PORTFOLIO OPTIMIZATION: A COMPARATIVE ANALYSIS BETWEEN TRADITIONAL AND MACHINE LEARNING

TAKUDZWA MAKUWA

STUDENT NUMBER: G19M0551

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTERS IN COMMERCE (FINANCIAL MARKETS)

DEPARTMENT OF ECONOMICS AND ECONOMIC HISTORY

RHODES UNIVERSITY, GRAHAMSTOWN

SUPERVISOR: PROFESSOR MESHACH AZIAKPONO

CO-SUPERVISOR: MR DELON TARENTAAL

ABSTRACT

Globally, there has been a rapid increase in machine learning (ML) integration within the financial field due to its ability to analyse big data and learn complex patterns to make informed decisions quickly compared to traditional practices. Empirical studies have investigated which traditional methods, such as mean-variance optimization, maximum Sharpe ratio optimization, risk parity, and equal-weighted portfolio, yield optimal returns when compared against each other and similarly when compared against ML methods such as linear regression, decision tree, random forest, and sector vector machine regression model. These studies reflect mixed results where one traditional method, such as mean-variance, outperforms the maximum Sharpe ratio portfolio and, in some cases, is the opposite. However, most literature agrees that implementing machine learning methods will provide higher risk-adjusted returns than traditional methods. While South Africa has surged in research regarding the use-case of machine learning since 2016, it has yet to explore ML's impact within the investment context fully. This study will investigate which traditional and ML methods will yield the highest composite score and which sector has consistently contributed to the optimal portfolio.

The study uses daily observations from 30 September 2019 to 30 September 2024, to construct 20 portfolios from randomly selected 27 stocks. In traditional methods, only daily closing prices were used, and in ML, daily opening, closing, high and low prices, and trading volume were used. This study used the following traditional methodologies: (i) Equal Weighted, (ii) Mean Variance, (iii) Maximum Sharpe Ratio, and (iv) Risk Parity optimization. Similarly, machine learning methods consist of (i) Linear Regression, (ii) Decision Tree Regression, (iii) Random Forest Regression, and (iv) Sector Vector Machine Regression models. The evaluation for the predictive analysis includes (i) Mean Absolute Percentage Error, (ii) R-squared, and (iii) Theil's U2. Additionally, the portfolio performance was evaluated by (i) Sharpe ratio, (ii) Tracking Error, (iii) Sortino Ratio, (iv) Information ratio, (v) Beta, and (vi) Treynor Ratio. Finally, a composite z-score ranking system was adopted to determine the optimal portfolio. Our analyses were conducted by executing a series of commands on Python via Google Colab.

In traditional methods, results indicated that the Equal Weighted Portfolio had the highest composite score. By contrast, ML tools enhanced each traditional method on 20 different portfolios and obtained an accuracy score between 75% and 98% for all predictive parameters. For ML optimization, it was determined that linear regression was the best ML model, and it was evident that the Mean-Variance portfolio outperformed the Equally Weighted, Mean-Variance, Maximum Sharpe Ratio, and Risk Parity portfolio. The study finds that ML outperforms traditional methods based on the aggregated score. The study's outcome recommends that South African investment professionals integrate ML into their investment strategies as it would be a rewarding opportunity.

DECLARATION OF WORK

This page declares that this thesis, which I now submit for the Master of Commerce in Financial Markets degree, is my own while completing the degree at Rhodes University. I, Takudzwa Makuwa, certify that this thesis has not been submitted for a degree in any other university, Technikon, or college; that is my original work.

Name: Takudzwa Makuwa

Signed: T. Makuwa

Date: 7 January 2025

ACKNOWLEDGEMENTS

I want to take this opportunity to thank my supervisors, Professor Meshach Aziakpono and Mr Delon Tarentaal, for allowing me to continue this research. The journey has been difficult, and I appreciate the kindness and patience. I would also like to thank the RU Economics Department for training me to be the best economist I can be. The Department has offered various research workshops throughout the year, and I found those to be insightful as they improved my research and writing skills. I also want to thank the Rhodes University Community Engagement Division for offering research-engaged workshops throughout the year; I found them immensely insightful. The advice, support, and continuous check-ins gave me a boost and the strength to continue.

I am grateful to my classmates for their endless support in the research. Thank you for answering my silly questions about how to run and execute models. Seeing the enthusiasm to finish and bag the degree was also encouraging. Even though I didn't sleep in the library in the 24-hour section almost daily, it was my home away from home. However, your determination and your persistence kept me going. A little birdy told me that I should start calling you "Fraudsters" as you continued to cook and bag the degree. I want to thank Mthokozisi S. for helping with the code for this paper—much appreciated, brother.

I want to thank my Breaking News with Boots Tk Fans and WhatsApp Community for their continuous support in my academic life. I know I have deprived you all during the last days of my submission, but your encouragement and support motivate me to finish this paper means a lot to me. Baie Danko. Finally, I would like to thank my parents and friends, who have continuously supported me throughout my university and academic career. You have inspired me to reach greater heights; I do not take it for granted. May you be blessed with many blessings.

Enkosi and Baie Danko!!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1. RESEARCH CONTEXT

Globally, there is an estimated 64% of investment managers that are currently employing or planning to employ machine learning in their asset management which will significantly rise to 71% by including the young professionals in the industry (Chartered Financial Analyst, 2024). The most common machine learning methods that are used in the industry consist of deep learning models and neural networks which are powerful algorithms (El Hajj and Hammoud, 2023). Chartered Financial Analyst (CFA) Institute (2024) provided a survey conducted by Invesco Global Systematic Investing Study (2023), which aimed to identify how investment professionals are using machine learning (ML) in the investment process. The study revealed that 86% of investors use it to optimize portfolio allocation and risk management, 76% develop and test investment strategies, and 55% monitor and adjust portfolio allocation, accordingly, pending market changes. The recent adoption of ML among investors was due to its outperformance compared to traditional methods. Additionally, CFA Institute (2024) indicated that from a sample of 119 investment professionals, on average, 75% of investors from institutional and wholesale indicated that ML offers better risk management, and 46.5% of investors suggest ML can improve portfolio diversification. However, there is limited research on whether ML can outperform traditional methods, specifically in South Africa. According to Ezugwu, Oyelade, Ikotun, Agushaka, and Ho's (2023) evaluation of 2761 machine learning-related documents in Africa, South Africa had published 23% of 2 468 articles between 1991 and 2021 using bibliometric analysis review. These include articles from all professions, such as healthcare, finance, and other sectors. By contrast, Buczynski, Cuzzolin, and Sahakian's (2021) literature review from examining 27 academic papers selected over 20-year period shows no indication of ML optimization with a distinct number of stocks to obtain portfolio performance subjected to the level of return at a given risk.

The core of machine learning being integrated into portfolio optimization and in the financial market system lies in the principle of Morden Portfolio Theory (MPT) which was introduced by Harry Markowitz (1952). The MPT aims to maximise returns while minimizing risk through portfolio diversification. Thus, MPT is somewhat linked to mean-variance optimization or portfolio as the technique is primarily driven by maximising returns and minimizing risk at the same time. On the contrary, this optimization technique is derived from the utility theory principle (Markowitz, 1958). For every stock invested by investors, each stock will exhibit a level of return and risk. There are many traditional methods of optimizing stocks as described by Gunjan and Bhattacharyya (2023) study. Gunjan and Bhattacharyya (2023) showed how it is possible to generate returns at a level of risk subjected to the investor goals in which traditional methods include: (i)Variance with Skewness (Samuelson, 1975), (ii) Mean-Absolute Deviation (Konno and Yamazaki, 1991), (iii) Value-at-Risk (Jorion, 1997), (iv)Minimax (Young, 1998), (v) Mean-Variance (Markowitz and Todd, 2000), (vi)

Conditional Value-at-Risk (Rockafellar and Uryasev, 2000). The main underlying characteristics of the traditional methods include technical abilities which often derive from simple mathematical derivations and require quadratic equations to determine portfolio weights.

To evaluate the performance of the portfolio, the most common practices according to several academic scholars such as Liu, Zhou, and Zhu (2020), Nagy and Benedek (2021), Vidal-García and Vidal (2022), and Hovhannisyan (2023) is the Sharpe ratio. The Sharpe ratio is the return per unit of standard deviation (Sharpe, 1966; Tarentaal, 2023a). The Sharpe ratio has been widely used as a benchmark of comparing various computed portfolios. Generally, investment professional prefers high Sharpe ratio as it indicates better investment returns relative to a given risk. Consequently, the covariance matrix constructed by Markowitz (1952) provided a pathway of Capital Asset Pricing Model and the Arbitrage Theory. When unpacking the entire MPT, it is well-known that there are two types of risk which are systematic and unsystematic risk (Munetsi, 2018). The systematic and unsystematic risks are associated with the market and actual asset respectively. To increase portfolio diversification, it is important to increase the number of stocks in the portfolio which narrows unsystematic risk overall (Elton, Gruber, Brown, and Goetzmann, 2009).

It is essential to recognize that there are several ways to construct a portfolio to satisfy the requirements of having a diverse portfolio. This can be done by selecting assets from different country, sector, and organization levels. Based on traditional literature, asset selection is driven by diversifying organization-level assets while neglecting sector and country-level diversification (Gupta and Basu, 2009). Diversifying assets at the sectoral and country levels has been seen as an additional feature of including it in the analyses of assets without being a primary decision indicator of stock selection.

South Africa offers an attractive investment opportunity due to its broad investment options, from equity to bonds to future indices and many other financial assets. However, the economy's performance only allows some sectors to perform at the same level (Muchaonyerwa and Choga, 2015). Stocks, particularly equities, tend to increase in value and trading volume when the economy is in the expansion phase; however, they perform worse during the contraction phase of the economy. Other types of financial assets, such as low-risk and dividend stocks, react similarly (Muriuki, 2003). Therefore, it opens a broad investment opportunity to various sectors that may yield different expected returns.

The Johannesburg Stock Exchange (JSE) currently has "more than 800 listed securities and approximately 400 listed companies, together with 60 equity market member firms. In addition, the JSE has comprehensive trading, statistical index, and non-index data nearing 200 Indices" (JSE Group, 2024a). It is the largest stock exchange in Africa and ranked 19th in the world by market capitalization (Kruger, 2024). The financial markets system within South Africa (SA) is unique compared to other countries. According to Industry Classification Benchmark, all listed companies in South Africa are grouped to three main sectors: Financials, Resources, and Industry. These three main sectors consist of the SA Resources classification, which is associated with basic materials and the energy sector. Secondly, the SA Financial classification consists of Financial and Real Estate assets. Finally, the SA Industrial classification caters to the remaining assets (JSE Group, 2024b). Moreover, if JSE's financial assets are further broken down, it will have 10 Industries, 20 Super-sectors, 45 Sectors, and 173 Subsectors (JSE Group, 2024c). Having a variety of financial market classifications to choose from, investors have too many choices to select from to maximize returns and minimize risks through sector diversification for their portfolio.

As noted earlier, there are two approaches to optimizing a portfolio, which can be described as traditional optimization methods and machine learning optimization. Traditional methods only capture and consider historical data when conducting analysis; however, they do not consider the changing market conditions (Botunac, Bosna, and Matetić, 2024). Some unforeseen rapid changes in the market include policy announcements, economic shocks, and other factors affecting stock prices (Škrinjarić and Orlović, 2020). By contrast, machine learning (ML) techniques help mitigate traditional optimization methods and limitations. ML includes anomalies, captures immediate changes in stock prices after changes in market conditions, and captures patterns of stock prices, which provides a proactive approach to portfolio optimization (Tacchini, 2024; Kang, Templeton, Kwak, and Um, 2024). ML makes it an appealing opportunity to enhance risk-adjust returns. However, the question of which approach performs better in portfolio optimization remains, especially given the recency of the ML approach.

Moreover, investors rely on two methods of investment decision, which include technical analysis, which focuses on historical trends and asset prices; however, the second method is the fundamental analysis of the company, which includes key financial performance metrics such as price-earnings ratio, marginal liquidity ratio, and many others. It is possible to integrate key financial performance metrics through coding using Python, which can help analyse information relatively fast compared to the traditional approach of calculating the returns and risks of assets (Zhang, 2023). Relatively, ML techniques are conducted in Python and other computational software, such as R programming; however, it depends on what is required to analyse and estimate, as each software has advantages and disadvantages.

1.2. RESEARCH OBJECTIVE AND QUESTIONS:

The primary objective of this research is to explore the potential portfolio performance by comparing portfolio optimization methods, including traditional and machine learning optimization methods, with a particular focus on sectoral analysis. The key question addressed in this study:

1.2.1. Which methodology- traditional or machine learning provides the best portfolio performance?

The above research question has underlying sub-research questions that include:

1.2.1.1. Which traditional method portfolio is the best in achieving portfolio diversification and performance?
1.2.1.2. Which machine learning method portfolio is the best in achieving portfolio diversification and performance?
1.2.1.3. Which sectors have demonstrated consistency superior results in contributing to the optimal portfolio?

1.3. RESEARCH METHODS

The study is based on a positivist research approach. For empirical analysis, the data will be obtained from Google Finance, which consist of 27 stocks, with three stocks representing one sector. Thus, implying 9 sectors are considered. The study uses daily closing price from 30 September 2019 to 30 September 2024, spanning five years with 1251 observations. Traditional methods will use daily closing pricing methods, while machine learning methods will use daily closing prices with low, high, open prices and trading volume. The study begins by providing overall summary statistics, such as the correlation and covariance of the stocks, which is essential for selecting a diverse portfolio. Based on the primary findings, the study employs portfolio optimization methods with an equally weighted portfolio as the benchmark against mean-variance, risk parity, and Sharpe ratio portfolios. Then, machine learning methods include decision-tree regression, random forest, sector vector machine, and linear regression methods. It is worth mentioning that the study will compare the traditional methods' optimal performance among the selected methods alone and, similarly, with machine learning optimal returns. Once an optimal traditional and machine learning approach has been selected, they will be compared against each other. The study will use python language as a computational tool to analyse the data.

1.4. ORGINASATION OF STUDY

This study is divided into six chapters. Chapter 2 will discuss the theoretical framework and empirical literature based on different optimisation methods ranging from traditional to machine learning with a key focus on sectoral analysis. Chapter 3 provides an overview of selected sectors' economic and financial performance and shed light on how machine learning has become an essential tool in data analysis. Chapter 4 describes the research methods and how the data will be analysed. Chapter 5 presents and discuss the results detail. Chapter 6 concludes the study.

CHAPTER 2: THEORETICAL FRAMEWORK AND EMPIRICAL LITERATURE REVIEW

2.1. INTRODUCTION

This chapter will unpack the theoretical framework of portfolio optimization, introduce different portfolio optimization techniques for traditional and machine learning models, and discuss the existing empirical literature. The chapter is organised as follows: section 2.2 discussed the Morden Portfolio Theory in-depth, section 2.3 outlines different types of traditional methods, section 2.4 unpacks the different types of machine learning methods, and section 2.5 reviews the empirical literature in depth. Section 2.6 will conclude this chapter.

2.2. MORDEN PORTFOLIO THEORY

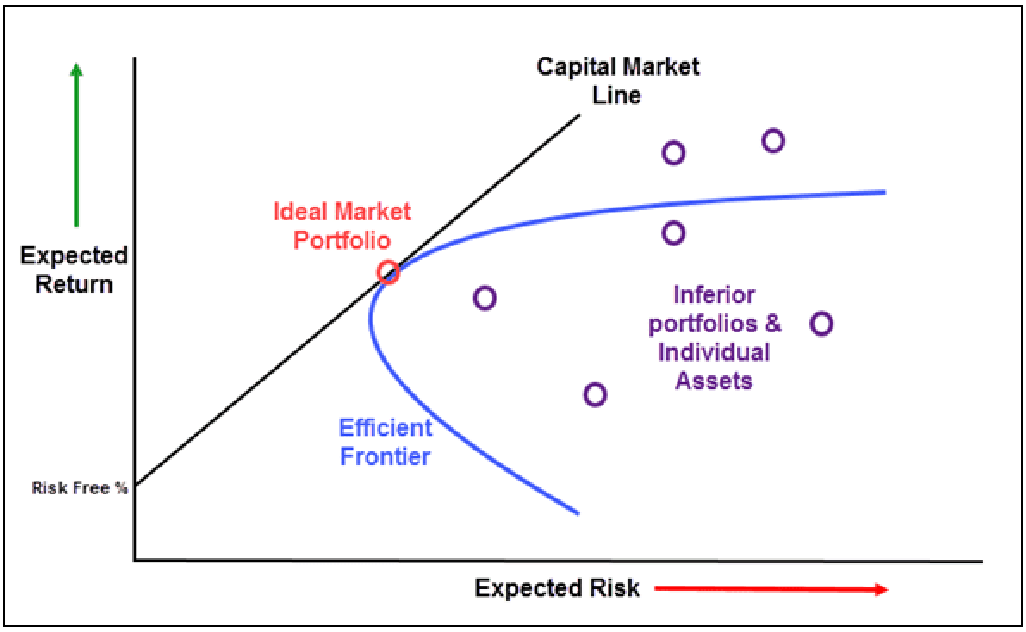
The Morden Portfolio Theory (MPT) refers to an investment theory that permits investors to construct an asset portfolio that maximizes returns at a given risk. Markowitz (1952) emphasizes the importance of diversification to achieve minimal risk with maximum returns. In MPT, risk can refer to volatility using "the expected return of the portfolio, the variance and standard deviation of the expected returns, as well as covariance and correlation between assets" (Ljungberg and Högstedt, 2021). Investors aim to achieve the best-expected returns from risky assets for a given portfolio. Markowitz (1952, 1959) shows that the total risk for a given portfolio is evaluated by the magnitude of asset variance and covariances in an investment pool. The metric gives a general view of how asset returns are dispersed around the expected return.

Utilizing the MPT risk-return ratios leads to indifferent efficient frontiers for both assets using covariance when developing an efficient frontier and the associated optimal portfolio (Zhao, Zhou, Palmar, and Feng, 2019). By contrast, the ratios have different interpretations that influence how investors' portfolios might increase their alphas in the market (Surtee and Alagidede, 2023). The shape of MPT efficient frontiers can determine the scenario above. Optimizing the objective function can construct an efficient frontier curve (Brandi and Santos, 2020).

Figure 2.1 represents an efficient frontier curve that has be constructed. The efficient frontier curve (EFC) shown by the blue line is consistent with diminishing marginal return to risk since every expected risk will increase by insignificant expected returns. Any inferior portfolios and individual assets that are above the efficient curve will mean these securities are undervalued, and investors can buy these securities at a low price with the expectation that their prices will increase (ZongMing, Koomson, and Guoping, 2017). Conversely, if the assets are below efficiency, it will imply that the securities are overvalued, and there is an expectation that the security price will decrease through its mechanism adjustment (Bechis, 2020). Any assets traded on the EFC would be an ideal market portfolio since the expected return and risk are traded fairly. The EFC is measured by beta (β). The slope in the above figure is the Capital Market Line (CML), which fluctuates depending on the risk sensitivity (Al Maiyahi, 2020).

The CML is a line shown by the black line in Figure 2.1 below that shows the trade-off between the expected risk-return ratio with an addition of risk-free rate. When the CML is tangent to EFC, it indicates that it is the point where there is the highest return at a given risk (Lee and Su, 2014). In other words, the CML is the Sharpe ratio of the market portfolio, and when the Sharpe ratio increases, the market performance also increases.

Figure 2.1: An Efficient Frontier Curve with Capital Asset Line



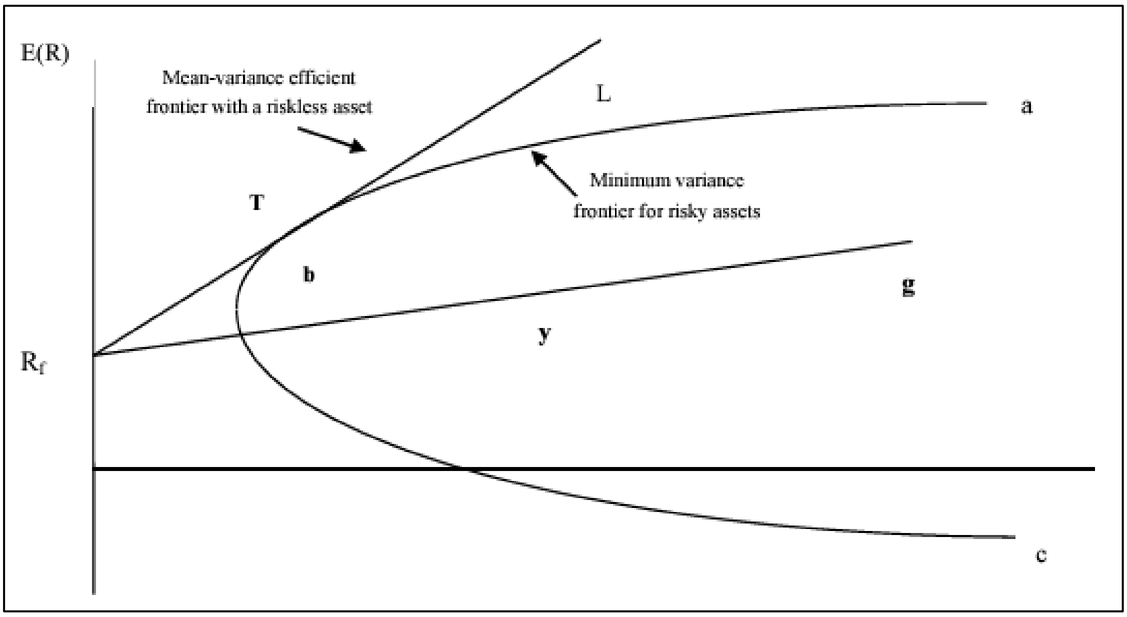
Source: (Al Maiyahi, 2020)

The significance of CML is drawn from the Capital Asset Pricing Model (CAPM) derivation, which was developed by William Sharpe (1964) and John Linter (1965), as it was a continuous work and a building block from Harry Marwitz (1959) (Rossi, 2016). The CAPM is a theoretical model that determines the trade-off between the expected return on assets and risk (Elbannan, 2015). The CAPM has ten assumptions, including the following: It does not consist of transaction costs, and assets are infinitely divisible (Elton, Gruber, Brown, and Goetzmann, 2009). No personal income tax is included, and the prices are not affected by the consistency of buying or selling assets (Elton et al., 2009). Furthermore, investors make decisions based on expected returns and risks of assets. Unlimited short sells, unlimited lending, and risk-free borrowing exist (Elton et al., 2009). Investors are more concerned about the mean and variance of returns of assets in addition to identical expectations. The last assumption is that all assets are marketable (Elton et al., 2009).

In reality, it is impossible to invest in all financial assets because there is always information that is withheld in which will benefit one party over the other. Additionally, the claim for transactions costs will be zero does not hold practically, due to investors being able to capitalize and benefit from misprice assets that requires transaction costs (Peng, 2021). Therefore, market can never be frictionless. Additionally, the assumption that implies that investors are rational and risk averse does not hold in reality due to investors sometimes investing in financial assets based on emotions as described in behavioural finance (Lai, and Stohs, 2021)

In Figure 2.2 below, the position of CML changes, shifting upwards and striking through the efficient frontier curve (EFC). It has already been established that the figure shows the relationship between expected returns and risks. The curve abc is known as the minimum variance frontier for risky assets. Point A is a trade-off between high expected returns and high-risk acceptance. However, at point T, investors will gain median expected returns at a lower level of risk. An investor may add risk-free borrowing and lend into the portfolio in the above figure. This leads to a positive diagonal straight line of R_f in 2 separate directions, to L or g. The risk-free rate does not contain risk; thus, a combination along the straight line is risk-free. The mean-variance efficient frontier with riskless assets is determined by a swing above the EFC and striking at the point of a single risky tangency portfolio T, which can be regarded as lending at a risk-free rate. In contrast, when the risk-free rate line cuts through to point g, a minimum variance frontier for risky assets is formed and lends at the risk-free rate.

Figure 2.2: A Diagram of Capital Asset Pricing Model



Source: (Elbannan, 2015)

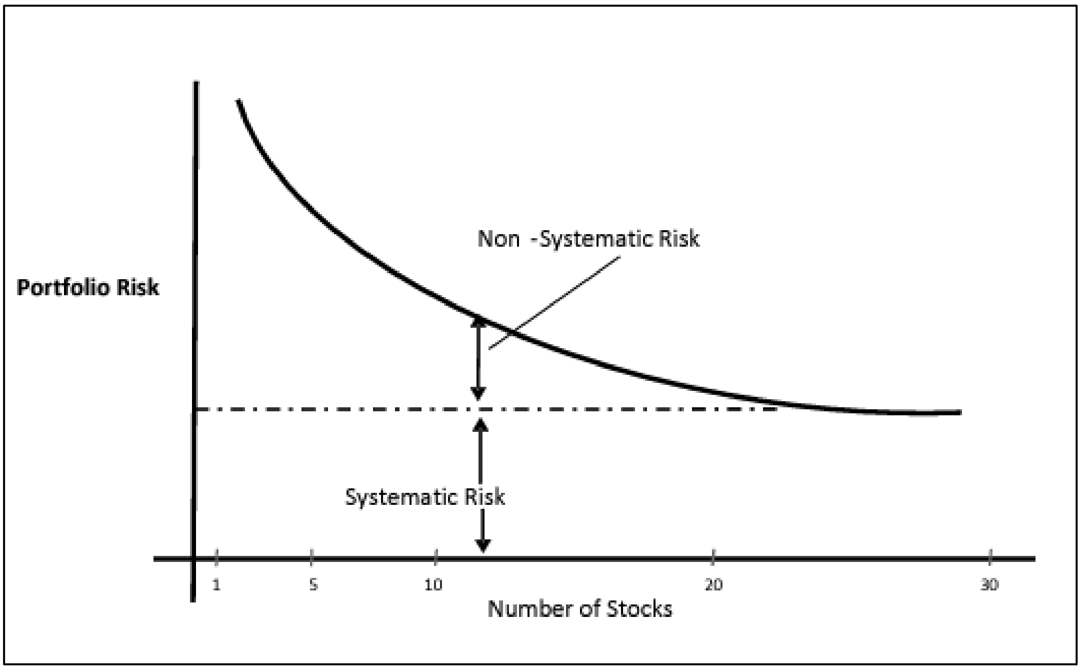
Several critics of CAPM include the lack of risk-free assets in the financial system (Levy, 2010). Secondly, there are no long-term investments as investors only focus on the term period in their expected returns. Thirdly, it is assumed that lending and borrowing rates are equal; however, there are different rates. Finally, betas are not static over time (Mansuri and Shah, 2022; Khandelwal and Chotia, 2022).

The Arbitrage Pricing Model (APM), an alternative to CAPM, relies on the law of one price where two sold assets cannot have the same price, established by Stephan Ross (1997). The APM differs from the

CAPM as the model consists of macroeconomic variables and substitutes the risk-free rate from the CAPM. The macroeconomic variables include gross domestic product (GDP), inflation rate, exchange rates, commodities prices, and market indices (Mark-Egart, 2020). The APM has several assumptions, including that assets with riskless opportunities do not exist, there is a linear relationship between macroeconomic variables and the expected return of stocks, and every economic variable can impact financial prices (Musa and Okologume, 2020).

The APM plays a significant role in facilitating active and passive management strategies. In an active management strategy, investors can track an index in which they will aim to gather stocks that mimic the movement of the index to outperform the market index (Ile, 2020). It would be ineffective for investors to buy all stocks as it will become relatively expensive due to transaction costs, and to mitigate this, investors would buy fewer stocks that mimic the index performance while minimizing costs (Stoilov, Stoilova, and Vladimirov, 2022.). Subsequently, the benefit of having a multi-asset index model over a single-asset index model is the ability to match influential returns, such as inflation, but have similar market risks. This provides a more comprehensive approach to portfolio construction. Unlike the active management strategy, which seeks to beat the market, the passive management strategy aims to mimic the same expected returns as the index (Estrada, 2006; Zuccheri, 2013; Souza, 2014.). In contrast, the transaction costs would be low due to the less frequent trading of stocks. Moreover, the exposure of investors to different stocks would widen, allowing for diversification without active decision-making (Fisher and Lie, 2004; de Pinho Azevedo, 2022)

Figure 2.3: Relationship Between Portfolio Risk Versus the Number of Stocks



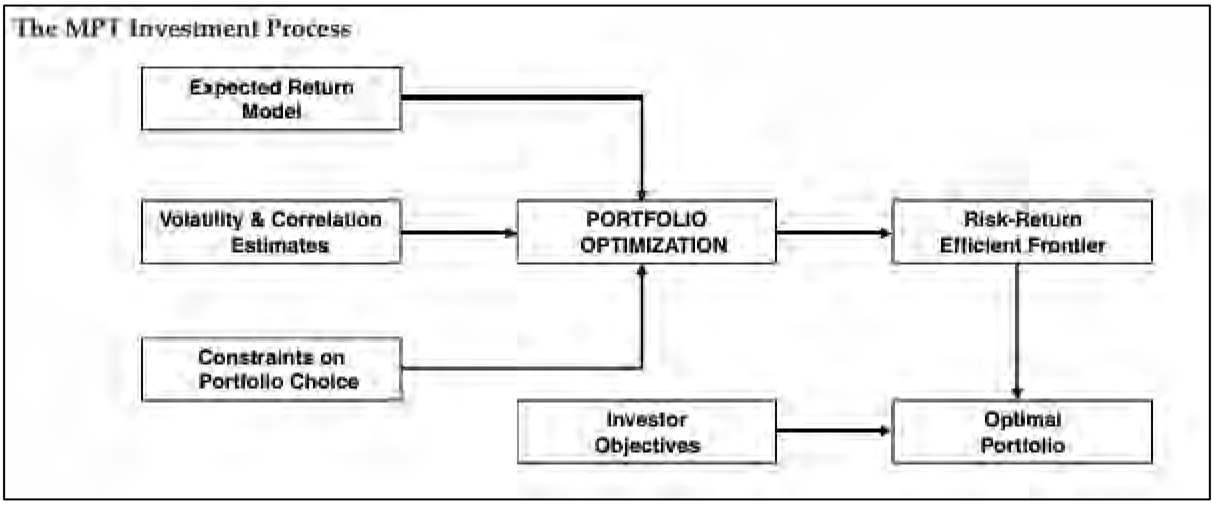
(Wafula, 2014)

Portfolio diversification is crucial for investors to maximize returns while mitigating risks. Two types of risks are at play to achieve diversification: systematic and unsystematic, as shown in Figure 2.3

above. The systematic risk is considered a market risk that cannot be decreased through diversification. Unsystematic risks are associated with risk reduction through diversification when the number of stocks increases in each portfolio (Wafula, 2014). To interpret the systematic risk, the beta(β) value is the appropriate measure, where $\beta=1$ implies that the systematic and overall market risks are equal. When $\beta>1$, the systematic risk is greater than the overall market, while $\beta<1$ implies that the systematic risk is lower than the overall market (Wafula, 2014).

The overall MPT investment process can be summarized in Figure 2.4 below, which shows that expected returns estimates, volatility and correlation estimates, and the constraints imposed on a portfolio due to investors' choice will lead to portfolio optimization. These three properties work concurrently to achieve portfolio optimization, a crucial objective in investment management. A portfolio optimization results in a risk-return efficient frontier due to having the highest expected returns at a given level of risk. However, an optimal portfolio is subject to investors' objectives and portfolios on the efficient frontier.

Figure 2.4: A Summary of The MPT Investment Process



Source: (Fabozzi, Gupta, and Markowitz, 2002)

2.3. TRADITIONAL METHODS OF PORTFOLIO OPTIMIZATION

2.3.1. WEIGHTINGS

In a portfolio of N risky assets is weighted subjected to minimizing risks and maximizing returns. A portfolio weight is the percentage of the stock that is found in the overall investment portfolio. The weight of a stock has significance influence on the overall expected return of the portfolio (Taljaard and Mare, 2021; Hanif, Hanun, and Febriansah, 2021). Weights are subjected to a summation of 1 or 100%. The weights will always be positive. Liagkouras and Metaxiotis (2018) state that the very small weightings of the portfolio do not significantly influence the portfolio's overall performance. By contrast, imposing a ceiling constraint restricts the exposure of abnormal behaviour of stocks. It could be an additional to the administrative and monitoring costs. Consequently, the number of assets in a single portfolio plays a role in the rate of return, as previously mentioned, and to mitigate this, cardinality constraints are imposed (Liagkouras, Metaxiotis, and Tsihrintzis, 2020). Cardinal constraints is the application of minimizing the number of assets in a portfolio (Cesarone, Scozzari, and Tardella, 2013). It consists of a lower bound and an upper bound. The lower bound avoids the unconventional behaviour of any asset, while the upper bound typically limits the additional administrative and monitoring costs (Cesarone et al., 2013; Liagkouras et al., 2018 and Liagkoura et al., 2020)

2.3.2. LAGRANGE-MULTIPLIER

Lagrange-Multiplier is "a technique for finding a maximum or minimum of a function subject to a constraint" (Salih, 2013). This technique was developed by Joseph-Louis Lagrange (1755), and at the time, the method was used to solve constraints through repetitive derivations. The intention was to optimize the objective function to meet the conditions of Kuhn-Tucker (Ferguson, 2004; He, 2017). Although the method is popular, sometimes the Lagrange multiplier does not work (He, 2017). The multiplier can work when there are equality constraints. However, it fails to work with inequality constraints subject to a few cases when a person quickly solves the first derivative of an equation in a closed form. By deduction, the inequality constraints are changed to equality constraints before using the Lagrangian method (Wah and Wang, 1999). Mathematically, the Lagrangian method has been set up by adding its multiplier to the objective function to differentiate the inequality and equality constraints, yielding a first-order partial derivative subjected to the weightings of the stock and the equality constant (Dowing, 2011). The equality constant is examined through the Kuhn-Tucker Theorem.

2.3.3. KARUSH KUHN TUCKER CONDITIONS FOR INEQUALITY CONSTRAINTS

Karush-Kuhn Tucker (KKT) was established in 1939 and later developed in 1951 by Harold Kuhn and Albert Tucker, respectively. KKT is used as a condition to determine the first derivative of an inequality constraint where it equals zero (Kasthuri, Vasanthi, Ranganayaki, and Seshaiah, 2011). For KKT to play an influential role in ensuring the rules of the constraints are met by local minimizers, it is vital to apply a principle of constraint qualifications that seeks to minimize risk and maximize returns. In a typical multi-objective optimization, the conditions are subjected to several constraints, including one or more. Assuming there is one constraint known to be positive, the KKT is determined to be a weak KKT; however, should it be in the case where all KKT constraints are positive, then the KKT is strong (Haeser and Ramos, 2020). It is with this view that KKT is used as an indicator of determining the minimum and maximum value through adjustments of constraints in which this theorem is sometimes referred to as the saddle point theory, as it is derived from the Langrage multiplier (Mesgarani, Aghdam, Beiranvand and Gómez-Aguilar, 2024).

The required conditions must be met for the Kuhn-Tucker conditions to be sufficient. Considering a maximization optimization method, the objective function should be concave, while the solution set should be convex. Similarly, in a minimum optimization method, the objective function should be a convex function, while the solution space should be a convex set (Kasthuri et al., 2011).

2.3.4. MEAN VARIANCE OPTIMIZATION

Mean-variance optimization (MVO) is an optimization that seeks to limit the risk at a given level of return or exploit returns as much as possible for a given level of risk (Pedersen, Babu, and Levine, 2021). These two approaches are associated with the minimum variance and maximum return portfolios. However, an additional equality constraint is added to ensure the given level of risk or return is achieved. The MVO has several assumptions, including but not limited to investors making decisions sorely based on the outcomes of the return at a given level of risk, and the time of investment is a single holding period due to everyone in the market having the same information (Nguyen, 2019).

There are a few advantages to MVO. Firstly, when estimating, the stocks that should be used must consist of small risk parameters used for ongoing portfolio decisions throughout the investment period (Gunjan and Bhattacharyya, 2023). In addition, the MVO is relatively simple to implement, fast, valuable for estimating optimal returns without a quadratic nature, and can be implemented for tracking errors (Gunjan and Bhattacharyya, 2023).

MVO has been subjected to criticism and limitations. Firstly, the portfolio can exhibit a few stocks, resulting in low diversification (Lagowski, 2022). A low diversification portfolio will result in an incorrect expected return, risk estimation, and misallocating stock weighting. Secondly, the MVO is very sensitive towards input parameter changes, often leading to constructing a different portfolio that can underperform (Lagowski, 2022). This implies that the model is not robust. Zhang, Li, and Guo (2018) consider time constraints to be a limitation in applying the principle in an investment setting due to the ever-changing market changes. At the initial time at t_0, the optimal portfolio may not be the same in the future at time t_1, and MVO does not consider that. Due to its static nature, there is no room for flexibility.

2.3.5. MAXIMUM RETURN OPTIMIZATON

The maximum return (MR) optimization aims to achieve the highest return while ignoring the level of risks unless there are two or more stocks with equal returns. This means that any stocks that provide the highest rate of return have a high concentration such that a stock with the lowest risk rate is selected to maximize the Sharpe ratio (Óskarsson, 2021). The assumption of MR shares the same views as the general assumptions of MPT, in which the model does not consider any transactional costs and taxes; investors are rational and independent, and investors share the same ideas and outcomes. In addition, the expected returns are distributed normally, while investors do not have any influence in the determination of demand within the economy; thus, there is no implication on the changes in stock market prices (Elkjaer, 2022).

2.3.6. SHARPE RATIO

The Sharpe ratio compares portfolio performance by the return per unit standard deviation. This can be shown by dividing the mean returns of a stock by its standard deviation. In this context, when the Sharpe ratio is large, it implies the more excellent performance of the portfolio. If there is a low Sharpe ratio, the portfolio is not producing returns at its optimal level. This is due to the compensation for each unit risk at different levels.

The Maximum Sharpe ratio (MSR) is a combination of the Minimum Variance and the Maximum Return portfolio, referred to as the extension of the variance optimization portfolio. In this portfolio, the same principle applies where investors seek to maximize returns while minimizing risk. The distinguishing aspect that sets it apart is the application of bi-objective optimization, which categories fall under the multi-objective optimization framework (Stampe and Revelsby, 2022). The framework involves an approach that solves several single-objective sub-problems where additional objectives are transformed into constraints (Rompotis, 2023).

There are limitations to the Sharpe ratio, and the main criticism is that using standard deviation as a metric value for the Sharpe ratio is not scientific. In addition, the Sharpe ratio instead provides a linear optimal outcome; however, it fails to capture risk when all assets have been placed on the efficient frontier curve (Yang, 2021). Thirdly, it does not consider how assets are correlated to one another, and fourthly, the ratio does not consider how to integrate into mutual funds (Yang, 2021).

2.3.7. EQUALLY WEIGHTED PORTFOLIO

Equally Weighted (EW) Portfolio is the most straightforward portfolio to construct where the weights of the stocks are equal. The benefits of the EW portfolio are diversification as the risk of the stocks has been minimized, which mitigates the overall portfolio volatility (Maillard, Roncalli, and Teïletche, 2010). Due to the nature of the 1/N rule of allocating the number of stocks equally, it serves as a better model than MVO since there is no need for complex calculations to estimate the expected returns and associated risks, leading to error reduction. Consequently, the EW portfolio will still perform well despite not knowing the risk or return, which refers to Knightian uncertainty (Battaglia and Leal, 2017). By contrast, an additional advantage is that the EW does not factor in market prices (Krueger and Wrolstad, 2013).

