Author Name: Molokwane Mpho Jacqueline

Student no: 0706756R

Master of Management in Finance & Investments Research

Name of Supervisor: Dr Jones Odei-Mensah

2024

Return Forecasting, Risk Modelling & Portfolio Construction in the Banking Industry of South Africa

Abstract

The integration of return forecasting, risk modelling, and portfolio construction is crucial for developing robust investment strategies. This research investigates the benefits of using advanced machine learning techniques, specifically Long Short-Term Memory (LSTM) and Support Vector Machine models, in predicting stock prices. Financial markets are dynamic and complex in nature and applying more advanced techniques yields more benefits on the predictive accuracy of models when using historical data. An essential part of these research is building multiple models with various statistical techniques and machine learning techniques to compare their performance. The best performing model based on selected performance metrics was chosen and used to predict future stock prices for portfolio construction. The findings of this research offer insights into the practical applications of advanced machine learning techniques for financial forecasting. By using advanced machine learning in return forecasting, we can assess associated risks comprehensively, and construct well-diversified portfolios which helps investors achieve superior performance and resilience in various market conditions. As financial markets evolve, ongoing research and development in these areas remains essential for achieving optimal investment outcomes and managing financial risks effectively.

Keywords: SA Banking Sector, Stock Forecasting, Machine Learning, LSTM, Portfolio Theory, Risk Modeling, Portfolio Design, Optimization.

Plagiarism Declaration

I **Mpho J. Molokwane** (Student Number: **0706756R**) am a student registered for Master of Management in Finance and Investments for 2024, hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above course is my own unaided work except where I have explicitly indicated otherwise.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

Signature: ____Mpho Jacqueline Molokwane__

Date: _____2021/05/30___________________

Acknowledgements

I am deeply thankful to everyone who supported me throughout the course of my master's studies, especially my family. I am thankful for the aspiring guidance, invaluably constructive criticism, encouragement, and friendly advice during this period. I am sincerely grateful to them for sharing their truthful and illuminating views on several issues related to my studies.

My profound appreciation goes to

1. Dr Jones Odei-Mensah for the help, support, and guidance throughout the duration of my Project.
2. My brother, Kgaogelo Gregory Molokwane for his immense encouragement, motivation and his highest believe in me.

Table of Contents

List of tables

List of figures

CHAPTER ONE: GENERAL INTRODUCTION

1.1 Introduction

The South African banking industry is crucial to the country's economy by providing financial services to individuals and businesses (Mafini, 2020). Over the years, the industry has undergone significant changes, including deregulation, consolidation, and technological advancements. These changes have brought new challenges and opportunities for the industry and investors. In this context, it is essential to understand the current situation and trends in the industry to make informed investment decisions and manage risks effectively. The South African banking industry is complex and has volatile macroeconomic factors that influence the economy's growth, and the economy has deteriorated over the past three years.

The banking sector show some impressive growth in 2022 against the backdrop of the COVID-19 pandemic. The key statistics such as combined headline growth was 16.1%, up from 2021 to a record of R100bn, ROE was 17.1% in 2022, up 2% from 15.9% in 2021, the net interest margin was 420 basis points (bps) against 410 basis points in 2021. The industry reported a credit loss ratio of 80 bps, up from 76 bps in 2021. The cost-to-income ratio declined by 53.6%, compared to 55.8% in 2021. The common equity tier ratio was 13.5%, compared to 13.4% in 2021 (PwC, 2023). The Non-Performing Loan (NPL) datasets show that a ratio of 5.2% was reported in June 2023, up by 1% from the previous month, May 2023. The highest reported report was 6.0% in November 2009, and the lowest was 2.1% in 2008 (SARB, 2023).

Table 1.1.1: South African Banking Industry Performance

	2022	2023
Profitability		
Return on Equity	14.48	14.9
Return on Assets	1.12	1.1
Cost to Income Ratio	57.72	56.68
Net Interest Revenue to Interest Earning Assets	3.91	4.09
Non-Interest Revenue to Total Assets	2.02	2.01
Operating Expenses to Total Assets	2.84	2.88
Profit/Loss (12Month) Rbn	101.74	104.8
Net Interest Income (12 Months) Rbn	201.36	228.79
Non-Interest income (12 Month) Rbn	139.96	149.57
Operating Expenses	196.99	214.44
Liquidity		
Liquid Assets held to liquid assets requirement	331.78	339.98
Short-term liabilities to total liabilities	59.58	58.73
Ten largest depositors to total revenue funding	21.19	18.77
Liquidity coverage ratio (%)	143.52	147.99
Net stable funding ratio (%)	116.76	117.49
Credit Risk		
Impaired advances (Rbn)	239	292.62
Impaired advances to gross loan and advances (Rbn)	4.63	5.47
Specific credit impairment to impaired advances	48.14	45.5
Portfolio credit impairment to gross loans and advances	1.2	1.14
Capital adequacy		
Total capital adequacy (%)	17.38	17.38
Tier 1 (T1) capital adequacy (%)	14.67	14.98
Common equity T1 capital adequacy (%)	13.43	13.3
Basel III Leverage Ratio		
Leverage ratio	6.47	6.57

South African Banking Industry Performance; Source (South African Reserve Bank 2023).
Table 1.1.1 presents the performance of the South African banking industry for October 2023. This period is vital because of the impact of the pandemic on the economy. The results show a significant improvement in the banking sector despite the pandemic.

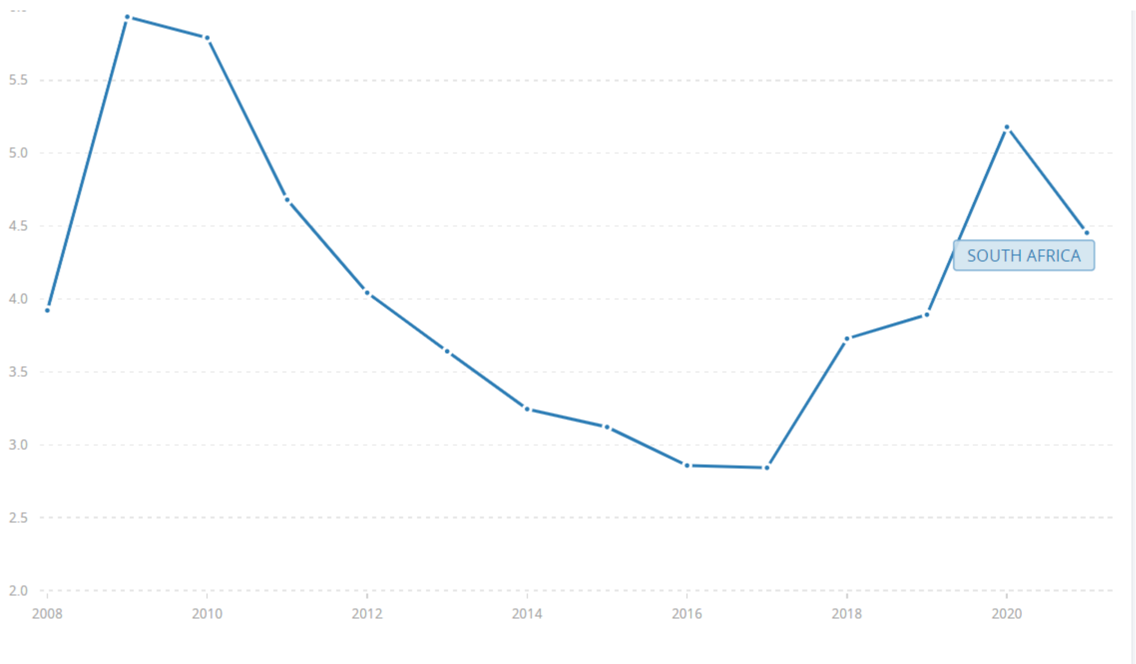


Figure 1.1.1: Growth Trends in South Africa Source: World Bank (2023)

Figure 1.1.1 shows the growth trend in South Africa. Some of the critical challenges to the South African economy include undependable electricity supply from Eskom, capital outflow, inflation of 7.7%, high interest rates with a repo rate of 7.75%, threats of sanctions, especially from the West because of the support to Russia, repricing of government debts, liquidity issues, greylisting, slow economic growth, and cyber-attack risk (SARB, 2023). From a historical perspective, the South African banking industry has undergone significant transformations. The industry was heavily regulated during the apartheid era, with limited access to credit for black South Africans (Mafini, 2020). The country's democratization in 1994 led to significant changes, including the deregulation of industry, the entry of new players, and increased competition. The industry has also undergone consolidation, with the top five banks accounting for over 80% of the sector's assets (SARB, 2021).

Despite the industry's growth and profitability, there is a need for investors and institutions to make informed decisions based on accurate forecasting and risk management. Kumbure et al., (2022) posit that there is a need for modern approaches to support decision-making in the industry. According to Milojević and Redzepagic (2021) introducing AI and ML potential and increasing the speed and agility of stock prediction, given real-world datasets, enable decision-makers and policy analysts to make quicker decisions, improve financial market performance, and attract new investors and players to address economic problems. Yadav and Khera (2023) assert that ARIMA is a valuable tool for predicting stock returns in emerging markets. Investors would require good returns on their investments. Machine Learning (ML) and ARIMA models are some of the tools at the disposal of investors. Both models used to the best of their capabilities may bring significant value to investors. Hybrid models are beneficial for stock return forecasts. This relatively innovative approach is relevant in the South African financial market. Additionally, research is needed in these areas to support scholars.

A review of recent literature shows that in South Africa, traditional methods are still widely used to study the stock market performance despite new technology supporting decision-making (Kumbure et al., 2022). Machine Learning has become essential in studying quantitative finance, and its application in South African financial research must be encouraged. The SA Banking and Financial institutions require a novel approach to risk management, stock prediction and forecasting, and credit risk assessment. Given the nature of the unique complexities of the SA banking industry, this thesis argues that the introduction of ML in the banking industry will significantly impact the sector's performance. Moreover, the ML approach has exceptional capabilities.

The banking sector has been recognizing the potential benefits of artificial intelligence (AI) and machine learning (ML) in recent years. AI and ML can be used to analyse vast amounts of data, identify patterns and trends, and make predictions with a high level of accuracy. This can help banks to improve decision-making, enhance customer experience, reduce risks, and increase efficiency. For example, banks can use AI and ML to detect fraud more effectively, personalize their services for customers, and automate tasks such as credit scoring and loan processing. Additionally, AI and ML can help banks to comply with complex regulations and to monitor and manage risks more efficiently. Overall, the application of AI and ML can help banks stay competitive, improve their bottom line, and provide better services to their customers. The theoretical postulates that underpin this study include modern portfolio theory, which emphasises the importance of diversification and efficient frontier analysis to achieve maximum returns with minimum risks (Markowitz, 1952; Michaud, 1989). By using advanced techniques for return forecasting, risk modelling, and portfolio construction, investors and institutions can achieve better outcomes in the South African banking industry.

This study aims to bridge the gap using techniques such as autoregressive integrated moving averages (ARIMA) and Machine Learning techniques such as Neural Networks to build robust financial models to forecast stock returns and manage risks effectively.

According to Sardar et al., 2023) Machine Learning models have gained widespread bacceptance in the financial industry. Several models, such as predictive modelling, have been used where machine learning algorithms are used to predict financial markets, such as stock price movements, currency exchange rates, or credit default probabilities. In addition, the risk assessment framework indicates that ML models can assess financial risk, including credit risk in lending, market risk in investment portfolios, and fraud detection in transactions. Furthermore, algorithmic ML trading supports several hedge funds and financial institutions using machine learning in their algorithmic trading strategies, where models analyse market data to make trading decisions (Schmitt, 2023).

The Efficient Market Hypothesis (EMH) and machine learning are two concepts in finance that have diverse viewpoints on how financial markets function and how investment decisions should be made. Moreover, the Efficient Market Hypothesis is a theory that proposes that financial markets efficiently integrate all available datasets and information into asset prices. It posits that it is difficult to reliably achieve above-average returns by analysing historical prices or other publicly existing information because such information is previously replicated in asset prices (Xin & Xin, 2023).

According to Laiyao (2023), Strong Form EMH and Insider Information indicate that even insider information cannot be used to manipulate the market consistently. Machine learning has the ability, in the context of insider trading, to detect unusual trading patterns or anomalies that may raise suspicion of insider trading to avoid market manipulation. In addition, applying ML in behavioural finance theory is imperative because BFT and ML are two distinct approaches within finance that can be combined to gain insights into investor behaviour, market anomalies, and decision-making processes. The motivation for using ML in the banking sector in South Africa is critical because ML can complement traditional models to support investors in decision-making. Moreover, behavioural finance is a field that explores how psychological factors and cognitive biases influence the behaviour of investors and financial markets. It departs from the traditional finance assumption that investors are perfectly rational and that markets are efficient (Zhao & Yang, 2023).

ML and portfolio optimisation have a relationship in which ML can optimise portfolios whilst considering investor behavioural factors, such as loss aversion or herding behaviour, to build portfolios that align with investors' psychological profiles (Wu et al., 2023). Integrating machine learning techniques with behavioural finance theory allows a deeper understanding of investor behaviour, market anomalies, and decision-making processes. It enables investors and financial professionals to make more informed investment choices and develop strategies that account for the impact of psychological factors on financial markets (Zhao & Yang, 2023).

The Banking industry in South Africa can improve performance by deploying ML. This is because machine learning proffers several advantages in return forecasting; firstly, feature selection, where ML models can automatically identify relevant features (predictors) from a wide range of financial, economic, and alternative data sources. This is particularly valuable when dealing with large datasets. Secondly, in nonlinear relationships, where financial data often exhibits nonlinear patterns and dependencies, Machine learning algorithms, including decision trees, random forests, and neural networks, can capture these nonlinear relationships, potentially leading to more accurate predictions. Thirdly, using time series analysis, where ML can manage time series data effectively, makes it suitable for modelling and forecasting financial time series, where past values are significant predictors of future returns. Fourthly, ensemble methods like bagging and boosting can improve forecast accuracy by combining multiple models, each trained on different subsets of data or with different hyperparameters. Fifthly, deep learning techniques (RNNs) and (LSTMs) can capture complex temporal dependencies in financial data, therefore being compatible with time series forecasting (Zaheer et al., 2023).

According to Sithole et al, (2025), the banking and financial industry in South Africa is behind in using modern financial instruments and techniques such as machine learning to manage risk and make decisions. The problem can be resolved using machine learning to complement traditional models, which have been at the forefront of industry practice for several decades. In addition, Gera et al, (2024) assert that scaling up AI in the banking industry is important in emerging markets to synchronize with other global economies. Although efficient markets, modern portfolios, behavioural finance, and capital asset pricing model theories have been utilised, ML is needed to help in return forecasting, risk modelling, portfolio construction, and portfolio optimisation. Moreover, using this novice instrument can significantly improve decision-making for investors and fund managers to ensure that the benefits accrue to all stakeholders (Wang et al., 2021). Emerson et al. (2019) highlight machine learning trends as an approach to prediction in the financial sector, especially regarding return forecasting, risk modelling and portfolio management. Machine learning is gaining traction in the financial sector, and computational power is the most significant factor behind adopting machine learning. Moreover, three areas of interest include linear regression, decision trees and artificial neural networks.

In the banking sector in South Africa, "Machine learning techniques, including regression, decision trees, neural networks, and more, can be applied to various financial tasks" (Cui, 2023, p. 19). Kumar & Mehta, (2019) posits that Machine learning techniques, such as Neural Networks, have emerged as promising tools for portfolio optimisation by enabling the discovery of nonlinear patterns in financial data. The application of 'Neural Networks' in portfolio optimisation has the potential to provide more accurate predictions and better risk management by incorporating nonlinear relationships and complex data structures. Thus, combining advanced techniques, such as the CLA algorithm and machine learning techniques, can enable investors and institutions to make more informed decisions in portfolio construction and risk management in the South African banking industry.

1.2 Problem Statement

Machine Learning has gained traction in the financial industry. It is important to investigate how these novel techniques can be used to complement other techniques such as Autoregressive Integrated Moving Average (ARIMA) and machine learning techniques such as Neural Networks in developing portfolios and managing risks in the South African banking sector from a theoretical standpoint. Diversification and efficient frontier analysis are essential concepts in modern portfolio theory that emphasise returns while minimising risk (Markowitz, 1952; Michaud, 1989). However, the assumptions that underpin classic portfolio theory may need to be revised, especially in highly complicated markets. Because of this, using more sophisticated methods may result in a more precise and nuanced understanding of the connection between risk and return, leading to improved decision-making (Kumar & Mehta, 2019).

The research aims to develop reliable financial models that can effectively manage risks and forecast stock returns by applying techniques such as ARIMA and ML methods such as neural networks. This thesis fills the void left by the need for more financial analytical tools. When applied to financial data, machine learning methods like Neural Networks may also make it possible to find nonlinear trends, resulting in more accurate forecasts and improved risk management (Kumar & Mehta, 2019). Adopting novel portfolio design and risk management approaches in the South African banking sector may enhance the knowledge of portfolio construction in its practical implementation. In today's rapidly changing financial landscape, South African banks must adopt novel portfolio design and risk management approaches. The traditional methods are no longer sufficient to keep up with the evolving market conditions. By embracing innovative strategies and technology banks can identify and manage risks better, leading to greater financial stability and improved performance (Sinha, 2024). Moreover, with heightened competition and regulatory pressures, banks that are ahead of the curve in terms of portfolio design and risk management are more likely to thrive in the long run. Therefore, the South African banking industry must stay competitive and resilient by adopting cutting-edge approaches to portfolio design and risk management.

The theoretical contributions of this work are noteworthy because of the significance of these advanced techniques. Investors and South African banking sector organisations may get better portfolio design and risk management results by integrating modern methods such as machine learning. The research results can potentially have real-world ramifications for South African banking sector investors looking to maximise profits while managing risks successfully.

This thesis argues that the disadvantages of traditional predictive models can be cured by machine learning, and the application of machines can provide a more accurate dataset for investment and policymakers in the South African financial industry.

1.3 Study Objectives

i. To assess the effectiveness of Machine Learning techniques in enhancing traditional models for return forecasting, risk Modeling, and portfolio construction by comparing predictive accuracy.
ii. To outline the methodology and quantify the benefits of implementing advanced machine learning techniques in portfolio construction, risk Modeling, and return forecasting by evaluating enhancements in model accuracy and risk assessment.

1.4 The Value and Relevance of the Study

This research contributes to the literature by comparing different studies on machine learning applications to a reliable tool for risk modelling, portfolio construction optimisation and return forecasting. Machine Learning idea tracks predicting factors of expected returns. The algorithms allow computers to uncover patterns in portfolio construction and financial market predictions. The rising use of computers and technology makes it possible to use machine learning and have real-time data and patterns without depending on programming instructions. As a result, the patterns are used to design decision trees that can perform regression and classification tasks that help portfolio managers make effective decisions. According to IBM, the term machine learning was first coined by Arthur Samuel to use data and computers to make business predictions (IBM Cloud Education, 2020). Over the years and with developments in artificial technology, it is now possible to track data in financial services and determine the level of risks that a company should take to optimise profits. Atsalakis and Valavanis (2009) reviewed the literature on the application of machine learning to forecast stock markets; most of the reviewed articles concentrated on single stock market index forecasting. There is a need for more studies in this area reported in South Africa because much research by then surveyed multiple and single stock behaviours. Previous machine learning predictions concentrated on predicting financial markets rather than the stock's quality or behaviour, which is a disadvantage when making predictions on investments and risks.
According to Ta et al., (2020), low-risk portfolio construction is another area of research where ML algorithms proffer statistical indications for making knowledgeable financial market decisions. Effective portfolio construction helps maximise the expected return for a given level of risk within a given timeframe while minimising risk for a given return, making

it a genuine decision-making tool for asset allocation and security selection. The use of Machine learning for forecasting, risk modelling and portfolio construction can greatly benefit the banking industry of South Africa, stakeholders and investors through the provision of enhanced return forecasting which ultimately improves investment decision, provides risk-adjusted return and helps with offering more tailored products. It can also assist in building optimised portfolios with efficient frontier optimization since it can incorporate economic data in analysis which is helpful for diversified markets and economies such as South Africa. Exposure to different sectors can potentially impact portfolio performance. Improved risk management can help banks achieve their regulatory requirements more accurately, enabling regulators to maintain a stable financial ecosystem. Improvement of products and services can help investors choose products that aligns with their financial goals and risk tolerance. The use of Machine learning will enhance risk management and portfolio construction which contributes positively towards economic stability, growth even amid market fluctuations. Overall, using machine learning presents opportunities to strengthen the South African banking industry's operational efficiency, financial stability and market competitiveness while also providing inclusivity, access to investments and economic reliance to stakeholders.

1.5 Limitations and Delimitations

The limitations of this study relate to the data used in the analysis and the focus of the study. The study relies on historical data, which may reflect something other than current market conditions or future trends. Therefore, the findings may only be applicable in some market conditions. Furthermore, the study focuses solely on the South African banking industry, and the findings may need to be more generalisable to other financial markets. Additionally, the study may only account for potential variables that could affect portfolio construction and risk management.

The delimitations of this study include a focus on using Machine Learning applications in the Banking industry and Neural Networks in portfolio construction and risk management in the South African banking industry. Therefore, the study needs to consider other methods or techniques for portfolio construction and risk management. The research focuses on a specific period, which may only capture some relevant market conditions or events. Finally, the study may only account for relevant factors, such as macroeconomic trends or regulatory changes, that could affect portfolio construction and risk management.

Despite these limitations and delimitations, the study provides valuable insights into using advanced portfolio construction and risk management techniques in the South African banking industry. Investors and institutions can achieve better portfolio construction and risk management outcomes in the South African banking industry by combining legacy techniques such as ARIMA and Neural Networks. The study's findings can have practical implications for investors seeking to maximise returns and manage risks effectively in the South African banking industry. Future research could build on this study's findings by exploring other advanced techniques or considering a broader range of market conditions and events.

1.6 Structure of the Study

This study is structured as follows: Chapter 1 introduces the study, including the research problem, the context of the study, the theoretical contributions, and the study's aims and objectives. Chapter 2 reviews the literature on portfolio theory, advanced portfolio construction and risk management techniques, and the South African banking industry. Chapter 3 presents the study's methodology, including the data sources, sample selection, and the techniques used in the analysis. Chapter 4 presents and analyzes the results, including data description, exploratory analysis, model forecasting, and portfolio construction. Chapter 5 provides a summary of the findings from Chapter 4, discusses research limitations, and recommends areas for future improvement.

