2023

An Online stock market recommender system using machine learning.

Muoki, Sharleen Mwikali
School of Computing and Engineering Sciences
Strathmore University

An Online Stock Market Recommender System Using Machine Learning

By
Sharleen Mwikali Muoki
090923

Master of Science in Information Technology

2023

An Online Stock Market Recommender System Using Machine Learning

By
Sharleen Mwikali Muoki
090923

Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Science in Information Technology at Strathmore University.

School of Computing and Engineering Sciences,
Strathmore University,
Nairobi, Kenya.

July, 2023

Declaration and Approval

Declaration

I hereby affirm that this work has not been previously submitted for the award of a degree by this or any other university. To the best of my knowledge and belief, this thesis contains no material that has been previously published or written by another person, except where due reference is made in the thesis itself.

Student's Name: Sharleen Mwikali Muoki

Sign: _____ Date: Monday 5th June 2023

Approval

The thesis submitted by the candidate, Name of Candidate (in black colour), has been thoroughly reviewed and approved for examination by the following individuals:

Dr. Esther Khakata,
School of Computing & Engineering Sciences,
Strathmore University.

Dr. Julius Butime,
Dean, School of Computing & Engineering Sciences,
Strathmore University.

Dr. Bernard Shibwabo,
Director of Graduate Studies,
Strathmore University.

Abstract

Investing in the stock market presented significant challenges for novice investors, primarily due to the overwhelming volume of available data. Consequently, novice investors often made irrational decisions and experienced unfavourable investment outcomes, particularly in the context of Kenya. This project aimed to address this issue by developing an innovative online stock market recommender system that utilised machine learning techniques. By leveraging these techniques, the system aimed to facilitate informed decision-making based on reliable data. Novice investors frequently encountered difficulties in accessing sufficient and well-organised information about the companies they intended to invest in, resulting in suboptimal returns. Traditional research methods often failed to adequately address the complexities and vastness of the stock market data. However, incorporating machine learning into the investment process held promise for analysing historical data, identifying patterns, and providing valuable insights to support informed decision-making. To comprehensively achieve the research objectives, a mixed-methods approach was employed, which integrated both quantitative and qualitative data collection and analysis in a sequential design. The Object-Oriented Analysis and Design (OOAD) technique was systematically and logically adopted to develop the software system. Additionally, the Dynamic System Development Methodology (DSDM) served as a guiding framework to address the identified problem and facilitate the development of the online stock market recommender system. This research project identified the information challenges faced by individual investors in the stock market, highlighting issues such as limited access to critical information, lack of necessary skills, and reliance on inaccurate media reports. Moreover, the study identified specific factors that significantly influenced stock market investments, including earnings per share, firm fundamentals, market trends, and news sentiment. To overcome these challenges, the project focused on developing an advanced online stock market recommender system that effectively leveraged machine learning techniques to provide personalised investment recommendations. The system underwent rigorous testing, demonstrating superior performance by offering accurate recommendations and comprehensive investment performance reports.

Keywords: Investing; Stock market; Novice investors; Poor investment outcomes; Online stock market recommender system; Machine learning techniques; Informed decision-making; Information challenges; Mixed-methods approach; Quantitative and qualitative data; Object-Oriented Analysis and Design (OOAD); Dynamic System Development Methodology (DSDM); Earnings per share (EPS); Market trends; News sentiment; Investment performance reports