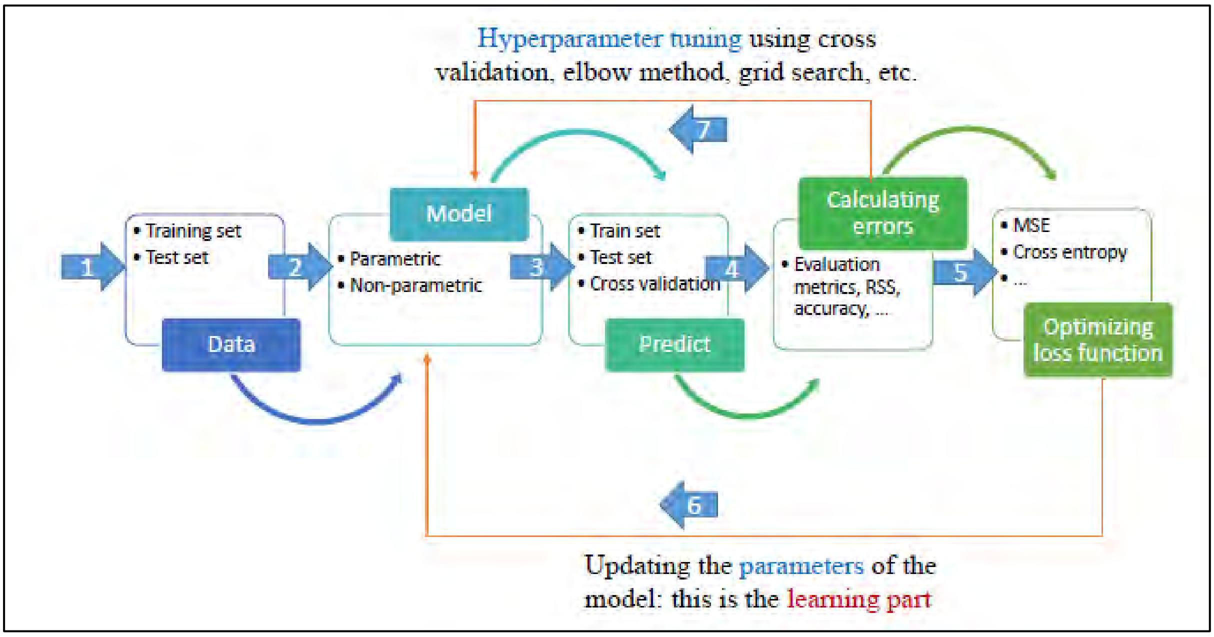
The disadvantage of having an EW portfolio is that when an investor invests in an underperforming stock, it will affect the entire portfolio, and to mitigate the issue is to remove the stock from the portfolio altogether (Wiecek, 2020). Another disadvantage of this portfolio is the implementation costs, which are significantly high because rebalancing the portfolio will prompt high management fees, including turnover. Subsequently, another limitation of the portfolio consists of lower liquidity of assets as it does not take into consideration business sizes in creating a portfolio (Štarolis and Krakauskas, 2015)

2.4. MACHINE LEARNING

2.4.1. INTRODUCTION

Machine learning (ML) is something familiar, as it started in the early 1940s when Pitts and McCulloch (1943) used a mathematical model technique to impressionist human thought processes using various neural network algorithms (Huntbach and Ringwood, 1999.). Hebb (1949) used neural networks to understand the behaviour of the brain (Kuriscak, Marsalek, Stroffek, and Toth, 2015). The first use of ML in finance was in 1982 when Apex created an algorithm program that gives their clients tax and financial advice (Apex); JP Morgan Chase (1987 and 1989) implemented the Personal Financial Planning System and Fico Score, which is used as an algorithm for credit scoring formulas (Brown, Nielson, and Phillips, 1990; Davenport, 2018 and Mantam 2020). ML can be described as an "algorithm that learns complex patterns in a high dimensional space without being specifically directed" (De Prado, 2018). For ML, one factor needs to hold by meeting criteria when ML is in use, and the factors can be "learns complex patterns," "learns in high dimension," and "learns without being specifically directed" (De Prado, 2018). ML differs from general programming because ML learns from the data, either structured or unstructured, and it identifies the relationship between two or more variables such that it can be used to make accurate forecasts. Unlike general programming, it heavily relies on traditional rule-based programming (Delgado Fernández-Valdés, 2023).

Figure 2.5: A Model Showing How Machine Learns.



(Tarentaal, 2023b)

There are three main types of ML, and it includes supervised learning, where it learns from training data to make predictions; unsupervised learning, where the algorithm is trained on unlabelled data to discover patterns, structures, or relationships; and reinforcement learning, which includes an agent that interacts with the environment as algorithms study the inputs of the agent. As a result, there will be feedback in the form of rewards or penalties (Hiran, Jain, Lakhwani, and Doshi, 2021). The machine learns systematically, as shown in Figure 2.5 above.

The process begins by training and testing the relevant data. This includes learning from the data and minimizing parameters from predicted outcomes versus actual outcomes. After learning from the model, the dataset is validated using a fine-tuning process, and the model's hyperparameters are optimized (Tarentaal, 2023b). Finally, the dataset from the trained model is evaluated in the testing phase. The data is then divided into a parametric or non-parametric model, which will be trained, tested, and cross-validated several times as part of reinforcement learning to provide error-free results accurately (Tarentaal, 2023b). The next step involves a process of error calculation using various evaluation metrics, adjusting parameters, and refining hyperparameters using cross-validation or test sets. The purpose of this nature is to minimize the loss function to give accurate results that can be predicted.

2.4.2. LINEAR REGRESSION

Linear Regression models consist of dependent variables assigned by the 'y' variable and independent variables assigned by the 'x' variables. The dependent variable estimates change when independent variables change as they are affected by changes; thus, they depend on one another (Maulud and Abdulazeez, 2020). In ML, the function of f(X, α) is unknown; thus, finding the function based on the limited data becomes its responsibility. It is unknown because the relationship between the dependent and independent variables is too complex to be explicitly defined (Chen, Zhang, Mehlawat, and Jia, 2021). The objective of why ML must estimate f is due to inference such that it becomes interpretable and able to make predictions. The error term in the function form model denotes the noises between the dependent and independent variables (Ding et al., 2023)

The model complexity increases as the number of degrees of the polynomial increases. As a result, the higher complexity of the model will lead to an increase of the variances and decreases the bias towards the model, becoming an overfitting model (Nelles and Nelles, 2001). By contrast, when the model is less complex, it will yield a high bias and low variance and becomes underfitting (Nelles and Nelles, 2001). To determine the minimized point in the test sample, one should determine the optimal degree of polynomial taking place.

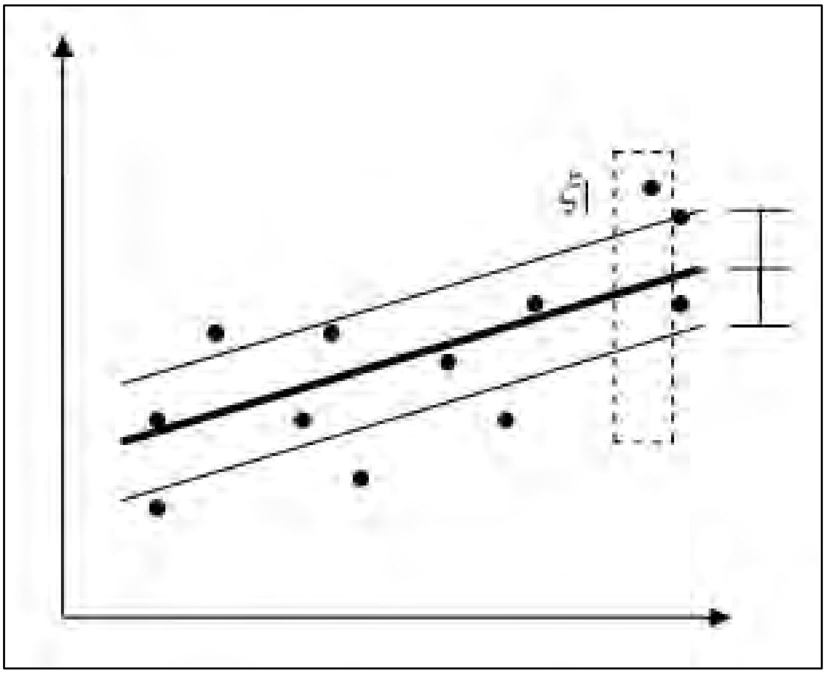
A typical polynomial function model can be classified as overfitting holds due to its complexity. A regularization is implemented for an algorithm to implement a less complex model, resulting in a higher bias and variance reduction (Fastrich, Paterlini, and Winker, 2015). The regularization is implemented by adding a penalty term. In this case, when the parameter of L increases, the penalty will increase. However, it will decrease errors in the training phase. The above model is a combination of the Least Absolute Shrinkage and Selection Operator (LASSO) and takes the L1 norm with Ridge Regression and takes the form of the L2 norm (Khazen, 2023). λ denotes a tuning parameter that controls the relative impact of the penalty term. As λ increases, the model becomes rigorous by lowering the value of the weights (L), and it is essential to select an optimal value of λ that can be used by finding it through cross-validation (Sun, 2018).

In a LASSO model, weights can be reduced to zero; however, Ridge Regression shrinks to zero but never to zero (Emmert-Streib and Dehmer, 2019). Contrary to the Ridge Regression model, the opposite will affect when the model complexity increases, the λ decreases, leading to variance decreasing. By using an Elastic Net model, a combination of LASSO and Ridge Regression model, one can get benefits of both models by having some weights zero and others minimizing closer to zero but never zero (Ren, Tang, and Chen, 2023)

2.4.3. SECTOR VECTOR MODEL – REGRESSION

Vapnik(1995) developed the Sector Vector Model (SVM), which studies a separate function of training datasets based on their classes. The study has placed SVM as a supervised learning model to solve scenarios as a classification or regression (Ukil and Ukil, 2007). An application of SVM to finance is the ability to detect market anomalies. Besides finding anomalies, SVM can generate abnormal returns and separate components that are considered the top 10% versus those in the bottom 10% with the help of the margins (Ahmed, Choudhury, and Uddin, 2017). The reasons mentioned suggest that SVM are better identifiers and predictors of financial market anomalies than linear regressions.

Figure 2.6: A Diagram of Sector Vector Regression Model



Source: (Basak, Pal, and Patranabis, 2007)

SVM consists of Regression, which aims to find the optimal separation between the distance training points that can be deviated. In addition, the main goal is to determine a hyperplane that has the highest number of data points against the required risk. From Figure 2.6 above, when a data point is observed outside the margin, it may be referred to as a slack. The optimization problem will arise by minimizing the sum of squared weights, which may refer to regulations (Basak and Patranabis, 2007.). On the contrary, minimizing data points from the original slacks becomes relatively costly, and an expectation is to be punished through the slack of C.

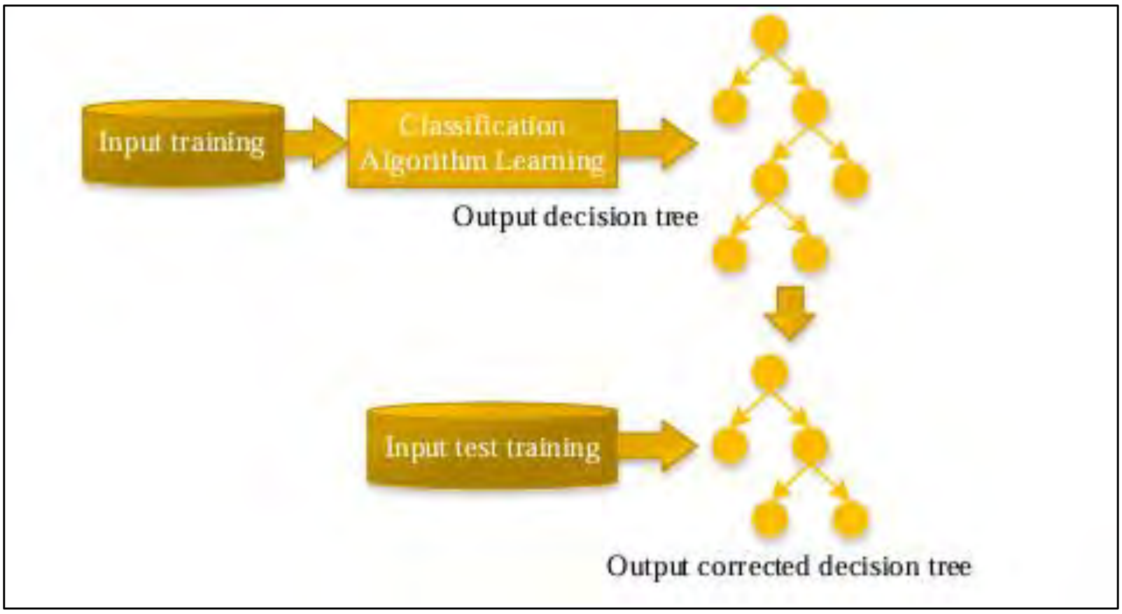
In a typical case, ξ_i, ξ_i^* are slack variables that are part of the system to consider the infeasible convex optimization problem and can be considered as the variance of the expected variables and required output (Still and Kondor, 2010; Kuang and Zhang, 2014; Zhao, Tao, and Zio, 2015; Yang, Zeng, and Xie, 2017). The parameter C is the penalty factor constant with a trade-off between the flatness and deviations of errors. The parameter ϵ controls the SRV behaviour as the loss function (Kuang and Zhang, 2014; Zhao et al., 2015). Thus, there are positive and non-zero samples referred to as Support Vectors (Houssein, Dirar, Abualigah, and Mohamed, 2022). The variables and constraints will be used to find the optimal solution by applying the Karush-Kuhn-Tucker (KKT) method, which is a prerequisite to finding the optimal solution for the global optimum (Chen, 2007; Yazdani, Babagolzadeh, Kazemitash, and Saberi, 2019; Yazdani, 2019).

2.4.4. DECISION TREES

Decision Trees are machine learning algorithms that can be used for classification and regression through a decision-making process designed to mimic human thought processes. The foundation of decision trees lies in the partition of data, which will be categorized subject to variables of input features (Abd, Jamaludin, Zainol, and Sembok, 2023). The partition can effectively be completed by optimally separating data into classes or predicted numerical values (Ramchandani, Paich, and Rao, 2017). Decision tree involves several stages: (i) obtaining data from a data source - the data is processed in a computational or machine learning algorithm. (ii) The data is converted into a training set, usually 70-30 split (70% training – 30% testing) data. This enables users to create inductive decision trees. (iii) When a decision tree has been created, the rules are extracted to be classified. (iv) Several rules will be created when rules are classified. (v) These rules are then applied to a practical application (Wichtlhuber, Strehle, Kopp, Prepens, Stegmueller, Rubina, Dietzel, and Hohlfeld, 2022)

To expand on the invention of the decision tree, the process begins at an initial model called the root note, in this case, the decision tree classification algorithm, see Figure 2.7 below, which shows how a decision tree is created and is used to predict future outcomes pending on various indicators of inputs. The root node is divided into splitting criteria based on either Gini impurity or Information gain for classification tasks (Nair, Mohandas, and Sakthivel, 2010). Classification can be divided into Class A or Class B. After selecting the best criteria, the dataset is divided into two subsets, which are associated with one outcome for each, which are connected by branches (Ren, 2024). This process is repeated until the maximum depth for the tree reaches a certain impurity level. The data can be used to make final predictions from the optimal outcome, and users can generate forecasted results (Ren, 2024).

Figure 2.7: Decision Tree Generating Process

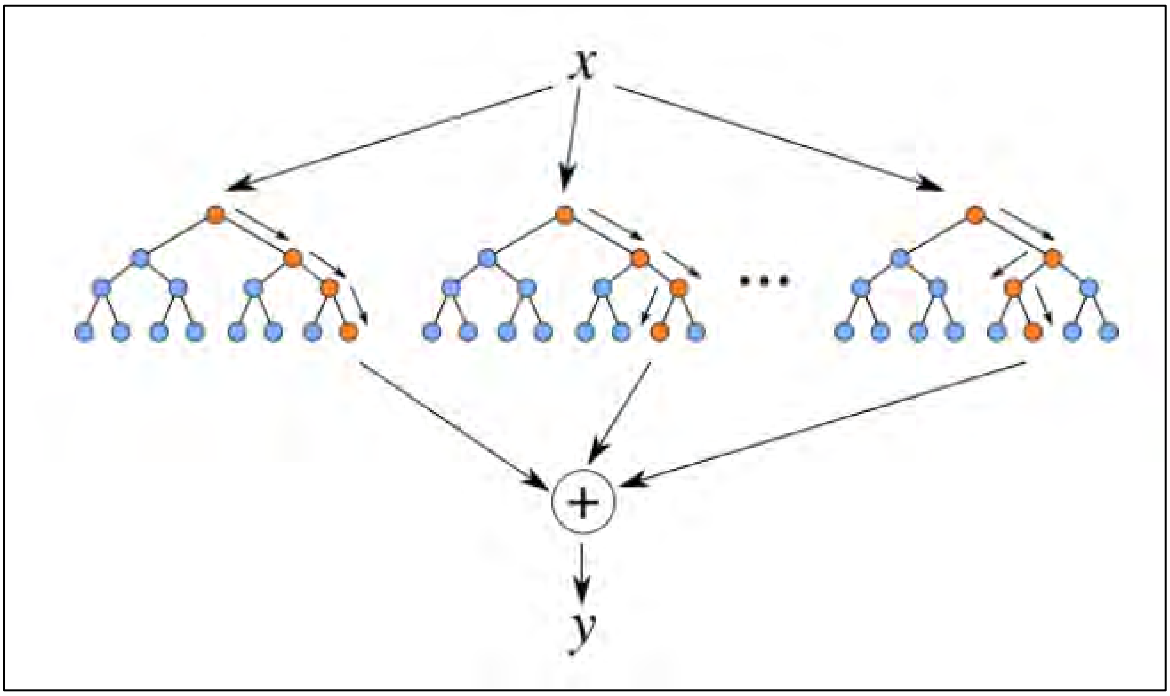


Source: (Ren, 2024)

2.4.5. RANDOM FOREST

A random forest combines multiple decision trees to increase forecast accuracy, decrease overfitting, and increase the robustness of machine learning algorithms (Basak, Kar, Saha, Khaidem, and Dey, 2019). Similarly, a random forest can be built by initiating a bootstrapping where there is a random selection of subsets of data that are selected with replacement – meaning that datasets can be repeated – in which the dataset is chosen from the same size as the original dataset that is created independently for each tree in the forest (Wynne, 2023). Multiple decision trees can be constructed from these bootstrapped subsets, where each decision tree is trained separately from a different bootstrapped subset. Creating a random forest going forward is like a decision tree until the maximum depth of data points has been reached (Mohandoss, Shi, and Suo, 2021). Thus, multiple independent trees are collected to ensemble as a random forest. See Figure 2.8 below. Combining the predictions and having a final prediction is essential when constructing decision trees. In a classification method, random forest is used by majority voting to select the final prediction from the individual trees (Zhu, 2020). In the regression method, the prediction is computed using the average or the mean of the individual tree's predictions.

Figure 2.8: A Random Forest Tree



Source: (Zhu, 2020)

Another critical component of random forest is featuring importance, which assesses the extent of input variables in predicting the outcome variables. The method is achieved by associating a score to each input variable with the highest impact on the model's prediction (Levantesi and Zacchia, 2021). Random forest uses residuals sum of squares or Gini index to evaluate the feature importance. A low score in feature importance indicates that it does not play an active role in making correct predictions and classification of data (Dewi and Chen ,2019). In essence, feature importance is used to prioritize and focus on the influential features of the model, which can reduce its dimensions and complexity.

2.5. EMPIRICAL REVIEW

2.5.1. TRADITIONAL METHODS

Our empirical review of traditional methods is structured by grouping similar conventional methods employed by different authors and determine whether they had similar results. Herzel, Nicolosi, and Stărică (2012) studied how including sustainable constraints will impact the overall portfolio. Their study covered 11 sectors: Basic Materials, Consumer Goods, Consumer Services, Financials, Healthcare, Industrials, Oil and Gas, Technology, Telecommunications, and Utilities. Their methodology included a classic mean-variance approach with sustainable constraints. The results indicated that socially responsible screening increases the slight loss of the Sharpe ratio (Herzel et al., 2012). Adhikari and Jha (2016) indicated that, in respective to which model performed better, the mean-variance portfolio outperforms the minimum variance portfolio where the 20 stocks that were randomly selected where the sectors were derived from Commercial Banks, Development Banks, Finance Companies, and Insurance Companies. Charles, Darné, and Fouilloux (2016) used risk-adjusted performance measures and tail risk measures to determine the impact of screening strategies on Economic, Social, and Governance (ESG) stocks in nine sectors. They discovered that higher concentration in some sectors leads to higher risk-adjusted returns.

Fagerholt and Aanonsen (2016) divided 80 stocks into ten sectors to determine the optimal model for portfolio optimization, with results implying that the risk parity portfolio performs better than the factor model and simulation of return distributions. Tupko, Tupko, and Vasil'Eva (2016) studied which model performs better between the mean-variance and the Copula-based portfolios by using stocks from sectors such as Financials, Chemical, Oil, and Electro energetics. The results indicated that the Copula-based portfolio performs better than the mean-variance portfolio. In the case where the mean-variance optimization model and threshold-accepting optimization model are compared by Masese (2017), he indicated that the thresholding model produces higher returns and performs better than the mean-variance model. The sectors of the investigation included manufacturing, construction, and automobiles. Additionally, Bora (2023) showed that downside risk optimization performs better than mean-variance optimization in all United States of America (USA) sectors.

AlHalaseh and Al Shawawreh (2024) categorized 99 companies into three sectors, namely, financial, industrial, and services, to compare which model returns the highest returns between the mean-variance model and the risk parity portfolio. Their paper showed that the risk parity portfolio outperforms the mean-variance model. Similarly, Auer (2024) selected 21 stocks from ten sectors, such as energy, financials, industrials, IT, health, utilities, real estate, communication, consumer discretionary, and consumer staples, to determine which traditional methods yield the highest returns. The paper concluded that maximum return generates more than the Sharpe ratio and minimum variance. Subsequently, Tung (2024) examines 30 Companies from the Taiwan Stock Exchange, which were divided into two sectors, namely technology and manufacturing, to determine which traditional method generates the highest returns. Tung (2024) illustrates that hierarchical risk parity yields more return than a mean variance and maximum diversification.

Liagkouras, Metaxiotis, and Tsihrintzis (2020) studied which stocks will generate the highest return between ESG and non-ESG stocks using a mean-variance optimization model across ten sectors. The results showed that ESG stocks hinder potential significant returns. Simply put, non-ESG stocks performed better than ESG stocks. Porage (2021) examined 15 stocks from OMXC25 index and these stocks constitute of ten sectors. It uses the likelihood ratio test to determine which traditional method performs better. The results suggest a statistically significant difference between the minimum variance frontier with ESG stocks and non-ESG stocks. Stampe and Revelsby (2022) studied one sector, which was the private equity sector, where 86 stocks were considered. The methodology for the study included aggregation of ESG scores, risk-adjusted performance, and linear regression analysis. The paper reached four conclusions: (i) ESG-integrated funds generate a higher Sharpe Ratio than non-ESG integrated portfolios, and (ii) non-ESG funds produce higher absolute returns than ESG funds. (iii) In the regression model, ESG funds produced a higher Sharpe ratio than non-ESG funds, resulting in high adjusted risk returns. (iv) ESG funds produced low beta risk scores, indicating symmetric risk absolute low, and (v) ESG integration funds were positive but not significant on funds risk-adjusted returns performance (Stampe and Revelsby, 2022).

May (2022) focused on seven sectors derived from the Johannesburg Stock Exchange among the Top 40 stocks to forecast returns based on portfolio optimization using XGBoost. In the study, May (2022) used several methodologies as a mechanism to forecast, which included equal-weighted, minimum variance, equal-weighted risk contribution, maximum decorrelation, inverse volatility, maximum diversification, Particle Swarm Optimisation (PSO), genetic algorithm, Long Short-Term Memory Networks (LSTM), random forest and Extreme Gradient Boosted Machines (XGBoost). The overall findings of the paper showed that the constructed portfolio by the author outperformed the equal-weighted and JSE benchmark index (May 2022). Rompotis (2022) used the capital asset pricing model, Fama and French five-factor model, Sharpe ratio, and Treynor ratios as part of their methodology. They showed that ESG ETFs outperform non-ESG stocks, and there was no significant relationship between ETFs' ESG ratings and their respective assets.

Senthilkumar, Namboothiri, and Rajeev (2022) derived stocks from ten sectors to unpack if portfolio optimization favours sector or broad market investments. Within the study, a comparison of traditional methods was conducted, and the results showed that Sharpe's model performed better than Markowitz's model. Useche, Martínez-Ferrero, and Gonzales (2023) used novel utility functions, mean-variance optimization with ESG elements, and summary statistics of optimization methods to investigate the performance of ESG stocks under their respective portfolios. A total of 79 stocks were used and divided into 11 sectors. The paper's findings implied that ESG fund companies produced higher returns and reduced risks than non-ESG fund companies. Xidonas and Essner (2024) used the multi-objective minimax optimization model to investigate whether ESG and non-ESG stocks will yield optimal returns from the Eurozone super-sector. The paper concluded that optimal ESG portfolios produce higher risk-adjusted returns than their respective market benchmarks. Bessler and Wolff's (2024) paper studied portfolio optimization in the context of sector return prediction models. The sectors for consideration included oil and gas, manufacturing, consumer goods and services, health care, technology, telecommunication, and financials. The result of the study showed that sector returns outperform asset allocation decisions more than past historical average returns and 1/N buy and hold portfolios (Bessler and Wolff, 2024).

2.5.2. MACHINE LEARNNING METHODS

Our empirical review of machine learning methods is structured by grouping similar machine algorithms methods employed by different authors and determine whether they had similar results.

Turner and Han (2009) explored how various economic climates can be used to allocate stocks efficiently through portfolio optimization. The paper used k-means clustering, gaussian kernel distance, fitted Q-iteration, and principal component analysis as their methodology and applied it to Russell 1000 stocks and 27 economic indicators, which were divided into 24 equal sectors. The results indicated that fitted Q-iteration yields a higher return than Gaussian kernel distance based on information ratio. All methods except principal component analysis were performed above the benchmark. Al-jomai (2014) used sedimental analysis, frequent pattern mining, clustering, and classification to develop the optimizer tool and increase stock returns. As part of the study, it was determined that there is a strong positive relationship between daily prices and news articles, which suggests that any significant event that has occurred and is publicised via the news will impact the changes in stock prices.

Bayramoglu and Basarir (2019) studied which model will yield better returns between machine learning models, such as the modelling multilayer perceptron and artificial neural networks, against traditional methods using 16 stocks over nine sectors. The results showed that artificial neural networks performed better than traditional methods and benchmarks. Additionally, De Franco, Margot, and Monnier (2020) determined that dimensional space accurately selects stocks that outperform compared to traditional strategies using the MSCI World Index and its stocks, divided into ten sectors. Sen, Dutta, and Mehtab (2021) used short-term memory (LSTM) and minimum risk portfolio to determine the most accurate model. It was demonstrated that LSTM is found to predict stock prices in the future accurately. A study was conducted by Kumar, Joseph, Muthukrishnan, Tulasi, and Varukolu (2021) to determine whether several cases and of which was which models outperformed the deep learning method and machine learning and machine learning methods and traditional methods and traditional methods against each other. Five sectors were considered: Metal, Pharmacy, IT, Banking, and Auto. The results showed that deep learning models provide better returns than machine learning methods, while machine learning methods perform better than traditional methods. The minimum variance portfolio also performed better than the optimal risk portfolio (Muthukrishnan, Tulasi, and Varukolu, 2021).

Simar (2023) selected two stocks from four sectors, IT, healthcare, financials, and industry, using machine learning to determine the optimal expected returns. The methodology used were Morden portfolio theory, random forest, and boosting methods. The paper concludes combining dividend and growth stocks yields greater returns than divided stocks alone. Similarly, Zhang (2023) found that machine learning methods generate higher returns than traditional methods from the six equities in six sectors. The primary machine learning method used was Monte Carlo simulation, and traditional methods included a maximum Sharpe portfolio and minimum volatility portfolio. Likewise, Lin and Taamouti (2023) showed that the conditional Sharpe ratio outperforms the minimum variance portfolio and equally weighted portfolio. At the core of the study, the methodology included applying a superimposable neural network model, portfolio formation, GARCH-based model, and back testing.

Muslim, Dasril, Harveend, and Muzanah (2024) used ten different stocks from seven sectors to determine which portfolio will yield the greatest return from the construction the author has implemented. The methodology included mean-variance optimization, linear programming portfolio, and summary statistics. The results indicated that the various portfolios were constructed, and portfolio D was the optional portfolio with over 80% in its Sharpe ratio. Dip Das (2024) factored six sectors: technology, healthcare, energy, banking, consumer, and cryptocurrency. The research findings outlined that the Bidirectional Gated Recurrent Unit (BiGRU) performed better than the Bidirectional Long Short-Term Memory (BiLSTM) and that traditional methods yield negative returns in the short run. Machine learning is more robust than traditional methods.

Fan and Michalski (2020) study how the impact of ESG can have on the overall performance of a portfolio based in Australia using LSTM and neural networks across ten sectors with over 423 stocks. The discovery was that the Sharpe ratio was improved by exploiting ESG scores with past returns and not including the non-related stocks with inferior performance. Additionally, an ESG integration drives financial performance by diverting from the governance dimension. Sen and Dutta (2021) considered five sectors, media, oil and gas, private banks, PSE banks, and reality, to optimize their portfolio. The methods that were conducted included Hierarchical Risk Parity (HRP), Hierarchical Equal Risk Contribution (HERC) such as Tree clustering, Optimal number of cluster identification, Top-down recursive bisection, Naïve risk parity within the cluster for weight allocation, and back testing method. The paper's main findings indicated that HRP yields better results than HERC portfolio except in one sector relative to the Sharpe ratio (Sen and Dutta, 2021).

Pritam, Mathur, Agarwal, Paul, and Mulla's (2022) paper focused on finding the best approach to constructing an optimal portfolio using P-index, Multi-criteria Volatility (MCV), clustering analysis, and fractional lion algorithm for rapid centroid estimation. Based on the results, the P-index fund outperformed three portfolios constructed across the six sectors involved. Jiménez-Preciado, Venegas-Martínez, and Ramírez-García (2022) selected 506 stocks across eleven sectors while considering 20 financial ratios, logistic regression, sector vector machine, random forest, gradient boosting, and artificial neural networks methodology. The deduction from the paper suggested that random forest yields the highest accurate score compared to ML algorithms.

Assael (2023) used 11 sectors across 2429 companies that were uniquely identified using a cross-validation scheme and gradient boosting model as their crucial methodology. The results showed that ESG scores had a negative impact on the price returns of small and large capitalization companies, with better ESG scores primarily associated with higher price returns. Dip Das, Bowala, Thulasiram, and Thavaneswaran (2023) selected 16 different stocks across five sectors: technology, healthcare, energy, banking, and consumer. To determine which model performs better, the authors considered bidirectional long short-term memory, bidirectional gated recurrent unit, and clustering: affinity propagation. The results showed that bidirectional gated recurrent unit outperforms bidirectional long short-term memory in dynamic and volatile market scenarios. Masongweni and Simo-Kengne (2024) used panel data regression techniques to understand the relationship between financial performance and ESG scores across four sectors. The paper's findings showed no direct relationship between the financial performance of South African firms and their overall ESG score.

2.6. CONCLUSION

This section provided the theoretical overview of portfolio optimization through modern portfolio theory (MPT). The MPT is driven by the relationship between return and risk of assets, where the relationship is positive. In the case of traditional methods, the paper unpacked and discussed their advantages and disadvantages in implementing a portfolio. The main optimization methods discussed were mean-variance optimization, maximum return optimization, Sharpe ratio, and equal-weighted portfolio. These techniques exhibit common characteristics, including but not limited to imposed transactional costs and taxes; investors are rational and independent, and investors share the same ideas and outcomes. In the context of machine learning, the section unpacked linear regression, sector vector machine, decision trees, and random forests. The common traits of these models are that they aim to reduce the penalty or increase their accuracy rate through multiple training and testing models.

The empirical literature suggests that many studies have been conducted in developed countries, particularly in the United States of America, on how traditional and machine learning methods enhance portfolio optimization. However, more studies are needed in Africa, particularly South Africa. Additionally, with a focus on sectoral diversification, most studies revealed that an appropriate number of sectors should be included, ranging from eight to eleven, on average. The timeframe for the study depended on the data points that will ensure that it will provide a better understanding of how the interplay of traditional and machine learning works. Daily data is used for a shorter period than Monthly data, where researchers would need to consider a longer timeframe. There have been many methodologies one can use when implementing from a machine learning point of view. This shows the variety one can use depending on mastery of the techniques and software. Despite the mechanics, the results based on the literature suggest that machine learning methods will provide higher returns than traditional methods.

CHAPTER 3: RESEARCH CONTEXT

3.1. INTRODUCTION

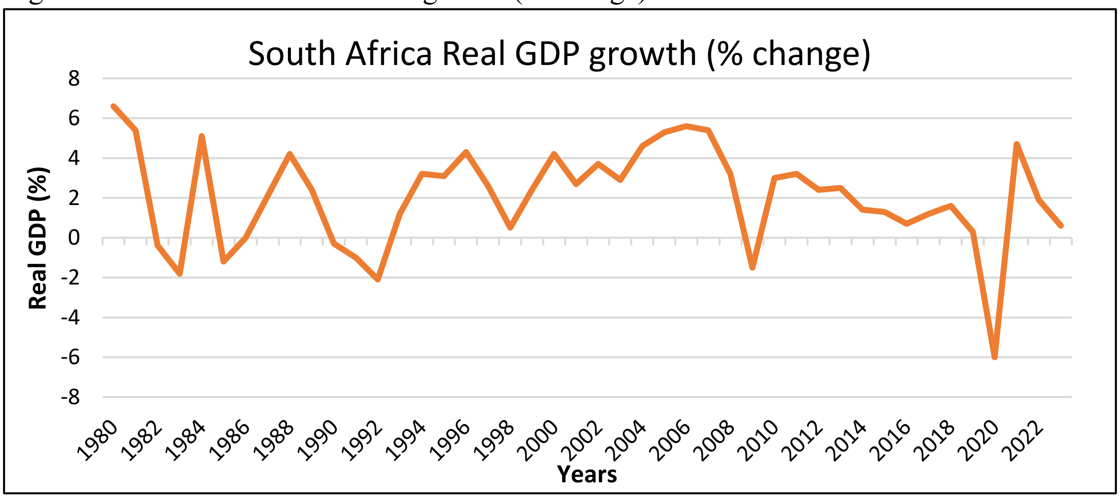
This chapter provides an overview of the economic and financial performance based on sectoral classes from pre-1994 to 2024 and 2019 to 2024, respectively. In addition, the chapter discusses the rise and the need to implement machine learning algorithms in the investment analysis process. The discussion highlights the benefits of machine learning over the traditional approach. In section 3.2, the chapter unpacks the economic sectoral performance as part of consideration in portfolio optimization. Section 3.3 presents an overview of economic growth between pre-1994 to 2024, and section 3.4 discusses an overview of financial performance from 2019 to 2024. Section 3.5 unpacks how machine learning algorithms and their integration into financial services have become relatively popular. Finally, section 3.6 concludes the chapter.

3.2. MOTIVATION OF UNPACKING ECONOMIC GROWTH PERFORMANCE

There is growing evidence that there is a relationship between business cycle and financial market performance. Jawadi, Ameur, Bigou, and Flageollet (2022) determined that there is a relationship between financial performance and economic performance and the business cycle helps to forecast a financial cycle. Muchaonyerwa and Choga (2015) determine there is a positive relationship between business cycle and stock market in South Africa. However. financial cycle tends to be longer than an ordinary business cycle as suggested by Farrell and Kemp (2020). In a typical expansion phase, stocks prices tend to increase in value while during recession phase stocks prices tend to move in the opposite direction (Choe, Masulis, and Nanda, 1993). It is with this view to unpack the historical and current economic climate trends of South Africa as it plays a key role on how portfolio selection will be determined to produce superior returns.

3.3. SOUTH AFRICA ECONOMIC PERFORMANCE

Figure 3.1: South Africa Real GDP growth (% change)



Source: (Author's own graph using International Monetary Fund (2024) data)

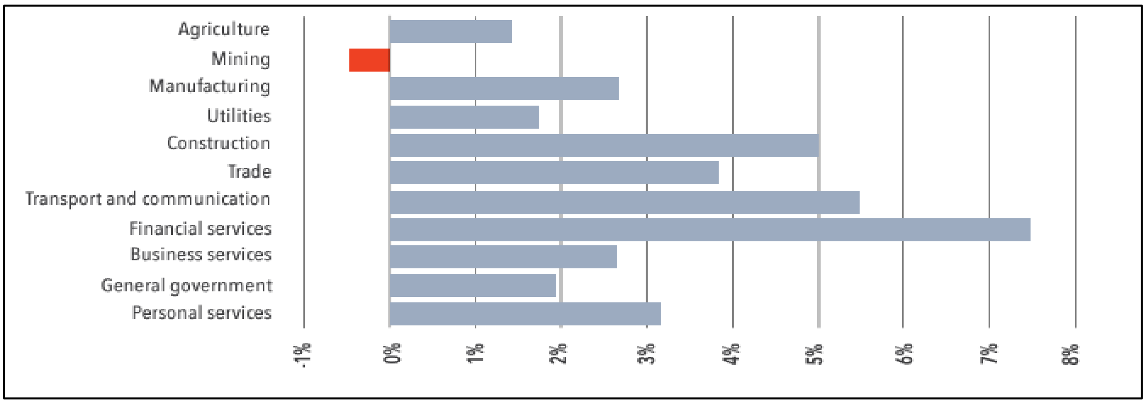
3.3.1. BEFORE 1994

South Africa experienced a consistent fluctuation in percentage change in Real Domestic Product (GDP) growth between -2% and 6%, as shown in Figure 3.1 above. Economic growth before 1994 was associated with political influence as South Africa was under apartheid by-laws. Between 1970 and 1980, economic growth was driven by capital and labour inputs, with the manufacturing sector playing an essential role, while technology was the worst-performing indicator of real GDP formation with a contribution of -0.5% (Faulkner and Loewald, 2008). In the early 1980s, South Africa adopted the tricameral constitution, which allowed the Indian and Coloured communities to participate in parliament proceedings, with white people having the majority voice, and this was running for over a decade. Despite the political tensions between 1970 and 1994, gross value was increased by an estimated 73% at a basic price. Over this period, the mining sector performed worse, while the insurance, business service, and real estate sectors heavily influenced economic growth (Faulkner and Loewald, 2008). On the contrary, the economic downturn faced by South Africa during the late 1980s and early 1990s was due to low investments and political uncertainty (Gondo, 2009). However, manufacturing was sector that restored economic growth in a positive outlook, while innovation in technology progression was neglected.

3.3.2. BETWEEN 1994 TO 2012

When the apartheid era was phased out, the government of South Africa took over 300 companies primarily dominated by the following sectors: energy, transport, telecommunications, and defence sectors (Hirsch, Levy, and Nxele, 2021). South Africa primarily focused on increasing employment, decreasing inequality, and reducing poverty to address socioeconomic challenges while promoting economic growth. The significant challenges in improving economic growth were maintaining or increasing investment within the country (National Planning Committee, 2020). However, this challenge was overcome when the economy was opened for global economic participation. From 1994 to 2008, South Africa experiences a positive outlook in real GDP percentage change ranging from 0 to 6%, as shown in the above Figure 3.1. However, due to the global financial crisis, the real GDP percentage slumped from 6% to -2% (International Monetary Fund, 2024). Despite facing a global financial crisis, the following sectors showed resilience in contributing to economic growth. The technology sector contributed as one of the fastest-growing sectors, averaging 9% between 1994 and 2012, followed by the financial service sector, contributing 7.5%. Agriculture contributed 1.4% towards gross value GDP, while the manufacturing sector and general government expenses contributed 2.7% and 1.9%, respectively (Department of Planning, Monitoring, and Evaluation, 2013). The mining sector is the only sector that experienced a decline of 0.4% in the gross value of GDP due to the decline of gold production in mining value from 50% to 16% between 1994 and 2012 (Department of Planning, Monitoring and Evaluation, 2013). See Figure 3.2 below.