1.7 Chapter Summary

This chapter has introduced the research problem of the study, which is the lack of financial analytical tools in the South African banking industry, and the aims and objectives of the study, which are to bridge the gap by using advanced techniques such as ARIMA and Neural Networks to build robust financial models to forecast stock returns and manage risks effectively. The chapter also discusses the context of the study, including the current situation and trends in the South African banking industry and the theoretical contributions of the study. Finally, the chapter has provided an overview of the dissertation structure, including a literature review, methodology, findings, discussion, and conclusion.

CHAPTER TWO: LITERATURE REVIEW

2.1 Introduction

The literature review examines the existing research on return forecasting, risk modelling, and portfolio construction. It critically evaluates the existing research and synthesises the findings to understand the current knowledge in the study area. In addition, the review provides a theoretical framework for the study and helps to identify gaps in the existing research.

2.2 Theoretical Landscape

A holistic view of the theoretical landscape of Modern Portfolio Theory (MPT), the Efficient Market Hypothesis (EMH), and Behavioral Finance provides a comprehensive understanding of how these concepts relate to each other in finance. According to Koumou (2020), MPT's key ideas focus on optimising investment portfolios to balance risk and return. The core concepts include an efficient frontier, in which a set of portfolios offers the maximum expected return for a given level of risk or the minimum risk for a given level of expected return. MPT emphasises that not all risks can be eliminated but that diversification can mitigate unsystematic risk. Diversification involves spreading investments across different assets to reduce risk. The systematic vs. unsystematic risk shows that MPT distinguishes between systematic risk (market risk) that cannot be diversified away and unsystematic risk (specific to individual assets) that can be diversified away.

The Efficient Market Hypothesis (EMH) key idea shows that EMH posits that financial markets are informationally efficient, meaning asset prices fully reflect all available information. It suggests that it is challenging to outperform the market consistently. Moreover, the critical forms of EMH where the weak form of EMH assumes that past price and volume data are already incorporated into current prices, making technical analysis ineffective. The assumption in the semi-strong form of EMH is that all publicly available information is reflected in stock prices, making fundamental analysis ineffective. Gomes (2023) asserts that the strong form argues that even insider information cannot be used to outperform the market consistently.

Almeida & Gonçalves (2023) assert that behavioural Finance key concepts explore how psychological biases and cognitive errors can influence investor behaviour and lead to market anomalies and inefficiencies. Several core concepts, such as investor biases, in which behavioural finance identifies various biases, such as overconfidence, loss aversion, herding behaviour, and anchoring, can lead to suboptimal investment decisions. The market anomalies show that behavioural finance explains persistent market anomalies, such as momentum and value effects, which are deviations from the predictions of traditional finance theories. Goyal et al. (2023) state that prospect theory is a critical concept in behavioural finance; prospect theory describes how individuals evaluate potential gains and losses, often deviating from rational utility maximisation.

The holistic view and connection between these theories show that MPT and EMH are complementary in some respects. MPT provides a framework for constructing diversified portfolios based on risk and return considerations, while EMH guides the understanding that asset prices incorporate all available information. Behavioural finance complements MPT and EMH by recognising that investor behaviour is not always rational and that psychological biases can lead to deviations from market efficiency. It explains why market anomalies persist and how investor sentiment can influence asset prices. However, investors and portfolio managers often integrate aspects of MPT, EMH, and behavioural finance to develop investment strategies. Sengupta and Mitra (2023) argue that MPT is needed for portfolio construction and risk management, acknowledge EMH's implications for market efficiency, and consider behavioural biases that may affect decision-making.

The interaction of these theories underscores the complexity of financial markets. While EMH provides a theoretical baseline, behavioural finance acknowledges that real-world market behaviour often deviates from this idealised model. MPT, in turn, offers a practical framework for constructing portfolios that navigate the nuances of market efficiency and investor behaviour.This rounded approach recognises that each theory contributes to a more nuanced understanding of finance, allowing investors and researchers to make informed decisions and navigate the complexities of the financial landscape.

2.3 A Critique of Theories

Modern Portfolio Theory (MPT), while influential and widely used in finance, has faced several critiques and evolving perspectives. The critical assumption of rationality shows that

MPT assumes that investors are rational and solely motivated by risk and return. Jang and Seong (2023) critique the assumptions of MPT and argue that this assumption does not accurately reflect real-world behaviour, as investors often exhibit emotional biases and irrational decision-making. Moreover, the evolving perspective in behavioural finance has emerged as a complementary field, acknowledging that investors only sometimes behave rationally. Modern portfolio theory has evolved to incorporate insights from behavioural finance, recognising that investor behaviour can deviate from pure rationality and impact asset prices. According to Karp & Van Vuuren (2019), the Efficient Markets debate shows that MPT relies on the Efficient Market Hypothesis (EMH), which assumes that markets are efficient and that all information is immediately reflected in asset prices. Delcey and Sergi (2023) assert that empirical evidence suggests that markets are only sometimes efficient, and anomalies persist.

In addition, the evolving perspective indicates that many practitioners now adopt a more nuanced view, recognising that markets can be partially efficient, and some anomalies can be exploited. This perspective has led to the development of alternative asset pricing models considering deviations from market efficiency. Lee et al. (2023) assert that the risk metrics critique shows MPT commonly uses standard deviation to measure risk, assuming it adequately captures all aspects of risk. Critics argue that standard deviation does not account for all types of risk, including tail risk and downside risk. The developing perspective indicates that alternative risk metrics, such as Value at Risk (VaR) and Conditional Value at Risk (CVaR), have gained prominence. These metrics offer a more comprehensive view of risk by considering extreme events and the potential for significant losses.

Furthermore, the correlation assumptions critique in which MPT assumes that correlations between assets remain stable over time, which is often not true during market crises. The financial crisis of 2008 highlighted the limitations of relying on historical correlations. Fahmy (2020) asserts that the budding perspective indicates that modern portfolio theory has adapted by incorporating time-varying correlations and dynamic asset allocation strategies that account for changing market conditions. The need to have more tangible assets and the critique in this theory is primarily based on a critique where MPT primarily deals with financial assets and may not adequately address the inclusion of tangible assets like real estate, commodities, or cryptocurrencies, which have unique risk-return characteristics.

Portfolio theory has expanded to incorporate a broader range of asset classes and alternative investments. According to Muldoon-Smith and Greenhalgh (2019), natural assets and alternative investments are expected in portfolio construction.
According to Schielzeth et al. (2020), the assumption of the normal distribution is a crucial supposition of MPT, where the assumption is that asset returns follow a normal distribution, which may not hold during extreme market events. The presence of fat-tailed distributions challenges traditional risk models. Mukherjee et al. (2022) state that MPT has evolved in response to various critiques and changing market conditions. It has embraced insights from behavioural finance, adapted to acknowledge market anomalies, incorporated alternative risk metrics, considered dynamic correlations, expanded to include a more comprehensive array of assets, and developed a more robust risk model.

According to Qiu et al. (2020), there is limited predictive power in critical models and the presence of biases and anomalies; it has been criticised for its limited ability to predict specific market behaviours accurately. However, critics contend that even if biases influence investors, the timing and magnitude of their actions are unpredictable.
The Modern Portfolio Theory critique has two main assumptions of constant parameters where MPT relies on stable parameters, such as expected returns and covariances, which can change over time, especially during extreme market conditions. Critics argue that the assumption of constant parameters may lead to suboptimal portfolio decisions. While sensitivity to inputs is another criticism of MPT's optimisation process, it is sensitive to the inputs, such as expected returns and risk measures. According to Xiao et al. (2020), small changes in these inputs can lead to significantly different portfolio allocations, raising questions about the robustness of MPT.

CAPM is a foundational tool for assessing "the risk and return trade-off" for individual assets within a well-diversified portfolio. It assumes that investors are rational, markets are efficient, and a linear relationship exists between an asset's risk (beta) and its expected return. According to Vergara-Fernández et al. (2023), the Capital Asset Pricing Model (CAPM) critique shows that simplistic assumptions of the model in which CAPM relies on simplifying assumptions, such as the risk-free rate and market risk premium, which can vary over time and may not accurately reflect market conditions.

2.3.1 Efficient Market Hypothesis (EMH)

The EMH is a theoretical framework that states that it is not possible to consistently achieve higher returns than the market average, as all available information is already reflected in current market prices (Fama, 1970). This theory could be relevant to the study, as it provides a basis for examining the effectiveness of machine learning algorithms in return forecasting and portfolio construction. One such model is the Adaptive Market Hypothesis (AMH), which seeks to bridge the gap between EMH and behavioural finance. The AMH posits that market efficiency is not a static concept but is shaped by market participants' evolving behaviours and expectations. According to the AMH, market efficiency can be seen as a dynamic process, fluctuating over time in response to changes in market participants and the information they use to make investment decisions.

The weak form of EMH assumes that all past trading information, such as historical prices and volume, is already reflected in current stock prices. Therefore, technical analysis (analysing historical price patterns) should not be advantageous in investment decisions. The Semi-Strong Form EMH asserts that all publicly available information is already reflected in stock prices. According to Noh et al. (2023), neither technical nor fundamental analysis of financial statements and other public information should consistently yield superior returns.

The Strong Form EMH is the most extreme form, portentous that all insider information is already reflected in stock prices. If EMH holds a substantial form, no market should be able to outperform the market, even with insider information consistently. According to Nti et al. (2020), the relationship between EMH and Machine Learning shows that the Weak Form of EMH and Technical Analysis, which comprises historical price data, are built into current market prices. Consequently, technical analysis can only beat the market occasionally. According to Bitetto et al. (2023), ML approaches can be applied to historical price data to find patterns and trends that may need to be evident through traditional technical analysis. However, whether this leads to consistent outperformance remains a subject of debate. Seabe et al. (2023) assert that the Semi-Strong Form EMH and Fundamental Analysis show that information available to the public has been incorporated into the stock prices, making it difficult to outperform the market using fundamental analysis. ML can be used to analyse vast amounts of data, containing unstructured datasets like news sentiment, to uncover insights not readily apparent through traditional fundamental analysis.

Ashtiani et al. (2023) asserts that while the EMH postulates that it is challenging to beat the market using publicly available information consistently, machine learning offers tools and techniques that can support the analysis of copious amounts of data and potentially uncover patterns and insights that traditional methods might miss. However, the debate on whether machine learning can consistently outperform EMH in practice continues, and it depends on numerous issues, including the quality of data, model design, and market conditions.

2.3.2 Modern Portfolio Theory (MPT)

MPT was a theory developed by Harry Markowitz in the 1950s. It opines that investors can optimise their portfolios by diversifying their investments across different assets to reduce risk while maximising returns (Markowitz, 1952). This theory could be relevant to the study as it provides a framework for examining the role of risk modelling and portfolio construction in the investment process. It has proven to be a widely accepted investment framework. Modern Portfolio Theory (MPT) and machine learning are two distinct financial approaches that can be used together to enhance portfolio management and investment decision-making. MPT was advanced by Harry Markowitz in 1952 as the foundational framework for portfolio construction and asset allocation. MPT is based on several fundamental principles. According to Habbab and Kampouridis (2023), MPT emphasises diversifying investments across various assets to reduce risk. The idea is that when investors hold a mix of assets with negative correlations, the overall portfolio risk can be minimised. Therefore, investors must hold a diversified portfolio to avert market risk.

Lee et al. (2023) asserts that there is the risk and return trade-off principle, in which MPT recognises that investors must create an equilibrium between risk and return. It measures risk as the volatility of returns and suggests that investors can enhance their portfolios to realise the most likely expected return for a specified level of risk. Ali et al. (2022) posits an approach whereby MPT presents the idea of the efficient Frontier, which embodies the set of portfolios that offer the peak expected return for an assumed level of risk or the lowest risk for a given level of expected return. These portfolios along the efficient Frontier are considered optimal to invest in to achieve the desired reward.

According to Nilsson and Zheng (2023), asserting the principle of asset allocation, the MPT encourages investors to allocate their assets across different asset classes. According to Habbab & Kampouridis (2023), the principle related to the correlation and covariance, and the MPT relies on statistical measures of correlation and covariance to assess the relationships between asset returns, which are crucial for calculating portfolio risk and expected.

The application of ML in portfolio management has come about because of several factors, such as its predictive analytics, where ML algorithms can analyse historical data and make predictions about asset returns, which can inform portfolio decisions. This includes regression models, time series analysis, and more advanced techniques like neural networks. The key concepts in behavioural finance that can be combined with ML include emotion and psychology. Behavioural finance theory recognises that investors often make decisions based on emotions like fear, greed, and overconfidence. According to Ashtiani and Raahmei (2023), emotional reactions can lead to irrational investment choices by investors. Cognitive biases are present in investment decisions, and behavioural finance can recognise innumerable cognitive biases that affect decision-making, such as bullishness, loss aversion, anchoring, and herding behaviour. Moreover, these cognitive biases can lead to deviations from rational behaviour.

Aithal et al. (2023) assert that market anomalies can be highlighted through behavioural finance, where anomalies or deviations from the efficient market hypothesis (EMH), such as momentum effects, value and growth investing, and underreaction or overreaction to news, are examined. According to Bihari et al. (2023), Prospect Theory supports risk aversion and uncertainty, and ML approaches can help support risk management. The relationship between ML and Behavioral Finance is essential. ML has been helpful in its application in behavioural finance to analyse vast amounts of data, detect patterns, and gain insights into investor behaviour. According to Dash et al. (2023), sentiment analysis uses ML**,** particularly NLP algorithms, to analyse news articles, social media sentiment, and financial reports to gauge investor sentiment and assess its impact on asset prices. This can be important in the banking industry to help managers in decision-making.

Market sentiment prediction integrates ML models to predict market sentiment based on historical data and news sentiment analysis. This can help investors anticipate market reactions to news events. Also, ML can be used in behavioural bias detection through the analyses and detection and quantification of behavioural biases in investor behaviour by analysing trading data, such as order flow and trading volume. According to Wu et al. (2023), predicting market anomalies is critical, and ML models can identify market anomalies and trends consistent with behavioural finance theories, such as momentum or contrarian effects. Besides, ML and risk assessment are essential, and ML can assess and predict investor risk preferences, helping financial professionals tailor investment recommendations to individual risk profiles.

The relationship between Behavioural Finance and Machine Learning shows that the two approaches can be integrated and used to model investor behaviour. ML can be used to build predictive models of investor behaviour based on behavioural finance theories. These models can help explain and anticipate deviations from rational behaviour. According to Goyal et al. (2023), sentiment analysis and trading strategies integrating sentiment analysis using machine learning can inform trading strategies that capitalise on market sentiment trends, often resulting from behavioural biases. According to Chou et al. (2023), risk management can be managed through ML to assess and manage risks associated with behavioural biases, allowing investors to make more informed decisions. Also, ML can support Robo-Advisors and Personalization in the financial industry, where ML-powered robo-advisors can personalise investment recommendations based on investors' behavioural tendencies and risk appetites.

2.4 Risk Modelling

Sahu et al. (2023) assert that risk modelling is a critical component of the finance industry, enabling financial institutions to assess and manage various types of risk, including credit risk, market risk, operational risk, and more risk modelling aims to forecast potential losses and develop strategies to mitigate or manage these risks. According to Khandani & Lo (2010), several methods have been developed for risk modelling, including traditional statistical methods, such as regression analysis and time series analysis, and more advanced techniques, such as machine learning algorithms. Machine Learning techniques have been increasingly applied in risk modelling because they can handle large datasets, capture complex relationships, and improve the accuracy of risk assessments.

There are significant markets and credit risks in the South African financial sector, and using ML can help assuage some of these concerns. Market risk primarily relates to losses occurring in the market where there is fluctuation due to several domestic and international factors, while credit risk results in consumers not fulfilling their contractual obligations. Investors are, therefore, interested in modelling, assessing, and forecasting risk to minimise losses in their investments. The literature shows that ML approaches in risk modelling are better than traditional models because they effectively solve hedging problems. Machine Learning provides robust tools to predict returns, manage risk, and construct portfolio returns. A multi-modality graph neural network is another machine learning approach used in risk modelling, financial forecasting, and risk management. Forecasting is essential for investors, financial institutions, and academics; therefore, using appropriate tools to produce the best results is necessary.

ML can perform three tasks to minimise market risk, namely, it can use qualitative datasets from news sources or social media, the dataset can be used to validate and test risk models and to use the dataset to produce accurate interest, exchange, bankruptcy, and insolvency rate reports. ML can data mine information from online platforms that can be used in decision-making. In addition, credit risk monitoring is essential for portfolio managers, and ML is helpful in this process. ML uses ANN and SVM to perform risk management, and the results are more reliable than traditional models.

According to Wang et al. (2018), Machine Learning algorithms have become increasingly popular in risk modelling because they can handle large amounts of data and identify complex relationships between variables. Chen and Liu (2009) assert that artificial neural networks (ANNs) have been used for risk modelling in the stock market, and support vector machines (SVMs) have been used for credit risk modelling. According to Goodfellow et al. (2016), deep learning algorithms based on artificial neural networks with multiple hidden layers can learn complex relationships between inputs and outputs, leading to more accurate predictions. On the other hand, Liu and Timmermann (2019) assert that reinforcement learning involves using an agent that learns from experience to make predictions about future risks.

Bengio and LeCun (2015) states that, despite the promise of machine learning and advanced techniques, there are still challenges associated with risk modelling. One of the main challenges is the issue of overfitting, where models that are too complex may fit the training data too well and not generalise well to new data.

According to Milojević and Redzepagic (2021), ML can be used to manage risk in the financial industry. ML models are used to model and predict market, credit, and operational risks. This helps the financial industry make informed decisions about portfolio management, risk mitigation, and capital allocation.

Furthermore, the prevalence of fraud detection in South Africa, where ML algorithms can identify unusual patterns in transaction data and detect potentially fraudulent activities in real time. To spot anomalies, they analyse transaction histories, geolocation data, and customer behaviour. Moreover, Customer Service and these Chatbots powered by natural language processing (NLP) and machine learning enhance customer service by quickly responding to customer inquiries and assisting with routine tasks like balance inquiries and transaction history checks.

There is a money laundering problem in South Africa, and ML has applications to support Anti-Money Laundering (AML). According to Chen et al. (2018), ML approaches can improve AML efforts by automatically identifying suspicious transactions and activities, reducing false positives, and streamlining compliance processes.

Bhattacharya et al. (2023) assert that ML is vital in credit risk assessment. Machine learning models, such as logistic regression, decision trees, and random forests, can analyse credit-related data (e.g., credit scores, income, payment history) to predict creditworthiness more accurately.

According to Cornwell et al. (2023), ML is essential for Market Risk Prediction. In these approaches, machine learning models can capture complex patterns in financial time series data, helping to improve market risk models, especially for assets with nonlinear behaviour. Other key areas include operational risk management. ML tools can analyse operational data, such as transaction records and logs, to identify patterns indicative of potential operational risk events or anomalies. In South Africa, fraud is becoming an increasing problem and fraud detection using ML, including supervised and unsupervised algorithms, can detect fraudulent transactions by identifying suspicious patterns or outliers in large volumes of transaction data.

Nazareth and Reddy (2023) state that liquidity risk management is vital. ML models can predict future liquidity needs based on historical cash flow patterns and market data, helping institutions manage liquidity risk more effectively. Also, the ML Model Validation is essential before it can be deployed so that ML models can validate and back-test traditional risk models to assess their performance and robustness.
Risk modelling quantifies the likelihood and impact of adverse events or changes in financial markets. The common types of risk modelling in finance include credit risk modelling, which assesses the risk of borrowers defaulting on loans or credit obligations. Models predict the probability of default (PD), loss given default (LGD), and exposure at default (EAD). Market Risk Behera et al. (2023) assert that modelling evaluates the potential losses in a financial portfolio due to fluctuations in market prices, such as equities, interest rates, currencies, or commodities.

Furthermore, operational risk modelling is critical in identifying and quantifying risks associated with internal processes, systems, people, and external events that could lead to financial losses. According to Xu et al. (2023), liquidity risk modelling analyses the risk of being unable to meet short-term financial obligations due to a lack of liquid assets. In South Africa, speedy fraud detection is essential in the financial industry, detects fraudulent activities in transactions and financial operations using anomaly detection and predictive modelling.

In sum, risk modelling is essential to investment management, helping investors understand and quantify the potential risks associated with their investments. Using machine learning and advanced techniques, such as ANNs, SVMs, deep learning, and reinforcement learning, can improve the accuracy and reliability of risk predictions. However, there are still challenges associated with risk modelling, such as overfitting and data snooping bias, which must be addressed to achieve more accurate and robust predictions.

2.5 Portfolio Construction

According to Emerson et al. (2019), portfolio construction uses a combination of return forecasts and risk models to optimise returns for a given investment under constraints. The literature has shown that applying ANN in the construction of portfolios produces significantly robust and accurate results compared to traditional models applied in the financial sector. Ozbayoglu *et al.* (2020) surmise that some of the insights from the literature show that portfolios weighted using deep machine learning have the potential to perform better than traditional models, simulation support model prediction, and cluster algorithm is essential in constructing diversified portfolios.

Jiang et al. (2017) assert that ML portfolio management framework models are helpful in resource allocation and decision-making. The modern stock markets necessitate quick decision-making for investors because of the large amount of data available to users. Additionally, Deep Learning Models (DLM) are even better because of the multiple layers embedded in the analysis.Portfolio decision-making requires managers to allocate resources to a more extensive set of assets so that the target portfolio meets specific objectives, such as mirroring an index while maximising Sharpe rations under given constraints. Mean variance is the theoretical foundation of portfolio analysis, a traditional method. ML has enhanced this technique to produce better yields in portfolio construction analysis. However, two problems exist, i.e., sensitivity and extensive time series data to make accurate estimates. ML can solve these problems because it can produce returns and estimates that are more reliable than traditional methods. Secondly, ML can produce a better out-of-sample portfolio because of its capacity to generate alternative portfolio weights. Moreover, evolutionary algorithms use elasticity to solve multifaceted asset allocation problems.

Portfolio construction is a crucial aspect of the finance industry, where investors and portfolio managers aim to build portfolios that balance risk and return to achieve specific financial objectives. Machine learning techniques have been increasingly applied to enhance portfolio construction by providing data-driven insights, improving asset allocation strategies, and optimising portfolio performance. According to Al-Muharraqi and Messaadia (2023), portfolio construction involves selecting an amalgamation of assets, such as stocks, bonds, or other investments, to create a diversified portfolio.