Table of Contents

List of Figures

List of Tables

Abbreviations/ Acronyms

A2C - Advantage Actor-Critic

AI - Artificial Intelligence

API - Application Programming Interface

BVPS - Book Value Per Share

BVS - Book Value of Shares

CNN - Convolutional Neural Network

CRUD - Create, Read, Update and Delete

CSS - Cascading Style Sheets

DDPG - Deep Deterministic Policy Gradient

DPG - Deterministic Policy Gradient

DPS - Dividend Per Share

DQN - Deep Q-Learning

DSDM - Dynamic System Development Methodology

EPS - Earnings Per Share

ERB - Ethics Review Board

ERD - Entity Relationship Diagram

G7 - Group of Seven

GAs - Genetic Algorithms

GBS - General Binary Search

GDP - Gross Domestic Product

HTML - Hypertext Markup Language

IPI - Industrial Production Index

IREC - Institutional Research Ethics Committee

JS - JavaScript

LogMCAP - Firm Size

LSTM - Long-Short Term Memory

MCDM - Multiple Criteria Decision Making

MVC - Model-View-Controller

NASDAQ - National Association of Securities Dealers Automated Quotations

NN - Neural Networks

OAA - One-Against-All

OAO - One-Against-One

OOAD - Object-Oriented Analysis and Design

PE - Price-to-Earnings Ratio

PPO - Proximal Policy Optimisation

RNN - Recurrent Neural Networks

ROE - Return on Equity

ROI - Rate of Interest

S&P 500 - Standard and Poor's 500

SQL - Structured Query Language

SSAD - Structured Systems Analysis and Design

SVMs - Support Vector Machines

TD Error - Advantage Function

TSE - Tehran Stock Exchange

UML - Unified Modeling Language

Definition of Terms

Group of Seven (G7) countries	The Group of Seven (G7) is an informal association of the world's most advanced economies. France, Germany, Canada, Italy, Japan, the United Kingdom, and the United States are among the member nations (*Canada and the G7 - Toronto*, n.d.).
Standard and Poor's 500 (S&P 500)	It is a stock market index that follows 500 domestically based publicly listed firms in the United States. Many investors regard it as the most accurate general indicator of the success of the American stock market (Kenton, 2022).

Acknowledgements

The successful culmination of this endeavour is a matter of great honour. I attribute my accomplishments to the meticulous supervision and invaluable assistance I have received, for which I express my profound gratitude. I would like to acknowledge the following individuals for their instrumental contribution to the triumph of this project:

First and foremost, I extend my sincere thanks to Dr. Esther Khakata, my supervisor, for her active involvement in guiding me throughout the process of documenting this project. Her genuine interest and provision of vital information were paramount in the development of a successful system. I am deeply indebted to her for her unwavering motivation and guidance.

I also wish to express my heartfelt gratitude to Dr. Ismael Ateya, the head supervisor of the Information System Thesis for the Master of Science in Information Technology in 2022-2023. His steadfast guidance and insightful suggestions throughout the project were invaluable. I am grateful for the consistent encouragement, assistance, and guidance provided by the faculty of the School of Computing and Engineering Sciences, which greatly facilitated the project's successful completion.

Furthermore, I would like to extend my appreciation to all faculty members, staff, non-staff contributors, and fellow students who played a significant role in transforming this documentation into a reality.

Lastly, I express my deep gratitude to my parents, whose unwavering love and guidance have been the foundation of my life and everything I strive to achieve.

Dedication

In deep appreciation, I dedicate this thesis project to my cherished family and friends, whose unwavering presence has been a constant source of inspiration and encouragement throughout my master's degree and the challenges life has presented. I am profoundly grateful for each and every one of you, as you have played a pivotal role in my journey.

Moreover, this project holds special significance for the upcoming generation of Kenyans. It is a testament to their enduring spirit and serves as a platform to enhance their understanding of the stock market. Their remarkable achievements have served as a driving force, motivating me to strive diligently for the successful completion of this endeavour.

Chapter 1: Introduction

1.1 Background

The stock market, a confluence of trading platforms for publicly traded corporations' shares, plays a pivotal role in the modern economic framework (Black& Gilson, 1998). For such corporations, achieving listing status on a trading exchange is a notable accomplishment that not only facilitates capital acquisition but also amplifies their reputation and operational infrastructure (*Benefits of Listing / Going Public*, n.d.). The stock exchange serves as an essential financial intermediary, coalescing capital-seeking firms and investors with surplus funds (Hayes, 2021).