Figure 3.2: GDP Gross Value Added Per Sector From 1994 to 2012

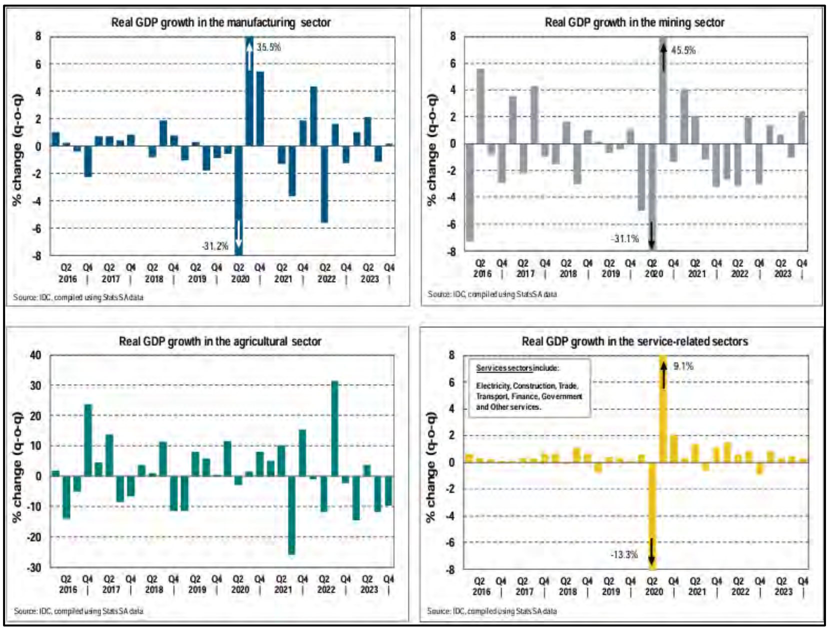


(Department of Planning, Monitoring and Evaluation, 2013)

3.3.3. BETWEEN 2013 TO 2024

From 2013 to 2018, the real GDP change in percentage declined from approximately 3% to 0%, as evident in figure 3.1. The main contributor to the slump in GDP was a result of load shedding, which had a significant impact on the operation of a business and a ripple effect on the overall economy. According to a South Africa Reserve Bank (2022), from 2014 to 2022, there was evidence of load-shedding showing a statistically significant negative impact on real GDP and its subsectors. The real GDP results indicated that manufacturing, storage and communication, transport, agriculture, forestry, fishing, and mining sectors were the most negatively impacted. In 2020, South Africa experienced a decline of 6% change in real GDP due to an economic shutdown because of the COVID-19 pandemic globally (International Monetary Fund, 2024). However, when the economy was reopened, the real GDP increased from -6% to approximately 4.5% in overall growth (International Monetary Fund, 2024).

Figure 3.3: Real GDP Growth in the Manufacturing, Mining, Agriculture and Services-Related Sector



Source: (Industrial Development Corporation, 2024)

Figure 3.3 above shows the selected sector's performance over the eight years from 2016 quarter 2 to 2023 quarter 4. The selected sectors are manufacturing, mining, agriculture, and services. The manufacturing sector had a minimal impact on real GDP growth, ranging from -2% to 2% between 2016 Q2 and 2020 Q1. Due to the pandemic, it experienced a loss of 31.2%, and when the economy reopened, the real GDP growth rose to 35.5%. Consequently, the sector experienced volatility in its real GDP after the pandemic till the end of 2023. Similarly, the mining sector experienced volatility in changes in real GDP growth ranging from -7% to 5.8% from 2016Q1 to 2020 Q1. Subsequently, it experienced the highest impact, decreasing to -31.1% during 2020 and increasing drastically to 45.5% in 2020 Q3. Overall, the sector faced fluctuations in real GDP ranging from -3.5% to 4%. Agriculture experienced significant changes in real GDP growth from approximately -13% to 23% from 2016 Q1 to 2021 Q2. The biggest takeaway in this sector is that despite the pandemic, the sector thrived, achieving 5% to 10% in growth towards the real GDP. Additionally, the most significant contribution was in 2022, when an increase of 30% in real GDP was experienced. In the services-related sector, from 2016 Q2 to 2020 Q1, experienced minimal impact on real GDP with less than 2%. However, during the pandemic, the sector experienced a decline of 13.3% in 2022 Q2 and then increased by 9.1% in the following quarter. Moreover, the service-related sector contributed to real GDP from less -than 1% to 2% between 2020 Q3 to 2023 Q4.

3.4. SOUTH AFRICA FINANCIAL PERFORMANCE

3.4.1. JOHANNESBURG STOCK EXCHANGE OVERVIEW

The Johannesburg Stock Exchange (JSE) came into existence on the 4th of November 1887 (JSE Group, 2024d). The current market capitalization is R10.94 billion as of October 2024 (Trading Economics, 2024a). Figure 3.4 below that shows JSE Market Capitalization from 2005 to 2024. JSE plays a crucial role in the economy, acting as an intermediary for buyers and sellers who wish to buy securities. Thus, JSE is licensed through the Financial Markets Act under the supervision of the Financial Sector Conduct Authority as the primary regulator and is underpinned by the Prudential Authority within the South African Reserve Bank (JSE Group, 2024e). JSE comprises "over 800 listed securities and approximately 400 listed companies, together with 60 Equity market member firms" (JSE Group, 2024f). These can be grouped into industrial classifications. The JSE constitutes eleven industrial classifications based on the Industry Classification Benchmark (ICB), called eleven sectors. These are: (i) Technology, (ii) Telecommunications, (iii) Healthcare, (iv) Financials, (v) Real Estate, (vi) Consumer Discretionary, (vii) Consumer Staples, (viii) Industrials, (ix) Basic Materials, (x) Energy, (xi) Utilities (JSE Group, 2024g).

Figure 3.4: JSE Market Capitalization

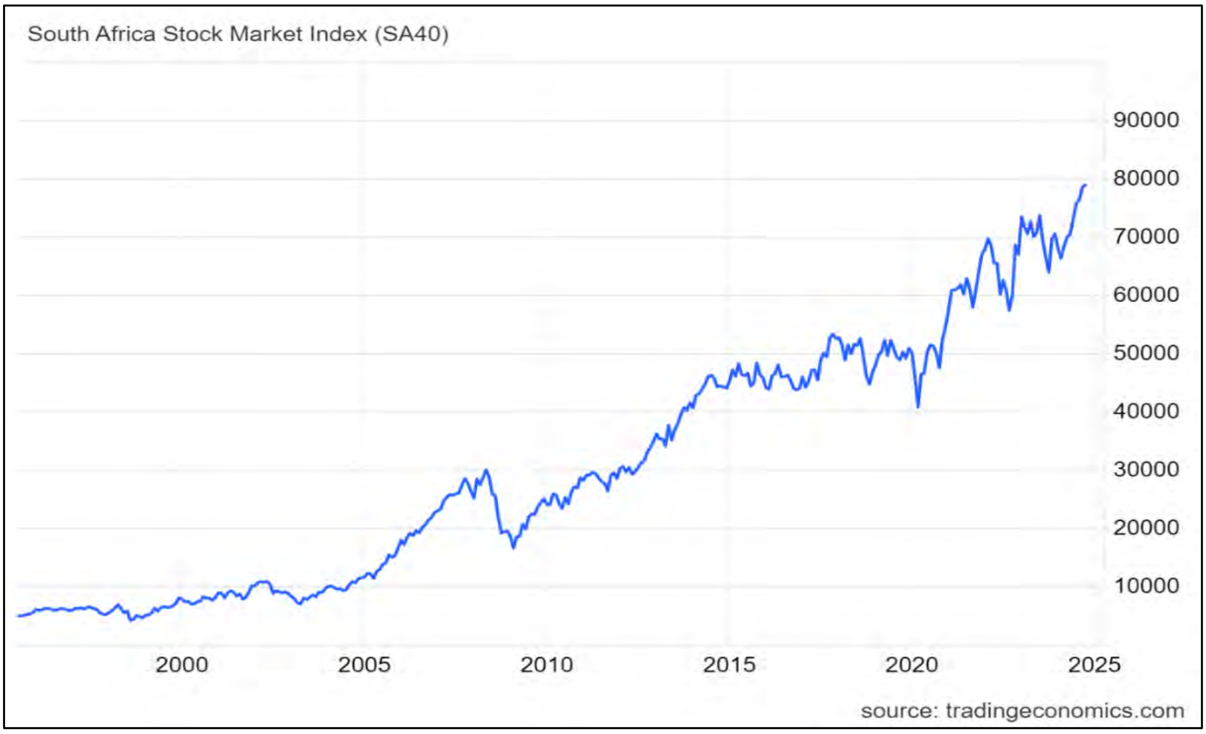
Source: (Trading Economics, 2024a)

3.4.2. PERFORMANCE OF JSE TOP 40

The JSE has established a benchmark as the FTSE/JSE Top 40 index, aggregating JSE Top 40 companies ranked by market value (London Stock Exchange Group, 2024). This benchmark serves as an overview of the current financial landscape in South Africa within the financial markets. Dating back to 30 June 1995, the JSE Top 40 index was valued at approximately 4779.621, and currently, to date, at the closing day of 24 October 2024, the share price is at 78881, which is approximately an increase of over 1550% in returns (Kotze, 2017; Trading Economics, 2024b). See the figure 3.5 below that shows the performance of the JSE Top 40 from 30 June 1995 to 24 October 2024. Between 1995 and 2008, the JSE index increased its returns over time to a value of approximately 31,315.34, the highest value recorded (Trading Economics, 2024b). The global financial crisis, with an excess increase in borrowing, made people unable to repay their loans. This eventually led to economic recession, as identified earlier (Majapa and Gossel, 2016). Due to the interconnectedness between economic growth and financial performance, the economic landscape had a ripple effect in the financial landscape, thus causing the Top 40 index to decrease in value to 16547.90.

After the global phenomenon, economic markets eventually recovered. Heymans and Heerden's (2014) paper studied the financial performance of JSE All-Share, a market capitalization-weighted index considering 99% of top companies, and JSE Top 40 against other global indices to establish if South Africa was the best place to invest in after the financial crisis using the period of January 2010 to December 2013. The study revealed that the JSE Top 40 and the JSE All-Share were the best two indices, on average, against other nine global indices such as the S&P 500, DAX, Nikkei 250, Dow Jones, and others. The pair also ranked the best two when it came to cumulating returns during the period, and it was among the top 4 based on high standard deviation. In this context, the risks of JSE All-Share Index was placed first. Similarly, from 2010 to 2013, JSE Top 40 and All-Share experienced the highest risk-adjusted returns as the top 2 indices compared to other selected global indices.

Figure 3.5: Performance of JSE Top 40 Index From 1995 to 2024



Source: (Trading Economics, 2024b)

Consequently, from the Figure 3.5 above, it is evident that the JSE Top 40 showed a continuous recovery in its asset prices, whereas the Top 40 index increased its prices between 2010 and 2015. From 2015 to 2019, there was some steadiness and a consolidation phase due to a combination of factors including but not limited to slow domestic economic growth, political uncertainty, and global outlook. Although not significant, the Top 40 index experienced a decline in its prices due to the COVID-19 pandemic as the economy was closed to save lives. After that, the Top 40 value has significantly increased to date. To put it into perspective, see the Table 3.1 below that shows the overall performance of South African markets, particularly the JSE Top 40 index. Within the last 5 years, from 2019 to 2024, the rate of return was at 55.39%; between October 2023 and October 2024 (1-year interval), the rate of return was at a staggering increase of 59,34%. In the short term, in the period of the last 1, 3, and 6 months, the rate of return of the index showed an increase of 4.27%, 16.16%, and 21.01%, respectively. In contrast, the average dividend yield from the top 40 companies was approximately 3.31%.

Table 3.1: The Overall Performance of South Africa Markets: JSE Top 40 Index

	Dividend	1 Month	3 Months	6 Months	1 Year	5 Years
JSE TOP 40 INDEX	3.31%	4.27%	16.16%	21.01%	59.34%	55.39%

Source: (TradingView, 2024)

3.4.3. JSE SECTORAL PERFOMANCE

Although, the JSE Top 40 and All-Share has performed relatively well over the years, it is important to note that these indices are primarily driven by individuals' assets which are group into sectors. As indicated earlier, that there 11 sectors within South Africa context. These 11 sectors play a role in driving the performance of the two indices specifically the All-Share indices as it covers all stocks available on the JSE exchange. A comparative analysis is made between all 11 sectors in the Table 3.2 below which shows the performance of the sector during short-term (1 to 6 months), medium term (1) and long-term (5 years) with the dividend yield gained within the current financial period.

Table 3.2: Sectoral Performance in South Africa As Per 21 October 2024

Sector	Dividend (%)	1 month	3 months	6 months	1 Year	5 Years
Technology	0.28%	11.86%	14.74%	14.60%	40.80%	66.86%
Tele - communications	4.45%	-2.92%	18.90%	16.38%	-2.13%	-13.21%
HealthCare**	2.85%	-0.285%	5.86%	21.22%	6.065%	20.91%
Finance	4.66%	1.48%	10.37%	28.36%	35.59%	49.02%
Real Estate	8.05%	-1.51%	11.35%	-	43.68%	-11.22%
Consumer non-durables	2.17%	7.25%	-7.35%	-3.94%	15.06%	137.30%
Consumer Staples	2.17%	-0.76%	0.92%	4.47%	12.23%	-1.90%
Industrials	2.16%	-3.54%	18.93%	73.37%	91.97%	86.50%
Basic Materials	-	9.80%	1.31%	-	-0.53%	9.94%
Energy	-	9.60%	-8.51%	-	-17.29%	38.51%

Notes: HealthCare is combined by Health Technology & Services from TradingView. In case where sectoral performance is not reported, the value is omitted in the table.
Source: (TradingView, 2024; Simple Wall Street, 2024)

3.4.3.1. SHORT-TERM: 1 TO 6 MONTHS

On average, the industrial sector has produced the highest rate of return of 29.59% within the last six months of 2024, followed by the Technology sector with an average of 13.73% and the finance sector, yielding a rate of return of 13.40%. The worst sector within this period was consumer non-durables, which experienced an average 4.04% decline, followed by consumer consumer staples increased, with an average of 1.54%. Subsequently, the rate of return of the JSE Top 40 Index, on average, over the last six months rose to 13.81%. This implies that all sectors except the industrial sector have outperformed the JSE Top 40 indices. This is not surprising considering that Minister of Finance Enoch Godongwana committed to increasing investments, as he announced during his Budget Speech in 2024, with an investment of R486.1bn in supporting public-private partnerships (National Treasury, 2024). This will increase the overall performance and efficiency of businesses, leading to higher revenues and potentially impacting the share prices to increase over time and in the coming years.

3.4.3.2. MEDIUM TERM: 1 YEAR

The best-performing sector over the medium term, in this case over 1 year period, is the industrial sector, which gained 91.97% in returns, followed by the real estate sector with a return of 43.68%, and the third best-performing sector is the technology sector, with a return of 40.80%. The worst-performing sector is utilities, with a decline of 32.67%, followed by the energy sector, with a decline of 17.29%, and followed by the technology sector, with a decline of 2.13%. Consequently, the rate of return of the JSE Top 40 Index, on average, over 1 year had a rate of return of 59.34%. Therefore, the utility sector is the only sector that performed better than the JSE Top 40 index compared to other sectors. The key to the success of utility sector performance was driven by the decline of load-shedding, which promoted an increase in the consumer confidence index, allowing a boost in the overall economy due to optimization from investor and business confidence (World Bank, 2024).

3.4.3.3. LONG TERM: 5 YEARS

The best-performing sector over the long term over a 5-year period is the utility sector, with a significant increase in the rate of return of 398.24%, followed by the consumer non-durables sector, with a 137.30% increase, and the industrials sector, with an 86.50% increase. The worst performing sector is the telecommunications sector, with a decline of 13.21%, followed by real estate, with a decline of 11.22%, and consumer staples, with a decline rate of return of 1.90% over the 5 years. The primary key to the rapid rise in the utility sector is businesses or firms transitioning to a green economy as part of the just-in-transition investment plan and independent power producer procurement program, which is aimed at decreasing carbon omissions mechanism and increasing renewable energy sources (The Presidency of the Republic of South Africa, 2024)

3.4.3.4.DIVIDEND YIELD

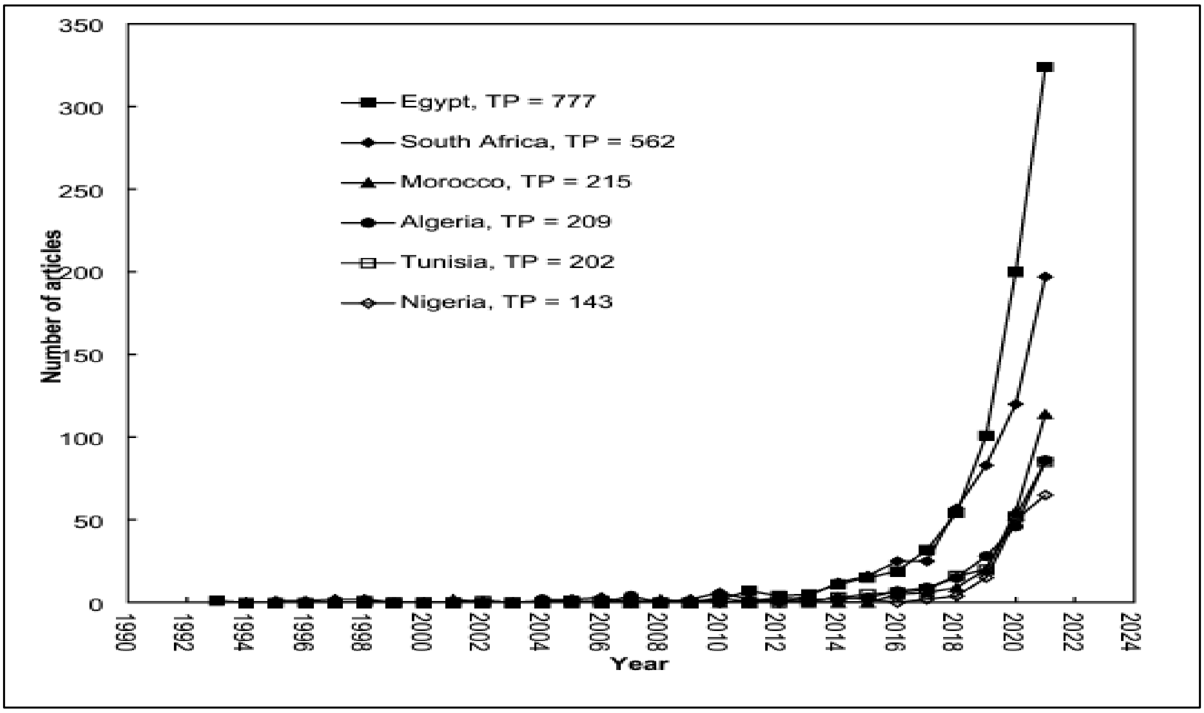
The dividend yield is the income earned from holding an asset. Overall, sectors show that holding an asset will generate an income between 0 and 10% over the financial period. The highest dividend accumulated is in the utilities sector, with an increase of 9.71%, followed by real estate, with an increase of 8.25%, and the finance sector, with an increase of 4.66%. The worst performing sector to receive income from holding an asset, although with a positive return, is technology 0.28%, followed by the industrial sector with an increase of 2.16%, and tied in the third worst sector in dividend yield is consumer non-durables and consumer staples with an increase of 2.17%. It is essential to consider growth stocks (performance-based stocks) and dividend stocks (income stocks) in having both stocks in a portfolio for better diversification when constructing a portfolio, as this will minimize risks.

3.5. ADAPTION OF MACHINE LEARNING IN AFRICA

3.5.1. MACHINE LEARNING IN AFRICA OVERVIEW

As the world moves into a fourth industrial revolution, there has been a rise in machine learning (ML) implementation in every field as it helps to address the limitations of traditional approaches such as pattern recognition, natural language processing, and many others. Ezugwu, Oyelade, Ikotun, Agushaka, and Ho's (2023) studies showed an overview of the rapid rise in machine learning research over 30 years between 1991 and 2021 in Africa using a bibliometric analysis review. See the Figure 3.6 below, which shows the number of articles that have been published over time. The selected countries, such as Egypt, South Africa, Morocco, Algeria, Tunisia, and Nigeria, have gained the most published articles in accredited journals. The rapid rise began with Egypt and South Africa in 2014, while other countries lagged. The other four countries started publishing ML articles from the period of 2016. At the time of this study, South Africa was the second-dominated country in Africa, with 562 published articles. These 562 published articles translate to 23% of the total number of articles evaluated against 2 468 articles.

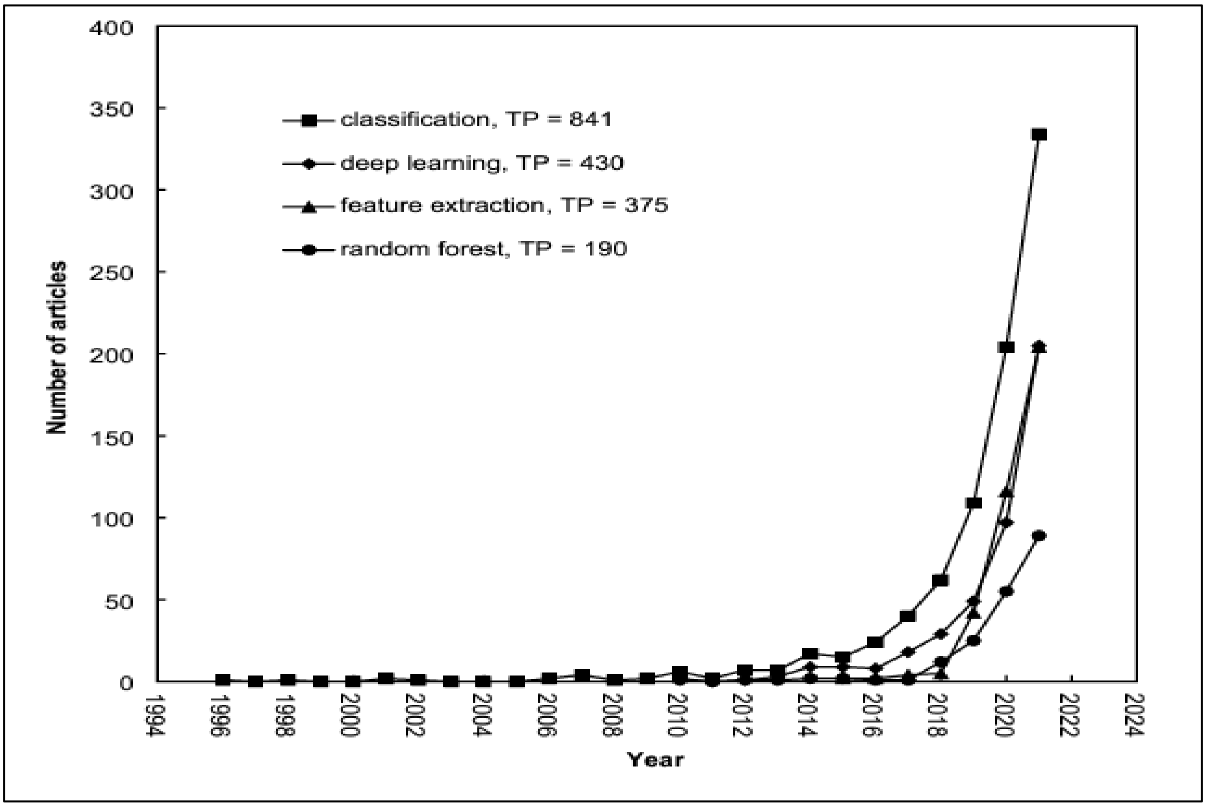
Figure 3.6: Number of Published Articles Between 1991 and 2021



Source: (Ezugwu et al, 2023)

Ezugwu et al. (2023) unpacked the key development trends of Africa's four most popular ML topics. The results show that classification (841 articles), deep learning (430), feature extraction (375), and random forest (190) are the most popular by 2021. Figure 3.7 below, shows the development trends over the 30 years. The classification method was the first method used around 2010, and it started gaining popularity in 2014 while other methods lagged. From 2018 until today, there has been an exponential increase in the number of published articles in Africa. It is also worth noting that these published articles were research articles from all career fields and not specifically the finance sector alone.

Figure 3.7: Number of Articles of the Developmental Trends of Most Popular ML Methods



Source: (Ezugwu et al, 2023)

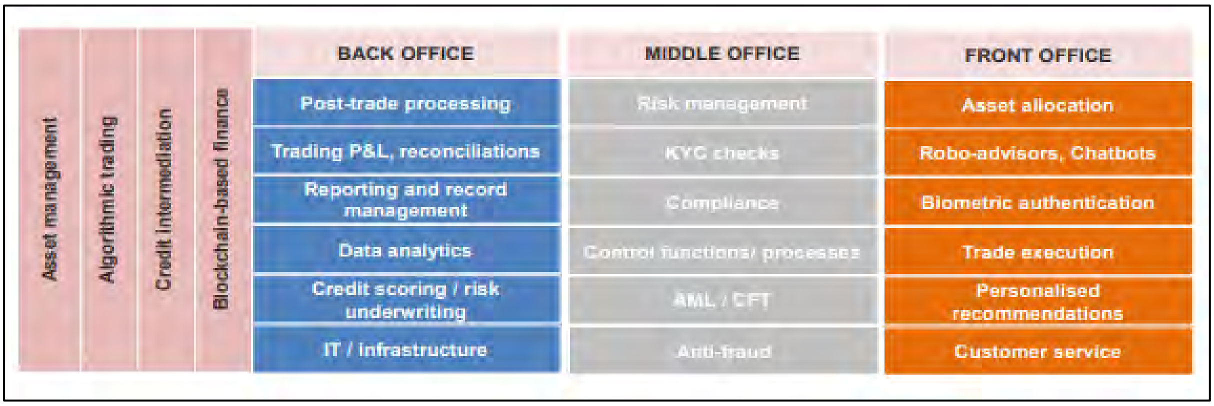
3.5.2. MACHINE LEARNING MARKET SIZE

There has been a recent increase in the machine learning market due to the rapid influx of technology use cases in work-related sectors in South Africa. In 2025, the market size is expected to increase to US$0.48bn (Statista, 2024). Forecasting machine learning market size between 2025 and 2030 is expected to increase on annual basis with expected growth of 34.72% yearly. The market size is driven by three sectors: the retail, finance, and the healthcare sector (Statista, 2024). With the increase in machine learning, the government of South Africa adopted the Artificial Intelligence National Plan as part of a strategic use case in the South African economy. This adoption is driven by following global trends to attract local and foreign investment opportunities in the country (Mashishi, 2023). In this venture, South Africa has prioritized four elements: predictive maintenance, logistic optimization and automated services, and diagnostic and analytical abilities (Mashishi, 2023).

3.5.3. MACHINE LEARNING IN FINANCE SECTOR

While there is rapid growth in machine learning (ML), financial services institutions use it to prompt signals and establish relationships in data that are too complex to comprehend at once. However, ML is used not only for decision-making but also for various financial market activities that include back, middle, and front offices; see Figure 3.8 below, which outlines some of the use cases of ML in the financial sector (OECD, 2021). Within the portfolio optimization, there is interconnectedness between these three offices where there is a need to analyse data (back office) to make an informed decision on how investors meet their investment goals based on their risk appetite, which is essential for risk management during asset allocation (front office)

Figure 3.8: Machine Learning Applications in Financial Markets.



Source: (OECD, 2021)

ML relies on big data, meaning the more extensive the dataset, the better the performance. Within asset management, ML has been used as a prominent feature for asset allocation through training and testing datasets, which improve data analysis and predictions over time. According to industry research by BarclaysHedge (2018), it was estimated that the adoption of ML within investment firms uses it for portfolio construction (58% indicating that), risk management (33%), and trading executions (27%) from 55 correspondence. Within their decision process, 19% of the companies fall into the category of 80% to 100% using ML to make their investment decisions, 6% of the companies rely on ML 60% to 80% most of the time, 28% of the companies rely on ML 40% to 60% most of the time, 28% of the companies rely on ML 20% to 40% most of the time and 19% of the companies rely on 1% to 20% on ML most of the time to make investment decisions.

Due to the nature of adopting ML in the financial service industry, there is a possibility of larger firms outpacing small firms in integrating ML in their investment process. This is due to the need for more resources, such as limited technology tools, and a need for employees with the required skills to succeed in the role. On the contrary, using ML algorithms may lead to herding behaviour and one-way markets (OECD, 2021). The upside risk of using ML algorithms is that they do not consider emotions, thus eliminating potential behavioral bias, which eliminates the potential for greed and fear. In addition, it reduces the dependence on human intervention and thrives with massive datasets after repeated tasks to reduce risk and maximize returns (Sanlam, 2022). On a macro level, implementing ML can help to reduce operational costs by 22% by 2030 (OECD, 2021).

In South Africa, ML has been considered the most useful technology in the companies' survey based on Ernst and Young's (2018) firm, with 67% of the company's utilizing ML in their field of work from 24 people surveyed from South Africa from over 100 people from different regions. Within ML, the most common was supervised ML, which relies on structured data to find patterns and interpret new data variables. In addition, the case study by Ernst and Young (2018) showed a rise in investment in ML from two transactions in 2008 to 171 transactions by 2018 from the selected countries for the study. 134 out of 171 transactions were from South Africa, of which 31 were associated with mergers and acquisitions. Thus, South Africa received $1 658 million in investments towards ML during the ten years. Moreover, a study by Price Waterhouse and Coopers (2022) indicated that 90% of institutional investors from a pool of 250 participants integrate ML in their portfolio optimization process to maximize returns while minimizing risks. However, the drawback of implementing ML is inclusive of market expectations in the investment process.

Alex Forbes (2024) conducted a similar study on the use case of ML in the investment analysis process from a pool of 20 participants among 24 asset managers. The study's main findings indicated that 34% of the investors indicated they have integrated ML algorithms as part of the investment decision-making process. The low integration of ML in the investment analysis process varies to concerns of data quality and availability (15.79%), interpretability and transparency of ML (15.79%), lack of skilled personnel (15.79%), regulatory concerns (7.89%) and high implementation concerns (7.89%). In addition, 66% of the asset managers fail to understand the positive outcomes of using ML in the investment decision-making process, whereas 24% of the respondents indicated that they included ML due to its competitive nature compared to the traditional methods. In comparison, 20% of the investors indicated that the inclusivity of ML is due to the improvement in risk-adjusted returns. Moreover, the study revealed that 37.5% of the asset managers will expand their operations by including ML in the investment process, and 16.67% will maintain their current form of using ML in the investment process as they are generally happy. However, there are open to embrace more technological advancement within the style. Finally, 8.33% of the asset managers plan to expand the use of ML.

3.6. CONCLUSION

This chapter analysed the economic growth performance from 1994 to 2024 and financial performance from 2019 to 2024 and discussed the rise of machine learning algorithms in Africa. This chapter forms the foundation of our study. The chapter indicated that sectors' performance affected economic growth during a prescribed period. An increase in economic sectoral performance contributed to an increase in economic growth. Similarly, financial sectoral performance contributed to internal shocks in which economic factors played a role. Thus, there is a link between financial performance and economic performance.

In addition, there has been a rapid increase in machine learning algorithms in which classification, deep learning, feature extraction, and random forest methods were the most popular by 2021. While there has been increased research output when conducting analysis, there have been mixed results on embracing machine learning in the industry. Some financial market experts support integrating machine learning to increase risk-adjusted returns and provide competitive advantages. However, experts argue that there needs to be more data quality and availability, interpretability and transparency, skilled personnel, regulatory measures, and high implementation costs. This leads to the next chapter, which unpacks the comparison of traditional and machine learning methods, which will provide the highest expected returns over time.

CHAPTER 4: DATA AND METHODS SECTION

4.1. INRODUCTION

In Chapter Two, the study discussed the theoretical framework and literature review on portfolio optimization on both traditional and machine learning, while Chapter Three detailed the economic and financial performance of different sectors in South Africa. Additionally, Chapter Three shows the relevance of machine learning in daily activities and how it is becoming relevant in the financial sector. It is worth reminding that the primary goal of this research is to identify which model between traditional approaches and machine learning model would yield the highest risk-adjusted expected returns with sub-goals of identifying how many stocks will be selected to produce an optimal portfolio and which sectors the stocks come from. Secondly, which traditional method would yield the highest risk-adjusted returns, and thirdly, which machine learning method would yield the highest risk-adjusted returns? With that in mind, this chapter provides the analytical framework that will be used for this study. Section 4.2 discusses the method of gathering the data. Section 4.3 presents a brief overview of the descriptive analysis. Section 4.4 outlines the empirical analysis in depth. Section 4.5 provides the evaluation criteria used for this study. Section 4.6 concludes the chapter.

4.2. DATA AND SOURCE

The variables used for the study consist of 27 stocks. These 27 stocks were chosen randomly from the most active stocks, similar to the approach to the study used by Boya (2024) with three stocks per sector for the purpose of the sectoral analysis. The choice of three stocks per sector is consistent with Zaimovic, Omanovic, and Arnaut-Berilo (2021) who demonstrated that "an optimal number of stocks that constitute a well-diversified portfolio does not exist for whatever market, period nor investor". The author's reached the conclusion after evaluating 150 papers between 1993 and 2021 in their study on how many stocks are required for an optimal equity portfolio diversification. South Africa has 11 sectors namely: Basic Materials, Consumer Discretionary, Consumer Staples, Energy, Financial Services, Health Care, Industrial, Real Estate, Technology, Telecommunications, and Utilities. All variables used for the study are detailed in the Annexture Table A3. In this study, the Energy and Utilities sectors were omitted. Energy sector stocks were omitted as there were more than ten missing data points from the available stocks. In contrast, the utility sector was omitted as, at the time of writing, no stocks in the JSE were defined in this sector.

The data for the analysis covers a period of 30 September 2020 to 30 September 2024 using daily closing prices, making 1251 daily observations for each stock selected. Subsequently, when conducting machine learning methods, the data used is daily closing prices, low prices, high prices, open prices, and trade volume, which also consist of 1251 daily observations. Based on systematic review of 50 literature reviews papers from 1998 to 2020, Prasad and Seetharaman (2021) prescribed that the various price variables and volume are the most important variables when using machine learning. The period from 30 September 2020 to 30 September 2024 was selected because most of the literature that uses daily prices for analysis has an average of five years the analysis, as shown in Tables A1 and A2 annexure. The data was retrieved from Google Finance via Google Excel and downloaded for further analysis. Consequently, the data is expressed in its original state. Additionally, the All-Share Index historical prices are obtained for the same period of study from Investing.com website.

4.3. OVERVIEW OF DESCRIPTIVE ANALYSIS

The study will provide summary of the descriptive data which includes the mean, median, maximum and minimum value, standard deviation, skewness, kurtosis, Jarque Beta, and number of observations in a table format. This summary table will provide an overview of the type of data that will be analysed and its appropriateness to capture trends. This data will be executed from daily returns view, which is a similar to the approach Sen, Dutta, and Mehtab (2021) performed over their training dataset on daily prices. Likewise, Bessler and Woff (2024) provided a descriptive summary statistic based on monthly observations.

For this method, a table will be generated using Python programming language which display the same output parameters and associated results. Additionally, the study will employ correlation analysis and produce covariance matrix to capture the relationship between assets and their movement. This is instrumental for portfolio diversification. The ideal portfolio is to have mixture of opposite relationships movements of assets to avoid excessive losses should assets lose their values. An additional feature to the table is a visually heatmap on the correlation of the assets as shown by Das, Bowala, Thulasiram, and Thavaneswaran (2024) in their execution of portfolio optimization on actual data and predicted data. However, in this study, a heatmap will be shown based on the actual data on both sector and individual stocks correlation. The author will include also provide covariance heatmap for completeness.

4.4. EMPERICAL ANALYSIS

4.4.1. TRADITIONAL METHODS

4.4.1.1. EQUALY WEIGHTED PORTFOLIO

In a portfolio of N risky assets is weighted subjected to minimizing risks and maximizing returns. A portfolio weight is the percentage of the stock that is found in the overall investment portfolio. The weight of a stock has significance influence on the overall expected return of the portfolio (Taljaard and Mare, 2021; Hanif, Hanun, and Febriansah, 2021). In this context, weights are denoted as w_s which referred to the weight of stock s-th stock within a portfolio of stocks for s=1, 2, …., t where t represents the total number of stocks in the portfolio. Weights sum up to 1 or 100%. Equally Weighted (EW) Portfolio is the easiest portfolio to construct where the weights of the stocks are equal. The weighting of the stocks is determined by the number of stocks in a portfolio. The portfolio calculation follows AlHalaseh and Shawawreh (2024) process:

$$w_i = \frac{1}{s}\ for\ i = 1,2,\cdots,s \tag{4.1}$$

When the weights are known, parameter such as portfolio return (μ_{ew}), portfolio variance (σ_{ew}^2), portfolio standard deviation (σ_{ew}) and the Sharpe ratio (SR_{ew}) can be obtained by the following: Initially, portfolio return is found by multiplying the weightings of assets with the mean of stock s.

$$\mu_{ew} = w\bar{\mu}_s = [w_1 \quad w_2 \quad \cdots \quad w_s]\begin{bmatrix}\mu_1\\ \mu_2\\ \vdots\\ \mu_s\end{bmatrix} \tag{4.2}$$

$$\mu_{ew} = \frac{\mu_1 + \mu_2 + \cdots + \mu_s}{s}$$

The result of the portfolio return will be used at the end of derivations when computing the Sharpe ratio for the equally weighted allocation. Next, the portfolio standard deviation is determined. The portfolio standard deviation can be obtained by finding the variance of the portfolio and then using the square root of the result. Mathematically, the output of portfolio variance is the product of the weightings of stocks s with the vector covariance matrix and the transpose of the weighting of stock s. The result will yield:

$$\sigma_{ew}^2 = wVw^T = [w_1 \quad w_2 \quad \cdots w_s]\begin{bmatrix}\sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1s}\\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2s}\\ \vdots & \vdots & \ddots & \vdots\\ \sigma_{1s} & \sigma_{2s} & \cdots & \sigma_s^2\end{bmatrix}\begin{bmatrix}w_1\\ w_2\\ \vdots\\ w_s\end{bmatrix} \tag{4.3}$$

$$\sigma_{ew}^2 = \frac{\sigma_1^2}{s^2} + \frac{2}{s^2}\sigma_{12} + \cdots + \frac{2}{s^2}\sigma_{1s} + \frac{\sigma_2^2}{s^2} + \cdots + \frac{2}{s^2}\sigma_{2s} + \frac{\sigma_s^2}{s^2}$$

$$\sigma_{ew} = \sqrt{\frac{\sigma_1^2}{s^2} + \frac{2}{s^2}\sigma_{12} + \cdots + \frac{2}{s^2}\sigma_{1s} + \frac{\sigma_2^2}{s^2} + \cdots + \frac{2}{s^2}\sigma_{2s} + \frac{\sigma_s^2}{s^2}}$$

From the output, the mean of our equally weighted stock is divided by the portfolio standard deviation to obtain the Sharpe ratio of the equally weighted portfolio. Deriving the Sharpe ratio will result in the following:

$$SR_{ew} = \frac{\mu_{ew}}{\sigma_{ew}} \tag{4.4}$$

$$SR_{ew} = \frac{\mu_1 + \mu_2 + \cdots + \mu_s}{s\sqrt{\frac{\sigma_1^2}{s^2} + \frac{2}{s^2}\sigma_{12} + \cdots + \frac{2}{s^2}\sigma_{1s} + \frac{\sigma_2^2}{s^2} + \cdots + \frac{2}{s^2}\sigma_{2s} + \frac{\sigma_s^2}{s^2}}}$$

The equally weighted portfolio Sharpe ratio will serve as a benchmark against other optimization traditional methods. Should other portfolio methods yield Sharpe ratio that is lower than equally weighted Sharpe ratio will be immediately disregarded for consideration.