The primacy of portfolio construction includes risk management, where building a portfolio minimises risk while achieving a desired level of return. This involves diversifying investments to spread risk across different assets. Asset allocation determines the appropriate allocation of assets to achieve specific investment objectives and risk tolerance. This may involve deciding on the portfolio's percentages of stocks, bonds, and other asset classes. According to Hambly et al. (2023), return maximisation is the creation of portfolios that aim

to maximise returns given the acceptable risk level. Portfolio optimisation is where risk and return are balanced by selecting the best combination of assets to achieve specific goals.
ML in the Portfolio construction approach can assist in asset selection. Machine learning models can analyse historical data, company financials, and other relevant information to identify undervalued or overvalued assets. Natural language processing (NLP) can help assess sentiment from news articles and reports. According to Sahu et al. (2023), risk assessment can be examined through machine learning to help quantify and manage portfolio risk by modelling potential scenarios, using techniques like Monte Carlo simulations to assess the impact of various market conditions.

Text analysis can be used to optimise resource allocation in the financial industry. Better estimates are derived from ML compared to traditional techniques. ML techniques can be used to construct portfolios with specific performance parameters. ML can input data to determine expected returns and variances in portfolio estimation, which results in more than accurate portfolio output. Deep Neural Networks (DNN) are pertinent to portfolio construction and return forecasting in the financial markets. Machine Learning can manage risk prediction and examine risk measures in the financial markets. In the application of Machine Learning, investors can construct socially responsible portfolios in emerging markets where there are challenges with global warming and climate change, given that traditional models have deficiencies in decision-making and portfolio construction.

2.6 Portfolio Optimisation

According to Lee et al. (2023), asset Allocation shows that ML can optimise asset allocation by considering historical performance, correlations, volatility, and factors influencing asset returns. Reinforcement learning algorithms can adjust asset weights dynamically as market conditions change. Factor-based investing is where ML can assist in identifying and integrating factors (e.g., value, momentum, size) that contribute to portfolio returns. Factor models can be used to construct factor-based portfolios. According to Faridi et al. (2023), portfolio rebalancing is where ML algorithms can automate portfolio rebalancing decisions, ensuring that the portfolio maintains the desired asset allocation over time while considering transaction costs.
Valuation theories revolve heavily around the concept of portfolio optimisation. Markowitz (1952) initiated more research toward developing a scientific approach to managing multi-asset portfolios. The classical model developed by Markowitz in 1952 uses the connection

between the mean and variance of the returns. Markowitz's (1952) expected variance of returns (EV) rule inferred an acceptable diversification, which depends upon the amount and diversity of securities across different asset classes to reduce variance. Companies operating in the same sector often exhibit a high degree of covariance among themselves because of how they react to comparable economic events, which prompts a more significant risk presentation. Therefore, an investor favours an MV-effective portfolio when a certain amount of risk results in the highest possible anticipated return (Campbell et al., 1997). If the unsystematic risk is ultimately to be mitigated or eliminated, the equity allocation must exhibit the best distribution of uncorrelated stock holdings. This is the explicit end aim. Machine learning techniques can enhance various aspects of portfolio optimisation in several ways, such as Asset selection, in which ML models can analyse historical data, fundamental factors, and alternative data sources to identify promising assets and assess their potential returns and risks. According to Lee et al. (2023), risk assessment utilises ML to provide more accurate risk assessment by modelling potential scenarios, analysing correlations, and accounting for tail risks. It can also incorporate macroeconomic indicators and market sentiment data. The concept of a coherent asset with desirable features is presented by Artzner et al. (1999) about risk measurements related to the tail of the distribution. CVaR is the cohesive risk indicator, a select group of writers referred to as anticipated shortfall. Much research has been done on different risk measurements and average risk models, but there is still much debate over which risk measure is the most appropriate. In addition, it persisted in casting doubt on the theoretical soundness of the mean risk models.

Regardless of the amount of research conducted on risk indicators, the subject of which risk measure is the most accurate is still up for debate.

The efficient frontiers developed by Markowitz in 1952 describe how investors pick portfolios that maximise return for a given degree of risk as measured by the return variance. ML can support portfolio optimisation in that it can be applied to optimise investment portfolios by considering historical data, expected returns, risk profiles, and constraints. According to Jang & Seong (2023), ML supports algorithmic trading because of its potential in that ML is often used in algorithmic trading strategies, where models analyse market data to make trading decisions. One of the most critical takeaways from this research is that investments should not be made only because of the remarkable characteristics of each asset's return and variance. The market experts must consider the link between a portfolio with the same return and a lower risk than a portfolio that disregards these relationships.

Since the early work by Markowitz (1952), a significant amount of research has been done on optimisation models that attempt to manage the exchange between risk and return in the financial industry.

2.7 Artificial Intelligence and Machine Learning in Finance

Machine learning is a subfield of artificial intelligence that involves the development of algorithms that can automatically improve their performance on a specific task through experience. Over the years, machine learning has been applied to various fields, including image and speech recognition, natural language processing, and predictive analytics.

Ahmed et al. (2022) show that AI and ML have become popular in finance. Since 2015, the number of papers published has increased significantly, but a limited number of studies are coming from emerging markets. Decision-making in emerging markets can significantly improve when AI and ML are used. Key research themes in AI & ML include stock price prediction, portfolio management volatility, risk modelling and liquidity in the market. Through AI and ML, traders, financial analysts, investors, and financial institutions can construct portfolios with particular risk and return profiles. This mechanism can ameliorate the problems of prior techniques used in the industry before the adoption of AI and ML. AI and ML are generally better at solving optimisation problems than other traditional methods used in the industry for decades. Algorithms can be programmed to execute trades efficiently with little human intervention. Due to uncertainty in the global market, risk management and compliance have become critical. The global market has become increasingly complex recently, and traditional models cannot produce accurate forecasts. AI and ML can utilise various sources to make decisions in complex environments.

The computer processing power has more than tripled over the past decade, making ML able to solve complex problems in the financial industry. Decision-making in the financial industry relies heavily on information processing, and ML can increase the amount of information processing to enable investors to make quick and better decisions.

ML uses several techniques to model risk, construct portfolio such as "artificial neural networks (ANNs), cluster analysis, decision trees and random forests, evolutionary (genetic) algorithms, least absolute shrinkage and selection operator (LASSO), support vector machines (SVMs), and natural language processing (NLP)". Neural networks or deep learning have become more popular than the other techniques described above.

A study of the South African stock market return by Gupta & Modise (2013), using the macroeconomic variable to predict stock returns, asserted that applying older models used in stock prediction, such as regressive models, is vulnerable to data mining. Machine learning can address this problem, and improved data capability can significantly increase the prediction accuracy in decision-making. Machine learning provides real-time predictive analysis, which is valuable compared to other models.

In Thakkar & Chaudhari (2021), three machine learning models are presented, i.e., information features and models which can predict stock returns, risk/returns, index, portfolio management, profitability, and revenue. Machine learning can make relatively accurate forecasts and risk management, unlike traditional models, because of these capabilities. Using a combination of both linear and nonlinear models has the potential to provide robust predictions in emerging highly volatile and complex markets, such as South Africa. Stock markets are usually volatile, so accurate predictions are difficult to ascertain. Diverse datasets and information are necessary to give investors more reliable information on which to base their decision-making. Wrong decisions on the investor's part can lead to harmful consequences. According to Thakkar and Chaudhari (2021), data fusion has the potential to combine financial and economic data to provide forecasters and investors with more reliable information.

One of the main advantages of machine learning is its ability to handle large amounts of data and to identify complex relationships between variables. For example, decision trees, random forests, and popular machine learning algorithms have been used for credit risk modelling (Chen & Liu, 2009). According to Goodfellow et al. (2016), another popular type of machine learning algorithm is artificial neural networks (ANNs), which are modelled after the structure of the human brain and can be trained to make predictions based on input data. According to Goodfellow et al. (2016), in recent years, deep learning algorithms, which are based on artificial neural networks with multiple hidden layers, have become increasingly popular due to their ability to learn complex relationships between inputs and outputs. Bengio and LeCun (2015) state that despite the advantages of machine learning, there are still challenges associated with its use. One of the main challenges is the issue of overfitting, where models that are too complex may fit the training data too well and not generalise well to new data. Another challenge is the risk of data snooping bias, where researchers may select models based on their performance on a specific dataset, leading to over-optimistic results.

Machine learning has become an essential tool for data analysis and predictive modelling due to its ability to handle large amounts of data and to identify complex relationships between variables. Popular machine learning algorithms, such as decision trees, random forests, ANNs, and deep learning algorithms, have been applied to various tasks with promising results. However, there are still challenges associated with machine learning, such as overfitting and data snooping bias, which must be addressed to achieve more accurate and robust predictions.

ML supports stock price prediction in that machine learning models can analyse historical stock price data, news sentiment, and other factors to predict future stock prices. Models like regression, ARIMA, or more advanced deep learning models can be used. In addition, the risk assessment can be assessed using ML because its potency in helping assess and manage financial risk is significant. According to Ibrahim et al. (2023), ML models can be constructed to predict credit, market, or operational risks. According to Yi et al. (2023), ML can aid in fraud detection because machine learning can identify unusual patterns in financial transactions that may indicate fraud or anomalies. According to Melina et al. (2023), ML supports risk management in financial institutions, and ML models can be used to assess and manage portfolio risk. This includes techniques like Value at Risk (VaR) modelling, stress testing, and scenario analysis.

ML has value application in portfolio optimisation in which ML optimisation algorithms can help construct portfolios that align with an investor's risk and return objectives. These algorithms can consider constraints, transaction costs, and various optimisation goals. Moreover, ML can be used to analyse alternative data sources, such as social media posts, news sentimentality, and satellite metaphors, to gain an understanding that traditional financial data may not capture. This is important given the abundant datasets traders need to use to make accurate decisions. In addition, ML supports market sentiment analysis whereby **s**entiment analysis using (NLP) can gauge market sentiment and assess the potential impact of news and events on asset prices. Moreover, MPT and ML can be integrated in several ways to support investors in return and risk estimation. ML can enhance return and risk estimation by analysing large datasets and capturing complex asset relationships, providing more accurate inputs for MPT.

Ciampiconi et al. (2023) assert that ML optimisation techniques can optimise portfolios focusing on explicit objectives, such as maximising returns or risk, whilst accounting for complex constraints and transaction costs. ML approaches and risk management are compatible in the financial industry. ML can be deployed to improve risk management by providing more sophisticated risk models that incorporate tail risk, extreme events, and nonlinear dependencies. According to Bluwstein et al. (2023), ML techniques support alternative data integration in the decision-making of analysts and investors. ML can process alternative data sources to identify potential investment opportunities or risks that traditional financial datasets cannot provide to investors and analysts.

According to Dezhkam and Manzuri (2023), MPT and machine learning can complement each other in portfolio management. Machine learning techniques can enhance MPT by providing better risk and return estimates, optimising portfolios more effectively, and incorporating alternative data sources for more informed investment decisions.
According to Fieberg et al. (2023), integrating machine learning can lead to more sophisticated and data-driven portfolio management strategies. Return forecasting is a fundamental task in finance, as it involves predicting the future returns of assets or portfolios. ML approach techniques have gained acceptance because they can analyse large datasets, capture complex relationships, and improve predictive accuracy.

2.8 Challenges of Machine Learning

According to Song et al. (2023), there are several challenges, such as overfitting, where complex machine learning models are prone to overfitting, especially when the number of predictors is high relative to the sample size. Regularisation techniques can help mitigate this issue. In addition, there are challenges with data quality where ML models are sensitive to data quality and outliers. Dataset cleaning and preprocessing and cleaning of financial data are crucial. Furthermore, market dynamics in the financial markets are influenced by various factors, including geopolitical events, economic indicators, and investor sentiment.
Grosso et al. (2023) assert that machine learning approaches may struggle to capture the impact of rare and unexpected events such as black swan events. Model Interpretability shows that ML approaches, such as deep neural networks, are often seen as "black boxes" with limited interpretability. This can be a concern for financial professionals who require transparent models.

According to Sonkavde et al. (2023), ML significantly improves return forecasting by offering the ability to capture complex patterns, automate feature engineering, and handle extensive and diverse datasets. However, addressing challenges related to overfitting, data quality, market dynamics, and model interpretability is essential when applying machine learning to return forecasting in the financial domain. According to Whang et al. (2023), data quality challenges exist in the ML approach because ML models are sensitive to data quality and completeness, and high-quality, well-structured data is essential for accurate modelling. Data collection and quality challenges in deep interpretability, where some ML machine learning models, like deep neural networks, can be challenging to interpret. Interpretability is crucial for regulatory compliance and model validation. The model validation in ML models requires rigorous validation to ensure they meet regulatory requirements and are suitable for the specific risk assessment task.

The problem of overfitting is common in machine learning, where complex ML learning models are prone to overfitting, especially when trained on small datasets. Ahmadi et al. (2023) assert that regularisation and cross-validation techniques are essential to mitigate overfitting. According to Brauneck et al. (2023), regulatory compliance challenges in financial institutions must ensure that machine learning models comply with regulatory standards, such as Basel III, for credit risk modelling or stress testing requirements.

2.9 The Value of Machine Learning in the Financial Industry

According to Loke et al. (2023), machine learning offers significant opportunities to enhance portfolio optimisation in the finance industry by providing data-driven insights, improving asset allocation strategies, and optimising risk-return trade-offs. However, when integrating machine learning into the portfolio optimisation process, careful attention to data quality, model interpretability, and regulatory compliance is essential.

Nazareth and Reddy (2023) assert that ML plays a significant role in risk modelling in the finance industry by improving the accuracy and sophistication of risk assessments. However, financial institutions must carefully manage data quality, model interpretability, and regulatory compliance when integrating machine learning into their risk management processes.

Asset allocation: Machine learning models can optimise asset allocation by considering historical asset performance, correlations, volatility, and macroeconomic indicators. Reinforcement learning algorithms can adjust asset weights over time based on market conditions. Factor-Based Investing: Machine learning can assist in identifying and incorporating factors (e.g., value, momentum, quality) that contribute to portfolio returns. This can lead to constructing factor-based portfolios that aim to outperform market benchmarks.

Portfolio Rebalancing where ML can automate portfolio rebalancing decisions to maintain desired asset allocations, considering transaction costs and market conditions. Risk Management, where Behera et al. (2023) contend that ML can proffer more sophisticated risk management techniques, such as Value at Risk (VaR) modelling, stress testing, and scenario analysis, which can be integrated into portfolio construction strategies.

Loke et al. (2023) assert that portfolio optimisation is a critical component of the finance industry, where investors and portfolio managers aim to construct portfolios that maximise returns while managing risk. Machine learning techniques have gained popularity by providing data-driven insights, improving asset allocation strategies, and optimising portfolio performance.

According to Han & Li (2023), portfolio optimisation involves selecting a combination of assets (e.g., stocks, bonds, commodities) that maximises expected returns or achieves specific financial objectives while managing risk. The primary objective of portfolio optimisation includes risk Management, where a portfolio is constructed that balances risk and return, ensuring that risk is diversified across different assets to reduce overall portfolio risk. According to Aithal et al. (2023), asset allocation determines the optimal allocation of assets to achieve specific investment goals, risk tolerance, and time horizons. Return maximisation is a designing portfolio that aims to maximise returns given a predefined level of acceptable risk. Diversification spreads investments across various asset classes to reduce exposure to individual asset risk. Efficient Frontier identifies the set of portfolios that offers the highest expected return for a given level of risk or the lowest risk for a given level of expected return.

According to Bitetto et al. (2023), risk management in which ML can provide advanced risk management techniques, including Value at Risk (VaR) modelling, stress testing, and scenario analysis, can be integrated into portfolio optimisation strategies. Alternative Data Integration (ADI) uses ML to process and analyse alternative data sources, such as social media sentiment, satellite imagery, or credit card transaction data, to gain insights that traditional financial data may not capture.

ML has found numerous applications in the banking industry, revolutionising various aspects of banking operations, risk management, customer service, and fraud detection. The application of Guerra & Castelli (2021) is the application of ML in the financial industry, such as credit scoring and underwriting, in which ML is applied to analyse massive datasets to assess the creditworthiness of applicants more accurately. According to Mhlanga (2021), ML considers various factors, including credit history, income, employment history, and even alternative data sources like social media and transaction history.

According to Candanedo et al. (2018), customer segmentation and personalisation can be achieved via ML, which supports financial institutions in segmenting customers based on their behaviour and preferences. This enables personalised marketing campaigns, product recommendations, and tailored services. Predictive analytics and ML use historical data to forecast customer behaviour, loan default rates, and deposit flows. These insights inform business strategies, marketing efforts, and resource allocation. According to Liu et al. (2021), Automated Trading and Algorithmic Trading are supported by ML algorithms used in algorithmic trading to analyse market data, identify trading signals, and execute trades at optimal times. This enables banks to automate trading strategies and optimise execution. Shaukat et al. (2020). ML models detect fraudulent credit card transactions by analysing spending patterns, transaction locations, and customer behaviour. Real-time alerts can help prevent unauthorised transactions. Furthermore, ML assists portfolio managers in optimising asset allocation, rebalancing portfolios, and identifying investment opportunities by analysing market data, economic indicators, and historical performance. According to Ma et al. (2021), ML can detect and prevent cybersecurity threats by analysing network traffic patterns, identifying anomalies, and safeguarding sensitive customer data.

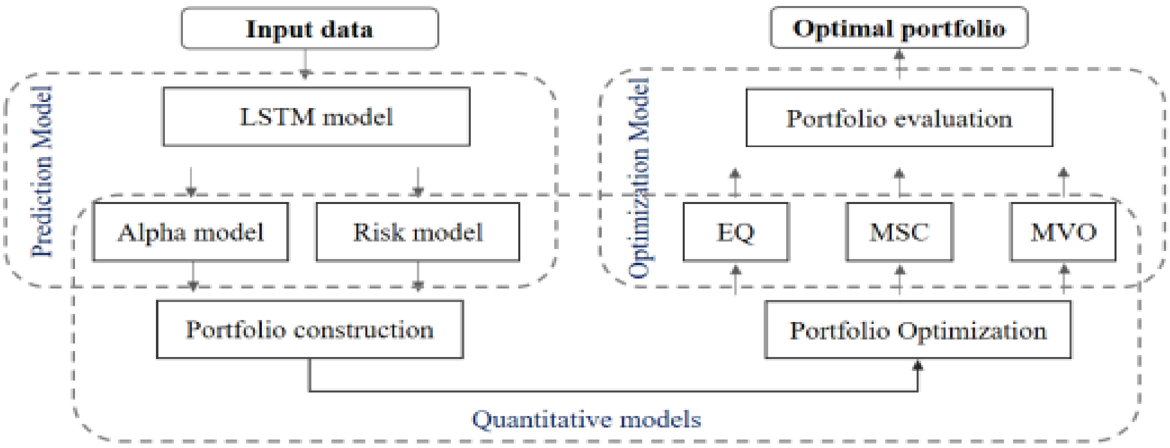
Munkhdalai et al. (2019) state that ML can streamline loan approval processes by automatically evaluating applicants' creditworthiness, income, and risk profiles, reducing manual underwriting efforts and speeding up loan disbursements. Lalwani et al. (2022) assert that ML models predict customer churn by analysing historical data and identifying factors that lead to customer attrition. Banks can then take proactive measures to retain customers. ML is significant in assisting banks to comply with regulatory requirements by automating data extraction, validation, and reporting processes. Leo et al. (2019) assert that ML reduces the risk of errors and ensures adherence to regulations. Hassani et al. (2020) ML supports

Voice recognition technology powered by machine learning, enhances customer authentication, and provides secure access to banking services through voice commands and biometric authentication.

CHAPTER THREE: METHODOLOGY

3.1 Introduction

This chapter outlines the methodology that is adopted to achieve the set research objectives. The study aims to uses quantitative techniques such as machine learning for analysis of financial markets and return forecasting of returns to model risk and use the outputs from the model as inputs for portfolio construction and optimization. The historical share price data for the selected top 4 banks in South Africa is used for the analysis. Empirical results from previous studies shows that forecasts differ significantly across stocks and have strong predictive power for actual returns (Lewellen, 2013). This property about stocks will challenge the use of machine learning to see if they can handle and differentiate risk well. Historical data available is collected and analysed and multiple prediction models are built to predict stock returns using time series and machine learning techniques such as Support Vector Regression (SVR), Random Forests or Gradient Boosting and Long Short-Term Memory (LSTM) networks. A representation of the overall process is shown below in Fig 3.1.



In constructing portfolios, the selection criteria are to choose stocks that outperforms the predicted result in terms of the uppermost predictable reoccurrence and lowest risk. An optimal stock allocation for the constructed portfolio is evaluated by the mean-variance optimization (MVO) technique and is used to evaluate the optimal stock allocation weights.

3.2 Research paradigm

A research paradigm entails a perspective or framework that shapes and guides the way researchers approach and conduct their studies (Chu & Ke, 2017). Therefore, as Saunders and Lewis (2017) indicate, the research paradigm encompasses a set of beliefs, assumptions, values, and methodologies that researchers adopt to generate knowledge and address research questions. Three main research paradigms, as Hirose and Creswell (2023), Dooley (2017), and Chu and Ke (2017) reveal, are positivism, interpretivism and pragmatism. Harrison et al. (2020) point out that positivism emphasizes objectivity, measurement, empirical observation, and the discovery of universal laws and causal relationships in undertaking research. Thus, it relies on quantitative methods and aims to uncover generalizable knowledge. According to Creswell and Hirose (2020), interpretivism focuses on interpreting and understanding social phenomena from the perspective of individuals and their subjective experiences. Therefore, it emphasizes qualitative methods such as observations, interviews, and textual analysis to explore meanings, contexts, and social constructions. Dooley (2017) demonstrates that pragmatism emphasizes usefulness and practicality in research. Thus, it seeks to combine elements from different paradigms to address research questions effectively.

This study aimed to use quantitative techniques such as time series analysis and machine learning models (e.g., ARIMA, SVR, Random Forests, Gradient Boosting, LSTM) to analyze financial markets and forecast stock returns. These quantitative techniques involve the collection and analysis of numerical data, which aligns with the positivist research paradigm Additionally, the adopted methodology use of economic indicators, historical data, and market data, was also seen as measurable and observable factors in the positivist paradigm. The research also refers to empirical results that demonstrate the predictive power of the forecasting models, further acknowledging the reliance on objective evidence and observation. Therefore, the researcher abodes with the assumptions and principles of positivism in undertaking this research.

3.3 Target population and sampling

The target entails a specific group or category of individuals, entities, or elements that a research study aims to investigate or draw conclusions about (Hirose & Crewell, 2023). Therefore, the target population of this study consisted of four South African banks namely Absa, FNB, Nedbank and Standard Bank.

