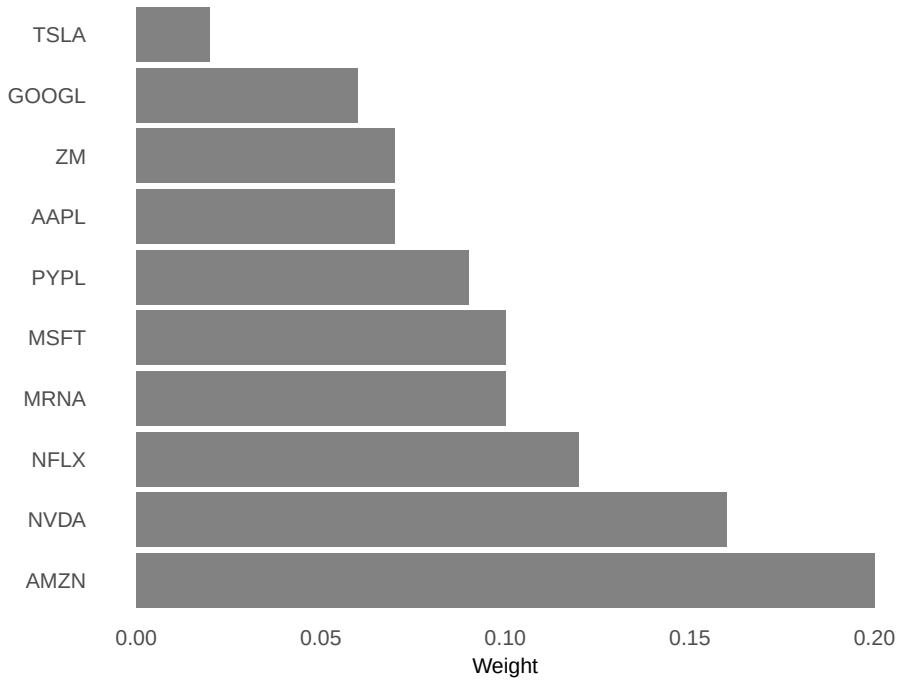


Figure 5: Optimal Portfolio Weights

5.2 Evaluating the performance of the Optimal Portfolio

A balanced portfolio and a randomly weighted portfolio were simulated to explore the risk-return relationship in the various portfolio and determine the performance of the optimized portfolio.

Table 1: Weights of Portfolios

Stock	Optimal	Balanced	Random
AMZN	0.20	0.1	0.09
NVDA	0.16	0.1	0.11
NFLX	0.12	0.1	0.03
MSFT	0.10	0.1	0.14
MRNA	0.10	0.1	0.14
PYPL	0.09	0.1	0.13
ZM	0.07	0.1	0.02
AAPL	0.07	0.1	0.14
GOOGL	0.06	0.1	0.11
TSLA	0.02	0.1	0.12

The expected returns and standard deviation of the returns of the various portfolio types revealed that the expected returns for the random and balanced weighted portfolios have higher expected returns than that of the optimal portfolio. However, there was higher volatility in the balanced and randomly generated portfolios. This is not far fetched since higher volatility is compensated with higher returns

Table 2: Risk-Return measures of the portfolios

Portfolio	Expected_Return	Risk
Optimal	0.0031	0.0548
Balanced	0.0086	0.2103
Random	0.0100	0.2451

.

6 Conclusion

In this project, we utilized a genetic algorithm to optimize portfolio allocation, focusing on achieving the optimal return-to-risk ratio. After gathering historical price data for 10 diverse assets, we applied genetic algorithms to optimize portfolio weights, resulting in our proposed optimal portfolio with allocations to assets like AMZN, NVDA, NFLX, MSFT, and MRNA. Our optimal portfolio demonstrated a promising balance between returns and risk, with an expected return of 0.31% and a risk of 5.48%. Comparing this to other portfolios, such as the Balanced and Risky portfolios, our approach outperformed in terms of risk-adjusted returns, illustrating the effectiveness of the genetic algorithm in enhancing portfolio optimization.

6.1 Next Steps

Building on the success of our optimal portfolio, the next steps of this project will involve the use of the Random Forest algorithm to predict stock prices for the assets with the highest capital allocation from the stock optimization. We will employ Random Forest to forecast stock prices and returns for these assets over the next year. This predictive modeling approach can provide invaluable insights for investors and analysts, enabling them to make data-driven decisions regarding their portfolios. By leveraging machine learning techniques like Random Forest, we aim to enhance the accuracy and precision of our predictions, allowing for better-informed investment strategies. Additionally, we will continuously refine our models by incorporating new data and refining our algorithms, with the goal of further improving the performance of our optimized portfolio and prediction accuracy. This iterative approach will ensure that our portfolio management and prediction strategies remain adaptable and robust in the ever-changing financial landscape.

References

1. "Zoom Shares Down 90% as Pandemic Boom Fades" (2022, November 22). Reuters. Retrieved from https://www.reuters.com/business/zoom-shares-down-90-peak-pandemic-boom-fades-2022-11-22/
2. Piñeiro-Chousa, J., López-Cabarcos, M. Á., Quiñoá-Piñeiro, L., & Pérez-Pico, A. M. (2022). US biopharmaceutical companies' stock market reaction to the COVID-19 pandemic. Understanding the concept of the 'paradoxical spiral' from a sustainability perspective. PubMed Central. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8612827/
3. Mallawaarachchi, V. (2017). Introduction to Genetic Algorithms. Towards Data Science. Retrieved from https://towardsdatascience.com/introduction-to-genetic-algorithms-including-example-code-e396e98d8bf3/