4.4.1.2. MEAN VARIANCE PORTFOLIO

The purpose of the mean-variance portfolio method is to minimize the risk of assets in each portfolio while ignoring the level of rate of return (Chen, Zhang, Mehlawat, and Jia, 2021). In a classical mean-variance optimization model, the objective function is:

$$\sigma_s^2 = wVw^T \tag{4.5}$$

$$\sigma_s^2 = \begin{bmatrix} w_1 & w_2 & \cdots w_s \end{bmatrix} \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1s} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1s} & \sigma_{2s} & \cdots & \sigma_s^2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_s \end{bmatrix}$$

Where σ_s^2 is the variance of the portfolio s, w represents the weights of the stocks subject to w=1, 2, …, m. V is the vector covariance matrix and w^T is the transpose of the weights. The above objective function is obtained by multiplying the weights row vector with the variance-covariance matrix, resulting in a matrix of 1xs dimension. The product of the variance-covariance matrix and a matrix of 1xs will be multiplied by the transpose of the weights row vector to produce a scalar or one matrix.

If the level of return is given, it will yield the following result:

$$\min s_{mvo}^2 = w_1^2\sigma_1^2 + 2w_1w_2\sigma_{12} + \cdots + 2w_1w_s\sigma_{1s} + w_2^2\sigma_2^2 + \cdots + 2w_2w_s\sigma_{2s} + w_s^2\sigma_s^2 \tag{4.6}$$

$$\therefore \max -s_{mvo}^2 = -(w_1^2\sigma_1^2 + 2w_1w_2\sigma_{12} + \cdots + 2w_1w_s\sigma_{1s} + w_2^2\sigma_2^2 + \cdots + 2w_2w_s\sigma_{2s} + w_s^2\sigma_s^2)$$

$$subject\ to:$$

$$\mu_{mvo} = w_1\mu_1 + w_2\mu_2 + \cdots + w_s\mu_s = \alpha \rightarrow w_1\mu_1 + w_2\mu_2 + \cdots + w_s\mu_s - \alpha = 0$$

$$w_1 + w_2 + \cdots + w_s = 1 \rightarrow w_1 + w_2 + \cdots + w_s - 1 = 0$$

$$w_1 \geq 0 \rightarrow -w_1 \leq 0$$

$$w_2 \geq 0 \rightarrow -w_2 \leq 0$$

$$\vdots$$

$$w_s \geq 0 \rightarrow -w_s \leq 0$$

By contrast, when the level of risk is given, the output will yield to the following result:

$$\max \mu_{mvo} = w_1\mu_1 + w_2\mu_2 + \cdots + w_s\mu_s \tag{4.7}$$

$$subject\ to:$$

$$s_{mvo}^2 = w_1^2\sigma_1^2 + 2w_1w_2\sigma_{12} + \cdots + 2w_1w_s\sigma_{1s} + w_2^2\sigma_2^2 + \cdots + 2w_2w_s\sigma_{2s} + w_s^2\sigma_s^2 - \beta = 0$$

$$w_1 + w_2 + \cdots + w_s = 1 \rightarrow w_1 + w_2 + \cdots + w_s - 1 = 0$$

$$w_1 \geq 0 \rightarrow -w_1 \leq 0$$

$$w_2 \geq 0 \rightarrow -w_2 \leq 0$$

$$\vdots$$

$$w_s \geq 0 \rightarrow -w_s \leq 0$$

The above optimization problem is solved by following Boya (2024) principle of using python language using a special solver of solver.qp function which is associated with CVXOPT library package. The contribution to this study is when the model selects the stocks for the portfolio; the portfolio should have a minimum number of stocks, and it will be set to 10 minimum stocks in the portfolio to promote diversification.

4.4.1.3. MAXIMUM SHARPE RATIO PORTFOLIO

The Maximum Sharpe ratio (MSR) is a combination of the Minimum Variance and the Maximum Return portfolio, referred to as the extension of the variance optimization portfolio. In this portfolio, the same principle applies where investors seek to maximise returns while minimising returns. The distinguishing aspect that sets it apart is the application of bi-objective optimization, which categories fall under the multi-objective optimization framework (Stampe and Revelsby, 2022). The framework involves an approach that solves several single-objective subproblems where additional objectives are transformed into constraints (Rompotis, 2023).

The process is as follows:

$$\max SR_{msr} = \frac{w_1\mu_1 + w_2\mu_2 + w_3\mu_3 + \cdots + w_s\mu_s}{\sqrt{w_1^2\sigma_1^2 + 2w_1w_2\sigma_{12} + \cdots + 2w_1w_s\sigma_{1s} + w_2^2\sigma_2^2 + \cdots + 2w_2w_s\sigma_{2s} + w_s^2\sigma_s^2}}$$

$$\begin{gathered} subject\ to: \\ w_1 + w_2 + \cdots + w_m = 1 \rightarrow w_1 + w_2 + \cdots + w_m - 1 = 0 \\ w_1 \geq 0 \rightarrow -w_1 \leq 0 \\ w_2 \geq 0 \rightarrow -w_2 \leq 0 \\ \vdots \\ w_s \geq 0 \rightarrow -w_s \leq 0 \end{gathered} \tag{4.8}$$

In this context, when the Sharpe ratio is large, it implies the greater performance of the portfolio. If there is a low Sharpe ratio, the portfolio is not producing returns at its optimal level. This is due to the compensation for each unit risk at different levels (Ngo, Nguyen, and Van Nguyen, 2023).

4.4.1.4. RISK PARITY PORTFOLIO

There are somewhat similarities between equal weighted portfolio and risk parity portfolio, where equal weighted portfolio focusses on having equal distribution of weights to all stocks. However, with risk parity portfolio works on having marginal contribution to portfolio risk of each stock to be equal (Fagerholt and Aanonsen, 2016). In this portfolio, the paper applies Fagerholt and Aanonsen (2016) partials work towards optimization concept in evaluating marginal contribution of risk of a stock. The marginal contribution of risk to each stock is evaluated by the following formula:

$$MC_s = \frac{w_s}{\sigma_{rp}} \sum_{j=1}^{m} w_j \sigma_{sj} \tag{4.9}$$

$$MC_s = \frac{w_s}{\sigma_{rp}} \begin{bmatrix} w_1 & w_2 & \cdots & w_m \end{bmatrix} \begin{bmatrix} \sigma_{s1} \\ \sigma_{s2} \\ \vdots \\ \sigma_{sm} \end{bmatrix}$$

$$MC_s = \frac{w_i}{\sigma_{rp}} (w_1\sigma_{s1} + w_2\sigma_{s2} + \cdots + w_m\sigma_{sm})$$

Where it is constraint is by:

$$MC_s = MC_{s+1}\ for\ i \in [1; m)$$

MC_s denotes marginal contribution of risk to stock s. Other variables have already been defined in previous sections. To ensure that there is marginal contribution of stock across all stocks selected, the weightings and risk will yield the following process:

$$MC_{m-1} = MC_m \tag{4.10}$$

$$\begin{aligned} w_{s-1}(w_1\sigma_{1(s-1)} + w_2\sigma_{2(s-1)} + w_3\sigma_{3(s-1)} + \cdots + w_{s-1}\sigma_{s-1}^2 \\ = s(w_1\sigma_{1s} + w_2\sigma_{2s} + w_3\sigma_{3s} + \cdots + w_s\sigma_s^2) \end{aligned}$$

$$\begin{aligned} w_{s-1}(w_1\sigma_{1(s-1)} + w_2\sigma_{2(s-1)} + w_3\sigma_{3(s-1)} + \cdots + w_{s-1}\sigma_{s-1}^2 \\ - w_s(w_1\sigma_{1s} + w_2\sigma_{2s} + w_3\sigma_{3s} + \cdots + w_s\sigma_s^2) = 0 \end{aligned}$$

$$\therefore MC_1 = MC_2 = MC_3 = \cdots = MC_{s-1} = MC_s$$

To optimize a risk parity portfolio, an objective function of Sharpe ratio is selected to maximise the return for this portfolio. The process will follow:

$$\max \mu_{rp} = w_1\mu_1 + w_2\mu_2 + \cdots + w_s\mu_s \tag{4.11}$$

$$subject\ to:$$

$$\begin{aligned} w_{s-1}(w_1\sigma_{1(s-1)} + w_2\sigma_{2(s-1)} + w_3\sigma_{3(s-1)} + \cdots + w_{s-1}\sigma_{s-1}^2 + w_s\sigma_{(s-1)s}) \\ - w_s(w_1\sigma_{1s} + w_2\sigma_{2s} + w_3\sigma_{3s} + \cdots + w_s\sigma_s^2) = 0 \end{aligned}$$

$$w_1 + w_2 + \cdots + w_s = 1 \rightarrow w_1 + w_2 + \cdots + w_s - 1 = 0$$

$$w_1 \geq 0 \rightarrow -w_1 \leq 0$$

$$w_2 \geq 0 \rightarrow -w_2 \leq 0$$

$$\vdots$$

$$w_s \geq 0 \rightarrow -w_s \leq 0$$

4.4.2. MACHINE LEARNING MODELS

This study will use four machine learning models for regression analysis which are Linear Regression Model, Decision Tree, Random Forest and Sector Vector Machine which follows study like Sen et al. (2022) and Jones (2023) methodology. These models analysed using Google Colaboratory (Python programming). All models will follow a training, testing and validation phase. The process of training phase comprises of the model learning the historical price data with the aim to minimize the difference between predicted output and the actual output. After training phase, we apply validation of our data in which is used to fine-tune and optimize the model's hyperparameters. Finally, we apply a testing phase where it is used to evaluate the final performance of the trained model. The data is split into 60-30-10 manually. The predictor's parameters of the models include opening, low and high prices, the volume of the stock and market capitalization. The target parameter will be the closing prices of the stock.

As mentioned in the previous chapter, the study will apply this methodology because machine learning takes different market data to be able to make prediction of future performance of a stock unlike the traditional method, where it only considers historical stocks prices. From predicting future performance of a stock, the models will then select stocks that will optimize the portfolio. In addition, a minimum of ten stocks should be selected in a portfolio thus, this constraint is applied to all models. The following section will describe machine learning methods in detail.

4.4.2.1. LINEAR REGRESSION MODEL

Assume the linear regression takes the following functional form:

$$y = f(X, \alpha) + \epsilon = f(X_1, X_2, \cdots, X_m, \alpha_1, \alpha_2, \cdots, \alpha_n) + \epsilon \tag{4.12}$$

Where y is the dependant variable and in ML is referred to the target variable, X is the independent variable and in ML is referred to input or feature variables. α denotes as an estimate or parameter and ϵ is the noise parameter. The linear regression model captures the relationship between dependant and independent variables which is assumed to be a linear relationship. However, our study consists of having opening, low and high prices, and the volume of the stock in which there are independent variables; this calls for applying a multi-linear regression model to our study. Therefore, the model of the study becomes in the form of:

$$y = \alpha + \beta_1 Open\_Price_1 + \beta_2 Low\_Price_2 + \beta_3 High\_Price_3 + \beta_4 Volume_4 + \epsilon \tag{4.13}$$

Where y is the forecasted closing daily price. The regression model will be runed via Google Colab using sklearn model selection, and linear regression model will be built with the aim to calculate closing daily prices. From the forecasted results, the author will create algorithm for stock allocation with the relevant weightings.

4.4.2.2. DECISION TREE REGRESSION MODEL

The decision tree model comes in two forms. Firstly, it can be a classification decision tree where it splits data based on categorial values or feature names, and secondly, it involves a regression decision tree where it splits data based on continuous data or numerical values. In this case, to conduct stock market prediction to choose stocks that will optimize the asset portfolio, the author will implement a decision tree regression model. This analysis will follow Burhan's (2023) method of application towards predicting the performance of decision tree regression in a stock exchange. This study will adopt a binary split process, deciding whether a stock should be selected based on its forecasted financial performance. The objective is to minimize the sum of squared deviations from the mean from each independent sample. To minimize the sample, the author will calculate the method of least squares used to solve the optimization problem, which is defined as:

$$Min_{w,b}\ MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - f_{L,c}(X_i)\right)^2 \tag{4.14}$$

Where L is a D-dimensional vector of parameters and c is a natural number and is the model that needs to be estimated. The model consists of several vector input features of X_i with one target variable of y_i.

Additionally, $\left(y_i - f_{L,c}(X_i)\right)^2$ is defined as the loss function that needs to be minimized over L and c. The minimization can be achieved by transforming the weights and the constant. The loss function has a quadratic formula consisting of a closed solution and a well-behaved derivative. The solution to this optimized problem is L* and c*. The above process will continue until the minimum node size is met (Burhan, 2023). This is a continuous partition or a sliding-window approach. When the process has been completed, the next step is to make predictions in which a path is followed from a root node (starting point); there will be a series of comparisons of values for the selected features associated with a branch. The process will continue until the ultimate lead node (endpoint) provides the final prediction value.

The model will be run via Google Colab through a Decision Tree Regressor model, which will be responsible for predicting the outcomes. The parameters that will be fine-tuned will be the number of feature variables, the number of decision trees, and the maximum depth of the tree. The parameters are selected by the default function of Google Colab. Below is Table 4.1 which shows parameters of the grid search. The parameter grid will consist of:

Table 4.1: Decision Tree Regression Parameter Grid Search

Maximum depth	[2, 3, 4, 5, 6]
Minimum sample of sub-tree split	[2, 5, 10]
Minimum sample leaves	[1, 2, 4]

Source: Computed by Author

4.4.2.3. RANDOM FOREST REGRESSION MODEL

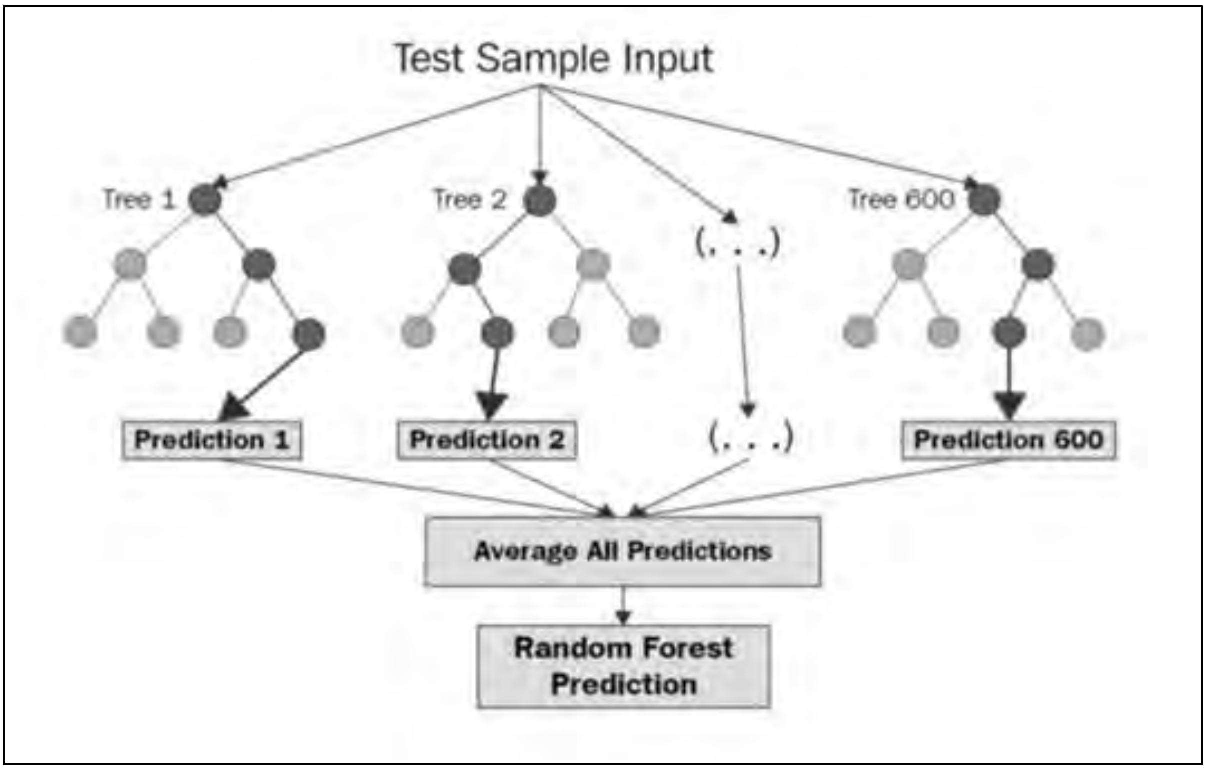
Burhan (2023) pointed out that having one decision tree model may result in less accuracy due to its high sensitivity through minor changes. A random forest, an assembly of decision trees, should be implemented to address this. Random Forests address the limitations of decision trees; thus, using random forests decreases the model from overfitting and enhances the stability of the model. In addition, a random forest is superior to a decision tree because it absorbs high dimensions of data and addresses concerns about having noisy data and missing values (Maniar, 2023). Therefore, it is a more reliable machine learning model. In this view, the author will employ random forest regression by Simar (2023) and Al Qassem (2022).

The method will carry out Simar's (2023) approach of using past and expected returns, which are paired randomly from the training sample with replacement. Due to the sample being trained by replacement, there is a continuous slide-window approach as the training set is repetitively processed. The next step is to split at each node of the tree using the best splitting feature after training the decision tree, which gives a predicted output of random forest. After establishing a random forest, the next step is calculating the final prediction by applying the average (mean) of each tree prediction. Subsequently, the method employs Al Qassem's (2022) prediction function of random forest, which yields:

$$T(x) = \frac{1}{n}\sum_{n=1}^{N} b_f + \sum_{s=1}^{S}\left(\frac{1}{n}\sum_{n=1}^{n} contr_i(p,q)\right) \tag{4.15}$$

Where T(x) is the predictive function of a random forest of T. The n denotes the number of decision trees in the forest. S represents the number of features $contr_i(p,q)$ denotes contribution features from q feature in vector p. Then b_f represents the bias from training set. This can display graphically in Figure 4.1.

Figure 4.1: A Random Forest for Test Sample Input



Source: (Simar, 2022)

This will be applied in Google Colab by following Simar's (2022) coding analysis using the Random Forest Regressor model from scikit-learn. Through coding, the model would be fine-tuned to optimize the performance. The parameters that will be fine-tuned will be the number of feature variables, the number of decision trees, and the maximum depth of the tree. Parameters are selected by default function from Google Colab. Below is Table 4.2 which shows parameters of the grid search. The parameter grid will consist of:

Table 4.2: Random Forest Regression Parameter Grid Search

Number of estimators	[100, 200, 300]
Maximum depth	[10, 20, 30]
Minimum sample of sub-tree split	[2, 5, 10]
Minimum sample leaves	[1, 2, 4]
Type of feature	'sqrt', 'log2'

Source: Computed by Author

4.4.2.4. SECTOR VECTOR MACHINE REGRESSION MODEL

The Sector Vector Machine (SVM) comes in two forms: SVM classifiers and SVM regression. Similarly, with the decision tree classifier, the SVM classifier aims to identify the size of the margin, which can impact misclassification. However, this study aims to predict stock prices so that the SVM regression model would be appropriate for this methodology. The author will employ methodologies like Mishra and Padhy (2019), who created their portfolio using predicted stock prices by support vector regression, and Ouyang (2022), who studied a portfolio strategy based on machine learning.

This mode aims to conduct a significant number of observations training in which the training observed values fall within the margin of error (ϵ), which is s, sometimes expressed as tolerance level. Additionally, the training observed values should lie within the margin of error region, and should the training data lie outside the region, they will be called slacks, which are expressed as ξ_i and ξ_i^*. Moreover, these slacks would be penalized by the regularization parameter. This allows to some degree of mis-predicted values. Thus, the optimization problem becomes:

$$Min_{w,b}\frac{1}{2}\left|\left|w\right|\right|^2 + C\sum_{i=1}^{I}(\xi_i + \xi_i^*) \tag{4.16}$$

$$\left(\sum_{k=1}^{K} w_k\phi(x_{i,k}) + b\right) - y_i \leq \epsilon + \xi_i^*$$

$$y_i - \left(\sum_{k=1}^{K} w_k\phi(x_{i,k}) + b\right) \leq \epsilon + \xi_i$$

$$\xi_i, \xi_i^* \geq 0\ \forall_i$$

Where b and w_i are variables that establish the hyperplane and are to be learned by the SVM. $\phi(x_{i,k})$ denotes kernel function which measures how two data points are similar. Kernal function determines which level of degree is a function is and if it can transform into a higher dimension of degree.

In Google Colab, the SVR model is imported from sklearn.svm and other relevant modules. The data is trained and tested using various predictive performance as outline in chapter 4.5.1. To optimize the model, hyper-parameters can be tuned in using GridSearch cross-validation through trial and error until the optimal solution is found. Hyper-parameters are selected by default function from Google Colab. Below is Table 4.3 which outlines the hyper-parameters of this method. The hypermeters are as follows:

Table 4.3: Sector Vector Machine Regression Hyper-Parameters Grid Search

Estimator kernel	'linear', 'rbf', 'poly'
Estimator C	0.1, 1, 10
Estimator gamma	'scale', 'auto', 0.01, 0.1, 1.0

Source: Computed by Author

From the training phase, stock prices are predicted. Using predicted prices, portfolio will be created subjected to the type of portfolio and assign with the relevant weightings to each stock.

4.5. EVALUATION CRITERIA

Evaluation of a portfolio would be done in two stages: predictive performance and optimal portfolio performance. This study will follow May (2022), Dip Das, Bowala, Thulasiram, and Thavaneswaran (2024), Jones (2023), and finally, Jerez and Kristjanpoller (2020) evaluation criteria of predictive performance and portfolio performance.

4.5.1. PREDICTIVE PEROMANCE

In the May (2022) study, the author used Mean Absolute Percentage Error (MAPE) to evaluate how accurate the model is when forecasting closing prices within a timeframe. Additionally, Dip Das et al. (2024) used MAPE, and R-Squared, while Jones (2023) used MAPE, and R-Squared. Similarly to Jerez and Kristjanpoller's (2020) paper; however, this paper will adopt Theil's U2 as a forecasting parameter.

The MAPE measures the absolute percentage of how predicted variables deviates away from the actual variables. It is given by:

$$MAPE = \frac{100\%}{N} \sum_{i=1}^{n} \left| \frac{\hat{y}_i - y_i}{y_i} \right| \tag{4.17}$$

The R-Squared is the measure of how proportional of variance in the dependent variable can be explained by the independent variable. It is given by:

$$R^2 = 1 - \frac{RSS}{TSS} \tag{4.18}$$

Where RSS is denoted as the sum of squares of residuals and the TSS denotes as the total sum of squares.

Finally, the Theil's U2 is measuring forecasting quality between two or more variables. It is given by:

$$U2 = \frac{\sqrt{\sum_{l=1}^{n} \frac{(y_i - \hat{y}_i)^2}{N}}}{\sqrt{\sum_{L=1}^{n} \frac{y_i^2}{N}} + \sqrt{\sum_{L=1}^{n} \frac{\hat{y}_i^2}{N}}} \tag{4.19}$$

The MAPE should also be relatively small where a lower MAPE denotes that the model is more accurate in its prediction. In contrast, a higher MAPE exhibits the opposite effect. The R squared is required to be a high value towards one, implying that the variable can explain another variable. In contrast, a low value towards zero means there is weak explanatory power between the two variables. Similarly, Theil's U2 should be 0, as zero indicates that the model has perfect forecasting ability or provides better results. In contrast, a higher Theil's indicates the model generates less accurate forecasting results.

4.5.2. OPTIMAL PORTFOLIO PERFOMANCE

This study will employ May's (2022) optimal portfolio performance, where the author used the Sharpe Ratio, Beta coefficient, Treynor ratio, and information ratio to determine the optimal performance of the portfolio. Following this approach, this study adds the Sortino ratio and tracking error to evaluate the portfolio's performance. In the previous section, the Sharpe ratio has already been defined; however, in the calculations, it is assumed to have a risk-free rate at 8.75% as the study is taking a rate of a 10-year government bond as of 30 October 2024 (National Treasury, 2024).

The Sharpe ratio is denoted as:

$$Sharpe\ Ratio = \frac{E(R_p) - R_f}{\sigma(R_p)} \tag{4.20}$$

Where E(Rp) is the expected return of the portfolio, R_f is the risk-free rate, σ is the standard deviation or risk of the portfolio.

The Tracking Error measures the difference between the performance of portfolio return and the benchmark over time. It is denoted as:

$$Tracking\ Error = \sqrt{\frac{\sum_{l}^{i} (R_p - R_b)^2}{N - 1}} \tag{4.21}$$

Where R_p is the portfolio return, R_b is the benchmark return overtime and N is the number of periods.

The Sortino Ratio measures each portfolio's risk-adjusted returns due to additional investment downside risk. The ratio is calculated as follows:

$$\text{Sortino}\ Ratio = \frac{R_p - R_f}{\sigma_p} \tag{4.22}$$

Information Ratio measures the performance of portfolio risk-adjusted returns based on a benchmark portfolio. Typically, if the portfolio can outperform the benchmark portfolio and by how much. The ratio is measured as:

$$Information\ Ratio = \frac{E(R_p) - E(R_b)}{\sigma(R_p - R_b)} \tag{4.23}$$

The beta coefficient measures the estimated volatility of portfolio returns relative to the benchmark portfolio. The coefficient is denoted by:

$$Beta = \frac{Cov(R_p, R_B)}{Var(R_B)} \tag{4.24}$$

Where Cov() measures the covariance between portfolio returns and benchmark returns while var() denotes variance of benchmark returns

Treynor Ratio measures the rate of return due to additional systematic risk experienced from the portfolio. The ratio is calculated as follows:

$$\text{Treynor}\ Ratio = \frac{R_p - R_f}{\beta_p} \tag{4.25}$$

Ideally, the Sharpe ratio should be greater than 0, where a ratio between 0 to 0.99 is considered a poor investment, and a ratio greater than 1-2 and 2+, respectively, is a good to excellent investment. A negative ratio denotes a portfolio performed below its risk-free rate. This also applies to the Sortino ratio. A low tracking error indicates that the portfolio performs similarly to its benchmark. High tracking error means the portfolio can either outperform or underperform the benchmark by a certain degree in its potential returns. Consequently, the information ratio should be greater than 0.4 to be considered a good outcome relative to its performance against the benchmark index. However, a lower 0.4 value may indicate greater volatility and high tracking error, leading to underperformance relative to the benchmark. A positive beta indicates that portfolio returns increase at a specific rate when benchmark returns increase. In contrast, a negative beta denotes that when benchmark returns increase, the portfolio returns will decrease and vice versa. A positive Treynor ratio indicates portfolio returns generate higher returns over the systematic risk it is experiencing. However, a negative ratio implies investors do not generate returns over the risk-free rate, thus generating low returns with very high risk.

4.5.3. COMPOSITE SCORE

The above six evaluation metrics mentioned in section 4.5.2 have different measurement units; thus, it would be difficult to determine the optimal portfolio considering all metrics. With this view, this study will adopt a composite scoring system to make all six metrics have the same unit of measurement. Empirical studies such as those by Lundh, Cardona, and Nielsen (2020) and Market (2022) used Z-score to compute composite scores. In Lundh et al. (2020) analysis, they computed by first standardizing all seven metrics: return of equity, return of assets, return on invested capital, operating profitability, gross profitability, debt to equity ratio, and investments. All seven metrics were averaged to determine the composite score. From these scores, they were ranked as the highest and best-performing stocks, while the lowest score was considered as the worst-performing stock. This is similar to the Market (2022) analysis, where the authors standardized forward prices to earnings, operating cash flows, and price-to-book ratio. The z scores obtained were weighted over the three scores to obtain the final composite score.

A Z-score is representing the number of datapoints that deviate from the mean of a sample or population. The formula of Z-score is:

$$z = \frac{x - \mu}{\sigma} \tag{4.26}$$

Where x is the value of interest, μ is the mean of the sample and σ is the standard deviation value.

This research will implement Lundh, Cardona, and Nielsen's (2020) and Market's (2022) methodology of ranking system. Firstly, the author will standardize all values in the following ways: in traditional methods, the author will group all metrics in their original unit of measurement. This means there will be six values: Sharpe ratio, Sortino ratio, Tracking Error, Information Ratio, Treynor Ratio, and Beta. A z score will be computed, and a weighted z score will be computed. Similarly to the machine learning approach, however, the author will sample all 20 values from each evaluation performance stemming from linear regression, decision tree, random forest, and sector vector machine regressions. Each machine learning method consists of an additional four portfolios, including equally weighted mean-variance, maximum Sharpe ratio, and risk parity portfolio. Thus, resulting in 20 values. Then, all variables are aggregated to determine a composite score. Mathematically using Equally Weighted Portfolio values as an illustration, it can be denoted as:

$$Composite\ Z - score = \frac{z(SR) + z(TE) + z(SR1) + z(IR) + z(B) + z(TR)}{6} \tag{4.27}$$

Where z is the z-score of SR is Sharpe ratio, TE is Tracking Error, SR1 is Sortino Ratio, IR is Information Ratio, B is Beta and TR is Treynor Ratio

4.6. CONCLUSION

This chapter details the analytical framework that will be applied to determine which model performs better than traditional and machine learning approaches. The chapter begins by describing the data that will be used and its preliminary descriptive statistics, which gives an overview of the data used to forecast stock returns. Four traditional approaches were discussed in depth: equally weighted, mean-variance, maximum Sharpe ratio, and risk parity portfolio. In addition, four machine models were discussed in depth: linear regression model, decision tree regression model, random forest regression model, and sector vector machine regression model. The evaluation process has been divided into three phases to ensure that the best portfolios are selected without bias. The phases begin by predicting stock prices and estimating six metrics to evaluate the machine learning model and the portfolio. Finally, a ranking system of the composite score is discussed to determine the optimal portfolio. Chapter five presents the findings on the performances experienced by traditional and machine learning methods

CHAPTER 5: RESULTS

5.1. INTRODUCTION

It is worth restating the primary objectives and the sub-goals for this study. In chapter one, the primary objective outlined aimed to explore the potential portfolio performance by comparing traditional and machine learning methods with a particular focus on sectoral analysis. The main objectives were sub-divided into to 3 sub-research goals. The first sub-research goal is to identify which traditional method portfolio is the best in achieving portfolio diversification and performance. The second sub-research goal is to identify which machine learning method portfolio is the best in achieving portfolio diversification and performance. To determine the best portfolio, the study implements six portfolio unit of measurements that will be defined by z-score and these z-scores will be ranked from ascending order as discussed in chapter four. The last sub-research goal is to identify which sectors have demonstrated consistent superior results in contributing to the optimal portfolio. In this case, the study will present the individual stock performances based on the level of return and risk.

Chapter two outline the existing literature review, chapter three discussed the research context, and chapter four indicated the methodologies that will be implemented for the study. This chapter presents and discusses empirical results derived from the application of methodologies that were discussed. Section 5.2 discusses the descriptive summary results, including summary statistics, correlation, and covariance analysis. Section 5.3 reports the various machine learning models' predictive results and capabilities. Section 5.4 discusses how 20 portfolios performed, and which machine learning and portfolio were the best. Section 5.5 shows which sectors and assets were the primary drivers of the portfolio performance and which sectors were not. Section 5.6 discusses the empirical results in depth by linking the results to theoretical and literature standpoints. Section 5.7 concludes the chapter.

5.2. DESCRIPTIVE SUMMARY

5.2.1. SUMMARY STATISTICS

As stated, earlier sample covers the period from 30 September 2019 to 30 September 2024. Using the stock prices, we calculated daily returns for each sector. Table 5.1 reports daily returns summary statistics for the nine sectors to give an overview of the sectors. These sectors include Technology, Telecommunications, Healthcare, Finance, Real Estate, Consumer Staples, Consumer Discretionary, Industrial, and Basic Materials. The data for the analysis resulted in 1250 daily observations. The technology sector experienced the highest average daily mean returns of 0.086%, followed by the Industrial and Basic Materials sector with daily average mean returns of 0.085% and 0.084%, respectively. The real estate and telecommunications sectors had the lowest daily mean average returns, 0.012%, and 0.009%, respectively

Table 5.1: Summary Statistics

	Mean	Median	Max	Min	Std. Dev.	Skewness	Kurtosis	Observations	Jarque-Bera
Tech	0.086%	0.013%	38.093%	-17.528%	2.379%	3.360	58.944	1250	181829.96 ***
Telecomm	0.009%	-0.009%	12.153%	-13.452%	2.106%	0.029	4.655	1250	1117.69 ***
Health Care	0.034%	0.022%	7.526%	-12.646%	1.681%	-0.439	6.214	1250	2032.21 ***
Finance	0.066%	0.042%	17.499%	-18.319%	2.017%	-0.184	13.456	1250	9354.94 ***
Real Estate	0.012%	-0.065%	17.533%	-19.421%	2.329%	0.163	17.308	1250	15474.05 ***
Consumer Staples	0.028%	0.024%	5.032%	-7.524%	1.172%	-0.172	3.223	1250	541.33 ***
Consumer Discret	0.051%	0.023%	7.707%	-10.063%	1.654%	-0.157	4.077	1250	861.94 ***
Industrial	0.085%	0.070%	11.170%	-10.967%	1.594%	-0.133	6.219	1250	1998.83 ***
Basic Materials	0.084%	0.057%	12.442%	-12.324%	2.089%	0.105	3.164	1250	517.88 ***

*Source: Compiled by Author from python results. Note *, **, *** denotes results are significant from 0 at 1%, 5% and 10% level, respectively. The symbol * indicates that the null hypothesis for sectors to exhibit normal distribution are rejected*

The real estate sector (-0.065%) also performed worse than the other sectors regarding median daily average returns. However, the industrial sector achieved the highest median daily average returns of 0.070%. In this examination period, at some point, the Technology sector produced the highest superior maximum daily average returns of 38.093%, and Consumer Staples experienced the lowest maximum daily average returns of 5.032%. Consequently, the best minimum daily average returns were derived from Consumer Staples at -7.524%, with the worst minimum daily average of -19.421% from the Real Estate sector. The daily risk (standard deviation) associated with all sectors was relatively low within the 1.5% to 2.4% range. Only the HealthCare, Finance, Consumer Staples, Consumer Discretionary, and Industrial sectors exhibit a negative skewness. All sectors displayed a Kurtosis greater than 3, suggesting that all sectors are normally distributed.

5.2.2. CORRELATION ANALYSIS

Table 5.2 below reports the correlation matrix for the sector for the entire period of study. Annexure B Figures B1 and B2 accompany this table. Figure B1 displays a heatmap of sector correlation analysis, which is a direct replicate of Table 5.2 However, Figure B2 presents a heatmap correlation analysis result for all 27 stocks for this study. From Table 5.2, as expected the correlation between the sector and itself is strong positive with a value of 1. From a close examination based on Figure B1, consumer discretionary and finance have the highest correlated value of 0.67, suggesting that they highly positively correlated sectors, followed by the correlation between real estate and finance at 0.56, suggesting a low positive correlated sector and a combination of industrial and finance sectors at 0.53. The technology and Telecommunications sector yielded the lowest correlation value at 0.10. In addition, from observing Figure B1, most combination sectors consist of blue-shaded blocks, suggesting that these sectors are negatively correlated, which is essential for portfolio diversification. Similarly, from the observation of Figure B2, there is a mixture of positive, moderate, and negative correlations between stocks. Thus, having a variety of stock correlations implies that when constructing portfolios, these portfolios will be able to balance returns and risks depending on investors' investment goals.

Table 5.2: Correlation Matrix

	Tech	**Telecomm**	**Health Care**	**Finance**	**Real Estate**	**Consumer Staples**	**Consumer Discret**	**Industry**	**Basic Materials**
Tech	1.00								
Telecomm	0.10***	1.00							
Health Care	0.16***	0.40***	1.00						
Finance	0.14***	0.48***	0.44***	1.00					
Real Estate	0.13***	0.44***	0.38***	0.56***	1.00				
Consumer Staples	0.13***	0.34***	0.36***	0.36***	0.31***	1.00			
Consumer Discret	0.16***	0.48***	0.44***	0.67***	0.50***	0.43***	1.00		
Industrial	0.17***	0.33***	0.35***	0.53***	0.46***	0.28***	0.50***	1.00	
Basic Materials	0.18***	0.28***	0.18***	0.27***	0.28***	0.18***	0.31***	0.25***	1.00

*Source: Compiled by Author from python results. Note *, **, *** denotes results are significant from 0 at 10%, 5% and 1% level, respectively.*

5.2.3. COVARIANCE ANALYSIS

Table 5.3 below shows the sector's annualized covariance matrices during the study period. Annexure B Figures B3 and B4 accompany this table. Figure B3 displays a heatmap of sector covariance analysis, which is a direct replicate of Table 5.3. However, Figure B4 presents a heatmap covariance analysis result for all 27 stocks for this study. The covariance matrix provides supplementary relationship results for the correlation matrix as an indicator that shows an overview of the direction of stocks and sectors' movement over time. The technology sector generated the highest covariance output with 0.153, followed by the telecommunications sector with 0.118, the basic materials sector at 0.116, and the finance sector at 0.108. These sectors experienced positive covariance, suggesting that the stocks among their respective groups tend to move in the same direction. For example, in the technology sector, should Naspers Ltd stock increase its share price or returns over time, the same effect will occur with Datatec Ltd stock and Prosus NV stock and, similarly, when returns decrease over time.

A combination of Consumer Staples and Technology showed the lowest covariance score of 0.009, suggesting a weak covariance between the two sectors. The combination is regarded as a strong, weak covariance sector, as indicated by the heatmap of Figure B3, which shows a dark blue block and, similarly, all blocks in blue. As shown in Figure B4, all stocks except Naspers Ltd and Capitec Bank Holdings Ltd have relatively high covariance scores. Only stocks are shown as red blocks in the heatmap where Naspers Ltd has the highest covariance score as it has a red-dark colour, while Capitec Bank has light red, suggesting it has a lower score than Naspers Ltd. The rest of the stocks are regarded as strong, weak covariance stocks due to their shade of blue. These stocks are close to 0, suggesting no linear relationship exists between all stocks except Naspers Ltd and Capitec Bank Holdings Ltd. In portfolio optimization, this means that these stocks do not affect one another whether they are added to the portfolio or not. In principle, the overall risk contributed to the portfolio would be minimized when these stocks are added, which is crucial for portfolio optimization.

Table 5.3: Covariance Matrix

	Tech	Telecomm	Health Care	Finance	Real Estate	Consumer Staples	Consumer Discret	Industry	Basic Materials
Tech	0.153								
Telecomm	0.013	0.118							
Health Care	0.016	0.036	0.074						
Finance	0.017	0.052	0.038	0.108					
Real Estate	0.018	0.056	0.038	0.068	0.146				
Consumer Staples	0.009	0.022	0.018	0.022	0.022	0.035			
Consumer Discret	0.016	0.043	0.031	0.058	0.049	0.021	0.071		
Industrial	0.017	0.028	0.024	0.044	0.044	0.013	0.033	0.066	
Basic Materials	0.023	0.031	0.016	0.029	0.035	0.011	0.027	0.021	0.116

Source: Compiled by Author from python results.

5.3. PREDICTION STATISTICS

In computing the prediction results, we first used the original data and forecasted the daily returns for the following day, resulting in 1250 data points from 1251. As described in Chapter Four, the methodology section, these daily returns were hyper-tuned by different values for different methods. When hyper-tuned variables are obtained, we obtain the forecasted prices in their original state and not daily returns in the following process. With these values, we optimized all the portfolios.