The specific sampling method employed in this study, since the study collected historical data from a reliable financial data source, is a non-probability sampling technique. In particular, a convenience sampling is used. Convenience sampling, according to Hirose and Creswell (2023), is a non-probability sampling technique where individuals or elements are selected based on their easy accessibility and proximity to the research setting. Therefore, this method is chosen because it works in selecting data sources based on their accessibility and relevance to the research objectives. In particular, the selection of data sources is driven by factors such as data reliability, availability, and the specific stocks or markets being analysed.

3.4 Data description and source

The study utilizes secondary data sourced from Yahoo Finance, a widely recognized platform for financial data, known for its comprehensive, up-to-date, and accessible financial market data. The dataset includes historical daily share prices, covering key financial metrics such as opening price, closing price, highest and lowest prices, trading volume, and adjusted close price for the four chosen banks in South Africa namely: Absa, Fnb, Nedbank, and Standard Bank. The collected data provides a reliable foundation for analyzing stock trends and patterns in time-series forecasting. The daily frequency data from January 2009 to January 2024, is selected to capture a complete market cycle, including periods of both growth and recession to provide a balanced view of market dynamics. This time span allows the model to experience a variety of market conditions, which is crucial for evaluating its robustness in real-world situations.

The analysis looks at historical data to make projections to inform the asset classes that will be used as inputs into portfolio construction and optimization. Data processing and exploratory data analysis (EAD) is carried out to understand and investigate any meaningful patterns, trends, and potential relationships and features that may impact share prices in the data through data visualization (Graphs & Charts).

3.5 Model Prediction

3.5.1 Arima Model

Auto-Regressive Integrated Moving Average (ARIMA) is a traditional time series forecasting model that uses historical values to forecast future values (Mahdavinejad *et al.*, 2018). The model was first introduced by Box and Jenkins in 1970 (Angra & Ahuja, 2017). The ARIMA model, according to Burbidge *et al.* (2001), is based on ARMA models which were formulated by the works of Yule, Slutsky and Yaglom. The key differences between these models are that ARIMA models convert non-stationary data to stationary data before working on it (Mondal, Shit & Goswami, 2014). The composite model of the time series is classified as ARIMA (p, d, q) where p represents the autoregressive parts, d represents the integrated parts and q represents the moving average parts of the data set. The ARIMA model is broken down and explained as follows:

Autoregression (AR) is a regression model that is based on dependencies between the observation and several lagged observations (p).

Integrated (I) measures the differences of observation at different times (d) to bring stationarity in the time series.

Moving Average (MA) is an approach that looks at dependencies between observations and the residual error terms when the MA is used for the lagged observations (q).

A mathematical representation of an AR model of order (p) is expressed by the linear equation below:

$$x_t = c + \sum_{i=1}^{p} \emptyset_i y_{t-i} + \epsilon_t \qquad (1)$$

Where x_t represents the stationary variable, the constant c and the terms in $\emptyset_i$ are autocorrelation coefficients at lags 1, 2……. P and the residuals denoted by $\epsilon_t$are the Gaussian white noise series with mean zero and variance $\sigma^2{}_\epsilon$.

An MA model of order q is expressed in the form below:

$$x_t = \mu + \sum_{i=0}^{q} \theta_i \epsilon_{t-i} \qquad (2)$$

Where μ represents the expectation of x_t, θ_i terms denote the weights applied to the current and prior values of a stochastic term in the time series with θ_0= 1. ϵ_t is taken as the white noise series with mean zero and variance $\sigma^2{}_{\epsilon}$. The two models above combine to form an ARIMA model of order (p, q) represented by the mathematical equation:

$$x_t = c + \sum_{i-1}^{p} \emptyset_i \; x_{t-1} + \epsilon_t + \sum_{i=0}^{q} \theta_i \epsilon_{t-i} \qquad (3)$$

Where $\emptyset_i \neq 0$, $\theta_i \neq 0$, and $\sigma^2{}_{\epsilon} > 0$ where the parameters p and q are AR and MA orders. The general form of the ARIMA model is represented by ARIMA (p, d, q) and is used to uncover patterns for better understanding and forecasting of time series data. Its capability of dealing with non-stationary data makes it superior to other traditional time series models and this is due to the "Integration" process in the model.

When there is seasonality in the time series data where short-run non-seasonal components contribute to the model, a seasonal ARIMA model will be estimated and will include both non seasonal and seasonal factors in a model. The model will be useful should we find an element of seasonality in the data during exploratory data analysis (EDA).
The general form of a seasonal ARIMA model is denoted by:
ARIMA (p, d, q) x (P, D, Q) S:

Where p is the non-seasonal AR order
d is the non-seasonal differencing
q is the non-seasonal MA order
P is the seasonal AR order
D is the seasonal differencing
Q is the seasonal MA order
S is the time span of repeating seasonal patterns.

3.5.2 Support Vector Regression Model

Support Vector Regression (SVR) is a machine learning algorithm used for solving regression problems (Verbeek *et al.*, 2020). SVR is based on the principles of Support Vector Machines (SVM), which are widely used for classification tasks (Kern *et al*., 2019). Therefore, in SVR, the goal is to find a regression function that predicts the target variable given a set of input features. SVR achieves this by constructing a hyperplane in a high-dimensional feature space, where the distance between the predicted values and the actual target values is minimized (Kamiri & Mariga, 2021). The algorithm also uses a kernel function to transform the input features into a higher-dimensional space, allowing for the nonlinear modelling of relationships between variables. Alsharif *et al*. (2020) emphasizes that SVR effectively handles nonlinearity and outliers in the data.

- Therefore, in this study the SVR model's objective is to find a function $f(x) = w^T x + b$ that has at most an ε-deviation from the actual targets y_i for all training points.

Given:

- x_i : input feature vector for the *i*-th sample
- y_i : actual target value for the *i*-th sample
- w : the weight vector
- b : $bias\ term$
- ε: margin of tolerance within which errors are not penalized

the SVR objective is to minimize:

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\varepsilon_i + \varepsilon_i^*)$$

Where:

- $\|w\|^2$ is the regularization term to minimise the model complexity
- C is a regularization parameter that allow the trade-off between maximizing the margin and minimizing the error.
- ε_i and ε_i^* are slack variables that allows for errors beyond the ε-margin.

The constraints for SVR are set up to ensure that most of the data points lie within the ε-margin.

$$y_i - (w^T x_i + b) \leq \varepsilon + \varepsilon_i$$

$$(w^T x_i + b) - y_i \leq \varepsilon + \varepsilon_i^*$$

$$\varepsilon_i,\ \varepsilon_i^* \geq 0$$

Here:

- ε_i and ε_i^* represents the positive and negative deviations from the ε: margin, respectively.

- The sum $\sum_{i=1}^{i}(\varepsilon_i + \varepsilon_i^*)$ penalized deviations that exceed the ε-tolerance.

SVR uses kernel function to optimize the function and once the problem is solved, the prediction function *f(x)* for a new input *x* can be written as:

$$f(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) \text{ K } (x_i, x\ ) + \text{b}$$

where:

- Only the data points with non-zero α_i and α_i^*(known as support vectors) contribute to the prediction, making SVR efficient and sparse.

regularization parameter C to balance error tolerance and model complexity. The dual formulation allows for the use of kernel functions which makes SVR powerful and The SVR aims to fit a function within an ε-margin , with a versatile in handling nonlinear data while controlling overfitting through regularization. Therefore, by solving the optimization problem, the SVR model aimed to find the appropriate regression function f(x) and the support vectors that define the margin.

3.5.3 Random Forest and Gradient Boosting

Random Forests and Gradient Boosting are both machine learning algorithms used for predictive modelling tasks (Yoo *et al.*, 2014). According to Reel *et al.* (2021), Random Forests is an ensemble learning method that combines multiple decision trees to make predictions. Each decision tree is trained on a random subset of features and produces an individual prediction. Therefore, the final prediction is determined by aggregating the predictions of all the trees. This approach reduces overfitting and improves robustness (Angra & Ahuja, 2017). On the other hand, Gradient Boosting is an iterative method that builds a sequence of decision trees, where each subsequent tree is trained to correct the errors of the previous trees. It focuses on minimizing the loss function by iteratively adding trees. Gradient Boosting excels in capturing complex patterns and achieving high predictive accuracy. Both Random Forests and Gradient Boosting are popular and widely used algorithms due to their ability to handle, handle missing data, and nonlinear relationships and provide feature importance rankings (Verbeek *et al.*, 2020). In this study, the Gradient Boosting model was adopted below is the simplified representation of the Gradient Boosting algorithm:

The first thing is to initialize the model with an initial prediction, usually the mean or median of the target variable.

Thus, for each iteration (m = 1 to M),

the following steps should be taken:

1ST: Compute the negative gradient of the loss function for the current model's prediction, using the training data.

2ND: Fit a weak learner (e.g., decision tree) to the negative gradient. This weak learner is typically a shallow tree with a limited number of nodes.

3RD: Update the model by adding the weak learner's prediction, multiplied by a learning rate (η), to the current model's prediction.

4TH: Repeat steps a-c until the desired number of iterations (M) is reached.

The final prediction is, therefore, obtained by combining the predictions of all the weak learners.

3.5.4 Long Short-Term Memory (LSTM) Networks

Long Short-Term Memory (LSTM) networks are a type of recurrent neural network (RNN) that excel at modelling and capturing time-dependent and sequential patterns in data (Kern *et al.*, 2019). Unlike traditional RNNs, LSTM networks address the vanishing gradient problem by incorporating memory gates and cells. Thus, these memory cells allow the network to selectively retain and forget information over long sequences, enabling them to capture long-term dependencies. LSTM networks are widely used in various tasks such as time series forecasting, natural language processing and sentiment analysis, where the understanding of long-term dependencies is crucial. Their ability to effectively process and learn from sequential data makes them a powerful tool for modelling and predicting complex temporal patterns (Kamiri & Mariga, 2021; Yoo *et al.*, 2014).

Given an input sequence $x = (x_1, x_2, ..., x_n)$, the LSTM computations can be represented as follows:

Initialization:

$c_0 = 0, h_0 = 0$

For each time step t from 1 to n:

1ST: Computing the input gate activation.

$i_t = \sigma (W_i \cdot x_t + U_i \cdot h_{t-1} + b_i)$

2ND: Computing the forget gate activation.

$f_t = \sigma (W\_f \cdot x_t + U\_f \cdot h_{t-1} + b\_f)$

3RD: Computing the output gate activation.

$o_t = \sigma (W\_o \cdot x_t + U\_o \cdot h_{t-1} + b\_o)$

4TH: Computing the candidate cell state.

$\tilde{c}_t = \tanh (W\_c \cdot x_t + U\_c \cdot h_{t-1} + b\_c)$

5TH: Computing the new cell state.

$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t$

6TH: Computing the new hidden state.

$h_t = o_t \odot \tanh(c_t)$

7TH: Generating the output prediction.

$y_t = \mathrm{softmax} (V \cdot h_t + b)$

In the above representation, σ represents the sigmoid activation function, tanh represents the hyperbolic tangent activation function, ⊙ denotes element-wise multiplication, and SoftMax is used for multi-class classification. W and U are weight matrices, b represents bias vectors, and V is the weight matrix for the output layer. The LSTM model's parameters are learned through backpropagation and optimization techniques such as gradient descent or variants like Adam.

3.6 Evaluation of Model Performance

The importance of model building is the ability to assess and evaluate the performance of the models in terms of their efficacy. Several important measures will be considered for the evaluation of each model's performance and model comparison purposes. Measures like R squared score (R2 score), Mean Squared Error (MSE), and Mean Absolute Error (MAE) are used to assess model accuracy and performance. The metrics offer a comprehensive understanding of the model's capacity to forecast outcomes on unobserved data with underlying patterns. R2 score is a mathematical measure that indicates the percentage of the dependent variable's variance that the model can account for.

A perfect prediction is achieved when the R2 score reaches a value of 1 and models are compared based on the percentage of the R2 score. A performing model can capture and explain the underlying patterns in the data and be able to predict unseen data accurately. The Mean Squared Error (MSE) is a popular metric used to assess an estimator's quality in its mean square error, a value near 0 indicates higher quality. The model learns well if the MSE is as low as possible, A high MSE on the test data indicates overfitting of the model. The Mean Absolute Error (MAE) is a measure that calculates the average magnitude of the errors in a set of predictions. It is the mean of the absolute differences between the observed and predicted values over the test sample, with each difference carrying equal weight. A high MAE indicates overfitting in the model and that the model is not able to generalize well on unseen data. Overall, the performance metrics help to determine the model's predictive accuracy and generalization to unseen data.

3.7 Asset Selection and Portfolio Optimization

This process will use the mean-variance method to construct, compute and plot the efficient frontier for the specified asset classes over a period of time. The efficient frontier depicts the set of optimal portfolios that offers the highest possible expected return relative to the portfolio’ level of risk. The mean-variance method was developed by Markowitz in 1952 and it’s a key part of the modern portfolio theory. It focuses on 5 key concepts to successfully select and optimise the portfolio, namely: Expected Return (Mean), Risk, Covariance and Correlation, Efficient Frontier and the objective.
The expected return of a portfolio is the weighted average of the expected returns of the individual assets in the portfolio
If

- R_p is the expected return of the portfolio
- W_i is the weight of the asset i in the portfolio

- R_i is the expected return of asset i

then :

$$R_{p=} \sum_{i=1}^{n} w_i R_i$$

where n is the number of assets in the portfolio.

Variance of a portfolio which is a measure of the risk in the portfolio. It accounts for the variances and the covariances between the assets .

If:

- ${\sigma^2}_p$ is the variance of the portfolio
- W_i and W_j are the weights of assets i & j
- $\sigma_{i,j}$ is the covariance between assets i & j (with $\sigma_{i,j}$ = ${\sigma^2}_i$ as the variance of asset i)

then :

$${\sigma^2}_p \ = \ \sum_{i=1}^{n} \sum_{j=1}^{n} w_i \, w_j \sigma_{i,j}$$

Alternatively, in matrix form

$${\sigma^2}_p = \ \boldsymbol{w}^T \sum \boldsymbol{w}$$

Where:

- W is the vector of assets weights
- Is the covariance matrix of asset return.

The Standard Deviation of a portfolio which is the square root of the portfolio variance:

$$\sigma_p = \sqrt{{\sigma^2}_p}$$

Optimization objective for mean-variance portfolio has two main objectives :

a) Minimise risk for a given expected return

$$\min_{w} \sigma_{p=} \sqrt{\boldsymbol{w}^T \ \sum \boldsymbol{w}}$$

Subject to:

$$\sum_{i=1}^{n} w_i R_i = R_{target}$$

$$\sum_{i=1}^{n} w_{i=1}$$

b) Maximize return for a given level of risk:

$$\max_{w} R_{p} = \sum_{i=1}^{n} w_i R_i$$

Subject to:

$$\sqrt{\boldsymbol{w}^T \sum \boldsymbol{w}} = \sigma_{target}$$

$$\sum_{i=1}^{n} w_{i=1}$$

Efficient Frontier , we can solve the optimization problems to obtain the efficient frontier by varying the target return R_{target} or Target risk σ_{target}. The curve on the risk-return plane that shows the optimal portfolios for each level of risk or return.

3.8 Data analysis

The data analysis plays a pivotal role, as Saunders and Lewis (2017) reveal, providing valuable insights into financial markets and facilitating the forecasting of stock returns. To achieve this, a range of techniques and procedures are employed. These includes machine learning algorithms, time series analysis, and statistical models such as Autoregressive Integrated Moving Average (ARIMA), Support Vector Regression (SVR), Random Forests, Gradient Boosting, and Long Short-Term Memory (LSTM) networks. By leveraging these quantitative techniques, stock price data which is numerical is analysed and hidden patterns, trends and potential relationships are uncovered to understand the impact they have on share prices. To aid in the exploration and comprehension of the data, various visualization techniques such as graphs and charts are used. These visual representations allow for a more intuitive understanding of the data, facilitating the identification of meaningful patterns and trends. By employing these data analysis techniques, valuable insights that informs asset allocation and optimization processes are generated, ultimately contributing to the achievement of the research objectives.

3.8 Chapter Summary

The chapter provides an overview of the methodology employed in the study. It discusses the use of quantitative techniques, such as machine learning and financial analysis, to analyse financial markets and forecast stock returns. The chapter highlights the collection of historical and the construction of prediction models using time series analysis and machine learning algorithms. It also addresses the research paradigm, target population, data sources, and data analysis tools utilized. Overall, the chapter presents a concise summary of the methodology and analytical approaches employed in the study.

Chapter Four: Results and Discussion

4.1 Datasets Description

The sets of data used for research is from Yahoo Finance which is a complete and trustworthy/reliable source of financial data. The cross-sectional and historical stock price data for the four selected banks in South Africa: Absa, FNB, Nedbank, and Standard Bank is collected from the finance python library for the period January 2009 to January 2024. The rows in the data set represent a trading day while the columns represent the stock performance for the day. The four datasets are downloaded separately, and a stock identifier variable is created in each data set before the merge.

There are 3832 rows in each dataset and 7 columns namely, Open, which represents the opening price of the stock, which is the highest or lowest stock price for the day, close represents the value of the last trade, adj close which is the value of the last stock trade after dividends and stock split, volume which represents the total number of shares that are traded on that particular day and lastly the stock name which identifies the stock amongst the 4 big banks. The final combined dataset for analysis consists of 15328 rows and 7 columns. below is a summary statistic of the data.

	stockname	count	mean	std	min	25%	50%	75%	max
High	**ABG_JO**	3 832	15 076	2 571	7 065	13 656	15 081	16 848	21 100
	FSR_JO	3 832	4 429	1 882	933	2 795	4 814	6 167	7 934
	NED_JO	3 832	19 251	5 187	7 062	14 445	20 229	23 046	31 650
	SBK_JO	3 832	13 926	3 386	6 250	11 090	13 676	16 575	27 410
Low	**ABG_JO**	3 832	14 663	2 542	6 330	13 326	14 671	16 402	20 582
	FSR_JO	3 832	4 314	1 834	849	2 732	4 690	6 018	7 649
	NED_JO	3 832	18 733	5 095	4 484	14 096	19 725	22 507	31 000
	SBK_JO	3 832	13 557	3 296	5 915	10 769	13 294	16 164	22 568
Close	**ABG_JO**	3 832	14 876	2 559	6 330	13 500	14 890	16 613	20 908
	FSR_JO	3 832	4 374	1 859	897	2 765	4 753	6 090	7 775
	NED_JO	3 832	19 004	5 137	6 560	14 300	19 956	22 796	31 300
	SBK_JO	3 832	13 746	3 334	6 015	10 950	13 498	16 349	22 741
Adj Close	**ABG_JO**	3 832	9 011	3 228	3 017	6 359	8 871	10 853	17 753
	FSR_JO	3 832	3 175	1 744	424	1 556	3 218	4 765	6 926
	NED_JO	3 832	12 094	4 577	2 841	8 308	12 412	15 858	20 952
	SBK_JO	3 832	8 996	3 616	2 692	5 768	8 475	11 981	19 354
Volume	**ABG_JO**	3 832	2 467 333	5 500 757	-	1 172 132	1 885 636	2 865 302	240 631 138
	FSR_JO	3 832	13 171 038	8 732 217	-	8 141 971	11 323 103	15 895 118	131 958 509
	NED_JO	3 832	1 243 434	1 046 613	-	654 478	1 021 912	1 530 524	27 844 067
	SBK_JO	3 832	4 015 108	2 315 719	-	2 606 410	3 536 889	4 875 151	26 808 710

The first step carried out in analysis is to clean the data by removing any duplicates across all variables for which 57 records are removed and a total of 15271 records remained, checking for missing data and dealing with non-completeness of records and just general anomalies checks to ensure that the data is reliable to carry out the analysis. The second step is to look at exploratory data analysis which checks the distributions of the stock price and other important variables in the data, to uncover patterns that may exist in the dataset, identify outliers that might impact the analysis and use credible methods to deal with outlier issues.

4.2 Data Exploratory Analysis

This section is focused on the analysis of the data to uncover and understand patterns in the data to help explain the behaviour of the stocks over a long-term period. Visualizing the data will help in identifying key trends, patterns and anomalies that contribute to the stock behaviour. The analysis below will provide valuable information about the stock’s historical performance. The Adjusted Close Price will be used for the analysis because the price considers corporate actions and other adjustments that may affect the historical prices of the stock. It is adjusted to reflect factors such as dividends, stock splits, rights offerings, and other corporate actions and provides a more accurate representation of the true value of the stock over time, especially for long-term analysis.

Figure 4.2.1: Line Graph for Closing Price Overtime

Fig 4.2.1 shows the final closing price for each Bank at the end of a trading day. The Closing price plays a crucial role in stock analysis and is considered the most accurate stock value for the day. Historical trend shows the stock performance over time which highlights the period of growth, stability, and fall since 2010. The period 2010-2016 shows a rising trend which represents a period of growth while the strong peaks and troughs, especially around 2018 indicate market instability brought on by individual events or the economy. A period of fall can be observed between 2020- 2021 which translates to a period of Covid 19. The graph above is very important for understanding the data as it shows the stock's volatility, trends and any cyclical behaviours which are crucial for building forecasting models.