Access to the stock market is broadly democratised, requiring no formal qualifications, thus making it readily accessible to anyone with disposable income. Historical precedence places the Dutch East India Corporation at the genesis of public corporations and stock exchanges, introducing the selling of shares and dividends to shareholders (Voigt, Parys & Yochim, 2022). Despite this, instances abound where investment decisions, influenced more by financial euphoria than by sound research, have precipitated market collapses owing to financial instability, unchecked speculation, and rising inflation (Hwang, Dow & Jones, 2021).

Understanding the intricacies of the stock market is of paramount importance for investors. Novice investors, particularly prone to speculation, often confront significant losses and reduced returns, underscoring the imperative of comprehensive research (Escobar & Pedraza, 2019). Over time, individual investors tend to acquire pertinent information, adjusting their behaviour to optimise investment outcomes. Yet, conducting such thorough research can be daunting, particularly for newcomers who need to parse relevant information for informed decision-making (Nicolosi, Peng & Zhu, 2009).

Investors frequently fall prey to irrational decisions, such as the pitfalls of buying high and selling low, adversely impacting long-term portfolio performance (Ahmad, 2020). Investment success is closely intertwined with human behavioural traits, market variables, and heuristic biases, such as complacency and anchoring (Hunter, 2012). As such, relying solely on intuition for investment strategy in a data-rich epoch is ill-advised. Instead, investment decisions should be anchored in reliable data, eschewing decisions driven purely by emotions or speculative tendencies.

Machine learning has recently emerged as a potent instrument in stock trading. Machine learning algorithms, capable of parsing voluminous historical stock market data, identify patterns and trends that may elude human analysts (Sengupta, Basak, Saikia, Paul, Tsalavoutis, Atiah, Ravi, & Peters, 2020). Employing machine learning techniques equips

investors with valuable insights into market behaviour, price trajectories, and potential investment opportunities.

The integration of machine learning into investment decision-making processes offers the potential to mitigate the influence of human biases and emotions, thereby enhancing investment performance (Trunk, Birkel, & Hartmann, 2020). Machine learning enables investors to scrutinise a broad spectrum of factors, such as financial statements, market news, social media sentiment, and macroeconomic indicators. This fosters more informed investment decisions. This thesis aims to delve into the application of machine learning in stock trading and examine its impact on investment decision-making, especially for novice investors, by weaving machine learning into the research methodology.

In conclusion, a comprehensive understanding of the stock market coupled with meticulous research forms the bedrock of successful investment decisions. While emotional investing and speculative tendencies can prove detrimental, machine learning presents untapped opportunities to elevate investment decision-making. By scrutinising historical data and identifying patterns, machine learning algorithms offer valuable insights, thereby curbing the impact of human biases and emotions. This alignment with machine learning techniques offers a promising horizon for new investors.

1.2 Problem Statement

New investors in the stock market often experience significant losses due to a lack of adequate and organised information about the companies in which they invest (Aduda, Oduor & Onwonga, 2012). This leads to poor investment decisions, resulting in suboptimal returns compared to benchmark indices. The challenge of inadequate information, misinformation, and analysis paralysis discourages new investors from participating in the stock market, potentially weakening the crucial link between companies seeking financing and investors with excess funds Adusumilli (2021). While traditional research methods provide information, they may not effectively address the complexities and vast amount of data available in today's stock market.

To rectify this issue and improve investment decision-making among new investors, there is a need to explore the role of machine learning (ML) in stock trading. Machine learning techniques have the potential to analyse large volumes of historical stock market data, identify patterns and trends, and provide valuable insights to assist investors in making informed decisions(Zhang, & Zhou, 2004; Thakkar & Chaudhari, 2021; Shah, Isah & Zulkernine, 2019). By incorporating machine learning into the investment process, new investors can access timely and organised information, potentially reducing losses and improving overall returns.

By addressing the challenge of poor investment decisions through the application of machine learning techniques, this research aims to bridge the information gap for new investors in the stock market and enhance their decision-making abilities. The investigation will explore the effectiveness of machine learning algorithms in analysing stock market data, assessing the impact of ML-driven insights on investment outcomes, and providing recommendations for incorporating machine learning into the investment strategies of new investors (Hajizadeh, Davari Ardakani & Shahrabi, 2010).