The prediction results for all four methods, including linear regression, decision tree regression, sector vector machine (SVM) regression, and random forest regression, are shown graphically in Figure 5.1 to Figure 5.4 and Table 5.4, respectively below. Figures 5.1 to 5.4 show a graphical representation of all methods of actual vs forecasted portfolio prices. In linear regression, the forecasted prices deviate slightly from the actual prices in different time frames, which suggests that forecasted prices are closely aligned with actual prices. For the decision tree, the forecasted prices present significant deviation from the actual prices in different time frames, which suggests that forecasted prices are not closely aligned with actual prices. The random forest forecasted prices deviate slightly from the actual prices in the first half of the time frame. Then, the forecasted prices present significant deviation from the actual prices in the second half of the time frame, which suggests that forecasted prices are not closely aligned with actual prices. SVM forecasted prices show slight deviation from the actual prices in different time frames, which suggests that forecasted prices are closely aligned with actual prices. Table 5.4 provides a more sectoral and overall predictive portfolio performance context for all methods.

Linear regression outperformed decision tree, SVM, and random forest regression in its mean absolute percentage error (MAPE). The overall performance for linear regression MAPE is 0.0233, and the poor predictor relative to the MAPE is the decision tree regression with 0.0677. Linear regression model performance is primarily driven by real estate and consumer discretionary sectors, which have the lowest MAPE of 0.0063, suggesting that forecasted prices can have up to a 0.63% average error of the actual prices. The basic material sector had the highest MAPE of 0.0107 on sectoral evaluation. Meanwhile, in decision tree regression, the lowest performing sector is consumer discretionary, which exhibits a value of 0.0094 for MAPE, suggesting that forecasted prices yield a 0.94% average error of the actual prices. The industry sector experienced the highest MAPE of 0.0552.

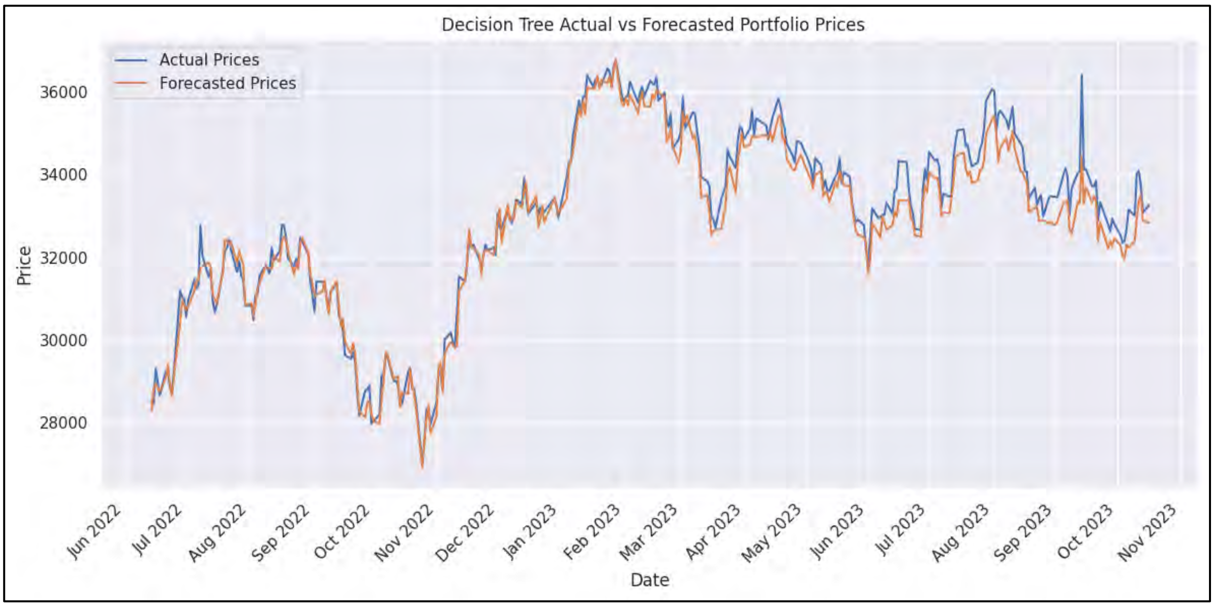
Random forest regression is the best predictive model relative to the model's explanation of the variance with R^2 of 0.9544, followed by SVM with R2 of 0.9472, followed by linear regression with R^2 of 0.8962 and the worst model is the decision tree regression with R^2 of 0.7728. An insignificant result from all variables is derived from the decision tree model and SVM within the industrial sector, which experienced the lowest R^2 with -0.0385 and -0.3678, respectively, indicating that the model does not explain any of the variances of the sector and therefore, is considered to perform worse than the baseline predictor. Likewise, random forest regression predictive model sectoral analysis revealed that the telecommunications sector experienced the highest R^2 with 0.9773, implying that the model can explain 97.73% of the variance of the sector. In decision tree regression, the basic material sector experienced the lowest R^2 with 0.3289, indicating that the model can explain 32.89% of the variance of the sector and, therefore, is considered a poor predictor. The technology sector experienced the highest R^2 with 0.9938, implying that the model can explain 99.38% of the variance of the technology sector.

Finally, Theil's U2 analysis revealed that the SVM overall portfolio is the best forecasting accuracy model with a value of 0.0072 compared to the other three methods. To shed some light on why the SVM overall portfolio performed better than the other methods, it is crucial to unpack the sectoral performance for each method. The SVM regression model showed that the real estate sector has the best forecasting accuracy compared to the naïve model. In contrast, the basic material sector has the worst forecasting accuracy, with 0.0041 and 0.0166, respectively. The portfolio has an average of 0.0071, implying it has outperformed the naïve model as it is close to 0, suggesting the high accuracy of the model. The second best-performing predictive model was linear regression. Linear regression indicated that the real estate sector has the best forecasting accuracy compared to the naïve model, while basic material has the worst forecasting accuracy with 0.0040 and 0.0260, respectively. The portfolio has an average of 0.0078, implying it has outperformed the naïve model as it is close to 0, suggesting the high accuracy of the model.

Figure 5.1: Linear Model - Actual vs Forecasted Portfolio Prices

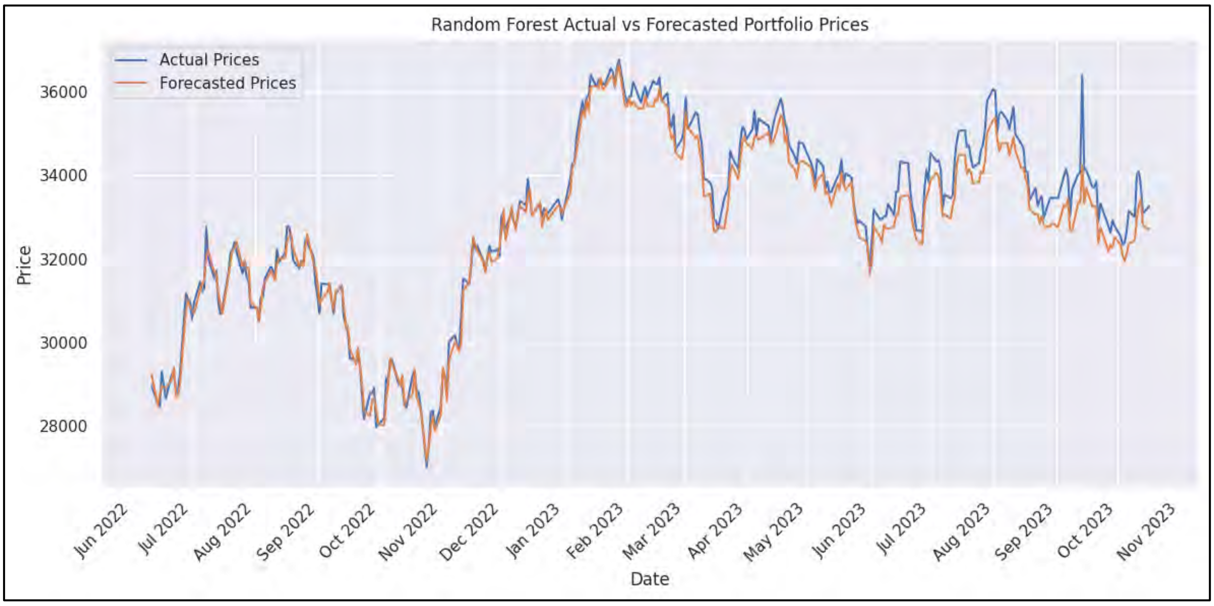
Source: (Aurthor's own graph using Google Finance data (2019-2024))

Figure 5.2: Decision Tree Model - Actual vs Forecasted Portfolio Prices



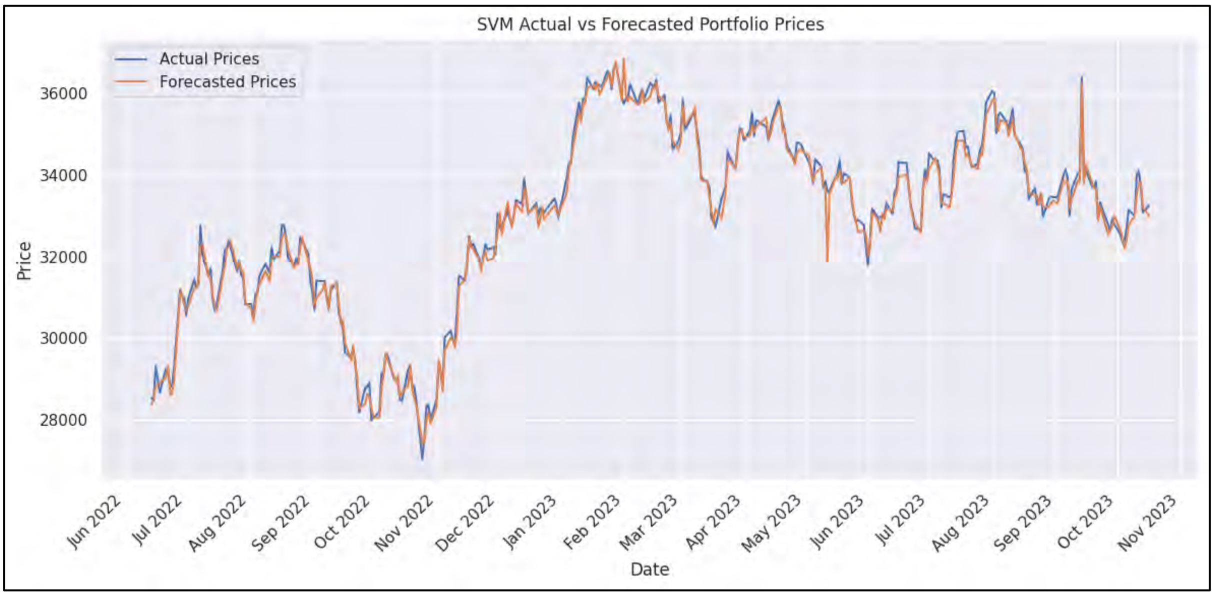
Source: (Aurthor's own graph using Google Finance data (2019-2024))

Figure 5.3: Random Forest Model - Actual vs Forecasted Portfolio Prices



Source: (Aurthor's own graph using Google Finance data (2019-2024))

Figure 5.4: Sector Vector Machine Model - Actual vs Forecasted Portfolio Prices



Source: (Aurthor's own graph using Google Finance data (2019-2024))

Table 5.4: Summary Regression Predictive Statistics

	Linear Regression			**Decision Tree Regression**			**Random Forrest Regression**			**Sector Vector Machine Regression**		
SECTORS	**MAPE**	**R^2**	**THEILS U2**	**MAPE**	**R^2**	**THEILS U2**	**MAPE**	**R^2**	**THEILS U2**	**MAPE**	**R^2**	**THEILS U2**
Technology	0.008	0.993	0.005	0.015	0.947	0.015	0.012	0.952	0.014	0.008	0.993	0.005
Tele - communication	0.007	0.990	0.004	0.014	0.961	0.009	0.010	0.977	0.007	0.008	0.989	0.005
Health Care	0.007	0.984	0.004	0.011	0.950	0.008	0.008	0.976	0.005	0.007	0.983	0.004
Financials Service	0.006	0.985	0.004	0.010	0.955	0.007	0.009	0.964	0.006	0.007	0.985	0.004
Real Estate	0.006	0.983	0.004	0.012	0.931	0.008	0.008	0.969	0.005	0.006	0.982	0.004
Consumer Discretionary	0.006	0.829	0.009	0.009	0.969	0.006	0.007	0.982	0.004	0.006	0.836	0.008
Consumer Staples	0.007	0.988	0.006	0.055	0.404	0.043	0.053	0.406	0.043	0.007	0.988	0.006
Industrial	0.010	0.981	0.006	0.055	-0.038	0.040	0.062	-0.367	0.044	0.013	0.962	0.008
Basic Materials	0.010	0.328	0.026	0.019	0.873	0.014	0.017	0.804	0.012	0.011	0.804	0.016
Overall Portfolio	**0.023**	**0.896**	**0.007**	**0.067**	**0.772**	**0.017**	**0.063**	**0.954**	**0.016**	**0.025**	**0.947**	**0.007**

Source: Estimated and complied by Author from python results

5.4. PORTFOLIO OPTIMIZATION RESULTS

The predictive analysis showed that all models are great forecasting models. Each predictive model has a distinct best predictive model. As discussed in the previous section, linear regression was the best predictor for MAPE, random forest regression was the best predictor for the coefficient of determination (R^2), and sector vector machine regression was the best predictor for Theils U2. Having identified the best predictive model, the next step was to obtain the forecasted prices. These forecasted prices have been optimized, and we obtained the performance of each portfolio for all evaluation metricises.

In traditional methods, four portfolios are evaluated: equally weighted, mean-variance, maximum Sharpe ratio, and risk parity portfolio. Similarly, machine learning methods, including linear regression, decision trees, random forest, and sector vector machine regression models, are evaluated against the four portfolios. Table 5.5 results below, summarising the composite scores for all methods and portfolios. In addition, refer to Tables C1.1, C2.1, C3.1, and C4.1 in the appendix, which provides a comprehensive overview of the portfolio performance over time.

Table 5.5: Composite Scores for All Methods and Portfolios

	Equally Weighted Portfolio	Mean Variance Portfolio	Maximum Sharpe Ratio Portfolio	Risk Parity Portfolio
Traditional Methods	0.5920	0.1996	0.1694	-0.9609
Linear Regression	0.6078	0.6613	0.6379	-0.1537
Decision Tree Regression	-0.1212	0.1717	0.1717	-0.9752
Random Forest Regression	-0.4303	-0.3242	0.1492	-1.5934
Sector Vector Machine Regression	0.5081	0.3324	0.3336	0.0217

Note: The composite score was computed by the author. In computation of composite score, all variables from evaluation were compiled and grouped for e.g. 20 variables of Sharpe ratio, Tracking Error, Sortino Ratio, Information Ratio, Beta and Treynor Ratio were grouped respectively by the same unit of measurement. Z-scores were computed for each of the 20 variables using sample deviation. Using the z-scores, the weighting average of 7 portfolio evaluation metrices are computed to determine the final composite score. See Z-scores in the appendices.
Source: Computed and estimated by author

5.4.1. TRADITIONAL METHODS RESULTS

The analysis revealed that when investing in an equally weighted portfolio, the expected returns will be 14.06% with 19.02% risk. The portfolio yielded a relatively low Sharpe ratio of 0.2791 and a tracking error of 0.1905. Additionally, the Sortino ratio experienced in the portfolio was 1.2260, with an information ratio of 0.1955. The ability to track the All-Share Index was subject to 0.2543 of beta. In contrast, the Treynor ratio was 0.2088. An aggregate composite score for the entire portfolio is 0.5920.

Following this process, the aggregate composite score for the mean-variance portfolio is 0.1996, followed by the maximum Sharpe ratio portfolio with 0.1694, and risk parity was the worst-performing portfolio, achieving a composite score of -0.9609. The risk parity negative portfolio performance is heavily contributed by the performance of its Sharpe ratio, Sortino ratio, information ratio, and Treynor ratio. All these ratios were negative with -0.3679, -3.5884, -2.2318, and -0.0715, respectively, suggesting the portfolio is experiencing low negative risk-adjusted returns, underperforming relatively to its downside risk, the portfolio is generating returns below the All-Share Index and there are negative returns experienced by the portfolio relatively to its All-Share Index returns.

5.4.2. LINEAR REGRESSION RESULTS

The mean-variance portfolio with a composite score of 0.6613 outperformed an equal-weighted portfolio with a composite score of 0.6078, the maximum Sharpe ratio with a composite score of 0.6379, and the risk parity portfolio with a composite score of -0.1537. The best-performing portfolio, mean-variance, has an excellent Sharpe and Sortino ratio, achieving a score above 2, with 2.3691 and 3.2900, respectively. Additionally, the portfolio is subjected to expected returns of 19.98% with an expected risk of 8.43%. An equal-weighted portfolio with 25.63% expected returns with 16.55% expected risk shows that the portfolio had a higher expected return relative to its risk than the mean-variance portfolio, and the overall composite score was lower. This is due to lower Sharpe ratio, Sortino ratio, and tracking error. However, the worst-performing portfolio using linear regression methodology is the risk parity portfolio. In comparison against all other three portfolios, the risk parity portfolio had the lowest expected returns and risk, achieving an optimal 15.85% returns at 12.57% risk. Moreover, the risk parity portfolio is bound to mimic the performance of the All-Share Index as indicated by the low tracking error, information ratio, and beta, which were very close to zero.

5.4.3. DECISION TREE REGRESSION RESULTS

Upon examining the performance of all portfolios, mean-variance and maximum Sharpe ratio portfolios have performed equally in all evaluation metrics, where no portfolio was better in one metric than the other. The composite score for both portfolios was 0.1717. The expected return for both portfolios is 9.52%, with an expected risk of 6.70%. The significant portfolio performance metric is how both portfolios have high volatility relative to the All-Share Index, with a beta set at 4.1647. This indicates that the portfolio will generate four times higher or lower excess returns than the All-Share Index returns, subject to movement towards a bullish or bearish trend. Although an equal-weighted portfolio has higher expected returns of 11.51%, the expected risk is higher than the returns with a level of risk of 12.09%. Likewise, an equal-weighted and risk parity portfolio achieved negative composite scores of -0.1212 and -0.9752, respectively. Equal-weighted portfolios performed moderately low in several evaluation metrics such as Sharpe ratio, tracking error, and Sortino ratio. Furthermore, the portfolio provides a fair, minimal excess return relative to its risk, as it has generated an information ratio of

0.1021. Unsurprisingly, the risk parity portfolio has performed the worst in this category, with low returns of 8.98% at 9.65% risk. Despite the low returns and risk, the contributing factors to the failure of the portfolio are the negative performance in its Sortino ratio, information ratio, and Treynor ratio. These results indicate that the portfolio is underperforming in its ability to generate excess returns against downside risk, underperformance against the All-Share Index, and lacks in receiving returns over its risk-free rate of 8.75%.

5.4.4. RANDOM FOREST REGRESSION RESULTS

Implementing a random forest regression model for portfolio optimization is not ideal, as three out of four portfolios achieved negative composite scores. The best-performing portfolio is the maximum Sharpe ratio with a composite score of 0.1492, which has outperformed the mean-variance portfolio with -0.3242. The mean-variance portfolio is the second-best, followed by an equal-weighted portfolio with a composite score of -0.4303. The worst-performing portfolio is the risk parity portfolio, achieving an underwhelming composite score of -1.5934. Both mean-variance and maximum Sharpe ratio portfolios had positive outcomes in their Sharpe ratio, tracking error, and Treynor ratio performance; however, both had negative outcomes in their information ratio, Sortino ratio, and beta performance. The mean-variance and maximum Sharpe ratio also have higher expected returns over the expected risk. In a mean-variance portfolio, the total expected returns have registered an all-time low of 3.93% with a very low risk of 3.17%. The maximum Sharpe ratio portfolio provides exceptional level reruns of 10% with 3.29% risk, indicating that the maximum Sharpe ratio portfolio outperforms the mean-variance portfolio. In contrast, the equal-weighted portfolio and risk parity portfolio have lower expected returns while having high expected risk experienced by the portfolio. In an equal-weight portfolio, the expected returns narrow to approximately 6% at 11.26% risk, and the risk parity portfolio yielded expected returns of 4.08%, which was more than double the level of risk of 8.32%.

5.4.5. SECTOR VECTOR MACHINE REGRESSION RESULTS

The best-performing portfolio is an equal-weighted portfolio with a composite score of 0.5081. The second-best performing portfolio is the maximum Sharpe ratio, with a composite score of 0.3336. The Maximum Sharpe ratio portfolio narrowly outperformed the mean-variance portfolio, achieving a composite score of 0.3324. The worst-performing portfolio is the risk parity portfolio, accomplishing a composite score of 0.0217. Reviewing each portfolio, they provide exceptional levels of returns, with all achieving above 15% expected returns. However, each expected return is associated with high expected risk, with the lowest at 8.06% and the highest at 15.06% experienced by the maximum Sharpe ratio and equal-weighted portfolio, respectively. Mean-variance and maximum Sharpe ratio both performed excellently in their Sharpe and Sortino ratios, above 2. In all portfolios, the expected returns are higher than the expected risks. However, all three portfolios expect risk parity portfolios to be prone to high volatility relative to the All-Share Index. The equal-weighted portfolio moves in the same direction as the All-Share Index by four units. However, the mean-variance and maximum Sharpe ratio move in the opposite direction of the All-Share Index by eight units. This suggests that an equal-weighted portfolio is suited for the economic expansion phase and the mean-variance and maximum Sharpe ratio portfolio are well suited for the economic contraction phase.

5.4.6. PORTFOLIO RANKING

In the previous section, we discussed all 20 portfolio performances by identifying which unit of measurement contributed to the underperformance and overperformance of the portfolios. We used each portfolio's unit of measurement to derive a composite z-score. The z-score comprises an aggregation of the performance relating to the Sharpe ratio, Tracking Error, Sortino ratio, Information ratio, Beta, and Treynor ratio. The next step is to rank all 20 portfolios based on their aggregated composite score to determine the best and worst-performing portfolios.

Below is Table 5.6, which displays the ranking of various portfolios based on their composite scores, which are ranked from the best (highest value) to the worst (lowest value) portfolios. The table further elaborates on the optimization method used, either traditional or machine learning methods, and the model type between linear regression, decision tree, random forest, or sector vector machine regression- however, it is only applicable to machine learning methods. Additionally, the type of portfolio is also mentioned.

Based on 20 different portfolios, it is evident that machine learning methods are ranked in the top 3 with the linear regression model under the mean-variance, maximum Sharpe ratio, and equally weighted portfolio, respectively. An equally weighted portfolio from the traditional method was ranked fourth. Machine learning methods followed from fifth to seventh place based on the SVM regression model. Under the SVM regression model, three portfolios fall into this rank range: equally weighted, maximum Sharpe ratio, and mean-variance portfolios. Eighth place is a traditional method of implementing a mean-variance portfolio. Joint in ninth place is the machine learning method with the same portfolio composite score: decision tree regression under maximum Sharpe ratio and mean-variance portfolios. The Maximum Sharpe ratio portfolio under traditional methods was ranked in 11th place. Ranked from 12th to 17th place are machine learning methods from different models, including random forest, decision tree, and linear regression models, and it includes all four portfolios. The last traditional method ranked portfolio is the risk parity, which was placed 18th, the worst portfolio from the traditional method side. The worst type of optimization is machine learning concerning decision tree and random forest regression models, ranked 19th and 20th, respectively. Both models have the same portfolio, a risk parity portfolio.

Table 5.6: Ranking of Portfolio Performances

Ranking Number	Type of Optimization	Model Name	Portfolio	Composite Score
1	Machine Learning	Linear Regression	Mean Variance	0.6613
2	Machine Learning	Linear Regression	Maximum Sharpe Ratio	0.6379
3	Machine Learning	Linear Regression	Equally Weighted	0.6078
4	**Traditional**	**-**	**Equally Weighted**	**0.5920**
5	Machine Learning	SVM Regression	Equally Weighted	0.5081
6	Machine Learning	SVM Regression	Maximum Sharpe Ratio	0.3336
7	Machine Learning	SVM Regression	Mean Variance	0.3324
8	**Traditional**	**-**	**Mean Variance**	**0.1996**
9	Machine Learning	Decision Tree Regression	Maximum Sharpe Ratio	0.1717
9	Machine Learning	Decision Tree Regression	Mean Variance	0.1717
11	**Traditional**	**-**	**Maximum Shape Ratio**	**0.1694**
12	Machine Learning	Random Forest Regression	Maximum Shape Ratio	0.1494
13	Machine Learning	SVM Regression	Risk Parity	0.0217
14	Machine Learning	Decision Tree Regression	Equal Weighted	-0.1212
15	Machine Learning	Linear Regression	Risk Parity	-0.1537
16	Machine Learning	Random Forest Regression	Mean Variance	-0.3242
17	Machine Learning	Random Forest Regression	Equal Weighted	-0.4303
18	**Traditional**	**-**	**Risk Parity**	**-0.9609**
19	Machine Learning	Decision Tree Regression	Risk Parity	-0.9752
20	Machine Learning	Random Forest Regression	Risk Parity	-1.5934

Source: Compiled by Author

5.5. PORTFOLIO PERFORMANCE

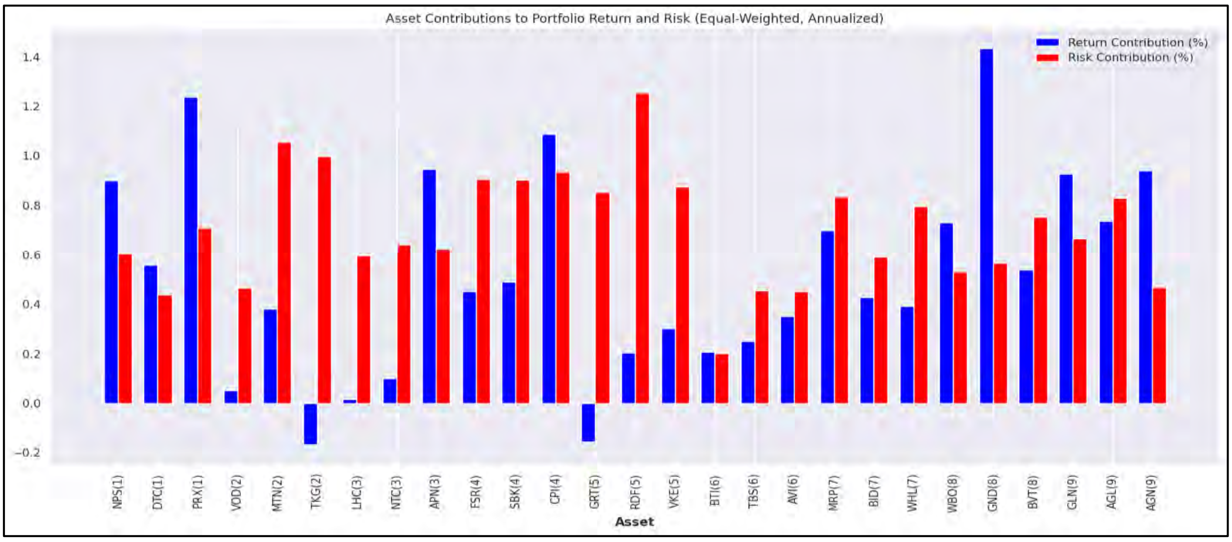
The previous section identified the best and worst-performing portfolios through the composite z-scoring system. The best-performing portfolio is the mean-variance portfolio using machine learning linear regression optimization technique, and for traditional technique, it is an equal-weighted portfolio. The worst performing portfolio for the machine learning technique is the random forest regressing with risk parity portfolio. Similarly, the risk parity portfolio is the worst-performing portfolio using the traditional method. From these results, in this section, we focus on what led to these portfolios being considered the best and worst-performing based on the optimization methods. Specifically, we analyse the relative contribution of each stock in the portfolio to the returns and risks associated with the portfolio.

For comprehensive results for this section, refer to Appendix C1.2 - 1.4, C2.2 - 2.4, C3.2 - 3.4, C4.2 - 4.4, and C5.2 – 5.4 which shows portfolio weightings of stocks for traditional, linear regression, decision tree regression, random forest regression and sector vector machine regression methods for equal weight, risk parity, mean-variance and maximum Sharpe ratio portfolio’s. Additionally, refer to Appendix D1.1 to D2.2, which provides the expected returns and risk of financial assets for best and worst portfolios using traditional and machine learning methods.

5.5.1. Traditional Method Best Performing Portfolio

The best-performing portfolio within the traditional methods was equal-weighted, having a composite score of 0.5920. An equal-weighted portfolio consists of all 27 stocks. The primary sectors driving the portfolio to have an expected return of 14.06% at 19.02% risk are technology, industry, and basic material sectors. Figure 5.5 shows assets' contribution to portfolio return and risk for the equally weighted portfolio. The best-performing stocks in this portfolio are Grindrod, having 1.45% expected returns at 0.57% risk; Prosus NV, with 1.24% returns at 0.71% risk; and Capitec Bank Holdings, performing with 1.09% total return at 0.93% risk. The worst performing stocks in this portfolio are Telkom SA, yielding expected returns of -0.17% at 1% risk; Growthpoint Properties Ltd, with an output of -0.16% expected return with 0.85% risk; and finally, Life Healthcare Group Holdings, a positive outlook of 0.02% expected returns from 0.60% risk. Lastly, all 27 stocks have an equal weighting of 0.0370.

Figure 5.5: Asset Contribution to Portfolio Return and Risk (Equal Weighted, Annualised)

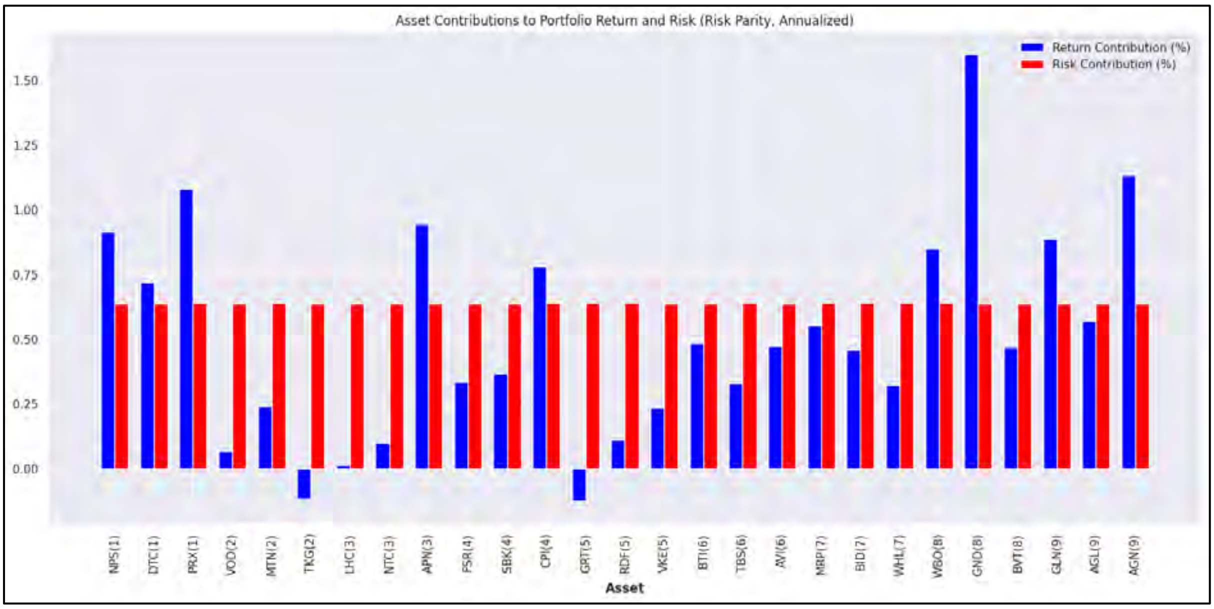


Note: The portfolio consists of all 27 stocks as described in Appendix Table B1 and all sectors have been selected.
Source: Own graph using Google Finance (2019-2024) data after optimizing portfolio’s

5.5.2. Traditional Method Worst Performing Portfolio

The worst-performing portfolio under the traditional method is the risk parity portfolio. This portfolio utilises all 27 stocks with the same marginal contribution to the portfolio's total risk. Having a composite score of -0.9609 and ranked 18th place, it is observed from Figure 5.6 below that technology, industry, and basic material sectors played a significant role in the outcome of portfolio return and risk. Figure 5.6 shows the relationship between returns and risk among the 27 stocks based on their contribution towards the optimal portfolio return and risk. Grindrod was the best-performing stock with 1.60% expected returns and a weighting of 0.0413, followed by AngloGold Ashanti at 1.13% expected returns with 0.0288 weightings and Prosus NV at 1.08% expected returns with 0.0324 weightings. Similarly, the worst-performing stocks in this portfolio are Growthpoint Properties, with 0.12% expected return and a weighting of 0.0291; Telkom SA, with 0.11% expected returns and a weighting of 0.0255; and Life Healthcare Group Holdings, with 0.02% returns and weighting of 0.0388. Additionally, all stocks in this portfolio have an average risk contribution of 0.64%.

Figure 5.6: Asset Contribution to Portfolio Return and Risk (Risk Parity, Annualised)



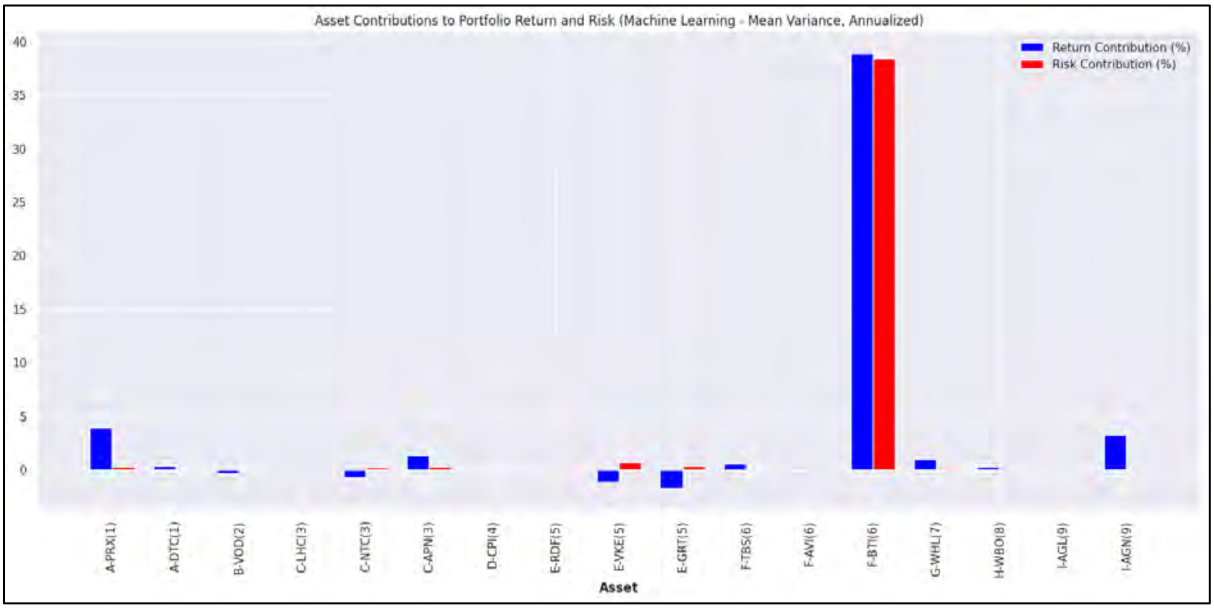
Note: The portfolio consists of all 27 stocks as described in Appendix Table B1 and all sectors have been selected.

Source: Own graph using Google Finance (2019-2024) data after optimizing portfolio's

5.5.3. Machine Learning Best Performing Portfolio

The best-performing portfolio is the mean-variance portfolio, achieving a composite score of 0.6613. The mean-variance portfolio does not take all stocks but considers all stocks from all nine sectors. The portfolio used three stocks from the healthcare service, real estate, and consumer discretionary sectors, two from the telecommunication and basic materials sectors, and stock from the remaining sectors. The main stock driving the portfolio to an exceptional 19.98% expected returns is British America Tobacco, with an impressive 38.93% expected returns at a very high expected risk of 38.44%, having a weighting of 0.1873; Prosus NV deriving 3.88% total expected returns at 0.27% total expected risk averaging 0.0288 weightings; and AngloGold Ashanti with 3.30% expected returns at 0.07% expected risk with 0.0920 weighting on the stock. The worst performing stocks in this portfolio are Growthpoint Properties, having -1.74% expected returns at 0.34% expected risk with a weighting of 0.1169 on the stock; Vukile Property Fund, performing with -1.16% expected returns at 0.73% expected risk with a weighting of 0.2168; and Netcare with relatively -0.68% of expected returns at 0.23% total expected risk with 0.0891 weighting on the stock.

Figure 5.7: Asset Contribution to Portfolio Return and Risk (Mean Variance, Annualised)

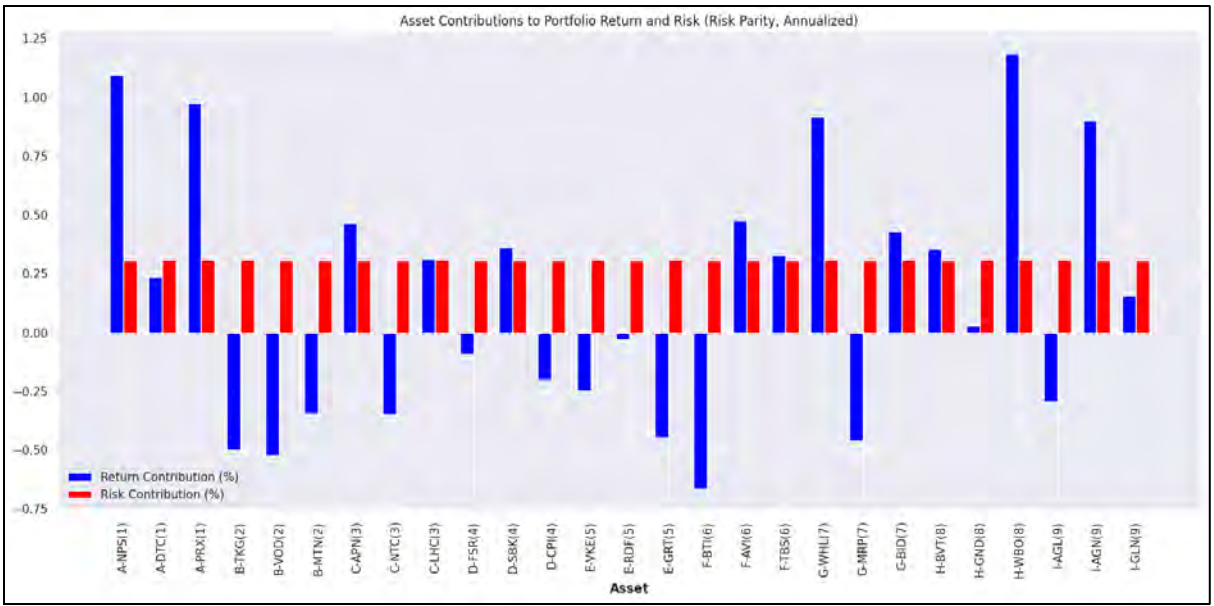


Note: The portfolio consists of all 17 stocks. These stocks are: Prosus NV, Datatec Ltd, Vodacom Group Ltd, Life HealthCare Group, Netcare Ltd, Aspem Pharmacare Holdings, Capetic Bank Holdings, Redefine Propeties Ltd, Vukile Property Fund Ltd, Growth Properties Ltd, Woolworths Holding, Avi Ltd,British America Tobacco, Tigers Brand, Wilson Bayly Holmes, Anglo American, Anglogold. Selected stocks consist of: 2 stocks from technology, 1 from telecommunication, 3 from healthcare, 1 from financial service, 3 from real estate ,3 from consumer discretionary, 1 from consumer staples, 1 from industrials, 2 from basic materials
Source: Own graph using Google Finance (2019-2024) data after optimizing portfolio's

5.5.4. Machine Learning Worst Performing Portfolio

The selected worst machine learning portfolio is a Risk Parity within the random forest model. This portfolio has all 27 stocks invested, and each stock has performed differently, contributing to a composite score of -1.5943. Examining further asset contribution to portfolio return and risk, it is observed that the technology and industry sectors played an active role in achieving positive portfolio expected returns. See Figure 5.8 which displays asset contribution to portfolio return and risk. Wilson Bayly Holmes had the highest expected returns with 1.19% at 0.31% risk, followed by Naspers Ltd with 1.10% at 0.31% risk, and the third highest asset contribution to portfolio return and risk is Prosus NV stock, achieving a return of 0.98% with just 0.31% risk. These stocks weighed 0.0601, 0.0405, and 0.0171, respectively. The negative impact on the portfolio's overall performance was heavily influenced by underperformance in the telecommunication and real estate sectors, which had returns for all stocks below 0%. The worst stocks in this portfolio are British America Tobacco, contributing -0.67% returns at 0.31% risk with a weighting of 0.0657, and Vodacom Group Ltd, achieving -0.52% returns at 0.31% risk with a weighting of 0.0203, and Telkom SA producing -0.50 returns at 0.31% risk with a weighting of 0.0293.