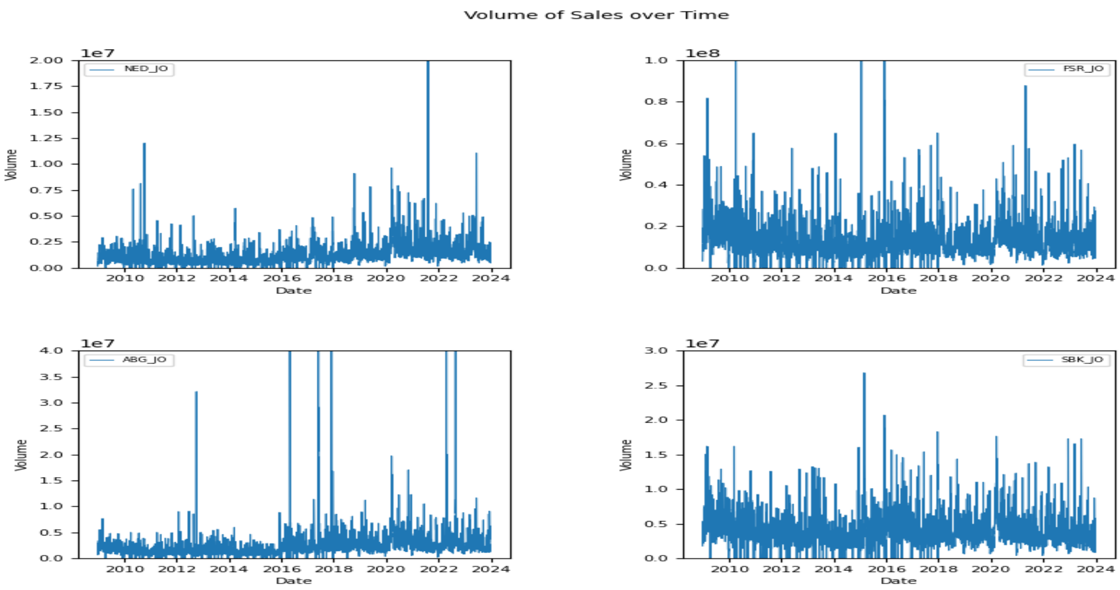


Figure 4.2.2 Volume of Sales Overtime

The daily volume of stock trading volume from 2009 to 2024 is presented in Figure 4.2.2 above. Trading volume is a good indicator of market activity and investor behaviour. The graphs show how many shares are exchanged each day. The periods of different trade volumes are displayed on this chart and show clear and more spikes in FNB and ABSA stocks. The spikes may represent significant news or market events that might have taken place and influenced the behaviour of investors.

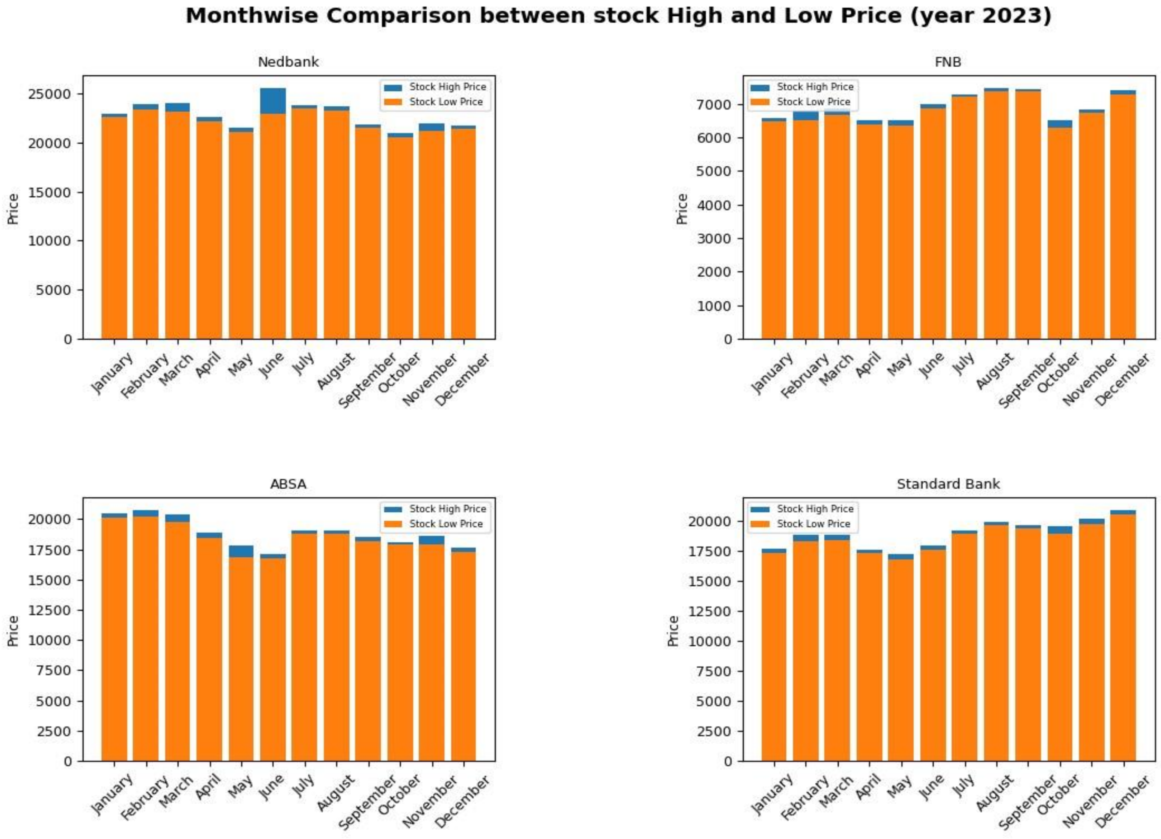


Figure 4.2.3. Bar Chart of High and Low Values of a Stock

A month-by-month comparison of the high and low values of each stock price for the recent year 2023 is shown in Fig 4.2.3. The bar is a representation of each stock price trading range for the month and shows how volatile the stock price was for that particular month. The view is very important for stock research as it gives information about the stability and possible danger of an investment. A case where a larger value is observed i.e. Nedbank's stock price in June 2023, shows uncertainty or noteworthy events or investors moods that may have impacted the company and essentially impacted the stock price.

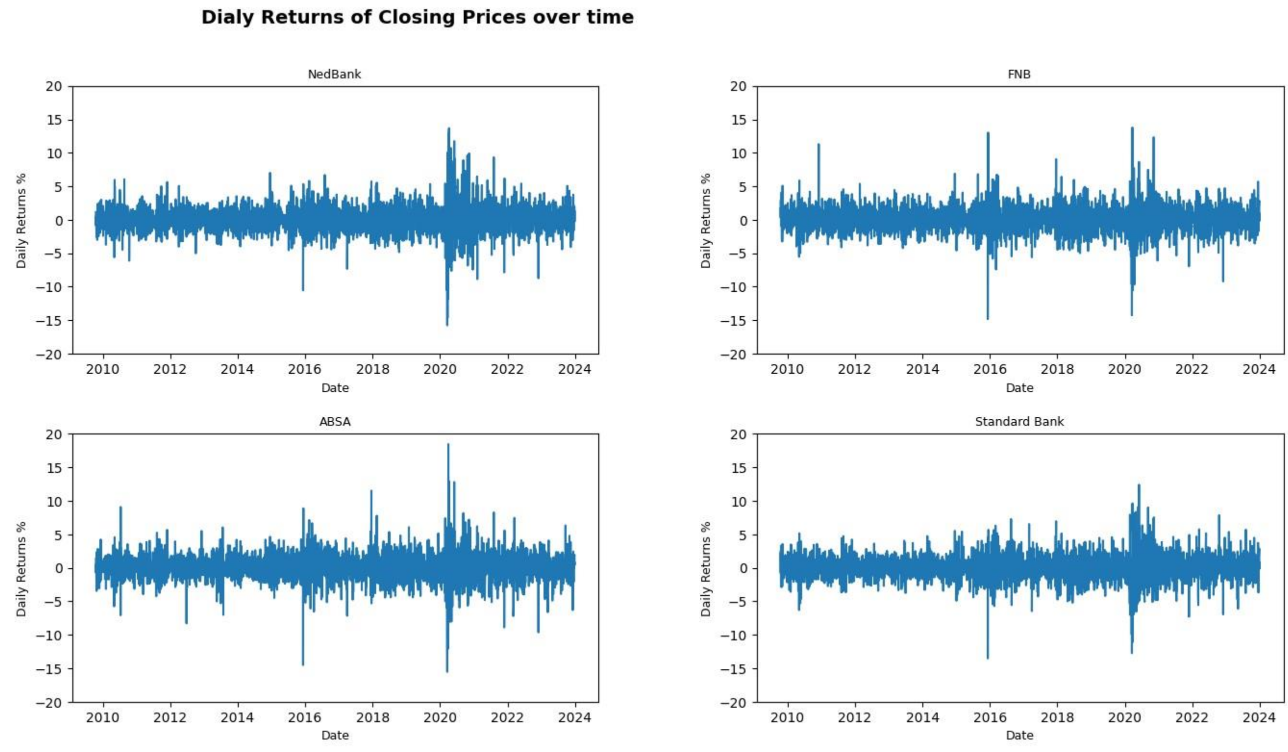


Figure 4.2.4. Daily Return of the Closing Price

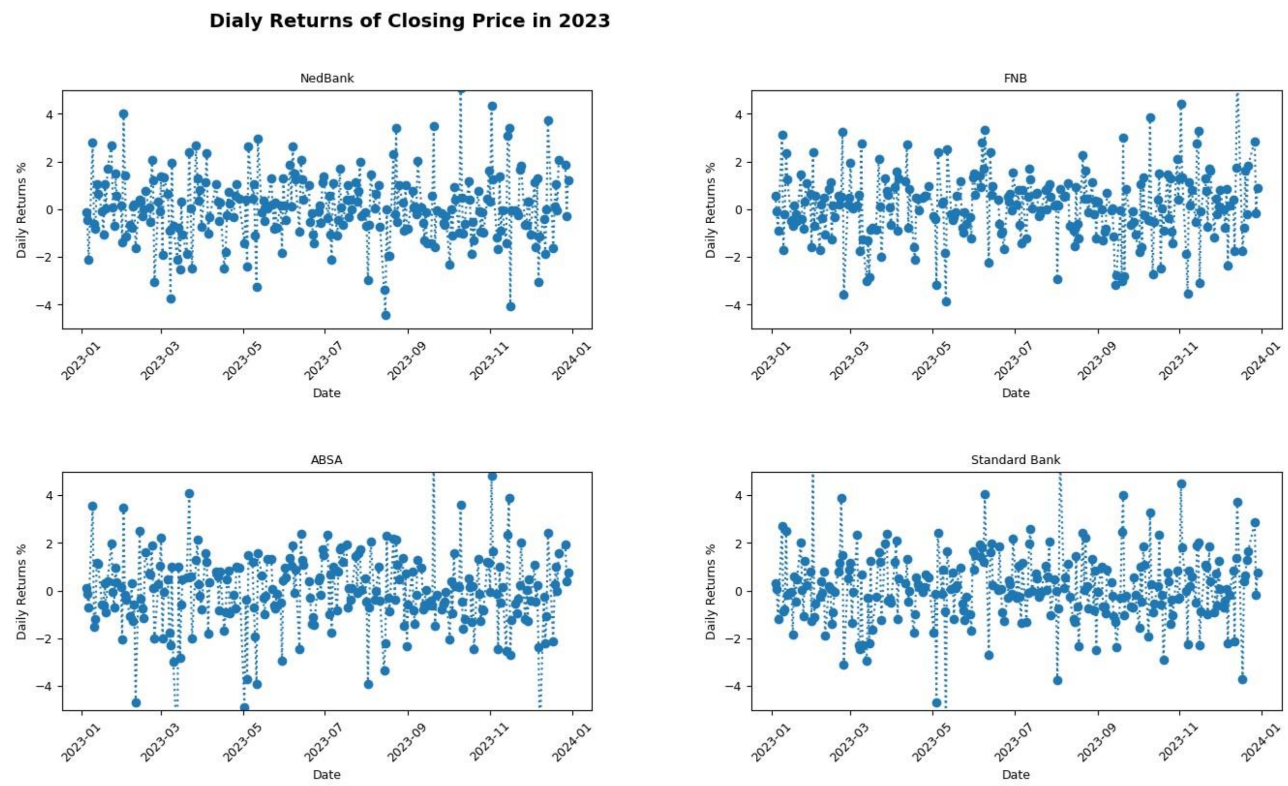


Figure 4.2.5. Daily Returns of Closing Price for 2023

Fig 4.2.4 and 4.2.5 show the daily stock returns over time and for the recent full year of 2023 respectively, the returns represent the percentage change in the price of a stock from one day to the next and measure the stock's volatility. It is an important aspect of stock profitability for investors. The daily return of closing price graphs gives a more comprehensive historical view which makes it easy to identify trends/patterns and abnormalities in the data. Both graphs show the daily return fluctuations which represent variation in stock returns.

The graphs show a high variation for 2016 and the Covid period which may indicate market sentiments as sudden changes happened affecting the stock return. The graph also helps in assessing the level of risk associated with the stock. If the daily returns are consistently high or low, it can help investors assess the level of risk they are willing to take. These visualizations are very important when it comes to risk assessment and decision-making for investors. The stock returns over time graph shows more stability than the recent year which indicates instability possibly resulting from the Covid period. Comparisons of daily returns of stocks help in evaluating the relative performance of investments. Stocks with higher daily returns may outperform those with lower returns over time. Overall, analysing daily returns provides valuable insights into the behaviour and performance of individual stocks and the broader market, helping investors make informed decisions about buying, selling, or holding their investments.

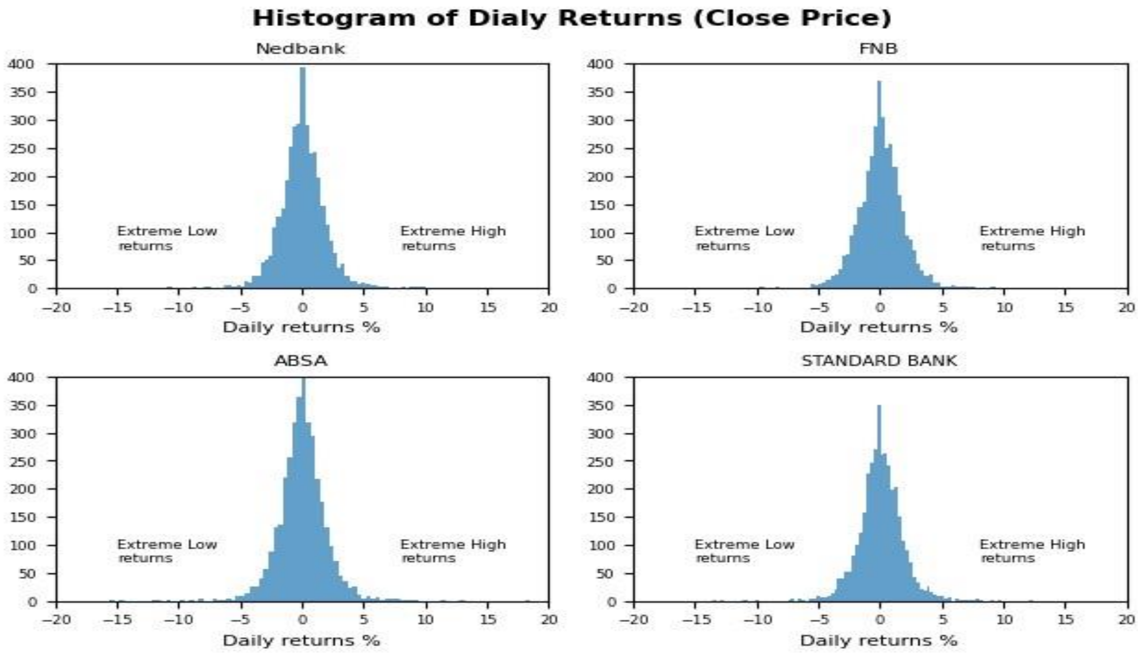


Figure 4.2.6. A histogram of Daily returns over time

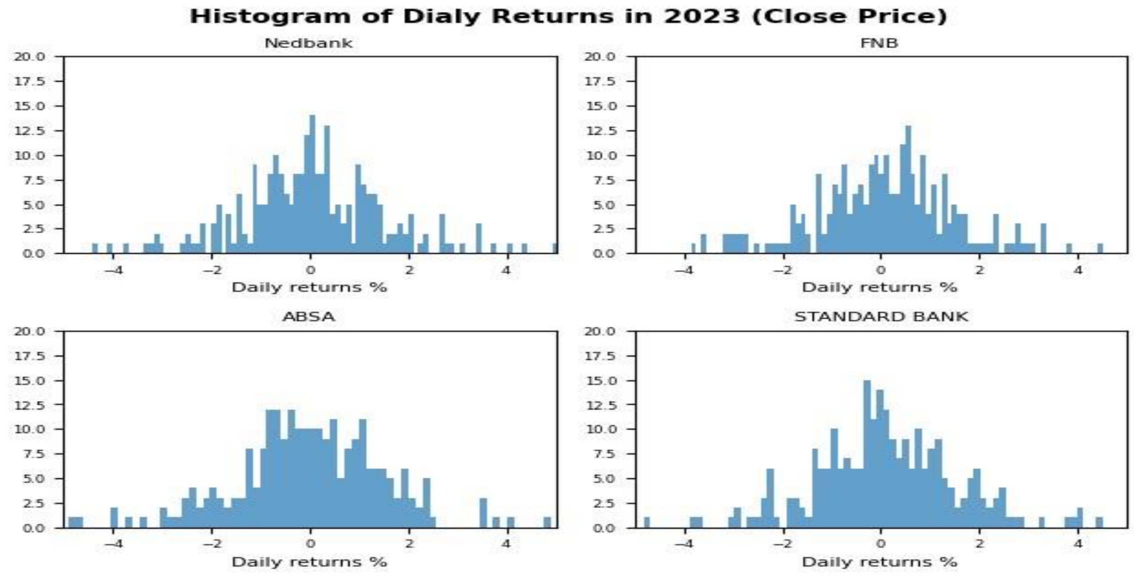


Figure 4.2.7 A Histogram of daily return for 2023

Fig 4.2.6 and 4.2.7 provide a visual representation of the distribution of returns for the stocks over some time. The distribution of returns gives insights into the magnitude and frequency of returns for the stocks. A symmetrical distribution centred around zero suggests that positive and negative returns are equally likely, while skewed distributions indicate a tendency towards either positive or negative returns. Overall, histograms of stock returns provide a visual summary of the distribution of returns, which is useful for understanding the risk, volatility, and central tendency of a stock's performance. It can inform investment decisions and risk management strategies.

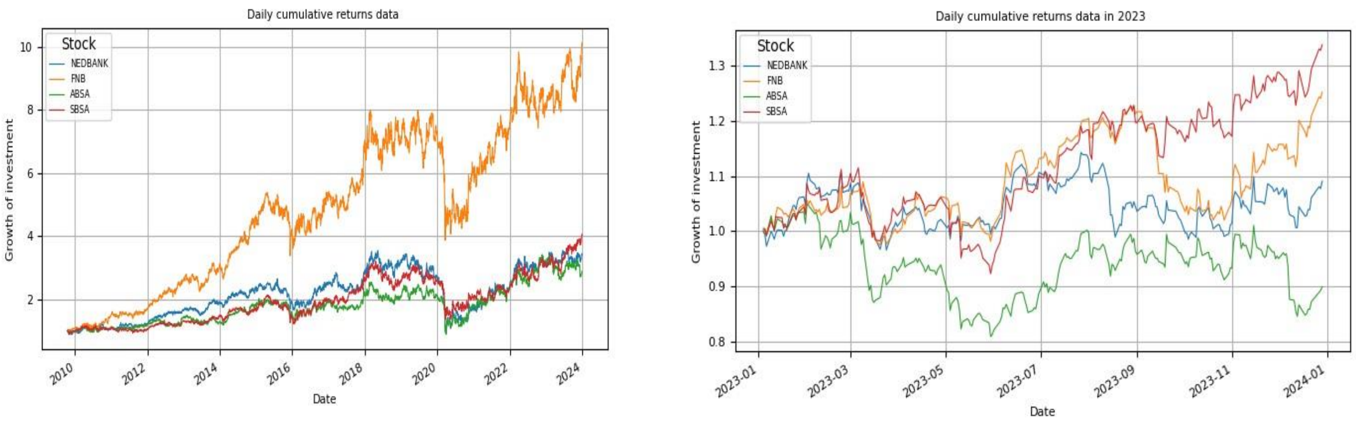


Figure 4.2.8 Daily Cumulative returns over time & for 2023

Figure 4.2.8 provides information about the total amount of return generated by an investment over a specified period. The graphs show that FNB stock has the highest growth generated as compared to the other stocks, proving that the stock has been performing better in comparison to other stocks. However, a declining performance can be observed during the Covid period for all the stocks. The cumulative returns views for the recent period (2023) provide insight into the volatility and risk associated with an investment. A high cumulative return can be associated with high volatility, this can be observed in Standard Bank stock as it is showing higher cumulative returns for the recent period than other stocks.

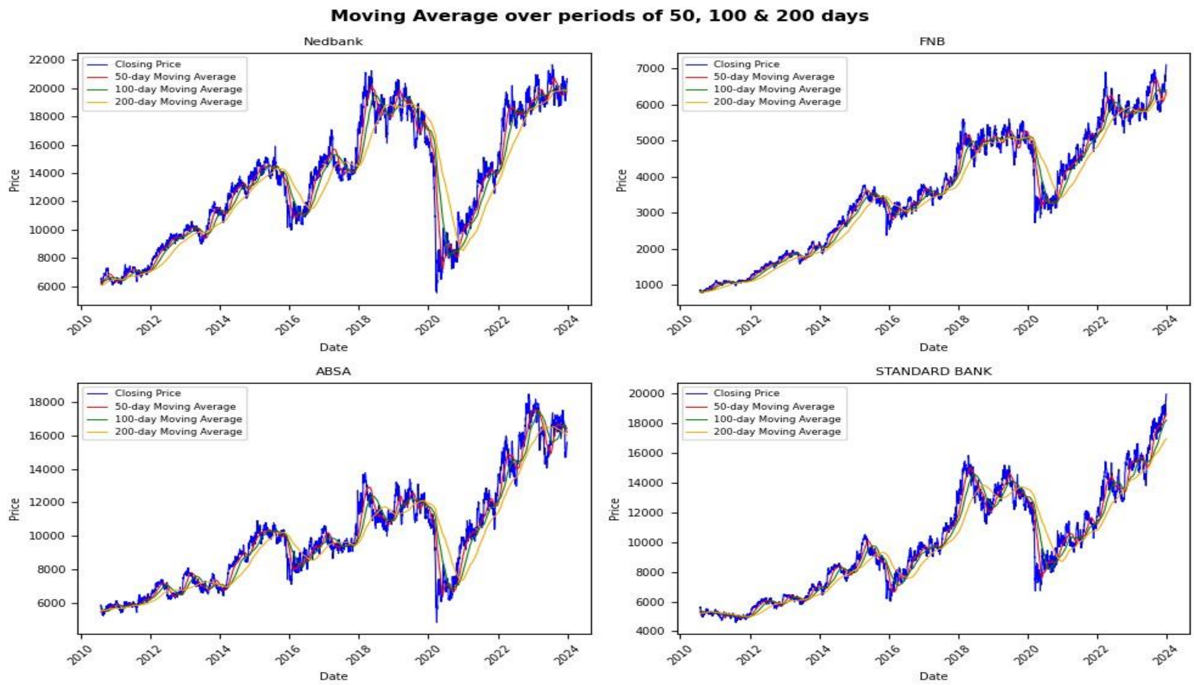


Figure 4.2.9 Line Graph of Moving Average Overtime

Figure 4.2.9 above shows the short-term (50 days); medium-term (100 days), and long-term (200 days) moving averages for the stocks plotted against the daily closing price at predetermined intervals. The graph reveals the long-term pattern and support & resistance levels in the stock. An upward-sloping trend can be observed since covid suggesting a bullish market sentiment. The narrowed gaps between the 50- & 100 days moving averages suggest decreasing volatility while the 200 days moving averages suggest a slight increase in volatility.