1.3 Research Objectives

1.3.1 Key Objective

The primary objective of this research project is to construct an online recommender system for the stock market, harnessing the power of machine learning. The intention behind this endeavour is to equip novice investors with reliable, simplified information to support informed trading decisions, thereby enhancing overall investment outcomes.

1.3.2 Specific Objectives

i. To conduct an in-depth investigation into the information challenges experienced by individual investors in the stock market.
ii. To determine the specific information requirements of individual investors in the stock market.
iii. To design advanced algorithms, models, and an architectural framework to support the development of an efficient online stock recommender system.
iv. To develop a functional prototype of a web-based stock recommender system.
v. To evaluate the performance and usability of the prototype platform in the context of providing stock trading information.

1.4 Research Questions

i. What are the information challenges faced by individual investors in the stock market?
ii. What are the specific information requirements of individual investors in the stock market?
iii. How can algorithms, models, and an architectural framework be designed to support the development of an efficient online stock recommender system?
iv. How can a functional prototype of a web-based stock recommender system be developed?
v. How can the performance and usability of the prototype platform be evaluated in providing stock trading information?

1.5 Justification

The intricacy and dynamic nature of the stock market often present challenges to individual investors, particularly those lacking the requisite information and skillset to make informed investment decisions. Such challenges can precipitate suboptimal performance and substantial financial losses. Consequently, gaining a comprehensive understanding of the information-related hurdles individual investors encounter in the stock market, as well as their specific information needs, is of paramount importance.

To mitigate these challenges and cater to the information requirements of individual investors, there is a compelling need to develop an online stock recommender system, underpinned by machine learning. Such a system would foster an environment conducive to informed trading practices predicated on empirical data and statistical analysis, rather than speculative impulses. Therefore, the conceptualisation and development of efficient algorithms, models, and architectural frameworks for this stock recommender system hold significant value.

Exploration of these research questions has the potential to yield insightful findings, instrumental in enhancing the success rates of individual investors in the stock market. This, in turn, may contribute positively to the overall market's health and robustness. With this focus, the research aligns with the thesis's theme of leveraging machine learning to improve investment decision-making for new investors.

1.6 Scope and Limitations

The purpose of this investigation was to develop a platform engineered to streamline access to pertinent information for individual investors within the stock market and to identify efficacious methodologies for communicating this information. It is imperative to underscore, however, that this exploration did not extend to potential impediments to stock market trading that may not hinge on information accessibility. Furthermore, due to the temporal constraints of this project, it was not feasible to delve into the actual trading process engaged in by investors.

The data underpinning this project were procured from multiple APIs, inclusive of Unicorn Data Services' Financial Data APIs, Twitter API, and Yahoo Finance API. Resource limitations necessitated a scope restriction in the data collection and analysis process, focusing on five selected companies.

Despite these limitations, the focus of this research remains aligned with the overarching theme of the thesis: the utilisation of machine learning to enhance investment decision-making for novice investors. This endeavour aims to provide meaningful insights

that could potentially bridge the information gap often faced by individual investors, thereby fostering more informed and effective investment strategies.

Chapter 2: Literature Review

2.1 Introduction

Investments serve as pivotal catalysts for both economic and social advancement, wherein the initial increment in invested capital has the potential to generate increased value through the mechanism of the investment multiplier (Busch, Bruce-Clark, Derwall, Eccles, Hebb, Hoepner, Klein, Krueger, Falko Paetzold, Scholtens & Weber, 2021). Consequently, investment promises significant returns while concurrently alleviating risks that could counterbalance missed opportunities and the anticipated depreciation in the value of invested funds due to inflation (Svensson, 2003). Hence, it is of vital importance that investors are well-versed with the elements influencing stock prices, enabling them to make judicious investment decisions. Although an array of factors bear upon stock prices, the final determinant of stock prices resides in market supply and demand. Key factors that influence stock prices include fundamental aspects such as profits and profitability, along with technical factors such as market momentum and investors' behavioural attributes (Shanmuganathan, 2020).