Figure 5.8: Asset Contribution to Portfolio Return and Risk (Risk Parity, Annualised)



Note: The portfolio consists of all 27 stocks as described in Appendix Table B1 and all sectors have been selected.

Source: Own graph using Google Finance (2019-2024) data after optimizing portfolio's

5.6. DISCUSSIONS

This study investigated which method would offer an optimal portfolio performance comparing traditional and machine learning methods. It is important to address the three sub-research questions to respond to this research question, which will help derive the solution. As Chapter One outlines, the first sub-research question focuses on identifying which traditional method portfolio will generate the highest composite score. Secondly, to exam and identify the machine learning method portfolio yielding the highest composite score. Based on the results of the two sub-objectives rank the optimisation methods based on composite scores. Finally, the third sub-research question focuses on the main sectoral drivers, consistently providing strong performance in the optimal portfolio selection. Although studies were conducted relating to the performance of various traditional and machine learning methods approaches, this study has not been conducted in the South African context.

Before addressing the research and sub-research questions for the study, it is important to understand the nature of the data collected and its behaviour over time. Considering this, the study first provided summary statistics, correlation analysis, and covariance performance between the nine sectors: Technology, Telecommunications, Healthcare, Finance, Real Estate, Consumer Staples, Consumer Discretionary, Industrial, and Basic Materials. On average, technology, industry, and basic material sectors were the top three performing sectors based on daily mean returns from 30 September 2019 to 30 September 2024. However, the financial performance of Simple Wall Street (2024) and TradingView (2024) on sectoral performance is partially agreed upon, as highlighted in Chapter Three. These financial platforms indicated that from 24 October 2019 to 24 October 2024, the best-performing sectors were Consumer non-durables, Industrials, and Technology. Basic material sectors were ranked the seventh best performing sector, according to the financial platforms. The difference in sectoral performance in this case is understandable as not all stocks were selected from all sectors, which does not fully capture the overall performance of their sector performance.

Based on daily returns correlation analysis, combinations of consumer discretionary with finance, real estate with finance, and industrial with finance sectors were among the moderately positively correlated. In contrast, the remaining sectors were low positively correlated sectors. Additionally, all sectors were statistically significant at the 10% level. These results are like those of Vengesai, Obalade, and Muzindutsi (2022), who utilized the All-Share Index, basic materials, consumer goods, financials, health industries, and Oil and gas sectors to find monthly returns correction. Their study revealed low corrections among sectors except the industrial sector, which had basic material, consumer goods, and financial sectors, between 1998 and 2018. Many combinations of low-sector correlations suggest that there is room for portfolio diversification among the selected stocks. A covariance matrix was considered to strengthen the argument for diversification opportunities. This paper showed relatively low covariance between the sectors except the technology and real estate sectors. This suggests that stocks in these sectors perform independently, and their performance does not affect the performance of other selected stocks indicating excellent portfolio diversification opportunities.

This study also explores machine learning forecasting accuracy for all methods, including linear regression, decision tree regression, random forest regression, and sector vector machine (SVM) regression. The underlying variables used to capture the various model predicates are mean absolute percentile error (MAPE), the R^2, and Theil's' U2. The overall portfolio MAPE ranged from 0.0233 to 0.0677, R^2 ranged from 0.7728 to 0.9544, and Theil's U2 ranged from 0.0071 to 0.0170. These overall portfolio results provide strong evidence of the ability of machine learning methods to correct forecast stock prices in the future. These results are consistent with May (2022) and Jones (2023).

In the May (2022) model evaluation, the author evaluated portfolio (JSE Top 40 stocks) prediction accuracy to determine the model's ability to forecast using historical prices. The study employed the XGBoost machine method, a similar tree-based structure of decision trees and random forests. This paper employed RMSE, Scatter Index, and MAPE as part of predictive models. The focus of the results of the May (2022) paper is on the MAPE results. The overall portfolio MAPE ranged from 0.0467 to 0.3044 for 2019 and from 0.0424 to 21.8379 for 2020. This paper's overall performance MAPE falls within both May (2022) results range. Similarly, Jones's (2023) study estimated 514 stocks from the Nasdaq-100, Dow Jones, and the S&P 500 predictive ability for building an optimized machine learning model. Their machine learning tool focus was linear and polynomial regression. Considering R^2's evaluation in his portfolio examination of predictive ability, the study found that the overall portfolio R^2 ranged approximately between 0.65830 to 0.96106, suggesting there is 65.83% to 96.11%

explanatory power of one variable with a target variable. These results are also like those of the current study as the range of R^2 of this study falls into the bracket Jones (2023) range.

Next, considering the first sub-research question, i.e., which traditional method portfolio is the best in achieving portfolio diversification and performance, the study revealed that the equal-weighted portfolio had the highest composite score, suggesting that it performs better than mean-variance, maximum Sharpe ratio, and risk parity portfolios. This conclusion was drawn from considering portfolio evaluation metrics, including Sharpe ratio, Tracking Error, Sortino ratio, Information ratio, Beta, and Treynor ratio. In principle, having an equal-weighted portfolio performing better than the mean-variance portfolio contradicts the Modern Portfolio Theory (MPT), which suggests the mean-variance portfolio should outperform equal-weighted portfolio due to "its complexity in with estimation errors of the covariance matrix and expected returns" (Cai and Schmidt, 2020). By contrast, there has been vast literature and studies indicating that equal-weighted portfolio outperforms value-weighted portfolios due to the high concentration of small-cap stocks investment, undervalued and dividend stocks, and the ability of equal-weighted portfolios having excess returns (see: Plyakha, Uppal, and Vilkov, 2015; Swade, Nolte, Shackleton, and Lohre, 2023). Despite these findings, Taljaard and Maré (2021) point out that an equal-weighted portfolio has been underperforming by 20% since 2002 in the South African market. They outlined their reasons for the underperformance including high concentration levels of market weightings, high correlation assets, and a higher turnover in index constitutes.

In addressing the second sub-research question, which machine learning portfolio is the best in achieving portfolio diversification and performance, this study revealed that the linear regression models with mean-variance portfolios have the highest score. Similarly, this optimal portfolio was drawn from the six-evaluation metrics used in traditional methods. The results suggest that linear regression outperforms decision tree regression, random forest (RF) regression, and sector vector machine (SVM) regression. Linear regression also outperformed all three regression models' predictive accuracy scores in all evaluation metrics except R^2, as shown in section 5.3.1. to 5.3.4. This suggests that most of the data collected exhibit a linear pattern, emphasizing the need for a robust predictive model that provides a pathway for an optimal portfolio. These results are parallel to Ma, Han, and Wang (2021), who used RF and SVR with three additional deep-learning models to create portfolios with return predictions. Their studies indicated that RF had the highest predictive score, leading to the best-performing portfolio under the mean-variance forecasting model. Additionally, their results were derived from six parameters, including information ratio, excess returns, maximum drawdown, standard deviation, and total return.

By contrast, it is important to note that the best predictive model does not indicate it would generate the best optimal portfolio, as Ngo, Nguyen, and Van Nguyen (2023) suggested. One clear example is the paper by Jones (2023), which examined portfolio optimization using linear and polynomial regression, as mentioned earlier. The key takeaway is that the best portfolio performance was ranked second in its predictive score ability with a coefficient of determination 0.95863 with expected return and risk of 23.96% and 14.27%, respectively, with a Sharpe ratio of 1.609. However, the best predictive score portfolio had a coefficient of determination of 0.96106, with expected return and risk of 28.38% and 26.02%, respectively, with a Sharpe ratio of 1.0892.

The third sub-research question addresses which sectors have consistently contributed to the optimal portfolio in traditional and machine learning methods. To unpack this research question, the author developed a breakdown of all financial assets' performances based on their level of returns and risk contribution to their respective portfolios. In the traditional best portfolio, the sectors that played in the instrumental optimal performance were technology, industrial, and basic material sectors, while the best machine learning portfolio were technology, consumer discretionary, and basic material sectors. Interestingly, the technology and basic materials sector tends to have the most consistent sectoral performance in this study for the best optimal portfolio. Also, these results are consistent with this study's summary statistics, where the technology and the basic material sectors were in the top 3 best sectors. The technology sector was also among the top 3 best-performing sectors in real-world financial data, as Simple Wall Street (2024) and TradingView (2024) indicated.

In addressing the third sub-research question, the contrast between the best and worst-performing portfolio sectoral performance is worth highlighting. Interestingly, in both cases of the worst portfolio performance in traditional and machine learning methods, technology and basic material sectors were the main contributors again in driving high expected returns at a certain level of risk. Although other sectors contributed to the portfolio's success despite being the worst portfolio compared to other models, these sectors were somewhat inconsistent, particularly in the machine learning methods.

Finally, regarding the main research question of which methodology between traditional and machine learning will provide the best portfolio performance, the results are based on the ranking system comprising each model's composite scores through a z-score system. This study found that machine learning provides the best portfolio performance. The method of execution comprised linear regression and mean-variance portfolio. The mean-variance portfolio consisted of all sectors and 17 stocks. Additionally, it had a composite score of 0.6613, which was derived from having an expected return of 19.98% with 8.43% expected risk. The output of the Sharpe ratio was 2.3691, with a tracking error of 0.5651, a Sortino ratio of 3.29, an information ratio of 0.3027, a beta of -2.6966, and a Treynor ratio of -0.0709. These findings are consistent vast literature confirming that machine learning methods outperform traditional methods (see for example: Dip Das, 2024). Additionally, the results of this study support view from some investment professionals who participated in various surveys from BarclaysHedge (2018), Ernst and Young (2018), Price Waterhouse and Coopers (2022), Alex Forbes

(2024), and Charted Financial Analyst (2024) as outlined in chapter 1 and 3 that machine learning methods exhibit better portfolio performance producing higher returns while minimizing risk and a more significant portfolio diversification.

5.7. CONCLUSION

This chapter provided the empirical results for the study to identify which traditional method (goal 1) or machine learning (ML) method (goal 2) yields the best optimal performance, and which sectors contributed to the success of the portfolio (goal 3). The evaluation for the study began by understanding the characteristics of the selected stocks in their sectors. This included summary statistics, correlation, and stock covariance. Results showed that the technology and basic material sectors performed the best. Most sectors exhibiting low correlation and covariance scores suggest that the portfolios selected for the study are well diversified. The study further examined the chosen ML methods, and the results showed that they can correctly forecast stock prices as the actual and predicted prices are closely aligned. Furthermore, it has been established that equal weighted portfolio outperforms mean-variance, maximum Sharpe ratio, and risk parity portfolios in traditional methods. By contrast, it was evident that linear regression outperformed decision tree, random forest, and sector vector machine regression in ML. Additionally, the mean-variance portfolio outperforms maximum Sharpe ratio, equal weighted, and risk parity portfolios in that order. These optimal portfolios were primarily driven by high performance experienced by technology and basic material sectors in traditional and ML methods.

CHAPTER 6: CONCLUSION

6.1. INTRODUCTION

Machine learning (ML) offers better benefits compared to traditional methods. Typically, traditional methods rely on historical prices to create portfolios subject to investor objectives. However, with ML methods, they can consume more information such as news and investor sediment, rapid market changes, different types of prices at the time, and other economic and financial indicators. Despite ML offering more significant opportunities, investment professionals have not hesitated to integrate machine learning into their portfolio optimization strategies. As Barclays (2018) indicated, only 58% of the surveyed investment professionals use ML for portfolio construction. Furthermore, a survey conducted by PwC (2022) among 250 participants from worldwide investment professions suggests that 90% of investment professionals use ML to increase returns while minimizing risk. However, Alex Forbes's (2024) survey indicated that approximately 66% do not implement ML in their investment analysis process due to numerous factors, including data quality and availability, lack of skilled personnel, and high implementation concerns. The mixed investment professionals' sentiments indicate that this study area should be explored further.

Chapter 1 sets the tone for our study, which explores the comparative analysis between traditional and ML methods in the context of portfolio optimization. The main research question of the study is to identify which method between traditional and ML methods yields the best optimal portfolio. The research question is divided into 3-fold. The first sub-research question is which traditional methods will yield the highest composite score. The second sub-research question is which ML method will yield the highest composite score. And based on the results of both sub-objectives ranked the portfolios and methods. Finally, the last sub-research identifies which sectors and stocks consistently contributes to the optimal portfolios.

Chapter 2 discusses the theoretical and empirical literature on portfolio optimization, which is implicated in traditional and ML methods while specializing in sectoral diversification. Chapter 3 provides the research context through an overview of various sectors' economic and financial performance. Chapter 4 unpacked the methodologies used in this study, and Chapter 5 discussed the results. Section 6.2 will conclude the entire study while summarizing the results. Section 6.3 will outline the implications of portfolio optimization. Section 6.4 will discuss the study's limitations, and finally, section 6.5 will suggest further research studies.

6.2. SUMMARY OF MAJOR FINDINGDS

Chapter 3 discusses various sectors' economic and financial performance and how ML has been integrated into academia and the financial sector. This chapter's primary focus was the research context, which provided an overview of sectoral performance over time. South Africa's economic growth has been volatile, ranging from -6% to 6% of its real Gross Domestic Price (GDP) growth from 1980 to

2024. The economic growth has been affected by the geopolitical tensions during specific phases of the study period, which has allowed it to be volatile. When observing the financial performance, the JSE Top 40 over the last five years showed a 55.39% return increase. Driving the JSE Top 40 performance is the consumer non-durables sector, producing a 137.30% increase in its returns, while the telecommunication sector was the worst in the last five years with -13.21% returns. The chapter also discussed the recent adoption of ML. In an academic setting, the number of articles published related to ML in South Africa was approximately 562 between 1991 and 2021. The popular ML methods in these articles were classification, deep learning, feature extraction, and random forest. Due to the rise of ML, it was also widely accepted by financial professionals, who began to use it as part of their investment strategies.

Chapter 4 outlined the methodologies for the study and provided explanations on why those specific methods were selected. The methodologies selected for the study were derived from the literature. The chapter explained why it was important to unpack the summary statistics. As mentioned, the study will employ mean-variance, maximum return, maximum Sharpe ratio, and equally weighted optimization methods within their relative portfolios for traditional methods. In machine learning, the paper utilizes linear regression, sector vector machine regression, decision tree regression, and random forest regression. The evaluation of the performance of the portfolios was divided into three categories. Firstly, the variables selected for the study were conducted to determine the predictive nature using three components: MAPE, R^2, and Theil's U2 coefficient. The second part of the evaluation was determining portfolio performance due to various financial characteristics, including Sharpe ratio, Tracking Error, Sortino ratio, Information ratio, Beta, and Treynor ratio. This leads to the third part of the evaluation, which aims to find the composite score. The composite score was determined by the weighted z-score of 20 variables, which helps to determine the portfolios that are best performing.

Chapter 5 revealed the results for summary statistics, predictive performance, and traditional and machine learning performance, which were ranked and finally showed the sectoral dominance in the best and worst-performing portfolios. Summary statistics showed that, on average, technology, industry, and basic material sectors were the top three performing sectors based on daily mean returns. These results were similar to real-world financial performance as the technology and basic materials sectors were the two of the top 3 best-performing sectors. Additionally, our correlation results showed that most sectors were positively correlated while sectors such as consumer discretionary with finance, real estate with finance, and industrial with finance sectors were among the moderately positively correlated. By contrast, this study also showed that the pairs of various sectors have relatively low covariance sectors except for technology and real estate sectors, which indicate excellent portfolio diversification opportunities. When comparing the predictive results, it was determined that linear regression has outperformed decision tree, random forest, and sector vector machine regression models' predictive accuracy scores in all evaluation metrics except R^2, suggesting the data exhibit a linear pattern. In comparison, decision tree regression accuracy was the worst as it tends to deviate from the actual values with a success rate of just over 75%.

In addressing sub-research questions, this study showed that an equal-weighted portfolio had the highest composite score when considering traditional methods, suggesting that it performs better than mean-variance, maximum Sharpe ratio, and risk parity portfolios. Secondly, when considering machine learning methods, the study showed that the linear regression models with mean-variance portfolios have the highest composite score, suggesting linear regression outperforms decision tree regression, random forest (RF) regression, and sector vector machine (SVM) regression. Lastly, when considering sectoral performance for traditional and ML approaches, technology, industry, and basic material sectors were regarded as driving sectors in the traditional best portfolio. The driving sectors for each portfolio are as follows: in ML's best portfolio, technology, consumer discretionary, and basic material sectors were at the top 3, while the worst portfolio performance in both methods indicates that technology and basic material sectors were the main contributors again in driving high expected returns at a certain level of risk.

Finally, this examination showed that machine learning provides the best portfolio performance over the traditional methods. The machine learning method that performed the best was the linear regression method using a mean-variance portfolio with a composite score of 0.6613. This portfolio performed relatively well, achieving the expected return of 19.98% with 8.43% expected risk.

6.3. IMPLTICATIONS OF PORTFOLIO DIVERSIFICATION AND RECOMMENDATIONS

There have been vast literature studies that use different machine learning algorithms to predict stock prices. However, there is a methodology gap when using linear regression, decision tree regression, sector vector machine regression, and random forest regression to optimize portfolios in the same study. The study recommends exploring these methodologies on different stock market exchanges in other countries for future studies. The study only used nine sectors and 27 stocks. As more data becomes available in future studies, it would be great to explore and revisit the research goals to determine if these results would be the same by adding stocks and sectors to the analysis. This ensures that the study and results remain relevant and that the model is still the best method. The analysis revealed a detailed stock performance, portfolio, and relevant weights. Portfolio management must identify which stock to invest in and which would lead to an optimal portfolio. As discussed, the technology sector was thriving in its performance of all portfolios. This suggests that adding technological stocks to a portfolio would be beneficial in increasing the expected returns consistently and potentially. There were other instances in our results were traditional optimization techniques outperformed machine learning optimization techniques. By integrating traditional and machine learning optimization techniques, investors would benefit from maximising returns at the best level of risk.

While the debate continues between investment professionals about ML in the investment process, it is important to note that whether an investor chooses to use the traditional or ML approach, they should do due diligence in selecting assets for their portfolio with strong research findings. Whether an investment professional chooses the traditional or ML method, the author would recommend using a hybrid approach to ensure that investment goals are met subjected to investors' goals and returns objectives at the appropriate risk tolerance of the client. As George Box (1979) once said, "All models are wrong, but some are useful."

6.4. LIMITATIONS OF THE STUDY

Integrating machine learning in portfolio management has increased rapidly. One aspect we have not covered in this study is integrating economic, social, and governance (ESG) stocks with non-ESG stock comparisons. Thus, studying ESG and non-ESG stocks for this topic is unknown. It would be a great addition to the investment field to understand which type of stock would be a good fit to optimize and which method would be the best when investing in these stocks. Secondly, there has been lack of data in the study. When researching for sectoral analysis, it was identified that South Africa recognises 11 sectors within the financial markets field, however for this study, there were two sectors that were omitted which were energy and utilities sectors. Energy sector stocks were omitted as there were more than ten missing data points from the available stocks. Utility sector was omitted as, at the time of writing, no stocks in the JSE were defined in this sector. For future studies, it would be favourable to add more stocks in the portfolios. Additionally, this study did not use any qualitative data to capture the stock performance by including data such as economic news and other relevant news.

6.5. AREAS OF FURTHER RESEARCH

The interplay between tradition and machine learning within portfolio performance has been established sectoral. However, it would be a great study to unpack portfolio optimization by primarily focusing on Environment, Social, and Governance (ESG) stocks and non-ESG stocks and comparing their performance between traditional and machine learning methods. Due to the nature of the thesis, it was not possible to provide a comprehensive analysis fully. Additionally, it would be a great initiative to include the maximum return portfolio in the study. For future studies, it would be favourable to add news sediments to the analysis to capture stock movements and reactions.

This current research can be deepened by identifying the entry and existing points for various stock prices to ensure that the level of return and risk is achieved. This can be done by effectively building trading signals for South African assets. In this regard, the existing ML models would be utilized to predict future stock prices of various stocks outside the study timeframe and dates. Additionally, incorporating the current news that affects stock movements within the investment process would be a great addition to ensure that all information has been captured in the model.

7. LIST OF REFERENCE

ABD RAHMAN, M.S., JAMALUDIN, N.A.A., ZAINOL, Z. AND SEMBOK, T.M.T., 2023. The Application of Decision Tree Classification Algorithm on Decision-Making for Upstream Business. *International Journal of Advanced Computer Science and Applications*, 14(8).

ADHIKARI, S. AND JHA, P.K., 2016. Applicability of portfolio theory in Nepali stock market. *NRB Economic Review*, 28(1), pp.65-92.

AGARWAL, S. AND MUPPALANENI, N.B., 2022. Portfolio optimization in stocks using mean–variance optimization and the efficient frontier. *International Journal of Information Technology*, 14(6), pp.2917-2926.

AHMED, M., CHOUDHURY, N. AND UDDIN, S., 2017, JULY. Anomaly detection on big data in financial markets. *In Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017,* pp. 998-1001.

AL-JOMAI, R., 2014. Portfolio Optimizer Tool (POT) Based on Trend Market Analysis and News Sentiment Analysis. (Doctoral dissertation, University of Calgary).

ALEXFORBES, 2024. The integration of artificial intelligence and machine learning in investment decisions and operations among asset managers in South Africa. [Online]. Available: https://myapi.alexanderforbes.co.za/content/download/investmentsmedia/communications?path=TheIntegrationOfArtificialIntelligence2024.pdf. [Access: 3 November 2024].

ALHALASEH, R.H. AND AL SHAWAWREH, F.K., 2024. Enhancing portfolio optimization: a comparative analysis of the mean-variance Markowitz model and risk-parity contribution strategies. *Corporate & Business Strategy Review*, 5(3), pp. 124–136.

AL MAIYAHI, S, S, 2020. Efficient Frontier Curve. [online]. Available: https://www.linkedin.com/pulse/efficient-frontier-curve-sultan-saif-al-maiyahi/. [Date Accessed: 20 February 2024].

AL QASSEM, L.M., 2022. Microservice architecture and efficiency model for cloud computing services. (Doctoral dissertation, Khalifa University)

ASSAEL, J., 2023. Machine learning for ESG data in the financial industry. (Doctoral dissertation, Université Paris-Saclay).

AUER, T., 2024. Regime-investing sector model performance in benchmark tilting; evidence in the US. (Master Theses, Aalto University)

BARCLAYSHEDGE, 2018. BarclayHedge Survey: Majority of Hedge Fund Pros Use AI/Machine Learning in Investment Strategies. [Online]. Available:

https://www.barclayhedge.com/insider/barclayhedge-survey-majority-of-hedge-fund-pros-use-ai-machine-learning-in-investment-strategies. [Date Access: 3 November 2024].

BASAK, S., KAR, S., SAHA, S., KHAIDEM, L. AND DEY, S.R., 2019. Predicting the direction of stock market prices using tree-based classifiers. *The North American Journal of Economics and Finance*, 47, pp.552-567.

BASAK, D., PAL, S. AND PATRANABIS, D.C., 2007. Support vector regression. *Neural Information Processing-Letters and Reviews*, 11(10), pp.203-224.

BATTAGLIA, T.K. AND LEAL, R.P., 2017. Equally weighted portfolios of randomly selected stocks and the individual investor. *Latin American Business Review*, 18(1), pp.69-90.

BAYRAMOGLU, M.F. AND BASARIR, C., 2019. International diversified portfolio optimization with artificial neural networks: An application with foreign companies listed on NYSE. *Machine Learning Techniques for Improved Business Analytics*, pp. 201-223. IGI Global.

BECHIS, L., 2020. Machine learning portfolio optimization: hierarchical risk parity and modern portfolio theory. (Master Theses, Libera Universita Internazionale Studi Sociali)

BESSLER, W, AND WOLFF, D, 2024. Portfolio Optimization with Sector Return Prediction Models. *Journal of Risk and Financial Management*, 17(6), p.254.

BLOOMBERG, 2024. FTSE/JSE Africa All Share Index. [Online]. Available: https://www.bloomberg.com/quote/JALSH:IND. [Date Access: 24 October 2024].

BORA, Y., 2023. Diversification and Risk Assessment using Data Science: Downside Risk vs. Mean Variance Optimization. *Iconic Research And Engineering Journals*, 6(12), pp.769-777.

BOTUNAC, I., BOSNA, J. AND MATETIĆ, M., 2024. Optimization of Traditional Stock Market Strategies Using the LSTM Hybrid Approach. Information, 15(3), p.136.

BOX, G.E., 1979. All models are wrong, but some are useful. *Robustness in Statistics*, *202*(1979), p.549.

BRANDI, H.S. AND DOS SANTOS, S.F., 2020. Measuring sustainable development goals: an application of modern portfolio theory on sustainability systems. *Clean Technologies and Environmental Policy*, *22*(4), pp.803-815.

BROWN, C.E., NIELSON, N.I. AND PHILLIPS, M.E., 1990. EXPERT SYSTEMS FOR PERSONAL FINANCIAL PLANNING. *Journal of Financial Planning*, 3(3).

BUCZYBSKI, W., CUZZOLIN, F. and SAHAKIAN, B., 2021. A review of machine learning experiments in equity investment decision-making: why most published research findings do not live up to their promise in real life. *International Journal of Data Science and Analytics*, 11, pp.221-242.

BURHAN, H.A., 2023. Comparison of Prediction Performances of Regression Models in Machine Learning: An Application on the Turkish Mercantile Exchange Wheat Index. *Nişantaşı Üniversitesi Sosyal Bilimler Dergisi*, 11(2), pp.602-623.

CAI, H. AND SCHMIDT, A.B., 2020. Comparing mean–variance portfolios and equal-weight portfolios for major US equity indexes. *Journal of Asset Management*, 21(4), pp.326-332.

CESARONE, F., SCOZZARI, A. AND TARDELLA, F., 2013. A new method for mean-variance portfolio optimization with cardinality constraints. *Annals of Operations Research*, 205, pp.213-234.

CHARLES, A., DARNÉ, O. AND FOUILLOUX, J., 2016. The impact of screening strategies on the performance of ESG indices. Working Papers hal-01344699, HAL. Available: https://ideas.repec.org/p/hal/wpaper/hal-01344699.html. [Date Access: 27 May 2024].

CHARTERED FINANCIAL ANALYST INSTITUTE, 2024. How machine learning is transforming the investment process. [Online]. Available: https://www.cfainstitute.org/insights/articles/how-machine-learning-is-transforming-the-investment-process. [Date Accessed: 17 December 2024]

CHOE, H., MASULIS, R.W. AND NANDA, V., 1993. Common stock offerings across the business cycle: Theory and evidence. *Journal of Empirical finance*, 1(1), pp.3-31.

DAS, J.D., BOWALA, S., THULASIRAM, R.K. AND THAVANESWARAN, A., 2023. Portfolio Diversification with Clustering Techniques. *2023 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 97-102. IEEE.

DAVENPORT, T.H., 2018. The AI advantage: How to put the artificial intelligence revolution to work. [e-Book]. Massachusetts Institute of Technology Press. ISBN electronic: 9780262350631. DOI: https://doi.org/10.7551/mitpress/11781.001.0001

DE FRANCO, C., GEISSLER, C., MARGOT, V. AND MONNIER, B., 2020. ESG investments: Filtering versus machine learning approaches. *Evolving Practices in Public Investment Management*, p.57-80.

DE PINHO AZEVEDO, A.R., 2022. The Performance of ETFs vs. Mutual Funds on the Euronext Amsterdam in the period 2010-2021. (Master Dissertation, Universidade do Porto)

DEPARTMENT OF PLANNING, MONITORING AND EVALUATION, 2013. Economic Transformation. [Online]. Available:

https://www.dpme.gov.za/publications/20%20Years%20Review/20%20Year%20Review%20Documents/20YR%20Chapter%204%20Economic%20Transformation.pdf. [Date Access: 17 October 2024].

DEWI, C. AND CHEN, R.C., 2019. Random forest and support vector machine on features selection for regression analysis. *International Journal of Innovative Computing, Information and Control*, 15(6), pp.2027-2037.

DING, Z., JI, B., YAO, H., CHENG, X., YU, S., SUN, X., LIU, S., XU, L., ZHOU, Y. AND SHI, Y., 2023. An Analysis of the Factors Affecting Forest Mortality and Research on Forecasting Models in Southern China: A Case Study in Zhejiang Province. *Forests*, 14(11), p.2199.

DIP DAS, J., 2024. Machine learning and data science application for financial price prediction and portfolio optimization. (Master Theses, University of Manitoba)

DOWNING, E. 2011. Schaum's outline series: Introduction to Mathematical Economics (3e). McGrawHill, ISBN 9780071762519

EL HAJJ, M. AND HAMMOUD, J., 2023. Unveiling the influence of artificial intelligence and machine learning on financial markets: A comprehensive analysis of AI applications in trading, risk management, and financial operations. *Journal of Risk and Financial Management,* 16(10), p.434

ELBANNAN, M.A., 2015. The capital asset pricing model: an overview of the theory. *International Journal of Economics and Finance*, 7(1), pp.216-228.

ELKJAER, C.J.,2022. ESG Investing Through the Lens of Portfolio Optimization. (Master Theses, Copehhagen University)

ELSHQIRAT, D.M., 2019. An empirical examination of the arbitrage pricing theory: Evidence from Jordan. *Journal of Studies in Social Sciences*, 18(2), pp.46-67.

ELTON, E.J., GRUBER, M.J., BROWN, S.J. AND GOETZMANN, W.N., 2009. Modern portfolio theory and investment analysis. (9th ed.) John Wiley & Sons.

ERNST & YOUNG GLOBAL LIMITED, 2018. Artificial Intelligence in Middle East and Africa. South Africa Outlook for 2019 and beyond. [Online]. Available: https://info.microsoft.com/rs/157-GQE-382/images/MicrosoftSouthAfricanreportSRGCM1070.pdf. [3 November 2024].

ESTRADA, J., 2006. Fundamental indexation and international diversification. Available at: Estrada, Javier, Fundamental Indexation and International Diversification. Available at SSRN: http://dx.doi.org/10.2139/ssrn.949162

ERNST & YOUNG GLOBAL LIMITED, 2018. Artificial Intelligence in Middle East and Africa. South Africa Outlook for 2019 and beyond. [Online]. Available: https://info.microsoft.com/rs/157-GQE-382/images/MicrosoftSouthAfricanreportSRGCM1070.pdf. [3 November 2024].

EZUGWU, A.E., OYELADE, O.N., IKOTUN, A.M., AGUSHAKA, J.O. AND HO, Y.S., 2023. Machine learning research trends in Africa: a 30 year overview with bibliometric analysis review. *Archives of Computational Methods in Engineering*, 30(7), pp.4177-4207.

FABOZZI, F.J., GUPTA, F. AND MARKOWITZ, H.M., 2002. The legacy of modern portfolio theory. *The journal of investing*, *11*(3), pp.7-22.

FABOZZI, F.J., MARKOWITZ, H.M. AND GUPTA, F., 2008. Portfolio selection. *Investment Management and Finance Management* 2(1), pp.45-78.

FAGERHOLT, A.L. AND AANONSEN, B.Ø., 2016. Risk Parity Stock Optimization Using Principal Component Quantile Simulation (Master theses, Norwegian University of Science and Technology).

FAN, J.H. AND MICHALSKI, L., 2020. Sustainable factor investing: Where doing well meets doing good. *International Review of Economics & Finance*, 70, pp.230-256.

FARRELL, G. AND KEMP, E., 2020. Measuring the financial cycle in South Africa. *South African Journal of Economics*, 88(2), pp.123-144.

FASTRICH, B., PATERLINI, S. AND WINKER, P., 2015. Constructing optimal sparse portfolios using regularization methods. *Computational Management Science*, 12(3), pp.417-434.

FAULKNER, D. AND LOEWALD, C., 2008. Policy change and economic growth: A case study of South Africa. International Bank for Reconstruction and Development/The World Bank.

FERGUSON, J., 2004. A brief survey of the history of the calculus of variations and its applications. DOI: https://doi.org/10.48550/arXiv.math/0402357.

FISHER, S.J. AND LIE, M.C., 2004. Asset allocation for central banks: optimally combining liquidity, duration, currency, and non-government risk. *Risk management for central bank foreign reserves*, pp.75-95.

GENUER, R., POGGI, J.M., TULEAU-MALOT, C. AND VILLA-VIALANEIX, N., 2017. Random forests for big data. *Big Data Research*, 9, pp.28-46.

GONDO, J., 2009. Financial development and economic growth: Evidence from South Africa: 1970-1999. In Annual Western Cape Economics Postgraduate Student Conference. Stellenbosch.

GUNJAN, A. AND BHATTACHARYYA, S., 2023. A brief review of portfolio optimization techniques. *Artificial Intelligence Review*, 56(5), pp.3847-3886.

GUPTA, R. and BASU, P., 2009. Sector analysis and portfolio optimization: The Indian experience. *International Business and Economics Research Journal*, 8(1), pp.119-130.

HAESER, G. AND RAMOS, A., 2020. Constraint qualifications for Karush–Kuhn–Tucker conditions in multiobjective optimization. *Journal of Optimization Theory and Applications*, 187, pp.469-487.

HEYMANS, A. AND VAN HEERDEN, C., 2014. A risk-adjusted evaluation of the JSE top 40 as an international investment option. *Journal of Applied Business Research*, 30(6), pp.1639-1654.

HERZEL, S., NICOLOSI, M. AND STĂRICĂ, C., 2012. The cost of sustainability in optimal portfolio decisions. *The European Journal of Finance*, 18(3-4), pp.333-349.

HIRAN, K.K., JAIN, R.K., LAKHWANI, K. AND DOSHI, R., 2021. Machine Learning: Master Supervised and Unsupervised Learning Algorithms with Real Examples (English Edition). Business Promotion Bureau Publications.

HIRSCH, A., LEVY, B. AND NXELE, M., 2021. Politics and Economic Policymaking in South Africa since 1994. *The Oxford Handbook of the South African Economy*, p.66-90.

HOUSSEIN, E.H., DIRAR, M., ABUALIGAH, L. AND MOHAMED, W.M., 2022. An efficient equilibrium optimizer with support vector regression for stock market prediction. *Neural computing and applications*, pp.1-36.

HOVHANNISYAN, R., 2023. Optimal allocation of pension funds. sensitivity analyses of the adjusted Sharpe ratio. *Economics, Finance and Accounting*, 1(11), pp.5-5.

HUNTBACH, M.M. AND RINGWOOD, G.A., 1999. Agent-oriented programming: From prolog to guarded definite clauses (No. 1630). Springer Science & Business Media.

Ile, H.S., 2020. Active share: implications for active portfolio management. (Doctoral Dissertation, University of Manchester)

INDUSTRIAL DEVELOPMENT CORPORATION, 2024. Key-trends-in-the-South-African-economy-April-2024. [Online]. Available: https://www.idc.co.za/wp-content/uploads/2024/04/Key-trends-in-the-South-African-economy-April-2024.pdf. [Date Access: 17 October 2024].

INTERNATIONAL MONETARY FUND, 2024. Real GDP Growth. [Online]. Available: https://www.imf.org/external/datamapper/NGDP_RPCH@WEO/ZAF?zoom=ZAF&highlight=ZAF. [Date Access: 3 October 2024].

JAWADI, F., AMEUR, H.B., BIGOU, S. AND FLAGEOLLET, A., 2022. Does the real business cycle help forecast the financial cycle? *Computational Economics*, 60(4), pp.1529-1546.

JEREZ, T. AND KRISTJANPOLLER, W., 2020. Effects of the validation set on stock returns forecasting. *Expert Systems with Applications*, 150, p.113271.

JIMÉNEZ-PRECIADO, A.L., VENEGAS-MARTÍNEZ, F. AND RAMÍREZ-GARCÍA, A., 2022. Stock portfolio optimization with competitive advantages (moat): A machine learning approach. *Mathematics*, 10(23), p.4449.

JONES K.M., 2023. Building an Optimized Stock Portfolio Using Machine Learning Models (Master theses, Savannah State University).

JSE GROUP, 2024a. List JSE. [Online]. Available: https://www.jse.co.za/list-jse. [Date Access: 5 November 2024].

JSE GROUP, 2024b. SA Sector. [Online]. Available: https://www.jse.co.za/sa-sector. [Date Access: 3 November 2024].

JSE GROUP, 2024c. Industry Classification Benchmark (ICB). [Online]. Available: https://clientportal.jse.co.za/technical-library/industry-classification-benchmark-(icb).[Date Access: 3 November 2024].

JSE GROUP, 2024d. Overview and History. [Online]. Available: https://group.jse.co.za/history. [Date Access: 17 October 2024].

JSE GROUP, 2024e. Group Overview. [Online]. Available: https://group.jse.co.za/group-overview#:~:text=The%20role%20of%20the%20JSE&text=As%20market%20infrastructures%2C%20they%20enable,listing%20and%20trading%20of%20securities. [Date Access: 17 October 2024].

JSE GROUP, 2024f. List on the JSE. [Online]. Available: https://www.jse.co.za/list-jse#:~:text=Our%20financial%20aggregation%20has%20more,index%20data%20nearing%20200%20Indices. [Date Access: 17 October 2024].

JSE GROUP, 2024g. ICB Industry. [Online]. Available: https://www.jse.co.za/services/indices/icb-industry. [Date Access: 17 October 2024].

KANG, M., TEMPLETON, G.F., KWAK, D.H. AND UM, S., 2024. Development of an AI Framework Using Neural Process Continuous Reinforcement Learning to Optimize Highly Volatile Financial Portfolios. Knowledge-Based Systems, 300(112017). DOI: https://doi.org/10.1016/j.knosys.2024.112017

KASTHURI, R., VASANTHI, P., RANGANAYAKI, S. AND SESHAIAH, C.V., 2011. Multi-item fuzzy inventory model involving three constraints: A Karush-Kuhn-Tucker conditions approach. *American Journal of Operations Research*, 1(3), pp.155-159.

KHANDELWAL, V. AND CHOTIA, V., 2022. Is There a Beta Anomaly? Evidence from India. *Annals of Financial Economics,* 17(04), p.2250020. DOI: https://doi.org/10.1142/S2010495222500208

KHAZEN, C., 2023. Forecasting Financial Returns using Machine Learning methods. (Bachelor's paper, Libera Università Internazionale degli Studi Sociali Guido Carli)

KOBETS, V. AND SAVCHENKO, S., 2022. Building an Optimal Investment Portfolio with Python Machine Learning Tools. *Information Technology and Implementation*, pp. 307-315.