4.3 Model Forecasting

4.3.1 Arima Modelling

Part of the requirements before fitting an Arima model is to check for stationarity in the data. ARIMA models require the time series data to be stationary, this means that the mean, variance, and autocovariance structure do not change over time. The use of statistical methods such as the Augmented Dickey-Fuller test will be used to assess the stationarity in the data, the results can be observed below.

Table 4.3.1.1 Arima Model: ADF Stationary Test Results

	ADJ Close		*Daily Returns*		
Stock	**ADF Statistic**	**P-value**	**ADF Statistic**	**P-value**	**Critical Values**
ABG_JO	-1.284455	0.636254	-12.357605	5.640099e-23	-2.862304
FSR_JO	-0.415572	0.907466	-13.073222	1.937686e-24	-2.862303
NED_JO	-1.544918	0.511186	-60.977410	0.000000e+00	-2.862298
SBK_JO	-0.047174	0.954453	-22.645380	0.000000e+00	-2.862299

Based on the results shown in table 4.3.1.1 of the ADF stationarity test, the null hypothesis is rejected since the time series (Adj Close) is non-stationary as the P-value is greater than 0.05. Daily returns will be considered for analysis as it exhibits stationary behaviour as seen in the table above. A stepwise check was run using a library in Python to determine the order of the best Arima Model. The data was split into 64/36 based on date and time. Time series data from 2009-2018 will be used for training the model while time series data from 2018 to present will be used for validation of the model. The time series trend for the train and test split is shown in Figure 4.3.1.1.

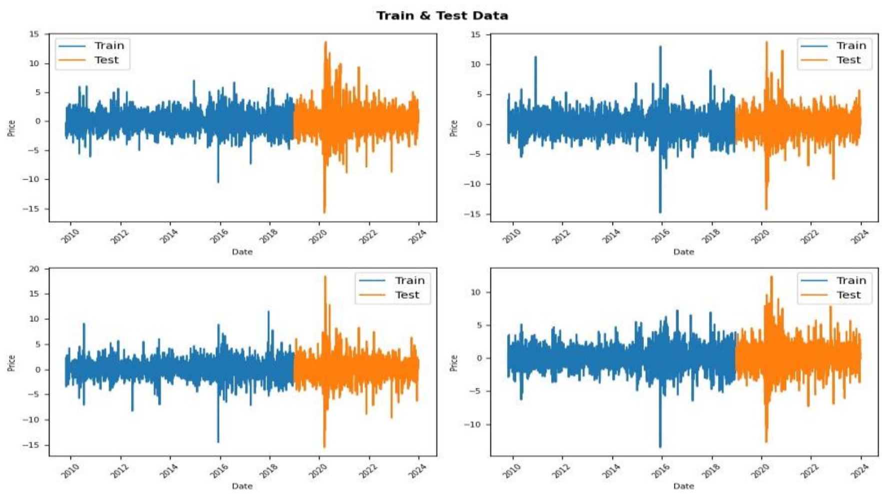


Figure 4.3 1.1 Train & Test Split Returns Trend

Arima model was trained using the following variables Volume, MA_50, MA_100, MA_200, Volatility and Momentum, lag_1, lag_2, lag_3, lag_4, lag_5 and Adj Close as the dependent variable. Creating lagged features generates new columns that contain the stock prices from previous days. The lagged features help the model learn patterns and dependencies in the data, which are crucial for time series forecasting. The model was fitted for each stock and the results of the performance of the model can be observed below in Table 4.3.1.2

Table 4.3.1.2 Arima Model Performance

Stock	Mean Squared Error	RMSE	MAE	R^2
NED_JO	0.87	0.93	0.55	0.86
FSR_JO	15.18	3.89	3.36	-2.368
ABG_JO	1.08	1.04	0.68	0.808
SBK_JO	0.67	0.82	0.54	0.856

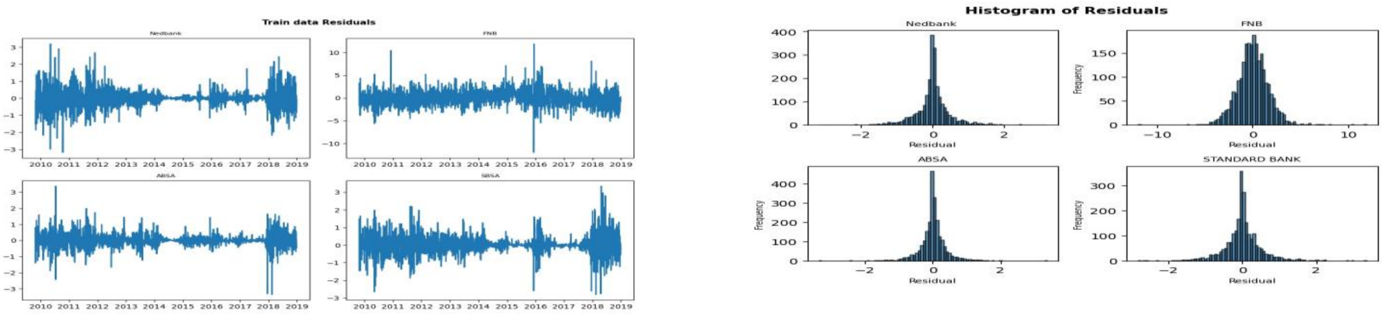


Figure 4.3.1.2 Train Residuals trend and histograms view on Daily Returns

The results above show poor performance of the Arima Models, especially for Fnb stock as it is very evident that the models are overfitting. This could be a result that the data was not scaled. The models also show the performance of traditional statistics models as compared to Machine learning models for forecasting. The residual trend and the histogram of residuals on the train data can be seen below. The patterns suggest poor capturing of the underlying patterns in the data. The R squared shows how much of the data can be explained by the model.

Figure 4.3.1.3 shows actual versus predicted daily returns, it is very evident of overfitting and poor performance of the model in predicting unseen data. Fnb Model shows worse performance and an inability to learn from the data. Although the model can capture the overall trend, it is however failing as it is predicting larger values than actuals and these could impact portfolio performance.

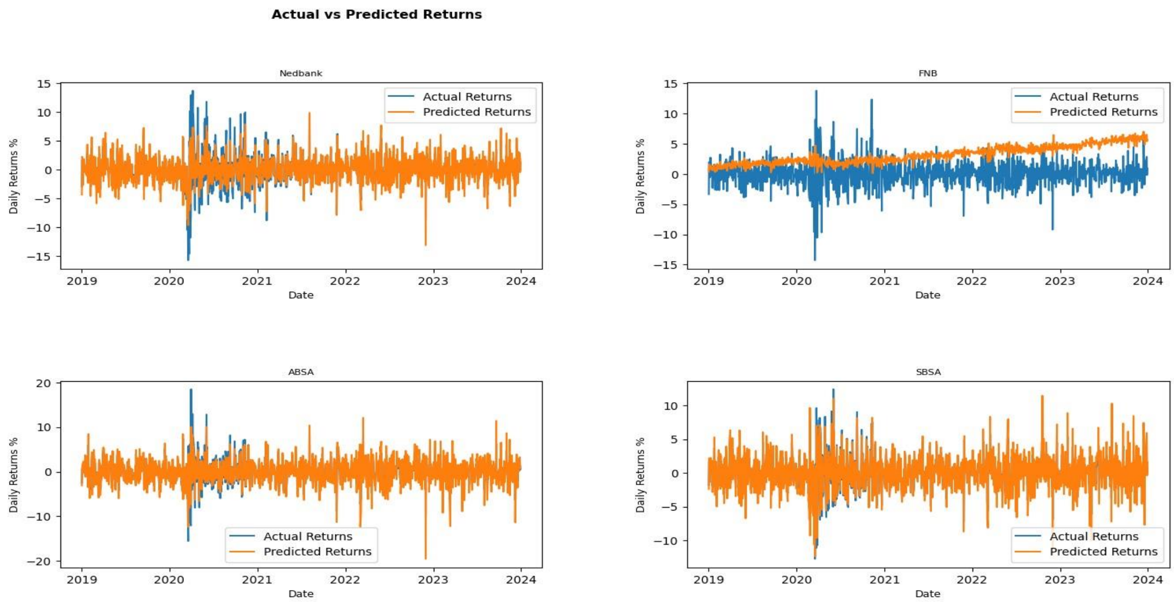


Figure 4.3 1.2 Arima Model: Actual Versus Predicted Daily Returns

4.3.2 Random Forest Modelling

A random forest was used to model the behaviour and trend of stocks, and the same variables used for Arima Modelling were used for all the stocks except for Fnb since it was observed that the models are struggling to capture the pattern in the data. The following variables were used MA_50, MA_100, MA_200 and lag_1 for the FNB stock model to try and improve the performance of the model. As seen below in Table 4.3.2.1 the Random Forest model is performing better than Arima on Fnb stock data.

Table 4.3.2.1 Random Forest: Model Performance

Stock	Mean Squared Error	RMSE	MAE	R^2
NED_JO	181893.409	426.490	277.835	0.990
FSR_JO	303000.265	550.455	365.232	0.682
ABG_JO	4914427.113	2216.851	1432.645	0.521
SBK_JO	949901.689	974.629	479.488	0.881

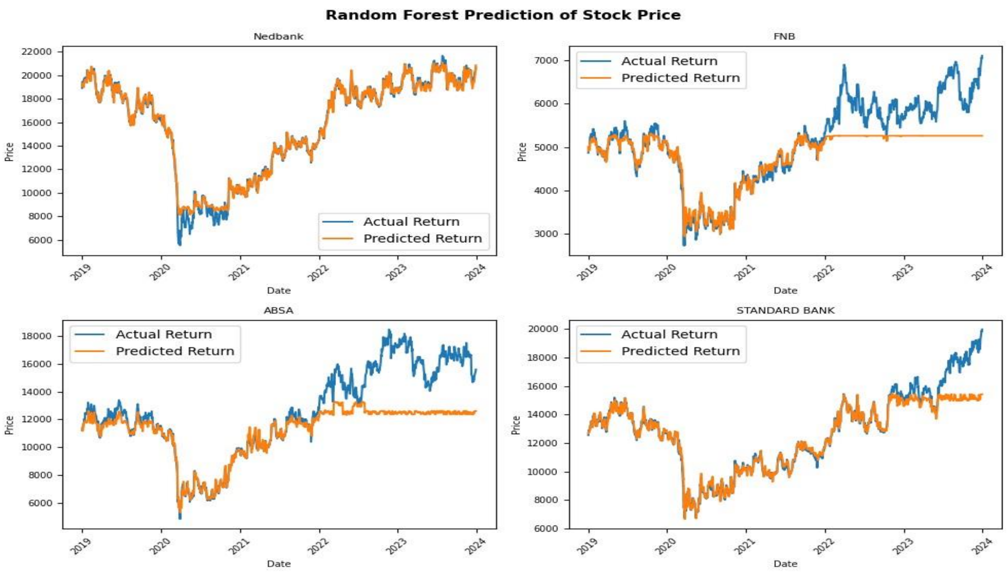


Figure 4.3.2.1 Actual versus Predicted Random Forest Model

Figure 4.3.2.1 shows the actual and the predicted values from the Random Forest model, it is evident that the Nedbank stock model is performing way better than the rest of the models. The models are performing poorly on unseen data which is evident that it is overfitting. Some of the important features can be seen in Figure 4.3.2.1 with just one variable as the important one which also proves that the model is not being able to generalize well.

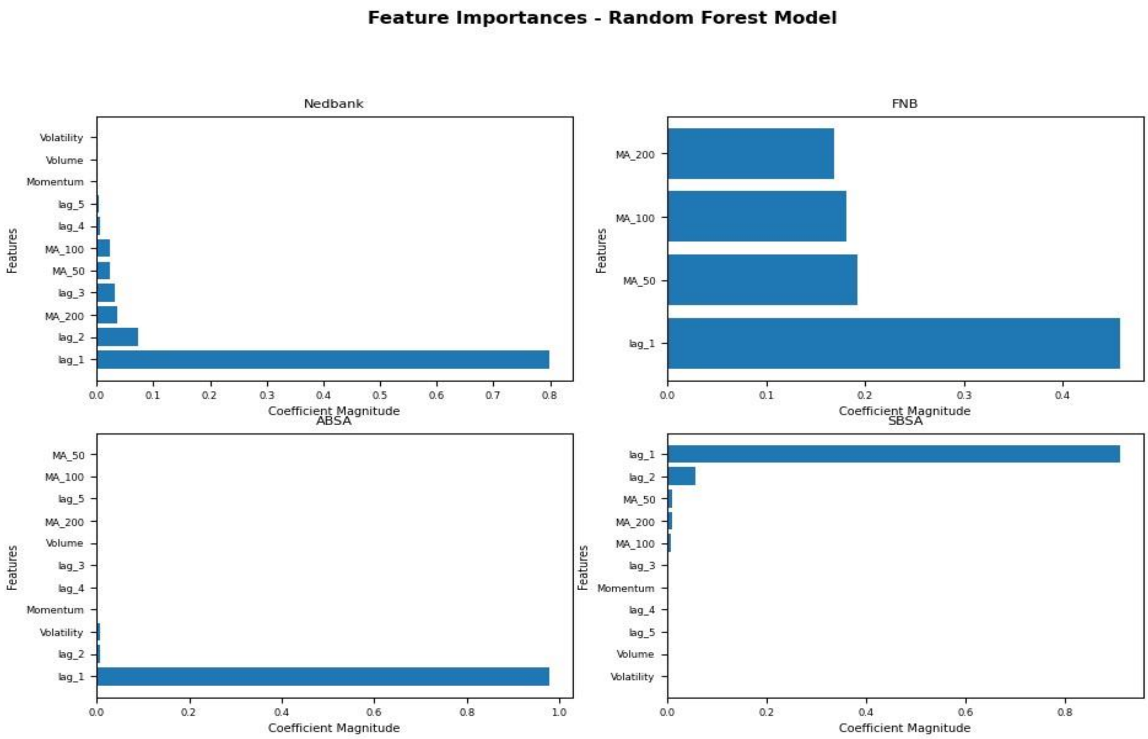


Figure 4.3.2.2 Random Forest Model Feature Importance

4.3.3 Gradient Boosting (XGBOOST) Modelling

A gradient boosting model was built with different parameters for the different models. The results can be seen in Figure 4.3.3.1 with a much better R squared as compared to Arima and Random Forest models. The R squared tells us how much of the data is explained by the model. The Nedbank Stock model performing better as observed previously with other models.

Table 4.3.3.1 Gradient Boosting: Model Performance

Stock	Mean Squared Error	RMSE	MAE	R^2
NED_JO	168776.180	410.824	239.613	0.990
FSR_JO	239765.487	489.659	347.913	0.749
ABG_JO	3612710.603	1900.713	1184.614	0.648
SBK_JO	1036646.993	1018.159	522.930	0.870

The graph below Figure 4.3.3.1 shows the actuals versus predicted from a gradient boost model showing better improvements as compared to the previous models built. Proving the power and capabilities of advanced machine learning techniques over traditional statistical models. Gradient boosting is an advanced random forest model and the distinction in performance for the two models can be seen by the prediction power on unseen data. More data Processing and parameter tuning are required to achieve the optimal results from advanced machine learning models.

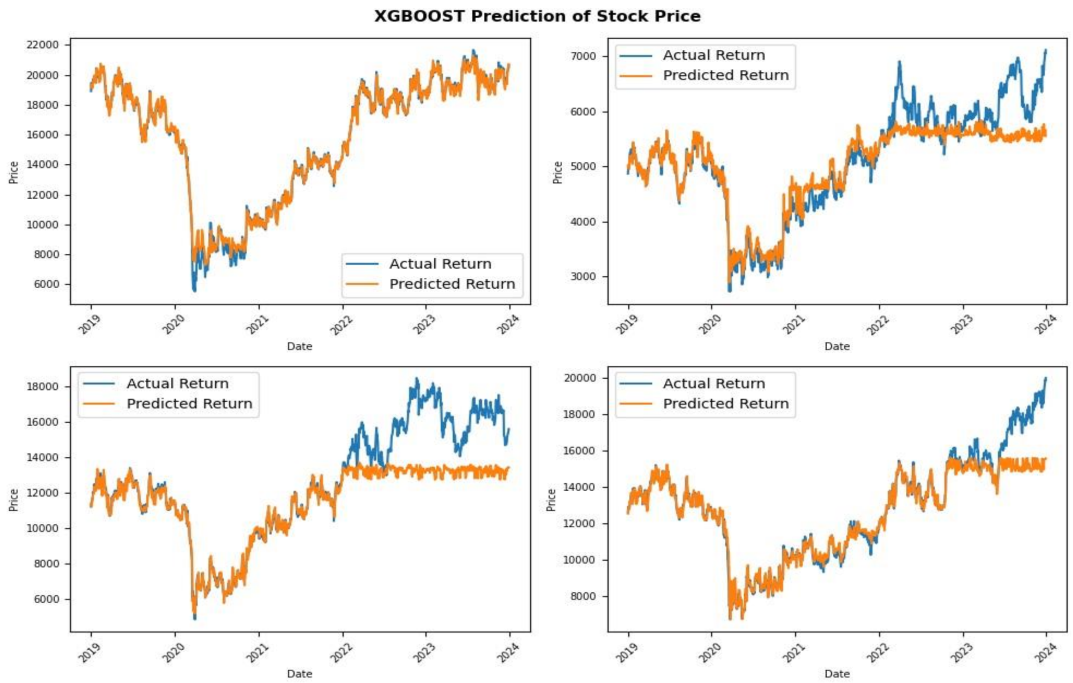


Figure 4.3.3.1 Gradient Boost Model Prediction versus Actual

4.3.4 Support Vector Regression Modelling

Support vector regression is an advanced and powerful technique that extends Support Vector Machines (SVM) to handle continuous outputs. The technique is good for its robustness to outliers, it is capable of dealing with outliers in the model. It also provides several hyperparameters to control model complexity and prevent overfitting and the evidence can be seen in table 4.3.4.1 below with a high R squared indicating the ability of the model to explain most of its underlying patterns.

Table 4.3.4.1 Support Vector Regression: Model Performance

Stock	Mean Squared Error	RMSE	MAE	R^2
NED_JO	151332.271	389.014	276.538	0.991
FSR_JO	27388.583	165.495	124.482	0.971
ABG_JO	30407.883	174.379	134.520	0.997
SBK_JO	43894.987	209.511	156.281	0.995

Support Vector Regression (SVR) can handle non-linear relationships between features and the target variable. Its ability to choose and customize kernels makes it a versatile model for various types of data. Its main objective is to minimize the generalization error by finding a balance between model complexity and the error margin, alternatively leading to great performance on unseen data. This can be observed in Figure 4.3.4.1 where the prediction on unseen data is extremely good as compared to the previous models. The model has great capabilities to learn from historical data and be able to accurately predict future occurrences.

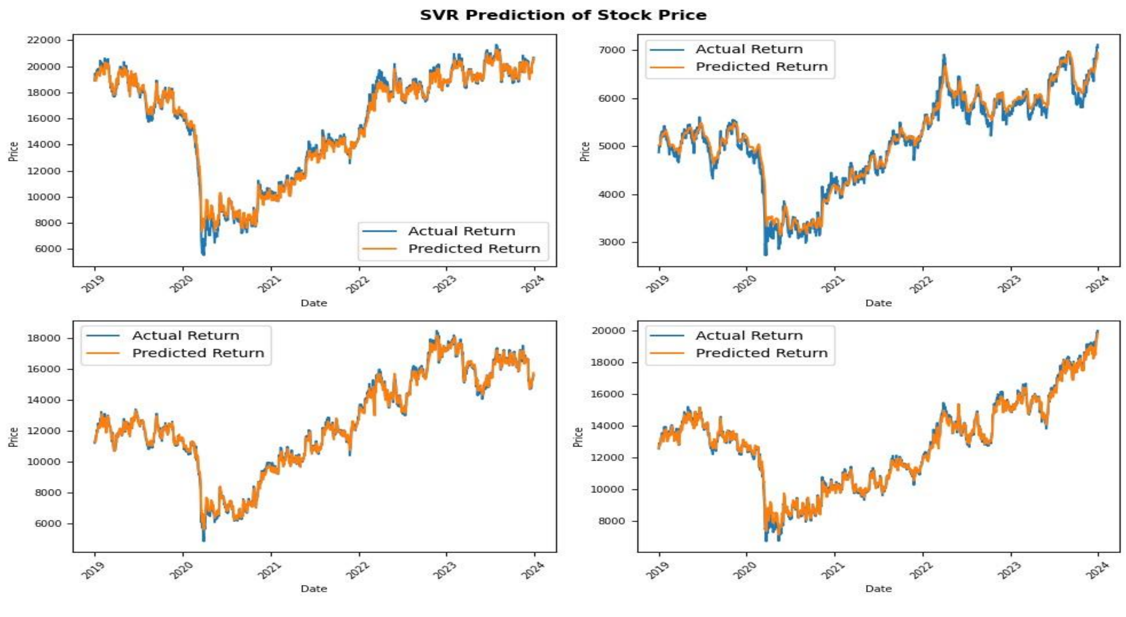


Figure 4.3.4.1 SVR predicted versus Actual.

4.3.5 LSTM Model Prediction

4.3.5.1 Data Preprocessing for LSTM

A sequential structure of stock price data was used for LSTM analysis. The training data was created in the preprocessing stage through an algorithm that trained 100 days' worth of data to predict the price on any given day. The dataset contains 100 days' worth of data from the previous day to be used to train the model and the adj closing price of the current day as the target variable. The process seeks to forecast the price of the stock on the 100th day so that the price from day 0 to 99 is in the training dataset and the price on the 100th day will be in the target dataset. The sequence will continue such that the price on day 250 will be in the train target data set while the price from day 149 to day 249 will be included in the train data set. The structure of the data for LSTM modelling will be set in this way to be able to capture the temporal and sequential nature of stock price data. Because of the great capabilities of LSTM to learn long-term dependencies in the data, The sequential structure allows the model to look back over a specified time window and capture trends and patterns that span across multiple time steps.