In order to glean an understanding of this literature review, it becomes necessary to first offer a synopsis of the primary notion being put forth and to actively involve the reader in synthesising the presented concept and the challenges encountered. This review utilises data and insights from esteemed academics and industry professionals to shed light on the factors influencing stock market prices on a global scale. Moreover, this review takes into account the informational needs of individual stock market investors worldwide, inclusive of those based in Kenya. It lays out a blueprint of the requisite information for informed stock market decisions, alongside a review of the existing stock market recommender systems, models, architectures, and algorithms. Ultimately, the review strives to identify the trajectory towards an ideal state of an investor's decision-making process, culminating in the conceptualisation of an optimal framework by the conclusion of this chapter.

2.2 Information Difficulties Faced by Individual Stock Market Investors

Individual stock market investors encounter a wide array of challenges, with a paramount one being the procurement of trustworthy and precise information. As Gärling, Kirchler, Lewis, & van Raaij (2009) delineate, individual investors often confront inequitable accessibility to information in comparison to institutional investors, which could detrimentally affect their investment decisions. Furthermore, novice investors may find it daunting to decipher financial information such as financial statements and may lack the necessary acumen to interpret market trends and economic indicators (Lin, 2016). Consequently, individual investors may resort to reliance on media coverage or rumours, the accuracy of which may be uncertain.

Additionally, behavioural biases of individual investors, such as herd mentality, can potentially engender irrational investment decisions. These biases may precipitate suboptimal investment decisions, leading to meagre returns or significant losses. Therefore, acknowledging the information-related challenges faced by individual stock market investors and devising efficacious solutions to address them is crucial. Such strategies could potentially aid investors in making more informed investment decisions, leading to increased returns and contributing to the overall vitality of the stock market (Talwar, Talwar, Tarjanne & Dhir, 2021).

Nevertheless, given the vast amount of information available, novice investors may seek assistance in sifting through the data to discern what is pivotal and relevant to a company's stock price or portfolio (Beattie & Murry, 2022). It becomes imperative in this context to exercise prudence and engage in exhaustive research prior to making investment decisions. Furthermore, investors should refrain from solely relying on the counsel of stockbrokers, who may harbour interests that diverge from those of their clients (Knauff, Budeck, Wolf & Hamburger, 2010).

2.3 Factors Affecting Stock Market Prices

The notable surge in global stock markets over the past few decades can be largely ascribed to the substantial inputs from emerging economies (Basher, Haug & Sadorsky, 2012). This accelerated growth has engendered considerable transformations in financial institutions and capital flows, leading to a multitude of variables influencing asset prices. These determinants can be broadly categorised into market and non-market variables, economic and non-economic factors, as well as corporate, trade, national, and international factors. Moreover, their assessment demands criteria for analysis at both the micro and macro levels.

Earnings per share (EPS) is positively correlated with stock returns, as Taani andBanykhaled, (2011) illustrate, denoting its critical role as a metric of corporate performance. The study indicated that an escalation in EPS correlates with a surge in stock prices, and the relationship between EPS and stock returns is more pronounced in firms with lower market-to-book ratios. Ackert and Athanassakos (2003) found that amalgamating EPS with other financial characteristics can facilitate future stock price prediction and result in superior returns. The study revealed that EPS, in conjunction with other financial measures such as price-to-earnings ratio, price-to-book ratio, and return on assets, positively correlates with future stock returns. Generally, a firm with a high EPS is deemed financially robust and more likely to yield returns to investors, potentially leading to an augmentation in stock prices.

A critical determinant of stock prices is commodity prices, which can influence companies' share prices on the stock market. Lee and Zeng (2011) discerned a significant correlation

between oil prices and real stock returns in the Group of Seven (G7) countries. The global economy hinges on commodities that are interchangeable with other similar items, and fluctuations in commodity prices can trigger a cascading effect on the pricing of other goods and services. Consequently, any perturbation in commodity prices, either positive or negative, may lead the stock market to undergo a similar shock, resulting in either economic growth or recession. With the escalating globalisation of commodity markets, commodity prices and global stock markets have grown increasingly interrelated (Fernando, 2022).