KOTZE, A., 2017. FTSE/JSE Top 40 index long-term returns. DOI: http://dx.doi.org/10.2139/ssrn.2978093

KRUGER, A, 2024. Here's why its global peers continue to outshine the JSE. Money Web. [Online]. Available: https://www.moneyweb.co.za/investing/heres-why-its-global-peers-continue-to-outshine-the-jse/. [3 November 2024].

KUANG, F., XU, W. AND ZHANG, S., 2014. A novel hybrid KPCA and SVM with GA model for intrusion detection. *Applied Soft Computing,* 18, pp.178-184.

KURISCAK, E., MARSALEK, P., STROFFEK, J. AND TOTH, P.G., 2015. Biological context of Hebb learning in artificial neural networks, a review. *Neurocomputing*, 152, pp.27-35.

LAGOWSKI, M., 2022. Portfolio Optimisation: An Empirical Study of The Hierarchical Risk Parity and Mean-Variance Methods. (Bachelor Degree, Lancaster University).

LAI, T.Y. AND STOHS, M.H., 2021. CAPM and Asset Pricing. *International Journal of Business*, 26(4), pp.105-118.

LEE, M.C., AND SU, L.E., 2014. Capital market line based on efficient frontier of portfolio with borrowing and lending rate. *Universal Journal of Accounting and Finance*, 2(4), pp.69-76.

LEVANTESI, S. AND ZACCHIA, G., 2021. Machine learning and financial literacy: An exploration of factors influencing financial knowledge in Italy. *Journal of Risk and Financial Management*, 14(3), pp.120-140.

LEVY, H., 2010. The CAPM is alive and well: A review and synthesis. *European Financial Management*, 16(1), pp.43-71.

LIAGKOURAS, K. AND METAXIOTIS, K., 2018. A new efficiently encoded multi objective algorithm for the solution of the cardinality constrained portfolio optimization problem. *Annals of Operations Research*, 267, pp.281-319.

LIAGKOURAS, K., METAXIOTIS, K. AND TSIHRINTZIS, G., 2020. Incorporating environmental and social considerations into the portfolio optimization process. *Annals of Operations Research*, pp.1-26.

LIN, W. AND TAAMOUTI, A., 2023. Enhancing Portfolio Resilience to Systemic Risk: A Neural Network Approach. (Paper, University of Liverpool).

LIU, Y., ZHOU, G. AND ZHU, Y., 2020. Maximizing the Sharpe ratio: A genetic programming approach. DOI: *http://dx.doi.org/10.2139/ssrn.3726609*

LJUNGBERG, A. AND HÖGSTEDT, A., 2021. Modern Portfolio Theory Combined with Magic Formula: A study on how Modern Portfolio Theory can improve an established investment strategy. (Bachelor Thesis, Linnaeus University).

LONDON STOCK EXCHANGE GROUP, 2024. FTSE/JSE Top 40 Index. [Online]. Available: https://research.ftserussell.com/Analytics/FactSheets/Home/DownloadSingleIssue?issueName=J200&isManual=False. [Date Access: 24 October 2024].

LUNDH, J., CARDONA, S.E. and NIELSEN, M.S., 2020. Quality Investing and Industry Heterogeneity. (Master Theses, Copenhagen University).

MA, Y., HAN, R. and WANG, W., 2021. Portfolio optimization with return prediction using deep learning and machine learning. *Expert Systems with Applications*, *165*, p.113973.

MAJAPA, M. AND GOSSEL, S.J., 2016. Topology of the South African stock market network across the 2008 financial crisis. *Physica A: Statistical Mechanics and its Applications*, 445, pp.35-47.

MANIAR, P.A., 2023. Stock Market Prediction and Portfolio Management with AI. *International Journal of Scientific Research in Engineering and Management* 7(8). DOI: https://doi.org/10.55041/IJSREM25440

MANSURI, N.S. AND SHAH, P., 2022. Dynamic Risk Factors in Expedition of Capital Asset Pricing Model for Measuring Required Rate of Return Intended for Equity Valuation. *Journal of Positive School Psychology*, 6(7), pp.1429-1436.

MANTA, O., 2020. Financing and Fiscality in the context of artificial intelligence at the global level. *European Journal of Marketing and Economics*, 3(1), pp.39-62.

MARK-EGART, D.B., 2020. Arbitrage Pricing and Investment Performance in the Nigerian Capital Market. DOI: http://dx.doi.org/10.2139/ssrn.3548208

MARKET, U.S., 2022. Regime-Based Factor Allocation. (Master Theses, Copenhagen University)

MARKOWITZ, H, 1952. Portfolio Selection. *The Journal of Finance,* 7(1), pp.77–91. https://doi.org/10.2307/2975974.

MARKOWITZ, H, M, 1959. Portfolio Selection: Efficient Diversification of Investments. *Cowles Foundation Monograph* 16, New York: John Wiley & Sons

MASESE, J.M., 2017. Portfolio optimization in the Kenyan stock market: a comparison between mean-variance optimization and threshold accepting (Doctoral dissertation, Strathmore University).

MASHISHI, A, A, 2023. AI National Government Summit Discussion Document. Department of Communications and Digital Technologies. [Online]. Available: https://www.dcdt.gov.za/images/phocadownload/AI_Government_Summit/National_AI_Government_Summit_Discussion_Document.pdf. [Date Access: 1 November 2024].

MASONGWENI, V.V. AND SIMO-KENGNE, B.D., 2024. The impact of sustainable investment on firm performance in South Africa. *South African Journal of Accounting Research*, pp.1-28.

MAULUD, D. AND ABDULAZEEZ, A.M., 2020. A review on linear regression comprehensive in machine learning. *Journal of Applied Science and Technology Trends*, 1(2), pp.140-147.

MAY, K., 2022. Forecast Based Portfolio Optimisation Using XGBoost (Doctoral dissertation, University of the Witwatersrand, Johannesburg).

MESGARANI, H., AGHDAM, Y.E., BEIRANVAND, A. AND GÓMEZ-AGUILAR, J.F., 2024. A novel approach to fuzzy based efficiency assessment of a financial system. *Computational Economics*, 63(4), pp.1609-1626.

MIKUČIAUSKAITĖ, P., 2024. Investigation of Equity through Deep Learning and Portfolio Optimization (Doctoral dissertation, Vilniaus Gedimino technikos universitetas.).

MISHRA, S. and PADHY, S., 2019. An efficient portfolio construction model using stock price predicted by support vector regression. *The North American Journal of Economics and Finance*, 50, p.101027.

MOHANDOSS, D.P., SHI, Y. AND SUO, K., 2021, JANUARY. Outlier prediction using random forest classifier. 2021 Institute of Electrical and Electronics Engineers 11th Annual Computing and Communication Workshop and Conference (CCWC), pp.27-33.

MUCHAONYERWA, F. AND CHOGA, I., 2015. Business cycles and stock market performance in South Africa. *Corporate Ownership and Control* 12(3), pp84-93.

MUNETSI, R.P., 2018. Testing the influence of herding behaviour on the Johannesburg Securities Exchange. (Master Theses, University of Western Cape)

MURIUKI, M.J., 2003. Portfolio Return Characteristics of Different Market Sectors at the Nairobi Stock Exchange (Doctoral dissertation, University of Nairobi).

MUSA, A.A. AND OKOLOGUME, H., 2020. Test for the Validity of the Arbitrage Pricing Theory (APT) in the Nigeria Banking Industry. *African Scholar Journal of Management Science and Entrepreneurship,* 19(7).

MUSLIM, M.A., DASRIL, Y., HARVEEND, M. AND MUZANAH, R., 2024. A Mean-Variance Approach to Portfolio Optimisation for Effective Stock Selection in the Malaysian Stock Market. *Journal of Technology Management and Business,* 11(1), pp.136-146.

NAGY, B.Z. AND BENEDEK, B., 2021. Higher co-moments and adjusted Sharpe ratios for cryptocurrencies. *Finance Research Letters*, 39, p.101543.

NAIR, B.B., MOHANDAS, V.P. AND SAKTHIVEL, N.R., 2010. A genetic algorithm optimized decision tree-SVM based stock market trend prediction system. *International journal on computer science and engineering*, 2(9), pp.2981-2988.

NATIONAL PLANNING COMMISION. Chapter 3 Economy and Employment. [Online]. Available: https://www.nationalplanningcommission.org.za/assets/Documents/NDP_Chapters/NDP%202030-CH3-Economy%20and%20employment.pdf [Date Access: 17 October 2024].

NATIONAL TREASURY, 2024. 2024 Budget Speech. [Online]. Available: https://www.treasury.gov.za/documents/national%20budget/2024/speech/speech.pdf. [Date Access: 24 October 2024].

NATIONAL TREASURY, 2024. Schedule of Domestic Debt Government Bond as at 31 October 2024. Available: https://investor.treasury.gov.za/Market%20Information/Forms/AllItems.aspx. [Date Access: 25 November 2024].

NELLES, O. AND NELLES, O., 2001. Model complexity optimization. *Nonlinear System Identification: Classical Approaches to Neural Networks and Fuzzy Models*, pp.157-201.

NGO, V.M., NGUYEN, H.H. AND VAN NGUYEN, P., 2023. Does reinforcement learning outperform deep learning and traditional portfolio optimization models in frontier and developed financial markets? *Research in International Business and Finance*, 65, p.101936.

NGUYEN, H., 2019. Portfolio Optimization Methods: The Mean-Variance Approach and the Bayesian Approach. (Honours Theses, University of Mississippi)

OECD, 2021. OECD Business and Finance Outlook 2021: AI in Business and Finance. OECD Publishing, Paris. Available: https://doi.org/10.1787/ba682899-en.

ÓSKARSSON, M., 2021. Back-testing portfolio risk management strategies (Master dissertation, Reykjavík University).

OUYANG, Z., 2022, MARCH. A Study of Stock Portfolio Strategy Based on Machine Learning. Atlantis Press. 2022 7th International Conference on Financial Innovation and Economic Development, pp. 79-87.

PEDERSEN, L.H., BABU, A. AND LEVINE, A., 2021. Enhanced portfolio optimization. *Financial Analysts Journal*, 77(2), pp.124-151.

Peng, S., 2021, December. The Validity of CAPM: A Critical and Conclusive Study with Empirical Evidence from the UK Security Market. 2021 *3rd International Conference on Economic Management and Cultural Industry (ICEMCI 2021)* (pp. 2237-2243). Atlantis Press.

PLYAKHA, Y., UPPAL, R. AND VIKOV, G., 2015. Why do equal-weighted portfolios outperform value-weighted portfolios. DOI: http://dx.doi.org/10.2139/ssrn.2724535.

PORAGE, C., 2021. Sustainability in portfolio optimization. [Online]. Available: https://www.diva-portal.org/smash/get/diva2:1549036/FULLTEXT01.pdf. [Date Access: 12 September 2024].

PRASAD, A. AND SEETHARAMAN, A., 2021. Importance of machine learning in making investment decision in stock market. *Vikalpa*, *46*(4), pp.209-222.

PRICE WATERHOUSE AND COOPERS, 2023. Asset and wealth management revolution 2023: The new context. [Online]. Available: https://www.pwc.com/gx/en/industries/financial-services/asset-management/publications/asset-and-wealth-management-revolution-2023.html. [Date Access: 3 November 2024].

PRITAM, K.S., MATHUR, T., AGARWAL, S., PAUL, S.K. AND MULLA, A., 2022. A novel methodology for perception-based portfolio management. *Annals of Operations Research*, 315(2), pp.1107-1133.

RAMCHANDANI, P., PAICH, M. AND RAO, A., 2017, AUGUST. Incorporating learning into decision making in agent-based models. *Cham: Springer International Publishing.* European Conference on Artificial Intelligence, pp. 789-800.

REN, Z., 2024. Systematic Design and Effectiveness Enhancement of English Teaching in Colleges and Universities Combining Decision Tree Modelling. *Applied Mathematics and Nonlinear Sciences*, 9(1).

ROMPOTIS, G.G., 2022. The ESG ETFs in the UK. *Journal of Asset Management*, 23(2), pp.114-129.

ROSSI, M., 2016. The capital asset pricing model: a critical literature review. *Global Business and Economics Review,* 18(5), pp.604-617.

SALIH, A, 2013. Method of Lagrange Multipliers. [Online]. Available: https://www.iist.ac.in/sites/default/files/people/Lagrange-Multiplier.pdf. [Date Access: 28 September 2024].

SANLAM, 2022. The Sanlam Artificial Intelligence Investment Capability. [Online]. Available: https://www.sanlam.com/productcatalog/SanlamFundFactSheets/SanlamFundFactSheets/Sanlam%20Ai%20Capability.pdf. [Date Access: 3 November 2024].

SEN, J., DUTTA, A. AND MEHTAB, S., 2021. Profitability analysis in stock investment using an LSTM-based deep learning model. *2021 2nd International Conference for Emerging Technology*, pp. 1-9.

SEN, J., JOSEPH, G., MUTHUKRISHNAN, K., TULASI, K. AND VARUKOLU, P., 2022. Precise Stock Price Prediction for Robust Portfolio Design from Selected Sectors of the Indian Stock Market. DOI: https://doi.org/10.48550/arXiv.2201.05570

SEN, J., MEHTAB, S., DUTTA, A. AND MONDAL, S., 2021. Hierarchical risk parity and minimum variance portfolio design on NIFTY 50 stocks. *2021 International Conference on Decision Aid Sciences and Application*, pp. 668-675.

SENTHILKUMAR, A., NAMBOOTHIRI, A. AND RAJEEV, S., 2022. Does portfolio optimization favor sector or broad market investments? *Journal of Public Affairs*, 22, p.e2752.

SIMAR, J., 2023. Enhancing estimation of expected returns in modern portfolio theory through machine learning. (Master Theses, University of Liège)

SIMPLE WALL STREET, 2024. South African (JSE) Market Analysis & Valuation. [Online]. Available: https://simplywall.st/markets/za. [Date Access: 21 October 2024].

ŠKRINJARIĆ, T. AND ORLOVIĆ, Z., 2020. Economic policy uncertainty and stock market spillovers: Case of selected CEE markets. *Mathematics,* 8(7), p.1077.

SOUTH AFRICA RESERVE BANK, 2022. December 2022 – Measures of electricity load-shedding. [Online]. Available: https://www.resbank.co.za/en/home/publications/publication-detail-pages/quarterly-bulletins/boxes/2022/December/Measuresofelectricityload-shedding. [Date Access: 17 October 2024].

SOUZA, L.A., 2014. Modern real estate portfolio management (MREPM): Applications in modern and post-modern real estate portfolio theory (MREPT/PMREPT). (Doctoral Dissertation, Golden Gate University).

STAMPE, M. AND REVELSBY, S.B., 2022. ESG Integration in the European Private Equity Industry: Does ESG integration provide superior risk-adjusted returns? (Master theses, Handelshøyskolen BI).

ŠTAROLIS, G. AND KRAKAUSKAS, T., 2015. Evaluation of passive investment strategies in Oslo stock exchange (Bachelor theses, ISM University of Management and Economics).

STATISTA, 2024. Machine Learning – South Africa. [Online]. Available: https://www.statista.com/outlook/tmo/artificial-intelligence/machine-learning/south-africa#analyst-opinion. [Date Access: 1 November 2024].

STILL, S. AND KONDOR, I., 2010. Regularizing portfolio optimization. *New Journal of Physics*, 12(7), p.075034.

STOILOV, T., STOILOVA, K. AND VLADIMIROV, M., 2022. Decision Support for Portfolio Management by Information System with Black–Litterman Model. *International Journal of Information Technology & Decision Making*, 21(02), pp.643-664.

SUN, M., 2018. Regression models with a universal penalized function and applications in economics and finance (Doctoral dissertation, University of Alabama Libraries).

SURTEE, T.G. AND ALAGIDEDE, I.P., 2023. A novel approach to using Modern Portfolio Theory. *Borsa Istanbul Review*, *23*(3), pp.527-540.

SWADE, A., NOLTE, S., SHACKLETON, M. and LOHRE, H., 2023. Why do equally weighted portfolios beat value-weighted ones? *Journal of Portfolio Management*, *49*(5).

TACCHINI, S., 2024. Forecasting stock market behavior using artificial intelligence (Master theses, Universitat Politècnica de Catalunya).

TALJAARD, B.H. and MARÉ, E., 2021. If the equal weighted portfolio is so great, why isn't it working in South Africa? *Investment Analysts Journal*, *50*(1), pp.32-49.

TARENTAAL, 2023a. Economics 318 Lecture 9 – Modern Portfolio Theory for Mathematical Economics course. Makhanda: Department of Economics, Rhodes University.

TARENTAAL, 2023b. Introduction to machine learning: course handouts for Ecos 408 lectures. Makhanda: Department of Economics, Rhodes University

THE PRESIDENCY OF THE REPUBLIC OF SOUTH AFRICA, 2024. Just transition to a low-carbon economy. State of the Nation Address. [Online]. Available: https://www.stateofthenation.gov.za/priorities/growing-the-economy-and-jobs/just-transition-to-a-low-carbon-economy#:~:text=Proposed%20projects%20for%20the%20Green%20Energy%20Transition%20plan&text=The%20Independent%20Power%20Producer%20(IPP,their%20own%20power%20from%20IPPs. [Date Access: 1 November 2024].

TRADING ECONOMICS, 2024a. JSE Market Capitalization. [Online]. Available: https://tradingeconomics.com/jse:sj:market-capitalization. [Date Access: 18 October 2024].

TRADING ECONOMICS, 2024b. South Africa Stock Market Index (SA40). [Online]. Available: https://tradingeconomics.com/top40:ind. [Date Access: 24 July 2024].

TRADINGVIEW, 2024. Stock Market South Africa. [Online]. Available: https://www.tradingview.com/markets/stocks-south-africa/sectorandindustry-sector/ [Date Access: 21 October 2024]

TUNG, L.C., 2024. Comparative Analysis of Hierarchical Risk Parity (Hrp) And Other Portfolio Optimization Methods in Taiwan Stock Market. (Bachelor Theses, Tallin University of Technology).

TURNER, C. AND HAN, J., 2009. Portfolio Optimization under Time-Varying Economic Regimes.

TUPKO, O., TUPKO, N. AND VASIL'EVA, N., 2016. Comparison analysis of copula-based and Markowitz portfolio methods. Технологический аудит и резервы производства, 4(2 (30)), pp.65-72.

UKIL, A. AND UKIL, A., 2007. Support Vector Machine. *Intelligent systems and signal processing in power engineering,* pp.161-226.

USECHE, A.J., MARTÍNEZ-FERRERO, J. AND ALAYÓN-GONZALES, J.L., 2023. Socially responsible portfolios, environmental, social, corporate governance (ESG) efficient frontiers, and psychic dividends. *Corporate Social Responsibility and Environmental Management,* 31(2), pp.1323-1339.

VENGESAI, E., OBALADE, A.A. and MUZINDUTSI, P.F., 2022. Country risk dynamics and stock market volatility: Evidence from the JSE cross-sector analysis. *Journal of Economics and Financial Analysis*, 5(2), pp.63-84.

VIDAL-GARCÍA, J. AND VIDAL, M., 2022. Modified Sharpe Ratio. DOI: http://dx.doi.org/10.2139/ssrn.4057728

WAFULA, F.J., 2014. The effect of diversification on portfolio returns of mutual funds in Kenya. (Doctoral dissertation, University of Nairobi).

WAH, B.W. AND WANG, T., 1999. Efficient and adaptive Lagrange-multiplier methods for nonlinear continuous global optimization. *Journal of Global Optimization*, 14, pp.1-25.

WICHTLHUBER, M., STREHLE, E., KOPP, D., PREPENS, L., STEGMUELLER, S., RUBINA, A., DIETZEL, C. AND HOHLFELD, O., 2022. IXP scrubber: learning from blackholing traffic for ML-driven DDoS detection at scale. *Proceedings of the ACM SIGCOMM 2022 Conference*, pp. 707-722.

WIECEK, P., 2020. Optimal Portfolio Selection: Comparison of Different Methods Based on Real Life Financial Data. DOI: http://dx.doi.org/10.2139/ssrn.3862506

WORLD BANK, 2024. The World Bank in South Africa. [Online]. Available: https://www.worldbank.org/en/country/southafrica/overview. [Date Access: 1 November 2024].

WYNNE, M., 2023. Long-term Stock Selection using Random Forest and LSTM Models for Fundamental Analysis. [Online]. Available: https://mlwynne24.github.io/pdf/Long-term%20Stock%20Selection%20using%20Random%20Forest%20and%20LSTM%20Models%20for%20Fundamental%20Analysis.pdf. [Date Access: 17 August 2024].

XIDONAS, P. AND ESSNER, E., 2024. On ESG portfolio construction: a multi-objective optimization approach. *Computational Economics*, 63(1), pp.21-45.

YANG, C., LIU, J., ZENG, Y. AND XIE, G., 2017. Prediction of components degradation using support vector regression with optimized parameters. *Energy Procedia*, 127, pp.284-290.

YANG, Z., 2021. Analysis on CAPM and Sharpe Ratio in Market Investment. *6th International Conference on Financial Innovation and Economic Development*, pp. 5-8.

YAZDANI, M., BABAGOLZADEH, M., KAZEMITASH, N. AND SABERI, M., 2019. Reliability estimation using an integrated support vector regression–variable neighborhood search model. *Journal of Industrial Information Integration*, 15, pp.103-110.

ZAIMOVIC, A, OMANOVIC, AND ARNAUT-BERILO, A, 2021. How many stocks are sufficient for equity portfolio diversification? A review of the literature. *Journal of Risk and Financial Management*, 14(11), p.551.

ZHANG, A., 2023. Portfolio Optimization of Stocks–Python-Based Stock Analysis. *International Journal of Education and Humanities*, 9(2), pp.32-38.

ZHANG, Y., LI, X. AND GUO, S., 2018. Portfolio selection problems with Markowitz's mean–variance framework: a review of literature. *Fuzzy Optimization and Decision Making,* 17, pp.125-158.

ZHAO, W., TAO, T. AND ZIO, E., 2015. System reliability prediction by support vector regression with analytic selection and genetic algorithm parameters selection. *Applied Soft Computing*, 30, pp.792-802.

ZHAO, Z., ZHOU, R., PALOMAR, D.P. AND FENG, Y., 2019. Portfolio optimization. [Online]. Available: https://faculty.sist.shanghaitech.edu.cn/zhao/pubs/ZhaoZhouPalomarFeng-SPMag-PortOpt.pdf. [Date Access: 16 May 2024].

ZHU, G., 2020. Transfer-Learning for Automated Seizure Detection Based on Electric Field Encephalography Reconstructed Signal. *Communications and Network*, *12*(04), pp.174-198.

ZONGMING, T., KOOMSON, P. AND GUOPING, D., 2017. Investment Risk and Returns: The Relationship Between a Stock and An Index Using the Modern Portfolio Theory. DOI: http://dx.doi.org/10.2139/ssrn.3014223

ZUCCHERI, A., 2013. On the Efficiency of the Benchmarks Used in the Asset Management. (Doctoral Dissertation, Università di Bologna)

8. APPENDICES

APPENDEX A

Table A1: Summary of Portfolio Optimization Method using Traditional Method – Different Sectors

Author and Year	Stocks Covered	Sector Covered	Country Covered	Period Covered	Frequency	Methodology	Findings and Conclusions
Herzel, Nicolosi & Stărică (2012)	Components of the S&P500 index	1) Basic Materials 2) Consumer Goods 3) Consumer Services 4) Financials 5) Healthcare 6) Industrials 7) Oil & Gas 8) Technology 9) Telecommunications 10) Utilities	United States of America	1 January 1993 to 31 December 2008	Quarterly	Classic Mean-Variance approach with sustainable constraints	Socially responsible screening increases small loss of Sharpe ratio. Socially responsible screening has a significant impact on market capitalization optimal portfolio.
Adhikari, S., & Jha, P. K. (2016)	20 stocks randomly chosen from NEPSE	1) Commercial Banks 2) Development Banks 3) Finance Companies 4) Insurance Companies	Nepal	1 April 2010 to 31 December 2014	Monthly	Mean-Variance Model Portfolio Frontier Efficient Frontier Minimum Variance Portfolio	Mean variance portfolio outperforms minimum variance portfolio
Charles, A., Darné, O., & Fouilloux, J. (2016)	MSCI, ECPI, ESI, STOXX, DJSI and FTSE4Good indices	1) Financials 2) Technology 3) Health 4) Consumer 5) Telecommunications 6) Industry 7) Energy 8) Materials 9) Utilities	United States of America, Belgium, Finland, France, Germany, Ireland, Italy, the Netherlands, and Spain.	1 September 2010 to 5 January 2015	Daily data	Risk-adjusted performance measures. Tail risk measures.	Higher concentration in some sectors leading to higher risk-adjusted returns

Fagerholt, A. L., & Aanonsen, B. Ø. (2016).	80 stocks from S&P100 index.	1) Energy 2) Materials 3) Industrials 4)Consumer Discretionary 5) Consumer Staples 6) Health Care 7) Financials 8)Information Technology 9)Telecommunication Service 10) Utilities	United States of America	1 January 1995 to 31 December 2015	Monthly	Factor Model Risk Parity Portfolio Optimization Simulation of Return Distributions	Risk Parity Portfolio performs better than factor model and simulation of return distributions
Tupko, O., Tupko, N., & Vasil'Eva, N. (2016).	13 different stocks	1) Financials 2) Chemical 3) Oil 4) Electro energetics	United States of America	1 January 2001 to 1 December 2005	Weekly	Copula-based portfolio Mean-variance optimization	Copula-based portfolio provides significantly higher return than Markowitz portfolio.
Masese, J. M. (2017).	21 different stocks	1) Manufacturing 2) Construction 3) Automobiles	Kenya	1 January 1998 to 31 December 2016	Daily	Mean Variance Optimization model Threshold Accepting Optimization model	Thresholding Accepts model produces higher returns and performs better than Mean Variance model
Liagkouras, Metaxiotis, and Tsihrintzis (2020)	UK Stock Market FTSE-100 corporate social responsibility index	1) Financials 2) Telecommunications 3) Healthcare 4) Technology 5) Consumer goods 6) Consumer services 7) Industrial 8) Energy 9) Basic materials 10 Utilities	England	January 2018 and May 2018.	Daily	Mean Variance Optimization.	ESG stocks hinder potential significant returns.
Porage (2021)	15 stocks of OMXC25 index	1) Health Care 2)Food, beverage, tobacco 3) Banks 4)Industrial goods and service	Denmark	January 2017 to December 2019	Weekly data	Likelihood Ratio test	Statistically significant difference between the minimum variance frontier with ESG stocks and non-ESG stocks.

		5) Utilities 6) Consumer products and service 7) Technology 8) Insurance 9) Energy 10)Construction and Material					
Stampe, and Revelsby (2022)	86 stocks	1) Private Equity Sector	Europe countries – undisclosed	31 December 2015 to 31 December 2020	Quarterly	Aggregation of ESG scores Risk Adjusted performance. Linear Regression Analysis	ESG integrated funds generate a higher Sharpe Ratio than non-ESG integrated portfolios. In regression model, ESG funds produced higher Sharpe ratio compared to non ESG funds, resulting to high adjusted risk returns.
Agarwal, S., & Muppalaneni, N. B. (2022).	Nifty 50 companies	1) Paint 2) Footwear 3) Chemicals 4) Banking 5) Insurance 6) Food 7) Petroleum	India	13 April 2015 to 13 April 2020	Daily	Mean–variance optimization Efficient frontier A developed machine learning algorithm	Author's machine learning model performed better Mean–variance optimization and Efficient frontier
May, K (2022)	JSE Top 40 stocks	1) Basic Materials 2) Industrials 3) Financials 4) Technology 5) Telecommunication 6) Health Care 7) Consumer goods	South Africa	1 January 2014 to 31 December 2020	Daily	Equal Weighted Minimum Variance Equal Weighted Risk Contribution Maximum Decorrelation Inverse Volatility Maximum Diversification Particle Swarm Optimisation (PSO) Genetic Algorithm	The constructed portfolio A by the author outperformed equal weighted and JSE benchmark index

						Long Short-Term Memory Networks (LSTM) Random Forest Extreme Gradient Boosted Machines (XGBoost)	
Rompotis (2022)	49 equities	1) ESG ETFs	England	31 December 2020 to 31 December 2021	Yearly	Capital Asset Pricing Model Fama and French Five Factor Model Sharpe ratio Treynor ratios	ESG ETFs outperform non ESG stocks and there was no significant relationship is between ETFs ESG ratings and their respective assets.
Senthilkumar, A., Namboothiri, A., & Rajeev, S. (2022).	191 companies	1) Automobile 2) Banking 3) Financial Services 4) FMCG 5)Information Technology 6) Media 7) Pharmaceuticals 8) Private Banks 9) PSU Banks 10) Reality 11) NIFTY 50	India	1 April 2013 to 30 October 2019	Monthly	Sharpe's Single Index Model with Elton's portfolio optimization Markowitz Portfolio Theory	Sharpe's model performs better than Markowitz's model.
Useche, Martínez-Ferrero, & Gonzales (2023)	28 companies on Chile's IPSA, 23 Companies on Colombia's COLIR, and 28 Companies on the S&P Lima General Index.	1) Consumer discretionary 2) Communication 3) Consumer Staples 4) Energy 5) Financial 6) Health Care 7) Industrials 8) Materials 9) Real Estate 10) Technology 11) Utilities	China, Colombia, and United States of America	1 January 2011 to 31 December 2019	Daily	Novel utility function Mean-variance optimization with ESG elements. Summary Statistics of optimization methods	ESG funds companies produced higher returns and reduced risks.

Bora (2023)	Top 10 stocks in every sector of the US stock market	1) All sectors in the USA	United States of America	1 January 2007 to 31 December 2022	Unknown	Downside Risk Mean Variance Optimization.	Downside Risk Optimization (DRO) performs better than Mean Variance Optimization (MVO) in all sectors
Xidonas & Essner (2024)	EURO STOXX 50, the DAX, the CAC 40 and the DJIA	1) Eurozone super-sector	Belgium, Finland, France, Germany, Ireland, Italy, the Netherlands, and Spain.	29 May 2016, to 30 May 2021.	Weekly Data	Multi-objective Minmax Optimization model	Optimal ESG portfolios produce higher risk adjusted returns compared to their respective market benchmarks.
AlHalaseh, R. H., & Al Shawawreh, F. K. (2024)	99 Companies from ASE	1) Financial 2) Industry 3) Services.	Jodan	2 January 2018 to 30 December 2022	Daily Closing Pricing	Summary Statistics Mean Variance Optimization Model Risk Parity Optimization	Risk Parity yields more return than Mean Variance
Auer, T. (2024).	21 different stocks	1) Energy 2) Financials 3) Industrials 4) IT 5) Health 6) Utilities 7) Real Estate 8) Communication 9) Consumer Discretionary 10) Consumer Staples	United States of America	1 January 2001 to 5 April 2024	Weekly data	Sharpe-ratio Maximum return Optimization Minimum variance Optimization	Maximum Return generates more return compared to Sharpe ratio and minimum variance
Bessler, W., & Wolff, D. (2024).	Overall S&P500 market	1) Oil and Gas 2) Manufacturing 3) Consumer Goods and Services 4) Health Care	United States of America	1 January 1974 to 31 December 2013	Monthly	Descriptive statistics for sector indices 19 predictive variables Blac- Litter Model Risk-Adjusted Performance	Sector returns outperform in asset allocations decisions more than past historical

		5)Technology and Telecommunication 6) Financials.				Sharpe Ratio portfolio	average returns and 1/N buy and hold portfolio
Tung, L (2024)	30 Companies from Taiwan Stock Exchange	1)Technology 2) Manufacturing	Taiwan	1 January 2007 to 31 December 2007	Daily Adjusted Closing Pricing	Mean-Variance Optimization Maximum Diversification Risk Parity Hierarchical Risk Parity	Hierarchical Risk Parity yields more return than Mean Variance and Maximum Diversification

Table A2: Summary of Portfolio Optimization Method using Machine Learning Method – Different Sectors

Author and Year	Stocks Covered	Sector Covered	Country Covered	Period Covered	Frequency	Methodology	Findings and Conclusions
Turner, C., & Han, J. (2009).	Russell 1000 27 economic indicators	24 equal sectors	United Kingdom	1 January 1980 to March 2009.	Monthly	k-means clustering Gaussian kernel distance Fitted Q-iteration Principle Component Analysis	Fitted Q-iteration yields higher return than Gaussian kernel distance based on information ratio They all perform above the benchmark, the first 3 methods.
Al-jomai, R. (2014)	30 stocks	1) Basic Industries 2) Capital Goods 3) Energy 4) Finance 5) Health Care 6) Miscellaneous 7) Public Utilities 8) Technology 9) Transportation	United States of America	1 April 2013 to 28 April 2014	Daily	Sentimental analysis Frequent pattern mining Clustering Classification	Strong positive relationship between daily prices and news articles
Bayramoglu, M. F., & Basarir, C. (2019).	16 stocks	1) Basic Industries 2) Capital Goods 3) Energy	United States of America	18 October 2007 to 13	Daily	Modeling of Multilayer Perceptron	Artificial neural networks performed better than

		4) Health Care 5)Consumer non-durables 6) Public Utilities 7) Technology 8) Transportation		October 2017		Artificial neural networks Traditional Portfolio	traditional methods and benchmark
Fan & Michalski (2020)	423-495 stocks	1) Energy 2) Financials 3) Industrials 4)Information Technology 5) Health Care 6) Utilities 7)Communication 8)Consumer Discretionary 9)Consumer Staples 10) Materials	Australia	1 February 1958 to 31 December 2016	Monthly	Long-short term memory Neural networks	The Sharpe ratio was improved by exploiting ESG scores with past returns and not including the non-related stocks which had inferior performance. An ESG integration drives financial performance by diverting away from governance dimension.
De Franco, Margot & Monnier (2020).	MSCI World Index	1) Energy 2) Financials 3) Industrials 4)Information Technology 5) Health Care 6) Utilities 7) Telecommunication 8)Consumer Discretionary 9)Consumer Staples 10) Materials	United States of America, Canada, Western Europe, Japan, Australia, New Zealand, Hong Kong, and Singapore	1 August 2009 to 31 March 2018	Monthly	A machine learning algorithm – dimensional space	Dimensional space accurately selects stocks that over performs compared to traditional strategies.
Sen, J., Dutta, A., & Mehtab, S. (2021).	45 stocks from NSE	1) Pharmaceuticals 2) Infrastructure 3) Realty 4) Media, 5) Public sector banks	India	1 January 2016 to 31 December 2020	Monthly	Long-and-short-term memory (LSTM) Minimum Risk Portfolio	LSTM is found to accurate predict stocks prices in the future

		6) Private sector banks 7) Large Cap 8) Mid Cap 9) Small Cap					
Sen, J., & Dutta, A. (2021).	NIFTY 50	1) Media 2) Oil & gas 3) Private banks 4) PSU banks 5) Realty	India	1 January 2016 to 29 October 2021	Daily	Hierarchical Risk Parity (HRP) Hierarchical Equal Risk Contribution (HERC): Tree clustering, Optimal number of cluster identification, Top-down recursive bisection, Naïve risk parity within the cluster for weight allocation Back Testing method	HRP yield better results than HERC portfolio except in one sector relative to Sharpe ratio
Sen, J., Joseph, G., Muthukrishnan, K., Tulasi, K., & Varukolu, P. (2022)	NIFTY50: NIFTY Index	1) Metal 2) Pharma 3) IT 4) Banking 5) Auto	India	1 January 2016 to 27 August 2021	Daily	Minimum Variance Portfolio Optimal Risk Portfolio Equal Weight Portfolio K Neighbouring Decision Tree Support Vector Machine Random Forest XGBoost Logistic Regression Long- and Short-Term Memory Network Convolutional Neural Networks Sector-wise results and analysis	Deep Learning models provide better returns than machine learning methods. Minimum variance portfolio performed better than optimal risk portfolio
Pritam, K. S., Mathur, T., Agarwal, S., Paul, S. K., &	SENSEX and NIFTY	1) Automobiles 2) Banking 3) IT 4) Oil 5) Pharmaceutical	India	1 March 2013 to 31 March 2021	Daily	P-index Multi-criteria Volatility (MCV)	Index fund outperformed on 3 portfolio that were constructed

Mulla, A. (2022)		6) Power				Clustering analysis Fractional lion algorithm for rapid centroid estimation	
Kobets, V., & Savchenko, S. (2022).	12 stocks	1) High technology 2) Microelectronics 3) Engineering 4) Banking 5) Finance	United States of America	1 January 2012 to 1 January 2022	Monthly	LSTM neural network Linear regression method	LSTM neural network forecast returns performed better than historical trends including during global recession
Jiménez-Preciado, A. L., Venegas-Martínez, F., & Ramírez-García, A. (2022)	506 stocks on S&P500	1)Information Technology 2) Industrials 3) Financials 4) Heath Care 5)Consumer Discretionary 6) Consumer Staples 7) Real Estate 8) Utilities 9) Materials 10)Communication Services 11) Energy	United States of America	Not disclosed	Not disclosed	20 Financial Ratios Logistic Regression Support Vector Machine Random Forest Gradient Boosting Artificial Neural Networks	Random Forest yield the highest accurate score compared to ML algorithms.
Mikučiauskaitė, P. (2023)	29 stocks	1) Consumer Staples	India, United States of America, Germany, France and Indonesia and China	1 March 2023 to 29 November 2023	Daily	Bibliometric analysis Multiple criteria decision making (MCDM) Proximity Indexed Value (PIV) Markowitz Mean–Variance (MV) Black-Litterman (BL) Recurrent Neural Network (RNN) Long Short-Term Memory (LSTM) Network	Machine learning methods yield better results than traditional methods

						MATLAB programming language.	
Assael, J. (2023)	2429 companies uniquely identified	1) Industrials 2) Financials 3) Energy 4)Academic & Educational Services 5) Basic Materials 6)Consumer Cyclicals 7)Consumer non-cyclicals 8) Healthcare 9) Real Estate 10) Technology 11) Utilities	Europe countries – undisclosed specific countries	1 January 2002 to 31 December 2019	Daily	Cross-Validation Scheme Gradient Boosting Model	ESG scores had a negative impact on price returns of small and large capitalization companies with better ESG scores are mostly associated with higher price returns.
Simar, J. (2023)	30 specific company stock based on EFT's	2 stocks from each sector: 1)IT 2) Healthcare 3) Industrials 4) Financials	United States of America	5 May 2014 to 5 May 2023	Daily	Morden Portfolio Theory Random Forest Boosting methods	Combination of dividend and growth stocks yields greater returns than divided stocks alone
Zhang, A. (2023)	6 companies	1) Finance 2)Automotive Securities 3) Industrial 4) Metals 5) Communication 6) Property	China	1 October 2020 to 1 October 2021	Daily	Monte Carlo simulation Maximum Sharpe Portfolio Minimum Volatility Portfolio	Machine learning methods generate higher returns than traditional methods
Lin, W., & Taamouti, A (2023)	50 companies	1) Mining 2) Manufacturing 3)Transportation 4) Wholesale 5) Retail 6) Finance 7) Service .	United States of America	1 January 1985 to December 2021	Monthly	SPNN model Portfolio formation GARCH-based model Back Testing	Conditional Sharpe ratio outperforms minimum variance portfolio, and equally weighted portfolio