The second step in data pre-processing was to scale the data to normalise the range. Neural networks work best with normalised data for faster model convergence during the training of the model and ensure that each feature contributes equally to the final prediction. A python Scikit-learn module, MinMaxScaler was used to scale the data to a range of 0 to 1. The output results were translated back to the original scale by inverting the scaling after predictions.

The final structure after pre-processing was a 3D array for dependent variables in the form of (2266, 100, 11) with 2266 representing the number of samples used for training, 100 denoting the time steps used and 12 representing the number of features included in the model. The target data set remains as a 2D array with (2266, 1) where 1 represents the daily target closing price.

4.3.5.2 Model Prediction

An LSTM model becomes optimal when hyperparameters are tuned essentially producing the best accuracy score. Tuning of hyperparameters includes changing the parameters and iterating over several possibilities to determine which hyperparameter setup offers the best predicted accuracy. It involves choosing the right activation function, the number of LSTM units and the number of layers. several models with different hyperparameters were trained for the four datasets and the assessment of R squared on the test data was considered and the results can be seen in table 4.3.5.1 below. An LSTM model with an activation function linear and units of 80 will be chosen as the final since it performs better.

Table 4.3.5.2.1 LSTM Model with different Hyperparameters

Stock	Model	Units	Activation Function	R2 Score Test
NED_JO	LSTM	80	Linear	0.991
FSR_JO	LSTM	80	Linear	0.955
ABG_JO	LSTM	80	Linear	0.957
SBK_JO	LSTM	80	Linear	0.981
NED_JO	LSTM	40	relu	0.953
FSR_JO	LSTM	40	relu	0.726
ABG_JO	LSTM	40	relu	0.844
SBK_JO	LSTM	40	relu	0.946
NED_JO	LSTM	120	Linear	0.992
FSR_JO	LSTM	120	Linear	0.949
ABG_JO	LSTM	120	Linear	0.970
SBK_JO	LSTM	120	Linear	0.986

The training loss graph in Figure 4.3.5.2.1 shows the training loss graph for the models. It shows greater loss values at the beginning and at some point, in time which suggests that predictions are not so accurate. It shows that the loss gradually drops showing the model's capacity to absorb information. Absa and Standard bank models show no convergence to an optimal solution based on the parameter chosen, so more training and parameter tuning will be required to improve the model even better. Increasing the epochs will show convergence at a certain iteration number, Nonetheless, the model still performs better. It is observed that increasing the units for a linear activation function does have insignificant effect on R squared.

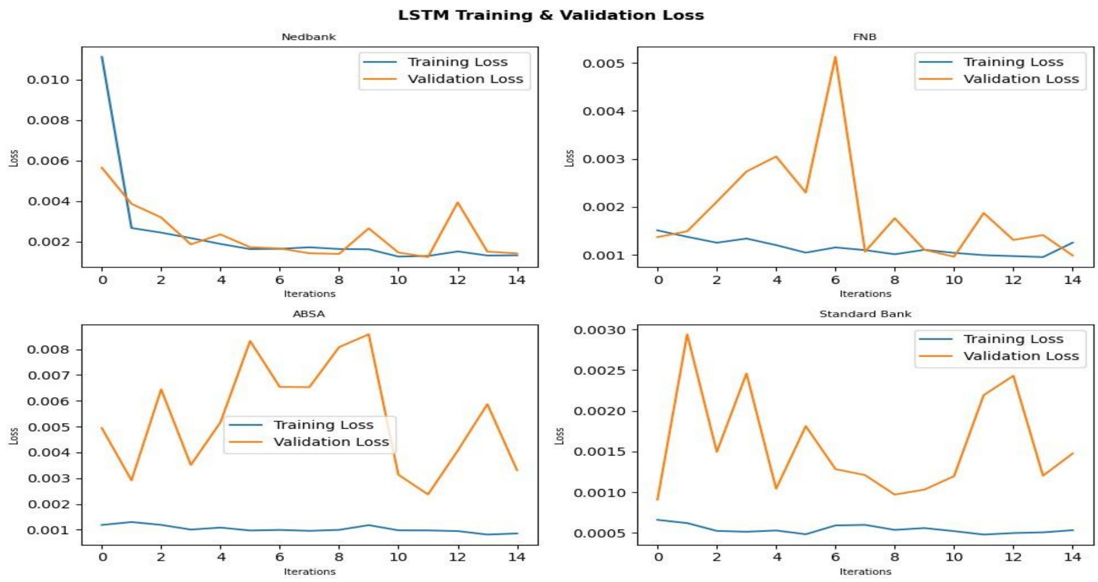


Figure 4.3.4.2.1 LSTM Training & Validation Loss

Assessing the LSTM model's Performance on test data is important to determine its ability to generalize well on unseen data. The results for the LSTM models are shown in Figure 4.3.5.2.2. The close adherence to the predicted values shows the ability of the model to generalize well. The models were able to capture the trend and the swings in the actual stock price. Although there are several points where the predicted and actual values disagree, this offers crucial information for when the model will need adjustments in the future impacted by market activities that the model could not account for.

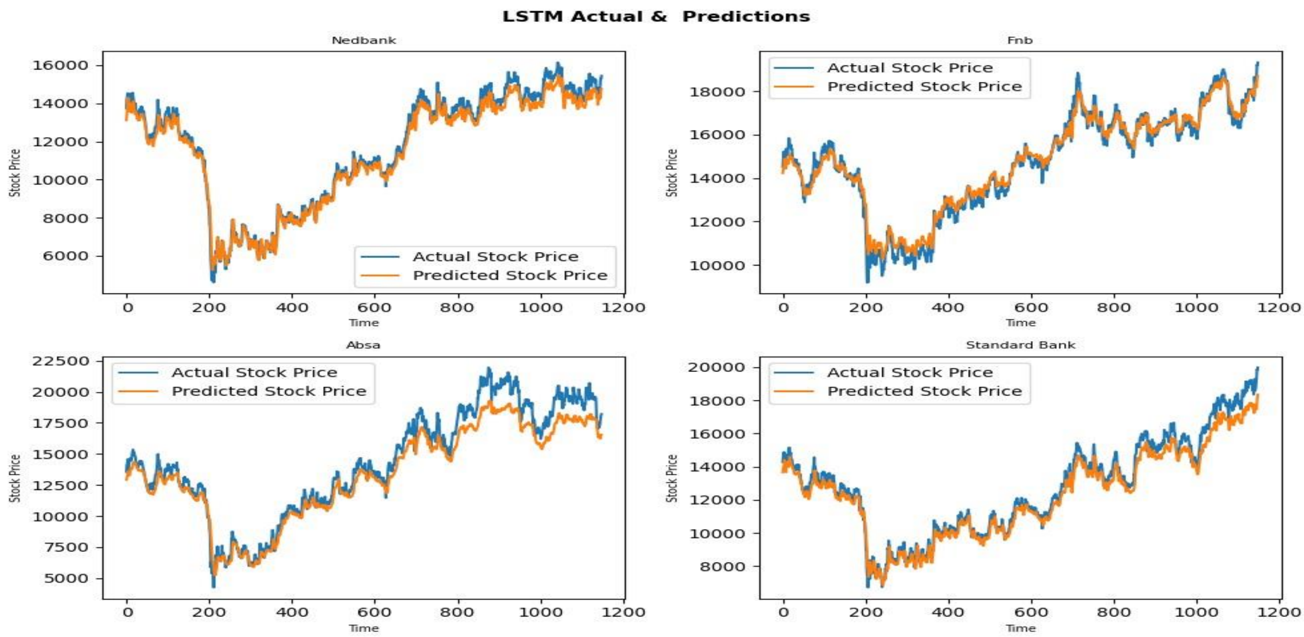
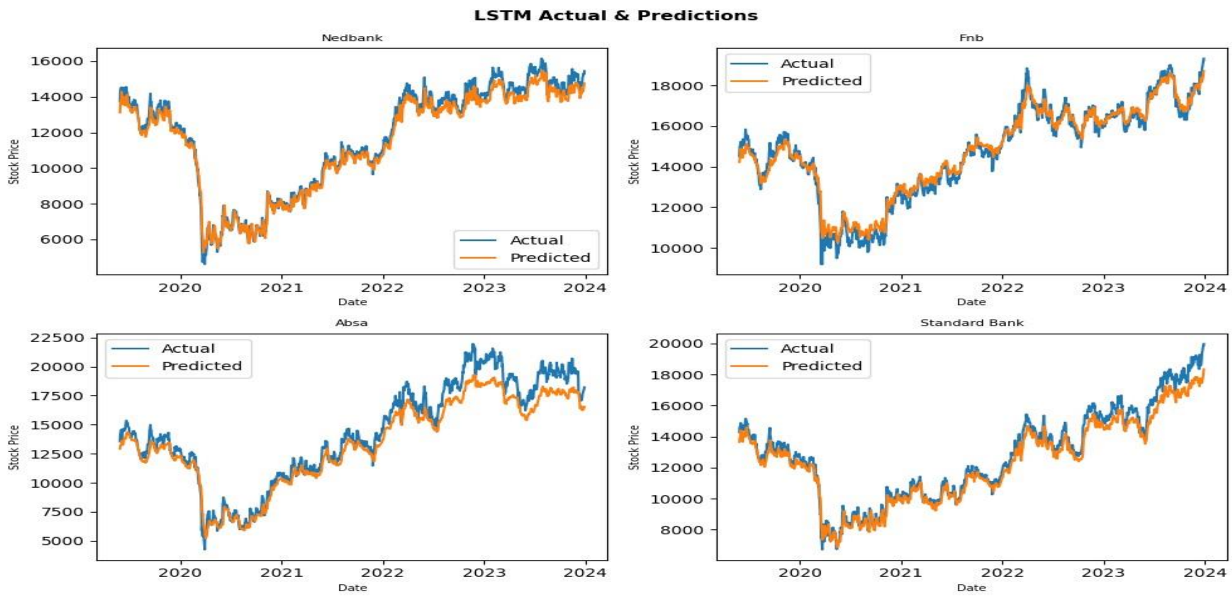


Figure 4.3.5.2.2 LSTM Prediction versus Actual



4.3.5.2.3 LSTM Prediction versus Actual

The overall model performance for all the various models built shows an outstanding performance on SVR and LSTM. An LSTM could do better with more hyperparameter tuning and more iterations over the data. Computational powers were a limit factor for deep learning of the LSTM. Because of its ability to learn from historical data and to predict future patterns, it will be considered the final chosen model that will be used to predict outputs that will be used for portfolio construction in the next step. The LSTM model is very useful as a tool for financial analysis and decision-making because of its capacity to maintain a constant level of prediction quality over various periods.

4.4 Portfolio Construction

4.4.1 Alpha and Risk Models

A comprehensive investment analysis and decision-making is crucial for any portfolio construction and optimalisation task. Selecting the right techniques to assist with portfolio construction yields more benefits and is essential for risk management. This research suggested the use of Alpha and Risk model to inform the stocks that will be used to construct an investment Portfolio. Alpha in the context of finance refers to the excess return of an investment relative to a benchmark index. It measures the performance of an investment strategy or portfolio after adjusting for market risk. A positive Alpha Value shows that the strategy generates higher return than the benchmark. Alpha is also used to evaluate the effectiveness of different forecasting models. A positive Alpha is always desirable. Assessing The analysis will use Sharpe Ratio for Portfolio performance measure since Alpha model is well suited when comparing portfolios against benchmark. Sharpe Ratio, Risk in the form of measuring volatility and expected return will be used as performance metrics for choosing the best portfolio.

A risk model is a quantitative tool used to assess and quantify the risks associated with investments or portfolios. It plays a major role for investors to understand the downside and volatility of investments through analysing various types of risks associated with portfolios. Risk models are very important for portfolio construction and optimization because there is a desire to construct risk-adjusted returns portfolios. Overall, risk models play a crucial role in investment analysis, portfolio management, and risk management as they provide investors with valuable insights into the risks associated with their investments and help them make informed decisions to achieve their investment objectives while managing risk effectively.

4.4.2 Portfolio Construction and Optimization Actual Returns

Historical Stock Price data from Jan 2023 to Jan 2024 will be used for portfolio construction and optimization, the chosen performance will be assessed, and the best portfolio chosen as the final one. The process will include comparing actual returns and predicted returns from the LSTM model to validate the model. The number of Trading days is assumed to be 252 and the process looks at building 5000 portfolios, where the optimal portfolio will be chosen based on the set performance metrics. The weight on the portfolio will be randomly selected in assessing the performance of Portfolios.

Table 4.4.2.1 Show the Portfolio performance metrics where stocks from major banks in south Africa namely, Nedbank, FNB, ABSA and Standard Bank were considered and assessed for Portfolio construction. The FNB and Standard Bank Stock were chosen with the respective weight in the portfolio of 0.119 and 0.881 respectively. The optimal portfolio will yield a 28% expected return with 23%, a 23% risk measure and 1.2 Sharpe ratio. The visual representation can be seen in figure 4.4.2.1 below.

Table 4.4.2.1 Optimal Portfolio Performance Metrics Actual Values

Optimal portfolio ['Nedbank','FNB','ABSA','SBSA']: [0. 0.119 0. 0.881]
Expected return, volatility and Sharpe ratio: [0.28816461 0.23996159 1.20087803]

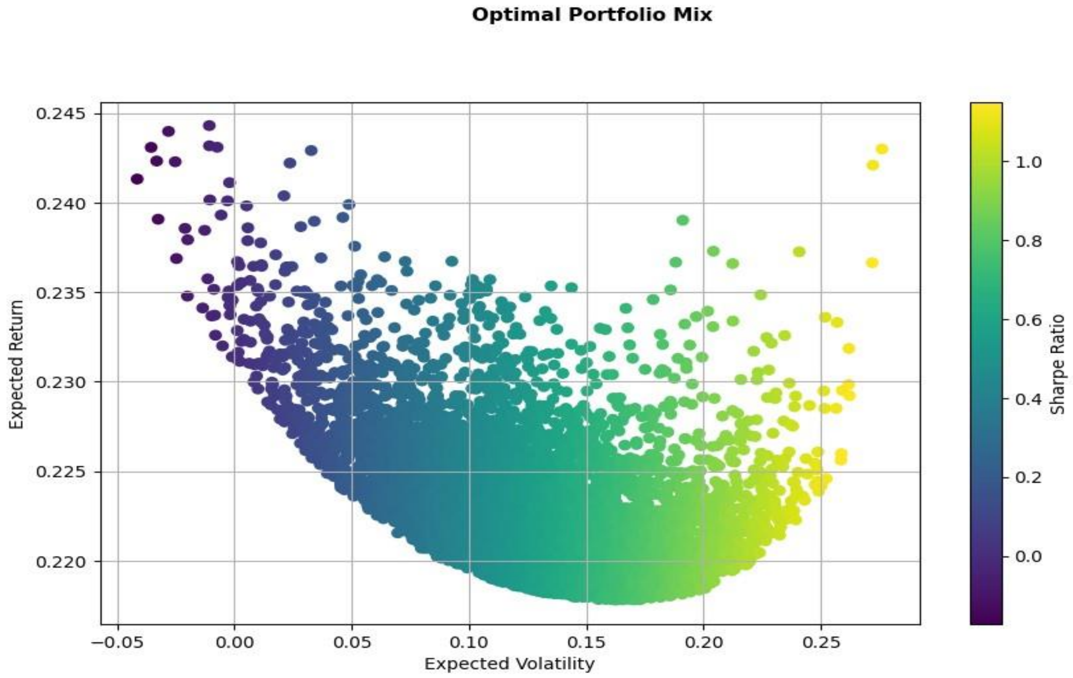


Figure 4.4.2.1 Visual Representation of Optimal portfolio mix Actual Returns.

Using Predictions from the LSTM model yields different metrics for an optimal Portfolio even though it selects the same stocks as with Actuals, With Predictions FNB Stock has more weight than Standard Bank in the portfolio with a 67% and 33% split. Expected return and Volatility are Low compared to the Actuals with Sharpe Ratio higher with predictions than Actuals. The Performance Metrics can be seen in Table 4.4.2.2 below.

Table 4.4.2.2 Optimal Portfolio Performance Metrics Predicted Values

Optimal portfolio ['Nedbank','FNB','ABSA','SBSA']: [0. 0.674 0. 0.326]
Expected return, volatility and Sharpe ratio: [0.17945088 0.11543929 1.55450437]

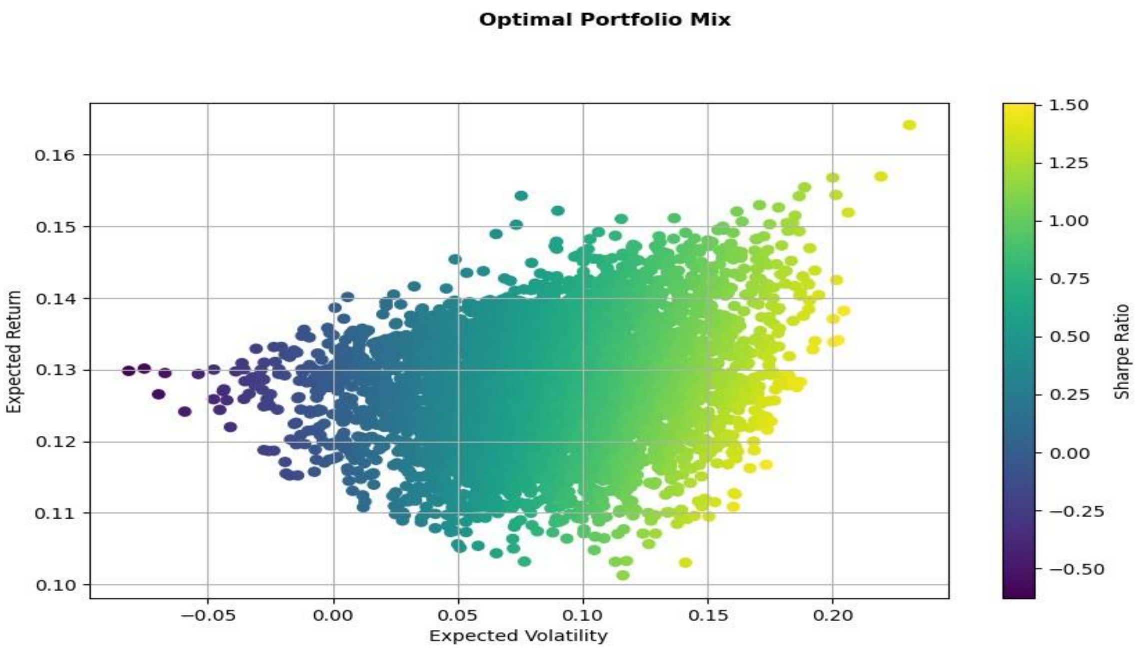


Figure 4.4.2.2 Visual Representation of Optimal portfolio mix Predicted Returns.

The table and figure above suggest that investors should consider investing in FNB and Standard bank shares since they yield and form part of the optimal portfolio from the forecasted stock returns. The average return for this portfolio mix is roughly 18% which represent the profit that an investor may earn on average by holding the portfolio over a certain period given the associated risk. This is the maximum return that is achievable for the chosen risk level of roughly 11% . The volatility gives an indication of how much the portfolios return might vary from the expected returns over time. A lower Volatility suggests a favourable risk-return profile. This portfolio will yield a higher return than its level of risk, this is an attractive balance for investors. A high expected return relative to a low volatility indicates that the portfolio construction is efficient due to a good diversification and an inclusion of low risk, high return assets. This portfolio will provide a higher potential reward for each unit of risk, which essentially makes it efficient to deliver returns without excessive risk exposure. The Sharpe ratio of 1.55 is observed which measures the risk-adjusted return. When we assume a risk-free rate, a higher expected return than volatility yield a higher sharper ratio which is what we observe in this optimal portfolio. The portfolio has potential to generate 1.55 units of return for every unit of risk it takes on after the risk-free rate adjustment. A Sharpe ratio of 1.55 gives indicates a good risk-adjusted performance, a higher return relative to volatility and a good benchmark for comparison. This type of portfolio is mostly attractive to risk- sensitive investors as it offers a solid balance of risk and return.

Chapter five: Conclusion

5.1 Summary of findings

Modelling techniques like traditional statistical techniques and Machine learning techniques have been developed to solve many Prediction/Forecasting problems. Traditional forecasting statistical methods are most used in finance. The introduction of machine learning in stock prediction has become a very popular alternative in forecasting because of its associated memory characteristics and its generalization capabilities. Although there are critics of machine learning techniques, there are far greater benefits to using machine learning than traditional techniques. The main aim of this study was to compare the prediction power of traditional and machine learning techniques in stock forecasting. A better prediction was achieved with more advanced machine-learning techniques like SVR and LSTM. The algorithms were tested rigorously for financial market trend predictions. This research concludes that both SVR and LSTM are more capable of uncovering and learning patterns in stock market data for future predictions. Differences between various algorithms and models were highlighted in the research including ARIMA modelling. The research provides a more comprehensive viewpoint based on the various approaches used in financial forecasting and its advantages.

Building the optimal portfolio from the different model's namely, support vector machine (SVR) and Neural network (LSTM) outputs produces very different results. Both Model's output into the risk model and portfolio construction chooses FNB and standard banks share price for portfolio maximization with different weightings. Comparison between the two optimal portfolios gives different expected returns, volatility and Sharpe ratio with SVR model outputs showing high values for expected return and volatility and a lower Sharpe return. A portfolio with a higher Sharpe ratio will be considered as the final optimal chosen portfolio since a higher Sharpe ratio is preferred.

5.2 Limitations of the study

a. **Data Quality and Quantity**: The data used is obtained from a public data source and may not be well maintained which could compromise the quality of the data.
b. **Cost of Data**: The data could be enriched with Companies financials and market data for better prediction and there was a cost to getting the data.
c. **Computational Costs and complexity**: advanced models require more parameter tuning and this is time consuming and computationally intensive.
d. **Feature Dependence:** The performance of many models is highly dependent on the features used. Poor feature selection can lead to poor model performance as seen with the less advanced machine learning techniques (Random Forest and Gradient Boosting) and statistical technique (ARIMA)
e. **Feature Engineering**: creating features to enhance the models can be time-consuming and requires domain expertise as one needs to create new variables and do thorough checks and test on the fit of the variables on the data.
f. Functions for Analysis: Not using functions to run the analysis was time Consuming as each model had to be built separately. To also have a clear and short python Code that produces required results.