The influence of news on stock prices constitutes another integral facet of the securities industry. As highlighted by Nandha, & Faff (2008), investor sentiment can be swayed by unexpected developments within a company, industry, or the global economy, and news can affect share prices positively or negatively. The interconnectedness among markets and economies worldwide, engendered by globalisation, implies that news in one country can exert a swift and substantial impact on investors in another. For instance, the publication of earnings reports can incite a rise or fall in stock prices based on investor reactions to the news. Moreover, Pagolu, Challa, Panda & Majhi (2016) determined that public sentiments about a company, expressed on Twitter via tweets, can significantly influence its share price. Therefore, analysing market mood can provide a valuable resource for investors and technical and analytical analysts to predict fluctuations in stock prices.

A comprehensive understanding of how commodities, news, market mood, consumer attitude, and earnings per share (EPS) affect stock market prices is a pivotal factor in devising effective investment strategies. Armed with a thorough grasp of the impact of these variables, investors and market analysts can make informed decisions about stock buying and selling, which can, in turn, contribute to the overall vitality of the stock market.

2.4 The Information Needs of an Individual Stock Market Investor

The information requirements of individual stock market investors may diverge, contingent on their investment goals, risk tolerance, and strategic investment approach (Häubl & Garnefeld, 2018). Generally, these investors seek credible, up-to-date information about companies, inclusive of financial statements, earnings reports, and news pertaining to performance, management, and competition (Chung & Lee, 2018). Additionally, economic data, industry trends, market analyses, and expert insights regarding specific stocks or industries are in demand (Revsine, Collins, Johnson, Winter, Mittelstaedt & Soffer, 2018). The objective is to secure access to reliable and timely information to make informed decisions about buying or selling stocks to achieve investment goals.

In past decades, individual investors grappled with obtaining adequate information without costly membership services. However, the advent of the Internet has revolutionised the

investment landscape by providing uncomplicated, free access to real-time data (Engin & Treleaven, 2019). The fundamental challenge lies in selecting the most pertinent data to effectively evaluate a given stock. As indicated by Engin and Treleaven (2019) research, the utilisation of risk and return information by investors plays a significant role in elucidating differences in investment decisions.

Pertinent information for making informed stock market decisions needs to be relevant, substantial, and deserving of consideration. It should be easily comprehensible, applicable, and replicable, with diagnostic variables, interventions with baselines, and outcomes embedded (Zacharakis & Meyer, 2000). Such material should demonstrate how the investor's risk tolerance is respected while meeting their goals and objectives.

A multitude of factors may be taken into account when analysing companies, and value investors with a comprehensive understanding of vital industry data can make superior decisions. In the quest for profitable investment opportunities, every stock analyst should take into account several critical data points (Shah, Isah & Zulkernine, 2019). These encompass factors like market mood, price variations in essential business commodities, and news.

2.5 Stock Market Recommender Systems

Considerable scholarly research has been dedicated to the deployment of decision support systems within the stock market. This body of work predominantly focuses on enhancing the accuracy of future value and pattern forecasts, generating buy or sell signals, and developing automated trading systems. However, the aspect of personalisation has been largely overlooked, resulting in a dearth of literature on the subject. Nonetheless, the global evaluation of available resources can be viewed as offering general recommendations.

According to Khedr, Salama, and Yaseen (2017), a comprehensive examination of various news categories, coupled with historical numerical values, enhances the understanding of stock market behaviour and yields strong predictive performance regarding future prospects. This methodology involves the evaluation of multiple types of daily news containing varying numerical values throughout the day. Stephan and Von Nitzsch (2013) assert that recommendations primarily rely on simple measures and are limited to a few specific firms. Additionally, online forums that provide information about a particular company may serve as valuable sources for prospective investors. Several research papers have employed natural language processing techniques to analyse financial news and discussions on social networks. Ravi and Ravi (2015) contribute to this field by exploring opinion mining and sentiment analysis approaches, while Eickhoff and Muntermann (2016) utilise logit models to shed light on the circumstances in which sentiment from social media content can be used to predict analytical judgments, and vice versa.