Dip Das, J., Bowala, S., Thulasiram, R.K. and Thavaneswaran (2023)	16 different stocks	1) Technology 2) Healthcare 3) Energy 4) Banking 5) Consumer	United States of America	1 January 1998 to 31 December 2022	Daily	Bidirectional Long Short-Term Memory Bidirectional Gated Recurrent Unit Clustering: Affinity Propagation	Bidirectional Gated Recurrent Unit outperforms Bidirectional Long Short-Term Memory in dynamic and volatile market scenarios.
Muslim, M. A., Dasril, Y., Harveend, M., & Muzanah, R. (2024).	10 different stocks	1) Energy, 2) Plantation 3) Technology 4) Financial services 5) Consumer products & services 7) Telecommunications & media.	Malaysian	1 January 2014 to 30 September 2020	Monthly	Mean Variance Optimization Linear Programming Portfolio Summary Statistics	Various portfolios were constructed, and portfolio D was the optional portfolio
Masongweni, V.V. and Simo-Kengne, B.D (2024)	120 JSE stocks	1)Consumer Discretionary and Staples industry 2)Energy and Industrials industry 3) Financial, Health and Real Estate industry 4)Technology and Telecommunication industry	South Africa	1 January 2015 to 31 December 2022	Unknown	Panel Data Regression techniques	There is no direct relationship between the financial performance of South African firms and their overall ESG score.
Dip Das (2024)	40 stocks and 25 cryptocurrencies	1) Technology 2) Healthcare 3) Energy 4) Banking 5) Cryptocurrency 6) Consumer	United States of America	1 January 1998 to 31 December 2022	Daily	Bidirectional Long Short-Term Memory (BiLSTM) Bidirectional Gated Recurrent Unit (BiGRU) Activation functions: ReLU, ELU, and Tanh Autoencoder Equal Weighted Portfolio Optimization (EW) Inverse Volatility Weighted Portfolio Optimization (IVW)	BiGRU performed better than BiLSTM Traditional methods yield negative returns in the short run ML more robust than traditional methods

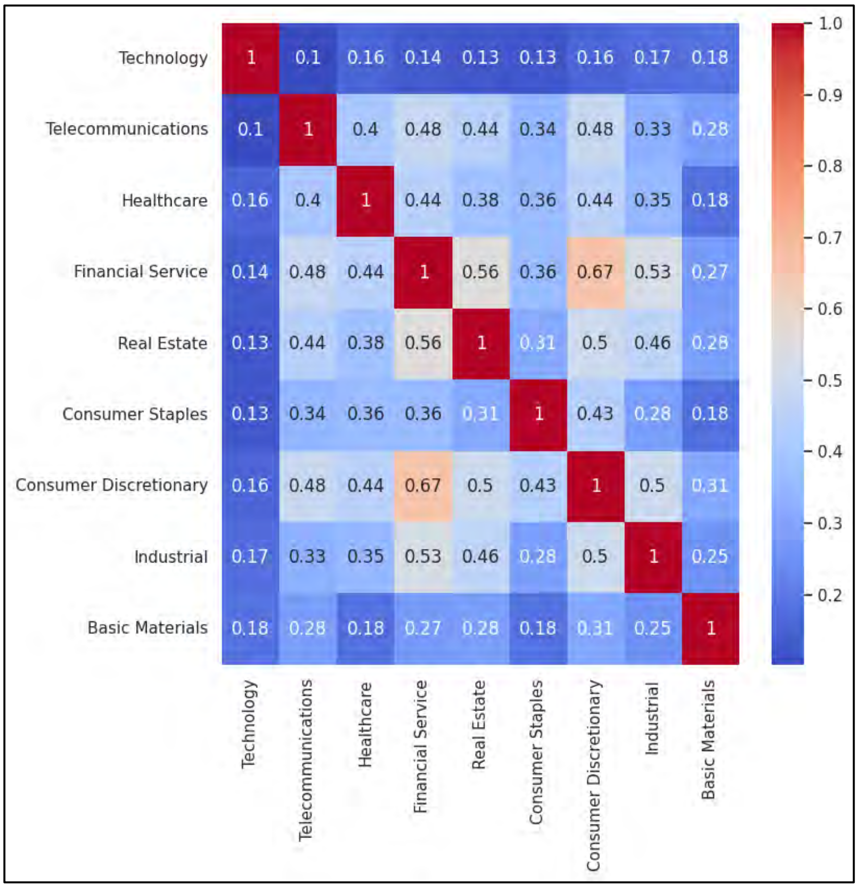
						Mean-Variance Optimization (MVO) Constrained Portfolio Optimization (CPO) Sharpe Ratio Optimization (SRO)	

APPENDIX B

Table B1: Names of 27 Selected Stocks for the Study

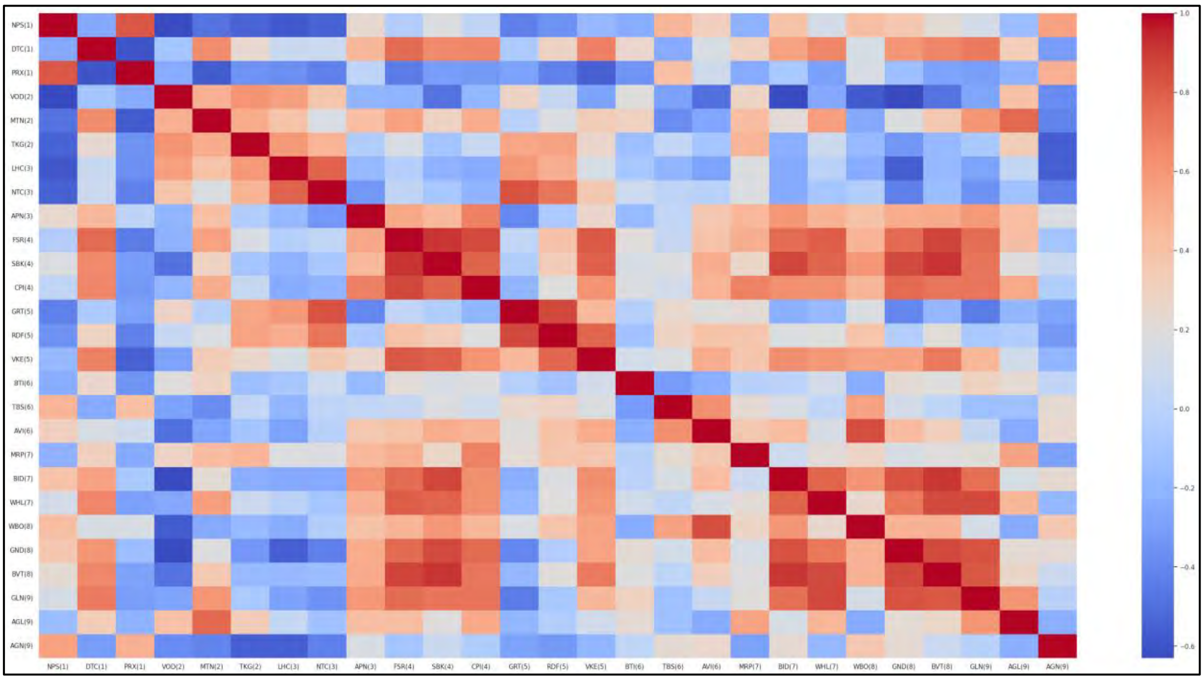
Symbol	Name	Sector	Sector Number
NPN	Naspers Ltd	Technology	1
DTC	Datatec Ltd	Technology	1
PRX	Prosus NV	Technology	1
VOD	Vodacom Group Ltd	Telecommunications	2
MTN	MTN Group Ltd	Telecommunications	2
TKG	Telkom SA SOC Ltd	Telecommunications	2
LHC	Life Healthcare Group Holdings Ltd	HealthCare	3
NTC	Netcare Ltd	HealthCare	3
APN	Aspen Pharmacare Holdings Limited	HealthCare	3
FSR	FirstRand Ltd	Financial Services	4
SBK	Standard Bank Group Ltd	Financial Services	4
CPI	Capitec Bank Holdings Ltd	Financial Services	4
GRT	Growthpoint Properties Ltd	Real Estate	5
RDF	Redefine Properties Ltd	Real Estate	5
VKE	Vukile Property Fund Ltd	Real Estate	5
MRP	Mr Price	Consumer Discretionary	6
BID	Bid Corporation Ltd	Consumer Discretionary	6
WHL	Woolworths Holdings Limited	Consumer Discretionary	6
BTI	British America Tobacco Plc Ads Common Stock	Consumer Staples	7
TBS	Tiger Brands Ltd	Consumer Staples	7
AVI	Avi Ltd	Consumer Staple	7
WBO	Wilson Bayly Holmes	Industrial	8
GND	Grindrod Ltd	Industrial	8
BVT	Bidvest Group	Industrial	8
GLN	Glencore PLC	Basic Materials	9
AGL	Anglo American PLC	Basic Materials	9
AGN	Anglogold Ashanti PLC	Basic Materials	9

Figure B1: Heatmap of Correlation of Sectors



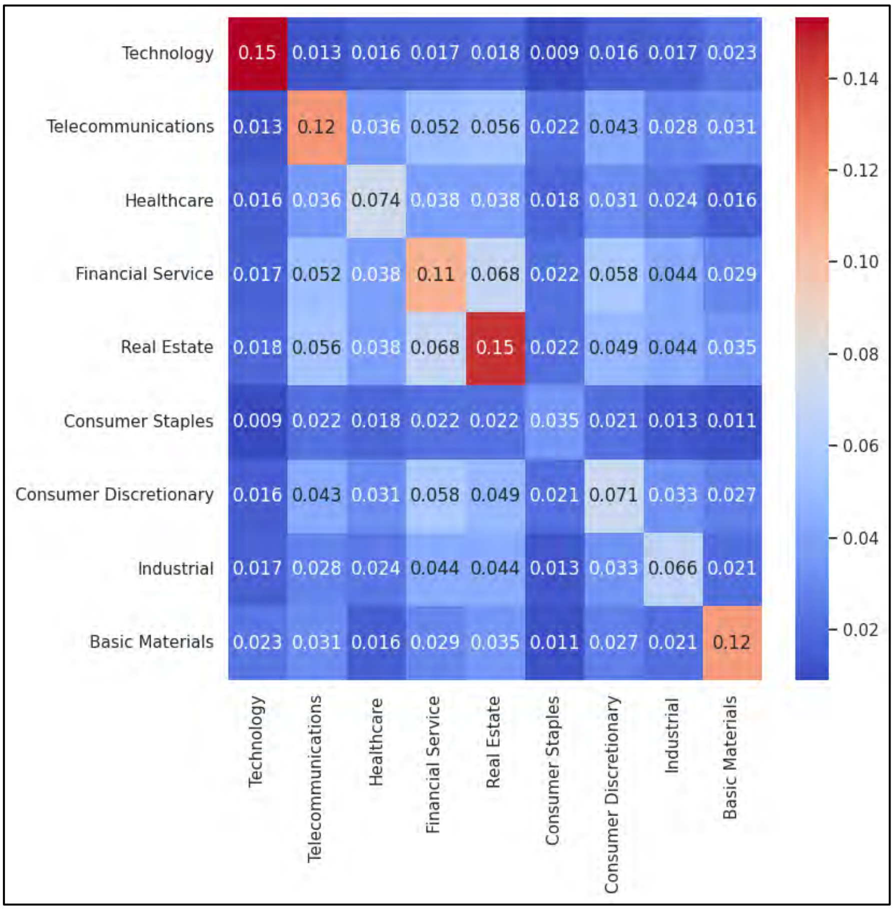
Source: Graphically represented by Author from Python results

Figure B2: Heatmap of Correlation of 27 Stocks



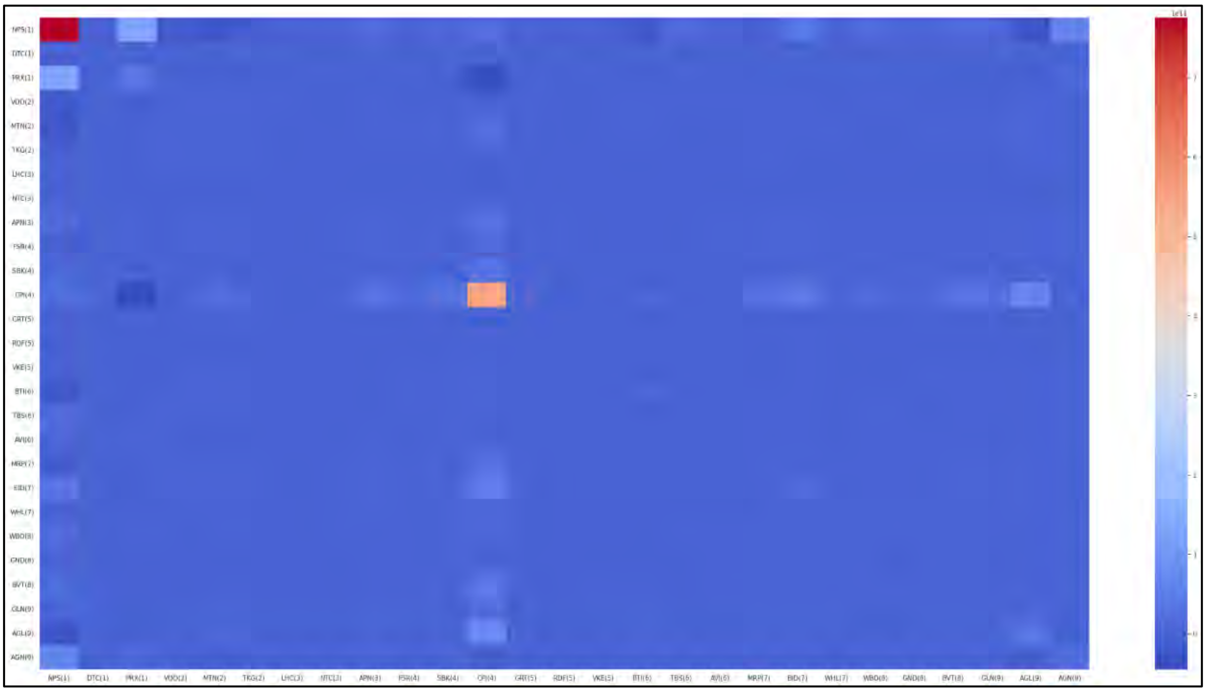
Source: Graphically represented by Author from Python results

Figure B3: Heatmap of Covariance of Sectors



Source: Graphically represented by Author from Python results

Figure B4: Heatmap of Covariance of 27 Stocks



Source: Graphically represented by Author from Python results

APPENDIX C

Traditional Methods

Table C1.1: Traditional Methods Results and Z-scores

Methods	Evaluation Metrics Results							
	Expected Returns	Expected Risk	Sharpe Ratio	Tracking Error	Sortino Ratio	Information Ratio	Beta	Treynor Ratio
Equally Weighted Portfolio	14.06%	19.02%	0.2791 (-0.16)	0.1095 (0.30)	1.2260 (0.57)	0.1955 (0.57)	0.2543 (0.82)	0.2088 (1.45)
Mean Variance Portfolio	15.06%	8.13%	0.7763 (0.76)	0.1196 (0.55)	1.1710 (0.54)	0.1070 (0.50)	-8.0042 (-0.85)	-0.0056 (-0.31)
Sharpe Ratio Portfolio	15.02%	6.89%	0.7723 (0.76)	0.1233 (0.64)	0.7926 (0.38)	0.0194 (0.42)	-8.1196 (-0.88)	-0.0042 (-0.30)
Risk Parity Portfolio	13.84%	17.28%	-0.3679 (-1.35)	0.0369 (-1.48)	-3.5884 (-1.49)	-2.2318 (-1.50)	0.7080 (0.91)	-0.0715 (-0.85)

Notes: Values in () are Z-scores and assume that the risk-free rate = 8.75%. Z-scores are computed by estimating the weighting of 20 variables of Sharpe ratio, Tracking Error, Sortino Ratio, Information Ratio, Beta and Treynor Ratio, respectively, which were grouped by the same unit of measurements.
Source: Estimated and complied by Author from Python results

Portfolio Allocation

Table C1.2: Equally Weighted Portfolio and Risk Parity Portfolio

Stock	Equal Weight	Risk Parity Weight	Stock	Equal Weight	Risk Parity Weight
NPS (1)	0.0370	0.0377	VKE (5)	0.0370	0.0288
DTC (1)	0.0370	0.0477	BTI (6)	0.0370	0.0867
PRX (1)	0.0370	0.0324	TBS (6)	0.0370	0.0489
VOD (2)	0.0370	0.0494	AVI (6)	0.0370	0.0498
MTN (2)	0.0370	0.0236	MRP (7)	0.0370	0.0294
TKG (2)	0.0370	0.0255	BID (7)	0.0370	0.0396
LHC (3)	0.0370	0.0388	WHL (7)	0.0370	0.0307
NTC (3)	0.0370	0.0370	WBO (8)	0.0370	0.0430
APN (3)	0.0370	0.0370	GND (8)	0.0370	0.0413
FSR (4)	0.0370	0.0275	BVT (8)	0.0370	0.0323
SBK (4)	0.0370	0.0277	GLN (9)	0.0370	0.0354
CPI (4)	0.0370	0.0266	AGL (9)	0.0370	0.0288
GRT (5)	0.0370	0.0291	AGN (9)	0.0370	0.0445
RDF (5)	0.0370	0.0205			

Source: Estimated and complied by Author from Python results

Table C1.3: Mean Variance Portfolio

Stock	Mean Variance Weight	Stock	Mean Variance Weight
DTC (1)	0.0390	RDF (5)	0.0333
PRX (1)	0.0124	VKE (5)	0.1811
VOD (2)	0.1049	BTI (6)	0.2336
LHC (3)	0.0403	AVI (6)	0.0498
NTC (3)	0.0251	WHL (7)	0.0354
APN (3)	0.0234	WBO (8)	0.0056

SBK (4)	0.0421	GND (8)	0.0151
GRT (5)	0.0778	AGN (9)	0.0810

Source: Estimated and complied by Author from Python results

Table C1.4: Sharpe Ratio Portfolio

Stock	Sharpe Ratio Weight	Stock	Sharpe Ratio Weight
DTC (1)	0.0390	RDF (5)	0.0649
PRX (1)	0.0123	VKE (5)	0.0022
VOD (2)	0.1044	BTI (6)	0.0405
TKG (2)	0.0009	AVI (6)	0.0781
LHC (3)	0.0404	WHL (7)	0.0390
NTC (3)	0.0249	WBO (8)	0.0048
APN (3)	0.0235	GND (8)	0.0150
SBK (4)	0.0418	ANG (9)	0.0810
GRT (5)	0.0888		

Source: Estimated and complied by Author from Python results

Linear Regression Method

Table C2.1: Linear Regression Results and Z score

Methods	Evaluation Metrics Results							
	Expected Returns	Expected Risk	Sharpe Ratio	Tracking Error	Sortino Ratio	Information Ratio	Beta	Treynor Ratio
Equally Weighted Portfolio	25.63%	16.55%	1.0204 (-0.41)	0.5089 (1.15)	2.3589 (0.90)	0.3285 (0.70)	0.5630 (0.28)	0.2999 (1.01)
Mean Variance Portfolio	19.98%	8.43%	2.3691 (1.13)	0.5615 (1.41)	3.2900 (1.43)	0.3374 (0.71)	-2.6966 (-0.63)	-0.0709 (-0.09)
Sharpe Ratio Portfolio	19.85%	8.43%	2.3543 (1.11)	0.5532 (1.37)	2.9925 (1.26)	0.3027 (0.66)	-0.8690 (-0.12)	-0.1946 (-0.46)
Risk Parity Portfolio	15.85%	12.57%	1.2613 (-0.13)	0.0408 (-1.17)	0.0628 (-0.39)	-0.0002 (0.15)	0.0136 (0.13)	0.1215 (0.48)

Notes: Values in () are Z-scores and assume that the risk-free rate = 8.75%. Z-scores are computed by estimating the weighting of 20 variables of Sharpe ratio, Tracking Error, Sortino Ratio, Information Ratio, Beta and Treynor Ratio, respectively, which were grouped by the same unit of measurements.
Source: Estimated and complied by Author

Portfolio Allocation

Table C2.2: Equally Weighted Portfolio and Risk Parity Portfolio

Stock	Equal Weight	Risk Parity Weight	Stock	Equal Weight	Risk Parity Weight
NPS (1)	0.0370	0.0542	VKE (5)	0.0370	0.0607
DTC (1)	0.0370	0.0177	BTI (6)	0.0370	0.0539
PRX (1)	0.0370	0.0278	TBS (6)	0.0370	0.0107
VOD (2)	0.0370	0.0251	AVI (6)	0.0370	0.0446
MTN (2)	0.0370	0.0424	MRP (7)	0.0370	0.0450
TKG (2)	0.0370	0.0274	BID (7)	0.0370	0.0476
LHC (3)	0.0370	0.0509	WHL (7)	0.0370	0.0273
NTC (3)	0.0370	0.0555	WBO (8)	0.0370	0.0355
APN (3)	0.0370	0.0356	GND (8)	0.0370	0.0389

FSR (4)	0.0370	0.0349	BVT (8)	0.0370	0.0416
SBK (4)	0.0370	0.0339	GLN (9)	0.0370	0.0230
CPI (4)	0.0370	0.0329	AGL (9)	0.0370	0.0391
GRT (5)	0.0370	0.0460	AGN (9)	0.0370	0.0122
RDF (5)	0.0370	0.0367			

Source: Estimated and complied by Author from Python results

Table C2.3: Mean Variance Portfolio

Stock	Mean Variance Weight	Stock	Mean Variance Weight
DTC (1)	0.0292	VKE (5)	0.2168
PRX (1)	0.0288	AVI (6)	0.0104
VOD (2)	0.0228	BTI (6)	0.1876
LHC (3)	0.0119	TBS (6)	0.0403
NTC (3)	0.0891	WHL (7)	0.0401
APN (3)	0.0545	WBO (8)	0.0062
CPI (4)	0.0140	AGL (9)	0.0034
GRT (5)	0.1169	AGN (9)	0.0920
RDF (5)	0.0359		

Source: Estimated and complied by Author from Python results

Table C2.4: Sharpe Ratio Portfolio

Stock	Mean Variance Weight	Stock	Mean Variance Weight
DTC (1)	0.0291	VKE (5)	0.2162
PRX (1)	0.0279	AVI (6)	0.0102
VOD (2)	0.0224	BTI (6)	0.1873
LHC (3)	0.0117	TBS (6)	0.0403
NTC (3)	0.0893	WHL (7)	0.0403
APN (3)	0.0541	MRP (7)	0.0032
CPI (4)	0.0118	WBO (8)	0.0066
GRT (5)	0.1167	AGL (9)	0.0033
RDF (5)	0.0372	AGN (9)	0.0922

Source: Estimated and complied by Author from Python results

<u>Decision Tree Regression Method</u>

Table C3.1: Decision Tree Regression Results and Z score

Methods	Evaluation Metrics Results							
	Expected Returns	Expected Risk	Sharpe Ratio	Tracking Error	Sortino Ratio	Information Ratio	Beta	Treynor Ratio
Equally Weighted Portfolio	11.51%	12.09%	0.2281 (-1.31)	0.2539 (-0.11)	0.3646 (-0.22)	0.1021 (0.32)	0.3347 (0.22)	0.0824 (0.37)
Mean Variance Portfolio	9.52%	6.70%	1.4208 (0.05)	0.1999 (-0.38)	0.1964 (-0.32)	0.0597 (0.25)	4.1647 (1.29)	0.0033 (0.13)
Sharpe Ratio Portfolio	9.52%	6.70%	1.4208 (0.05)	0.1999 (-0.38)	0.1964 (-0.32)	0.0597 (0.25)	4.1647 (1.29)	0.0033 (0.13)
Risk Parity Portfolio	8.98%	9.65%	0.9308 (-0.51)	0.0262 (-1.24)	-0.9031 (-0.94)	-0.8947 (-1.33)	0.0309 (0.14)	-0.7051 (-1.97)

Notes: Values in () are Z-scores and assume that the risk-free rate = 8.75%. Z-scores are computed by estimating the weighting of 20 variables of Sharpe ratio, Tracking Error, Sortino Ratio, Information Ratio, Beta and Treynor Ratio, respectively, which were grouped by the same unit of measurements.

Source: Estimated and complied by Author

Portfolio Allocation

Table C3.2: Equally Weighted Portfolio and Risk Parity Portfolio

Stock	**Equal Weight**	**Risk Parity Weight**	**Stock**	**Equal Weight**	**Risk Parity Weight**
NPS (1)	0.0370	0.0374	VKE (5)	0.0370	0.0418
DTC (1)	0.0370	0.0191	BTI (6)	0.0370	0.0497
PRX (1)	0.0370	0.0223	TBS (6)	0.0370	0.0674
VOD (2)	0.0370	0.0224	AVI (6)	0.0370	0.0255
MTN (2)	0.0370	0.0317	MRP (7)	0.0370	0.0748
TKG (2)	0.0370	0.0201	BID (7)	0.0370	0.0537
LHC (3)	0.0370	0.0419	WHL (7)	0.0370	0.0286
NTC (3)	0.0370	0.0248	WBO (8)	0.0370	0.1038
APN (3)	0.0370	0.0279	GND (8)	0.0370	0.0632
FSR (4)	0.0370	0.0265	BVT (8)	0.0370	0.0233
SBK (4)	0.0370	0.0282	GLN (9)	0.0370	0.0308
CPI (4)	0.0370	0.0262	AGL (9)	0.0370	0.0263
GRT (5)	0.0370	0.0267	AGN (9)	0.0370	0.0208
RDF (5)	0.0370	0.0350			

Source: Estimated and complied by Author from Python results

Table C3.3: Mean Variance Portfolio

Stock	**Mean Variance Weight**	**Stock**	**Mean Variance Weight**
DTC (1)	0.0237	WHL (7)	0.1446
PRX (1)	0.0115	BID (7)	0.1167
NTC (3)	0.0209	MRP (7)	0.0291
APN (3)	0.0125	GND (8)	0.0168
RDF (5)	0.0737	BVT (8)	0.1438
GRT (5)	0.0194	GLN (9)	0.0193
VKE (5)	0.0725	AGN (9)	0.0062
AVI (6)	0.0449	AGL (9)	0.0061
BTI (6)	0.0874		

Source: Estimated and complied by Author from Python results

Table C3.4: Sharpe Ratio Portfolio

Stock	**Mean Variance Weight**	**Stock**	**Mean Variance Weight**
DTC (1)	0.0237	WHL (7)	0.1446
PRX (1)	0.0115	BID (7)	0.1167
NTC (3)	0.0209	MRP (7)	0.0291
APN (3)	0.0125	GND (8)	0.0168
RDF (5)	0.0737	BVT (8)	0.1438
GRT (5)	0.0194	GLN (9)	0.0193
VKE (5)	0.0725	AGN (9)	0.0062
AVI (6)	0.0449	AGL (9)	0.0061
BTI (6)	0.0874		

Source: Estimated and complied by Author from Python results

Random Forest Regression Model

Table C4.1: Random Forest Regression Results and Z score

Methods	Evaluation Metrics Results							
	Expected Returns	**Expected Risk**	**Sharpe Ratio**	**Tracking Error**	**Sortino Ratio**	**Information Ratio**	**Beta**	**Treynor Ratio**
Equally Weighted Portfolio	6.00%	11.26%	-0.2446 (-1.84)	0.2319 (0.22)	-0.2677 (0.58)	-0.0519 (0.07)	0.2501 (0.20)	-0.1102 (-0.21)
Mean Variance Portfolio	3.93%	3.17%	1.2419 (-0.15)	0.1507 (-0.62)	-0.6525 (0.80)	-0.3690 (0.46)	-1.2145 (-0.21)	0.0586 (0.30)
Sharpe Ratio Portfolio	10.00%	3.29%	3.0437 (1.89)	0.2003 (-0.38)	-0.3895 (-0.65)	-0.1296 (-0.06)	-1.0209 (-0.16)	0.0406 (0.24)
Risk Parity Portfolio	4.08%	8.32%	0.4904 (-1.01)	0.0199 (-1.27)	-2.7741 (-1.99)	-1.9963 (-3.16)	0.0685 (0.15)	-0.8077 (-2.28)

Notes: Values in () are Z-scores and assume that the risk-free rate = 8.75%. Z-scores are computed by estimating the weighting of 20 variables of Sharpe ratio, Tracking Error, Sortino Ratio, Information Ratio, Beta and Treynor Ratio, respectively, which were grouped by the same unit of measurements.
Source: Estimated and complied by Author

Portfolio Allocation

Table C4.2: Equally Weighted Portfolio and Risk Parity Portfolio

Stock	**Equal Weight**	**Risk Parity Weight**	**Stock**	**Equal Weight**	**Risk Parity Weight**
NPS (1)	0.0370	0.0405	VKE (5)	0.0370	0.0334
DTC (1)	0.0370	0.0166	BTI (6)	0.0370	0.0410
PRX (1)	0.0370	0.0171	TBS (6)	0.0370	0.0634
VOD (2)	0.0370	0.0203	AVI (6)	0.0370	0.0304
MTN (2)	0.0370	0.0181	MRP (7)	0.0370	0.0235
TKG (2)	0.0370	0.0293	BID (7)	0.0370	0.0657
LHC (3)	0.0370	0.0234	WHL (7)	0.0370	0.0502
NTC (3)	0.0370	0.0294	WBO (8)	0.0370	0.0601
APN (3)	0.0370	0.0398	GND (8)	0.0370	0.1694
FSR (4)	0.0370	0.0242	BVT (8)	0.0370	0.0255
SBK (4)	0.0370	0.0239	GLN (9)	0.0370	0.0264
CPI (4)	0.0370	0.0226	AGL (9)	0.0370	0.0249
GRT (5)	0.0370	0.0384	AGN (9)	0.0370	0.0169
RDF (5)	0.0370	0.0255			

Source: Estimated and complied by Author from Python results

Table C4.3: Mean Variance Portfolio

Stock	**Mean Variance Weight**	**Stock**	**Mean Variance Weight**
DTC (1)	0.0049	WHL (7)	0.1079
TKG (2)	0.0064	BID (7)	0.0263
NTC (3)	0.0102	BVT (8)	0.1015
VKE (5)	0.0308	GND (8)	0.6173
GRT (5)	0.0159	GLN (9)	0.0106
AVI (6)	0.0106	AGN (9)	0.0094
BTI (6)	0.0482		

Source: Estimated and complied by Author from Python results

Table C4.4: Sharpe Ratio Portfolio

Stock	Sharpe Ratio Weight	Stock	Sharpe Ratio Weight
DTC (1)	0.0079	AVI (6)	0.0033
PRX (1)	0.0072	BTI (6)	0.0453
TKG (2)	0.0037	WHL (7)	0.1190
VOD (2)	0.0057	BID (7)	0.0338
APN (3)	0.0079	BVT (8)	0.1021
NTC (3)	0.0148	GND (8)	0.5868
FSR (4)	0.0021	GLN (9)	0.0092
VKE (5)	0.0193	AGN (9)	0.0248
GRT (5)	0.0071		

Source: Estimated and complied by Author from Python results

Sector Vector Machine Regression Model

Table C5.1: Sector Vector Machine Regression Results and Z score

Methods	Evaluation Metrics Results							
	Expected Returns	Expected Risk	Sharpe Ratio	Tracking Error	Sortino Ratio	Information Ratio	Beta	Treynor Ratio
Equally Weighted Portfolio	21.24%	15.06%	0.8291 (-0.62)	0.4414 (0.82)	1.7744 (0.57)	0.2791 (0.62)	4.7836 (1.46)	0.0261 (0.20)
Mean Variance Portfolio	18.26%	8.07%	2.2630 (1.01)	0.5085 (1.15)	3.2062 (1.38)	0.3412 (0.72)	-8.7882 (-2.33)	-0.0199 (0.06)
Sharpe Ratio Portfolio	17.76%	8.06%	2.2024 (0.94)	0.4811 (1.02)	2.8755 (1.19)	0.3015 (0.66)	-7.0576 (-1.84)	-0.0208 (0.06)
Risk Parity Portfolio	14.39%	11.92%	1.2075 (-0.19)	0.0383 (-1.18)	-0.1640 (-0.52)	-0.1602 (-0.11)	-0.0071 (0.13)	0.6346 (2.01)

Notes: Values in () are Z-scores and assume that the risk-free rate = 8.75%. Z-scores are computed by estimating the weighting of 20 variables of Sharpe ratio, Tracking Error, Sortino Ratio, Information Ratio, Beta and Treynor Ratio, respectively, which were grouped by the same unit of measurements.
Source: Estimated and complied by Author

Portfolio Allocation

Table C5.2: Equally Weighted Portfolio and Risk Parity Portfolio

Stock	Equal Weight	Risk Parity Weight	Stock	Equal Weight	Risk Parity Weight
NPS (1)	0.0370	0.0169	VKE (5)	0.0370	0.0360
DTC (1)	0.0370	0.0260	BTI (6)	0.0370	0.0516
PRX (1)	0.0370	0.0516	TBS (6)	0.0370	0.0111
VOD (2)	0.0370	0.0281	AVI (6)	0.0370	0.0444
MTN (2)	0.0370	0.0243	MRP (7)	0.0370	0.0279
TKG (2)	0.0370	0.0401	BID (7)	0.0370	0.0482
LHC (3)	0.0370	0.0480	WHL (7)	0.0370	0.0486
NTC (3)	0.0370	0.0517	WBO (8)	0.0370	0.0428
APN (3)	0.0370	0.0357	GND (8)	0.0370	0.0387
FSR (4)	0.0370	0.0354	BVT (8)	0.0370	0.0453
SBK (4)	0.0370	0.0326	GLN (9)	0.0370	0.0379
CPI (4)	0.0370	0.0334	AGL (9)	0.0370	0.0184
GRT (5)	0.0370	0.0565	AGN (9)	0.0370	0.0235
RDF (5)	0.0370	0.0453			

Source: Estimated and complied by Author from Python results

Table C5.3: Mean Variance Portfolio

Stock	Mean Variance Weight	Stock	Mean Variance Weight
PRX (1)	0.0303	VKE (5)	02225
DTC (1)	0.0247	GRT (5)	0.0622
VOD (2)	0.0294	RDF (5)	0.0579
LHC (3)	0.0163	BTI (6)	0.2017
NTC (3)	0.0544	TBS (6)	0.0416
APN (3)	0.0517	WHL (7)	0.0545
CPI (4)	0.0135	GND (8)	0.0368
SBK (4)	0.0099	AGN (9)	0.0927

Source: Estimated and complied by Author from Python results

Table C5.4: Sharpe Ratio Portfolio

Stock	Mean Variance Weight	Stock	Mean Variance Weight
PRX (1)	0.0292	RDF (5)	0.0572
DTC (1)	0.0240	BTI (6)	0.1997
VOD (2)	0.0283	TBS (6)	0.0414
LHC (3)	0.0156	MRP (7)	0.0013
NTC (3)	0.0563	WHL (7)	0.0548
APN (3)	0.0513	WHO (8)	0.0004
CPI (4)	0.0122	GND (8)	0.0368
SBK (4)	0.0087	AGN (9)	0.0912
VKE (5)	0.2233	AGL (9)	0.0030
GRT (5)	0.0648		

Source: Estimated and complied by Author from Python results

APPENDIX D

TRADITIONAL METHODS

Best Portfolio

Table D1.1: Expected Return and Risks of Assets

Stock	Expected Returns (%)	Expected Risk (%)	Stock	Expected Returns (%)	Expected Risk (%)
NPS (1)	0.8994	0.6052	VKE (5)	0.3034	0.8761
DTC (1)	0.5599	0.4383	BTI (6)	0.2076	0.2025
PRX (1)	1.2378	0.7094	TBS (6)	0.2506	0.4563
VOD (2)	0.0524	0.4672	AVI (6)	0.3529	0.4531
MTN (2)	0.3823	1.0546	MRP (7)	0.6986	0.8357
TKG (2)	-0.1666	0.9976	BID (7)	0.4280	0.5929
LHC (3)	0.0159	0.5976	WHL (7)	0.3919	0.7963
NTC (3)	0.1000	0.6400	WBO (8)	0.7329	0.5327
APN (3)	0.9463	0.6243	GND (8)	1.4358	0.5666
FSR (4)	0.4530	0.9056	BVT (8)	0.5403	0.7536
SBK (4)	0.4919	0.9037	GLN (9)	0.9281	0.6650
CPI (4)	1.0890	0.9350	AGL (9)	0.7354	0.8295
GRT (5)	-0.1554	0.8536	AGN (9)	0.9423	0.4694
RDF (5)	0.2047	1.2538			

Source: Estimated and complied by Author from Python results

Worst Portfolio

Table D1.2: Expected Returns and Risk of Assets

Stock	Expected Returns (%)	Expected Risk (%)	Stock	Expected Returns (%)	Expected Risk (%)
NPS (1)	0.9146	0.6401	VKE (5)	0.2356	0.6400
DTC (1)	0.7210	0.6400	BTI (6)	0.4857	0.6390
PRX (1)	1.0830	0.6416	TBS (6)	0.3306	0.6403
VOD (2)	0.0700	0.6396	AVI (6)	0.4746	0.6399
MTN (2)	0.2437	0.6403	MRP (7)	0.5550	0.6408
TKG (2)	-0.1148	0.6400	BID (7)	0.4586	0.6406
LHC (3)	0.0166	0.6395	WHL (7)	0.3246	0.6409
NTC (3)	0.0999	0.6396	WBO (8)	0.8504	0.6404
APN (3)	0.9463	0.6395	GND (8)	1.6029	0.6399
FSR (4)	0.3358	0.6397	BVT (8)	0.4710	0.6401
SBK (4)	0.3678	0.6397	GLN (9)	0.8879	0.6395
CPI (4)	0.7833	0.6406	AGL (9)	0.5724	0.6398
GRT (5)	-0.1222	0.6404	AGN (9)	1.1339	0.6396
RDF (5)	0.1135	0.6410			

Source: Estimated and complied by Author from Python results

MACHINE LEARNING METHODS

Best Portfolio

Table D2.1: Expected Returns and Risk of Assets

Stock	Expected Returns (%)	Expected Risk (%)	Stock	Expected Returns (%)	Expected Risk (%)
DTC (1)	0.3919	-0.0803	VKE (5)	-1.1567	0.7345
PRX (1)	3.8788	0.2698	AVI (6)	0.1470	0.0220
VOD (2)	-0.3561	0.0487	BTI (6)	38.9272	38.4378
LHC (3)	0.1341	0.0164	TBS (6)	0.5707	0.0594
NTC (3)	-0.6827	0.2276	WHL (7)	0.9957	0.1394
APN (3)	1.3358	0.2759	WBO (8)	0.2672	0.0142
CPI (4)	-0.0812	0.0406	AGL (9)	-0.0029	0.0143
GRT (5)	-1.7385	0.3366	AGN (9)	3.2965	0.0621
RDF (5)	0.0218	0.1336			

Source: Estimated and complied by Author from Python results

Worst Portfolio

Table D2.2: Expected Returns and Risk of Assets

Stock	Expected Returns (%)	Expected Risk (%)	Stock	Expected Returns (%)	Expected Risk (%)
NPS (1)	1.0954	0.3066	VKE (5)	-0.2472	0.3082
DTC (1)	0.2342	0.3084	TBS (6)	0.3302	0.3079
PRX (1)	0.9778	0.3087	BTI (6)	-0.6661	0.3079
VOD (2)	-0.5216	0.3079	AVI (6)	04773	0.3076
MTN (2)	-0.3445	0.3074	MRP (7)	-0.4621	0.3072
TKG (2)	-0.5008	0.3086	WHL (7)	0.9174	0.3086
APN (3)	0.4644	0.3076	BID (7)	0.4295	0.3082
NTC (3)	-0.3467	0.3079	BVT (8)	0.3551	0.3077
LHC (3)	0.3136	0.3086	WBO (8)	1.1869	0.3084
SBK (4)	0.3608	0.3080	GND (8)	0.0287	0.3089
CPI (4)	-0.2010	0.3079	AGL (9)	-0.2918	0.3086
FSR (4)	-0.0901	0.3076	AGN (9)	0.9005	0.3077
GRT (5)	-0.4488	0.3092	GLN (9)	0.1566	0.3079
RDF (5)	-0.0290	0.3080			

Source: Estimated and complied by Author from Python results