5.3 Areas of Improvement

The study on stock price prediction acknowledges limitations that can improved for future studies in order to reap the benefits of these advanced techniques like SVR and LSTM. Financial market unpredictability is also a major limitation as stock price is highly dependent on several factors such as political events, market sentiments and economic indicators which makes it difficult to accurately quantify and predict future stock prices. The market is sensitive to unexpected events for which models are not able to accurately capture. Market data such as economic indicators can be studied and included in the modelling process and assess the impact it will have on stock predictions. Creating functions in python for analysis will save time and avoid redundancy in the process, therefore creating a precise and clear

Python Code.

Further research on global economic indicators, news sentiments analysis and social media trends can be incorporated to enhance and enrich data for better prediction power of stock prices. Interpretability of models can also be improved so we can better understand the model structures in Python since most models are "black box" in nature which makes prediction a harder to understand. Using ensemble methods could also improve the study as multiple predictive techniques can be combined to mitigate weakness and improve accuracy. The research has made significant progress in applying machine learning to stock price prediction, but financial markets remain complex. Future research must address these limitations to advance the field and maximise machine learning in financial forecasting.

References

1. Ahmadi Abkenari, F., Milani Fard, A., & Khanchi, S. (2023). Hybrid Machine Learning-Based Approaches for Feature and Overfitting Reduction to Model Intrusion Patterns. *Journal of Cybersecurity and Privacy*, *3*(3), 544-557.
2. Ahmed, S., Alshater, M. M., El Ammari, A., & Hammami, H. (2022). Artificial intelligence and machine learning in finance: A bibliometric review. *Research in International Business and Finance*, *61*, 101646.
3. Aithal, P. K., Geetha, M., Dinesh, U., Savitha, B., & Menon, P. (2023). Real-Time Portfolio Management System Utilizing Machine Learning Techniques. *IEEE Access*, *11*, 32595-32608.
4. Almeida, J., & Gonçalves, T. C. (2023). A systematic literature review of investor behavior in the cryptocurrency markets. *Journal of Behavioral and Experimental Finance*, 100785.
5. Al-Muharraqi, M., & Messaadia, M. (2023, January). Implementing Machine Learning in Optimizing Stock Portfolios: A review. In *2023 International Conference On Cyber Management And Engineering (CyMaEn)* (pp. 500-504). IEEE.
6. Alsharif, M. H., Kelechi, A. H., Yahya, K., & Chaudhry, S. A. (2020). Machine learning algorithms for smart data analysis in internet of things environment: taxonomies and research trends. Symmetry, 12(1), 88.
7. Angra, S., & Ahuja, S. (2017). Machine learning and its applications: A review. In 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC) (pp. 57-60). IEEE.
8. Artzner, Philippe, Freddy Delbaen, Eber Jean-Marc, and David Heath. 1999. Coherent measures of risk. *Mathematical Finance* 9: 203–28.
9. Ashtiani, M. N., & Raahmei, B. (2023). News-based intelligent prediction of financial markets using text mining and machine learning: A systematic literature review. *Expert Systems with Applications*, 119509.
10. Behera, J., Pasayat, A. K., Behera, H., & Kumar, P. (2023). Prediction based mean-value-at-risk portfolio optimization using machine learning regression algorithms for multi-national stock markets. *Engineering Applications of Artificial Intelligence*, *120*, 105843.
11. Bengio, Y. & LeCun, Y. (2015). Deep Learning. *Nature*, *521*(7553), 436–444.
12. Bihari, A., Dash, M., Muduli, K., Kumar, A., Mulat-Weldemeskel, E., & Luthra, S. (2023). Does cognitive biased knowledge influence investor decisions? An empirical investigation using machine learning and artificial neural network. *VINE Journal of Information and Knowledge Management Systems*.
13. Bitetto, A., Cerchiello, P., & Mertzanis, C. (2023). Measuring financial soundness around the world: A machine learning approach. *International Review of Financial Analysis*, *85*, 102451.

14. Bluwstein, K., Buckmann, M., Joseph, A., Kapadia, S., & Şimşek, Ö. (2023). Credit growth, the yield curve and financial crisis prediction: Evidence from a machine learning approach. *Journal of International Economics*, 103773.
15. Box, G. & Jenkins, G. (1970). Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco.
16. Brauneck, A., Schmalhorst, L., Kazemi Majdabadi, M. M., Bakhtiari, M., Völker, U., Saak, C. C., ... & Buchholtz, G. (2023). Federated machine learning in data-protection-compliant research. *Nature Machine Intelligence*, *5*(1), 2-4.
17. Candanedo, I. S., Nieves, E. H., González, S. R., Martín, M. T. S., & Briones, A. G. (2018). Machine learning predictive model for industry 4.0. In *Knowledge Management in Organizations: 13th International Conference, KMO 2018, Žilina, Slovakia, August 6–10, 2018, Proceedings 13* (pp. 501-510). Springer International Publishing.
18. Chen, L. and Liu, Y. (2009). A Support Vector Machine Approach to Credit Risk Assessment. Expert Systems with Applications, 36(2), pp.2529-2536.
19. Chen, Z., Van Khoa, L. D., Teoh, E. N., Nazir, A., Karuppiah, E. K., & Lam, K. S. (2018). Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: a review. *Knowledge and Information Systems*, *57*, 245-285.
20. Chou, S. Y., Lin, C. W., Chen, Y. C., & Chiou, J. S. (2023). The complementary effects of bank intangible value binding in customer robo-advisory adoption. *International Journal of Bank Marketing*, *41*(4), 971-988.
21. Chu, H., & Ke, Q. (2017). Research methods: What's in the name? *Library & Information Science Research*, 39(4), 284-294.
22. Ciampiconi, L., Elwood, A., Leonardi, M., Mohamed, A., & Rozza, A. (2023). A survey and taxonomy of loss functions in machine learning. *arXiv preprint arXiv:2301.05579*.
23. Munkhdalai, L., Munkhdalai, T., Namsrai, O. E., Lee, J. Y., & Ryu, K. H. (2019). An empirical comparison of machine-learning methods on bank client credit, N., Bilson, C., Gepp, A., Stern, S., & Vanstone, B. J. (2023). Modernising operational risk management in financial institutions via data-driven causal factors analysis: A pre-registered report. *Pacific-Basin Finance Journal*, *77*, 101906.
24. Creswell, J. W., & Hirose, M. (2019). Mixed methods and survey research in family medicine and community health. *Family Medicine and Community Health*, 7(2).
25. Cui, Y. (2023). The Recent Progress and State-Of-Art Applications of CAPM Model. *Highlights in Science, Engineering and Technology*, *49*, 432-437.
26. Dash, B., Swayamsiddha, S., & Ali, A. I. (2023). Evolving of Smart Banking with NLP and Deep Learning. In *Enabling Technologies for Effective Planning and Management in Sustainable Smart Cities* (pp. 151-172). Cham: Springer International Publishing.

27. Delcey, T., & Sergi, F. (2023). The efficient market hypothesis and rational expectations macroeconomics. How did they meet and live (happily) ever after?. *The European Journal of the History of Economic Thought*, *30*(1), 86-116.
28. Dezhkam, A., & Manzuri, M. T. (2023). Forecasting stock market for an efficient portfolio by combining XGBoost and Hilbert–Huang transform. *Engineering Applications of Artificial Intelligence*, *118*, 105626.
29. Dooley, K. (2017). Simulation research methods. The Blackwell companion to organizations, 829-848.
30. Emerson, S., Kennedy, R., O'Shea, L., & O'Brien, J. (2019, May). Trends and applications of machine learning in quantitative finance. In *8th international conference on economics and finance research (ICEFR 2019)*.
31. Fahmy, H. (2020). Mean-variance-time: An extension of Markowitz's mean-variance portfolio theory. *Journal of Economics and Business*, *109*, 105888.
32. Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance*, *25*(2), 383–417.
33. Faridi, S., Madanchi Zaj, M., Daneshvar, A., Shahverdiani, S., & Rahnamay Roodposhti, F. (2023). Portfolio rebalancing based on a combined method of ensemble machine learning and genetic algorithm. *Journal of Financial Reporting and Accounting*, *21*(1), 105-125.
34. Fieberg, C., Metko, D., Poddig, T., & Loy, T. (2023). Machine learning techniques for cross-sectional equity returns' prediction. *OR Spectrum*, *45*(1), 289-323.
35. Gera, R., Chadha, P., Saxena, A., & Dixit, S. (2024). A Scientometric and Bibliometric Review of Impacts and Application of Artificial Intelligence and Fintech for Financial Inclusion. *In Artificial Intelligence, Fintech, and Financial Inclusion* (pp. 82-111). CRC Press.
36. Gomes, O. (2023). Behavioral economics and finance: a selective review of models, methods and tools. *Studies in Economics and Finance*, *40*(3), 393-410.
37. Goodfellow, I., Bengio, Y. & Courville, A. (2016). *Deep Learning*. MIT Press.
38. Goyal, P., Gupta, P., & Yadav, V. (2023). Antecedents to heuristics: decoding the role of herding and prospect theory for Indian millennial investors. *Review of Behavioral Finance*, *15*(1), 79-102.
39. Grosso, G., Lai, N., Letizia, M., Pazzini, J., Rando, M., Wulzer, A., & Zanetti, M. (2023). A fast and flexible machine learning approach to data quality monitoring. *arXiv preprint arXiv:2301.08917*.
40. Guerra, P., & Castelli, M. (2021). Machine learning applied to banking supervision a literature review. *Risks*, *9*(7), 136.
41. Gupta, R., & Modise, M. P. (2013). Macroeconomic variables and South African stock return predictability. *Economic modelling*, *30*, 612-622.

42. Habbab, F. Z., & Kampouridis, M. (2023). An in-depth investigation of five machine learning algorithms for optimizing mixed-asset portfolios including REITs. *Expert Systems with Applications*, 121102.
43. Hambly, B., Xu, R., & Yang, H. (2023). Recent advances in reinforcement learning in finance. *Mathematical Finance*, *33*(3), 437-503.
44. Han, J., & Li, Y. (2023). Asset Allocation Strategy Based on Announcements and Machine Learning--An approach in Chinese Market. *Highlights in Business, Economics and Management*, *5*, 251-263.
45. Harrison, R. L., Reilly, T. M., & Creswell, J. W. (2020). Methodological rigor in mixed methods: An application in management studies. *Journal of Mixed Methods Research*, 14(4), 473-495.
46. Hassani, H., Huang, X., Silva, E., & Ghodsi, M. (2020). Deep learning and implementation in banking. *Annals of Data Science*, *7*, 433-446.
47. Hirose, M., & Creswell, J. W. (2023). Applying core quality criteria of mixed methods research to an empirical study. *Journal of Mixed Methods Research*, 17(1), 12-28.
48. Ibrahim, A. A., Saeed, B. N., & Fadil, M. A. (2023). Forecasting Stock Prices with an Integrated Approach Combining ARIMA and Machine Learning Techniques ARIMAML. *Journal of Computer and Communications*, *11*(8), 58-70.
49. Jang, J., & Seong, N. (2023). Deep reinforcement learning for stock portfolio optimization by connecting with modern portfolio theory. *Expert Systems with Applications*, *218*, 119556.
50. Jiang, Z., Xu, D., & Liang, J. (2017). A deep reinforcement learning framework for the financial portfolio management problem. *arXiv preprint arXiv:1706.10059*.
51. Kamiri, J., & Mariga, G. (2021). Research methods in machine learning: A content analysis. International *Journal of Computer and Information Technology* (2279-0764), 10(2), 78-91.
52. Kern, C., Klausch, T., & Kreuter, F. (2019, April). Tree-based machine learning methods for survey research. *In Survey Research Methods* (Vol. 13, No. 1, p. 73). NIH Public Access.
53. Khandani, A. and Lo, A. (2010). What Happened to the Quants in August 2007? *Journal of Investment Management*, 8(4), pp.5-54.
54. Karp, A. and Van Vuuren, G., 2019. Investment implications of the fractal market hypothesis. *Annals of Financial Economics*, 14(01), p.1950001.
55. Koumou, G. B. (2020). Diversification and portfolio theory: a review. *Financial Markets and Portfolio Management*, *34*(3), 267-312.
56. Kumbure, M. M., Lohrmann, C., Luukka, P., & Porras, J. (2022). Machine learning techniques and data for stock market forecasting: A literature review. *Expert Systems with Applications*, *197*, 116659.
57. Laiyao, C. (2023). Risk and Legal Regulation of Algorithm Application in Insider Trading Supervision. *Technium Soc. Sci. J.*, *43*, 274.

58. Lalwani, P., Mishra, M. K., Chadha, J. S., & Sethi, P. (2022). Customer churn prediction system: a machine learning approach. *Computing*, 1-24.
59. Lee, J., Chang, J. R., Kao, L. J., & Lee, C. F. (2023). Financial Analysis, Planning, and Forecasting. In *Essentials of Excel VBA, Python, and R: Volume II: Financial Derivatives, Risk Management and Machine Learning* (pp. 433-455). Cham: Springer International Publishing.
60. Lee, M., Shin, M., Lee, J., & Jeong, K. (2023). Mutual Information Maximizing Quantum Generative Adversarial Network and Its Applications in Finance. *arXiv preprint arXiv:2309.01363*.
61. Lee, Y., Thompson, J. R., Kim, J. H., Kim, W. C., Fabozzi, F. A., Fabozzi, F. J., ... & Belanger, D. (2023). An Overview of Machine Learning for Asset Management. *The Journal of Portfolio Management*, *49*(9), 31-63.
62. Leo, M., Sharma, S., & Maddulety, K. (2019). Machine learning in banking risk management: A literature review. *Risks*, *7*(1), 29.
63. Lewellen, J. (2013). The cross section of expected stock returns. Working paper, *Dartmouth College*.
64. Li, J., Wang, X., Ahmad, S., Huang, X., & Khan, Y. A. (2023). Optimization of investment strategies through machine learning. *Heliyon*, *9*(5).
65. Liu, X. Y., Yang, H., Gao, J., & Wang, C. D. (2021, November). FinRL: Deep reinforcement learning framework to automate trading in quantitative finance. In *Proceedings of the second ACM international conference on AI in finance* (pp. 1-9).
66. Liu, Y. and Timmermann, A. (2019). Predicting Stock Returns Using Reinforcement Learning. Journal of Financial Economics, 132(2), pp.367-385
67. Loke, Z. X., Goh, S. L., Kendall, G., Abdullah, S., & Sabar, N. R. (2023). Portfolio Optimisation Problem: A Taxonomic Review of Solution Methodologies. *IEEE Access*.
68. Ma, Y., Han, R., & Wang, W. (2021). Portfolio optimization with return prediction using deep learning and machine learning. *Expert Systems with Applications*, *165*, 113973.
69. Mahdavinejad, M. S., Rezvan, M., Barekatain, M., Adibi, P., Barnaghi, P., & Sheth, A. P. (2018). Machine learning for Internet of Things data analysis: A survey. *Digital Communications and Networks*, 4(3), 161-175.
70. Markowitz, H. (1952). Portfolio selection. Journal of Finance, 7(1), 77-91.
71. Melina, Sukono, Napitupulu, H., & Mohamed, N. (2023). A Conceptual Model of Investment-Risk Prediction in the Stock Market Using Extreme Value Theory with Machine Learning: A Semisystematic Literature Review. *Risks*, *11*(3), 60.
72. Mhlanga, D. (2021). Financial inclusion in emerging economies: The application of machine learning and artificial intelligence in credit risk assessment. *International journal of financial studies*, *9*(3), 39.

73. Michaud, R.O., 1989. The Markowitz optimization enigma: Is ‘optimized’optimal?. *Financial analysts journal*, 45(1), pp.31-42.
74. Milojević, N., & Redzepagic, S. (2021). Prospects of artificial intelligence and machine learning application in banking risk management. *Journal of Central Banking Theory and Practice*, *10*(3), 41-57.
75. Mukherjee, A., Singh, A. K., Mallick, P. K., & Samanta, S. R. (2022). Portfolio Optimization for US-Based Equity Instruments Using Monte-Carlo Simulation. In *Cognitive Informatics and Soft Computing: Proceeding of CISC 2021* (pp. 691-701). Singapore: Springer Nature Singapore.
76. Muldoon-Smith, K., & Greenhalgh, P. (2019). Suspect foundations: Developing an understanding of climate-related stranded assets in the global real estate sector. *Energy Research & Social Science*, *54*, 60-67.
77. Munkhdalai, L., Munkhdalai, T., Namsrai, O. E., Lee, J. Y., & Ryu, K. H. (2019). An empirical comparison of machine-learning methods on bank client credit assessments. *Sustainability*, *11*(3), 699.
78. Nazareth, N., & Reddy, Y. Y. R. (2023). Financial applications of machine learning: A literature review. *Expert Systems with Applications*, 119640.
79. Noh, Y., Kim, J. M., Hong, S., & Kim, S. (2023). Deep Learning Model for Multivariate High-Frequency Time-Series Data: Financial Market Index Prediction. *Mathematics*, *11*(16), 3603.
80. Ozbayoglu, A. M., Gudelek, M. U., & Sezer, O. B. (2020). Deep learning for financial applications: A survey. *Applied Soft Computing*, *93*, 106384.
81. Qiu, J., Wang, B., & Zhou, C. (2020). Forecasting stock prices with long-short term memory neural network based on attention mechanism. *PloS one*, *15*(1), e0227222.
82. Reel, P. S., Reel, S., Pearson, E., Trucco, E., & Jefferson, E. (2021). Using machine learning approaches for multi-omics data analysis: A review. Biotechnology Advances, 49, 107739.
83. Saunders, M., & Lewis, P. (2017). *Doing research in business and management.* Pearson.
84. Sahu, S. K., Mokhade, A., & Bokde, N. D. (2023). An Overview of Machine Learning, Deep Learning, and Reinforcement Learning-Based Techniques in Quantitative Finance: Recent Progress and Challenges. *Applied Sciences*, *13*(3), 1956.
85. Sardar, I., Akbar, M. A., Leiva, V., Alsanad, A., & Mishra, P. (2023). Machine learning and automatic ARIMA/Prophet models-based forecasting of COVID-19: Methodology, evaluation, and case study in SAARC countries. *Stochastic Environmental Research and Risk Assessment*, *37*(1), 345-359.
86. Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allegue, H., Teplitsky, C., ... & Araya-Ajoy, Y. G. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in ecology and evolution*, *11*(9), 1141-1152.

87. Schmitt, M. (2023). Deep learning in business analytics: A clash of expectations and reality. *International Journal of Information Management Data Insights*, *3*(1), 100146.
88. Seabe, P. L., Moutsinga, C. R. B., & Pindza, E. (2023). Forecasting cryptocurrency prices using LSTM, GRU, and bi-directional LSTM: a deep learning approach. *Fractal and Fractional*, *7*(2), 203.
89. Sengupta, S., & Mitra, S. (2023). How Does the Behavioural Finance Theories Impact the Investor Decision Making: An Empirical Study. *International Journal of Banking, Risk & Insurance*, *11*(2).
90. Shaukat, K., Luo, S., Varadharajan, V., Hameed, I. A., & Xu, M. (2020). A survey on machine learning techniques for cyber security in the last decade. *IEEE access*, *8*, 222310-222354.
91. Sithole, S., Bakama, E. M., & Mushavhanamadi, K. (2025). A Review of Blockchain and Cryptocurrency in the South African Banking.
92. Song, Y., Tang, X., Wang, H., & Ma, Z. (2023). Volatility forecasting for stock market incorporating macroeconomic variables based on GARCH-MIDAS and deep learning models. *Journal of Forecasting*, *42*(1), 51-59.
93. Sonkavde, G., Dharrao, D. S., Bongale, A. M., Deokate, S. T., Doreswamy, D., & Bhat, S. K. (2023). Forecasting Stock Market Prices Using Machine Learning and Deep Learning Models: A Systematic Review, Performance Analysis and Discussion of Implications. *International Journal of Financial Studies*, *11*(3),
94. Thakkar, A., & Chaudhari, K. (2021). Fusion in stock market prediction: a decade survey on the necessity, recent developments, and potential future directions. *Information Fusion*, *65*, 95-107.
95. Vergara-Fernández, M., Heilmann, C., & Szymanowska, M. (2023). Describing model relations: The case of the capital asset pricing model (CAPM) family in financial economics. *Studies in History and Philosophy of Science*, *97*, 91-100.
96. Wang, J., & Chen, Z. (2023). Exploring Low-Risk Anomalies: A Dynamic CAPM Utilizing a Machine Learning Approach. *Mathematics*, *11*(14), 3220.
97. Whang, S. E., Roh, Y., Song, H., & Lee, J. G. (2023). Data collection and quality challenges in deep learning: A data-centric ai perspective. The VLDB Journal, 32(4), 791-813.
98. Wu, Z., Yang, L., Fei, Y., & Wang, X. (2023). Regularization methods for sparse ESG-valued multi-period portfolio optimization with return prediction using machine learning. Expert Systems with Applications, 120850
99. Xiao, C., Xia, W., & Jiang, J. (2020). Stock price forecast based on combined model of ARI-MA-LS-SVM. Neural Computing and Applications, 32, 5379-5388.
100. Xin, L., Xin, K., & Xin, H. (2023). On Laws of Thought—A Quantum-like Machine Learning Approach. *Entropy*, *25*(8), 1213.

101. Xu, X., Xiong, F., & An, Z. (2023). Using machine learning to predict corporate fraud: evidence based on the gone framework. *Journal of Business Ethics*, *186*(1), 137-158.

102. Yadav, M. P., & Khera, A. (2023). Forecasting the stock return of emerging economies: an empirical study based on ARIMA. *International Journal of Public Sector Performance Management*, *11*(4), 451-466.

103. Yoo, C., Ramirez, L., & Liuzzi, J. (2014). Big data analysis using modern statistical and machine learning methods in medicine. *International Neurourology Journal*, 18(2), 50.

104. Yi, Z., Cao, X., Pu, X., Wu, Y., Chen, Z., Khan, A. T., ... & Li, S. (2023). Fraud detection in capital markets: A novel machine learning approach. *Expert Systems with Applications*, 120760.

105. Zaheer, S., Anjum, N., Hussain, S., Algarni, A. D., Iqbal, J., Bourouis, S., & Ullah, S. S. (2023). A Multi Parameter Forecasting for Stock Time Series Data Using LSTM and Deep Learning Model. *Mathematics*, *11*(3), 590.

106. Zhao, Y., & Yang, G. (2023). Deep Learning-based Integrated Framework for stock price movement prediction. *Applied Soft Computing*, *133*, 109921.