Various machine learning models have been examined for stock price prediction, including genetic algorithms (GAs), support vector machines (SVMs), neural networks (NNs), one-versus-one (OAO) techniques, and one-versus-all (OAA) neural network approaches. However, it has been observed that the data source significantly influences the performance of these models. Li, Wang, Dong, Wang, Deng, and Zhu (2011) conducted a study that demonstrated a single-source baseline system underperforming in comparison to a system that combined information from multiple sources, as corroborated by cross-validation and independent testing.

2.5.1 Models and Frameworks

Extensive scholarly inquiry has been devoted to exploring models and frameworks in recommender systems, with a focus on identifying and addressing obstacles and enablers, providing implementation guidance, and integrating components for collective evaluation (Crockett, 2017). The progression of recommender system research has witnessed the evolution from basic techniques like nearest neighbour and matrix factorisation to more advanced approaches such as deep learning and graph topology-based high-order connection identification. This project aims to present a comprehensive review of the existing literature by examining the various models and frameworks discussed in prior research, in order to establish an optimal framework for addressing the identified problem.

Yujun, Jianping, and Yimei (2016) developed a stock recommendation algorithm that utilises large-scale net inputs. This algorithm effectively excludes stocks with minimal value, resulting in improved forecast accuracy. It enables the recommendation of higher-value stocks to targeted clients and investors, thereby enhancing stock returns. The model achieves computational efficiency by employing innovative algorithms like the Binary Search Algorithm, effectively meeting the investment demands of real-life investors. Experimental results demonstrate that the proposed stock recommendation model outperforms money flow-based models. Additionally, it offers the advantage of enhancing investment returns for target investors by eliminating equities with poor or negative investment returns. Figure 2.1 illustrates the process of the proposed model, which comprises the user clustering stage and the stock recommendation stage. Typically, investors with similar qualities exhibit comparable investing interests.

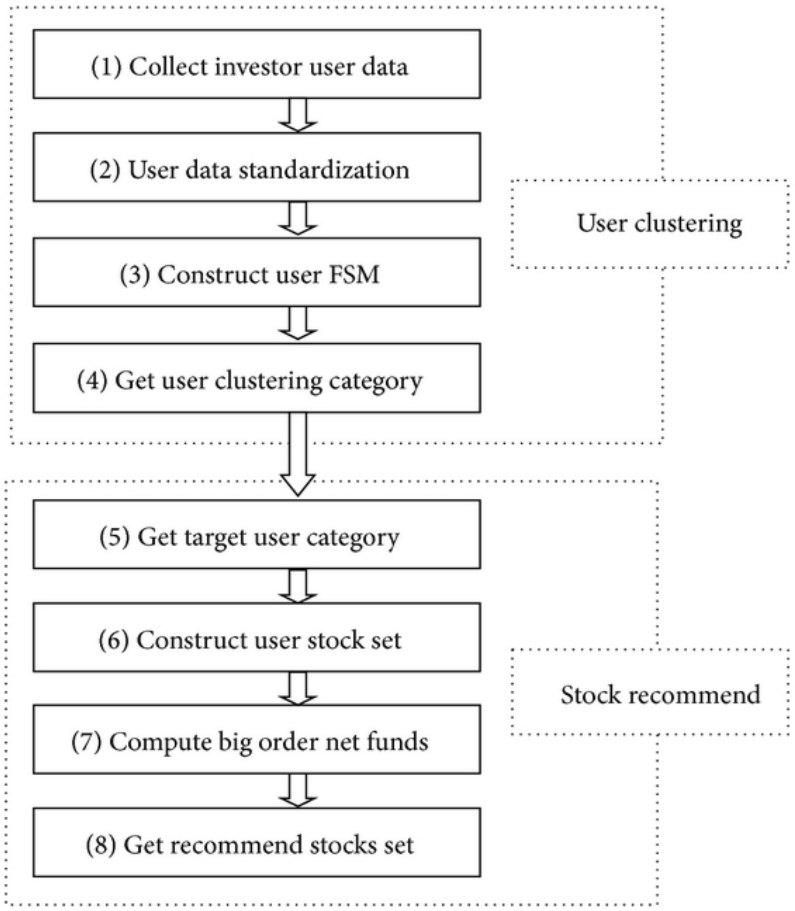


Figure 2.1: The Process Framework of The Proposed Model (Yujun, Jianping & Yimei, 2016)

Wang and Mishra (2017) presented an innovative stock market forecast and recommender system in their study, which incorporates user-friendly features. This system serves as a valuable tool for investors, providing them with informed recommendations regarding whether to sell or buy equities in the upcoming period. To enhance the coherence and comprehensibility of the data, the study employs information granulation techniques, transforming the original time series data into more meaningful and interpretable granules. Moreover, the authors propose an effective inhomogeneous segmentation strategy for prediction. The experimental results, depicted in Figure 2.2, clearly demonstrate the efficacy of the proposed approach, as it significantly improves prediction accuracy and facilitates profitable investment opportunities.

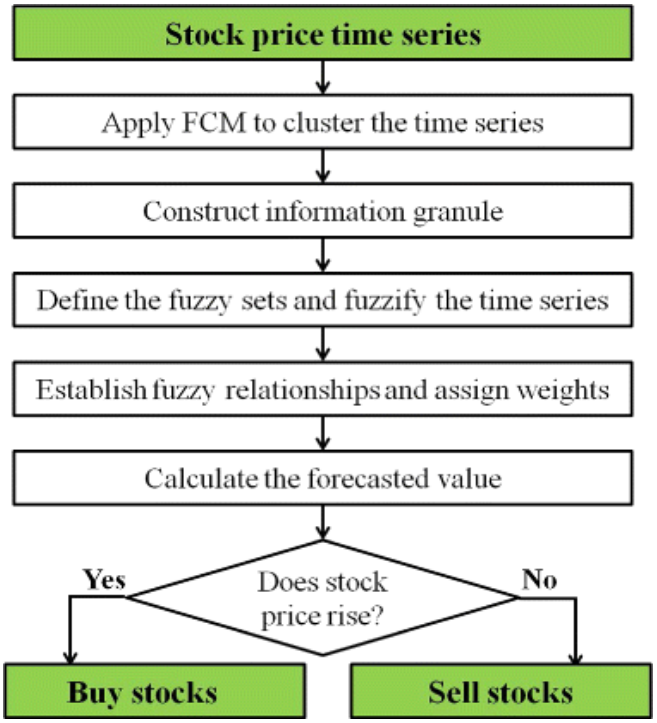


Figure 2.2: The Stock Trading Prediction and Recommendation System (Wang & Mishra, 2017)

Boonpeng and Jeatrakul (2016) undertook a comprehensive study to explore various multi-binary classification experiments, specifically employing the One-Against-One (OAO) and One-Against-All (OAA) Neural Network approaches. The study's findings indicate that the OAA technique exhibits superior performance compared to other methods, such as Genetic Algorithms (GAs), Support Vector Machines (SVMs), and Neural Networks (NNs), when applied to multi-binary classification tasks. In the OAA approach, each class is systematically compared against all other classes. In the context of stock market analysis, the classes correspond to buying, selling, and holding data, as exemplified in Figure 2.3.

In their study, Jeatrakul and Wong (2012) introduced the utilisation of the One Against All (OAA) approach to transform K machine learning instances with binary classes into a singular machine learning instance featuring K classes. These K classes encompass buying, holding, selling, and remaining data, as visually depicted in Figure 2.4. The authors highlight the advantage of the OAA approach, as it requires a reduced number of binary classifiers compared to previous methodologies.

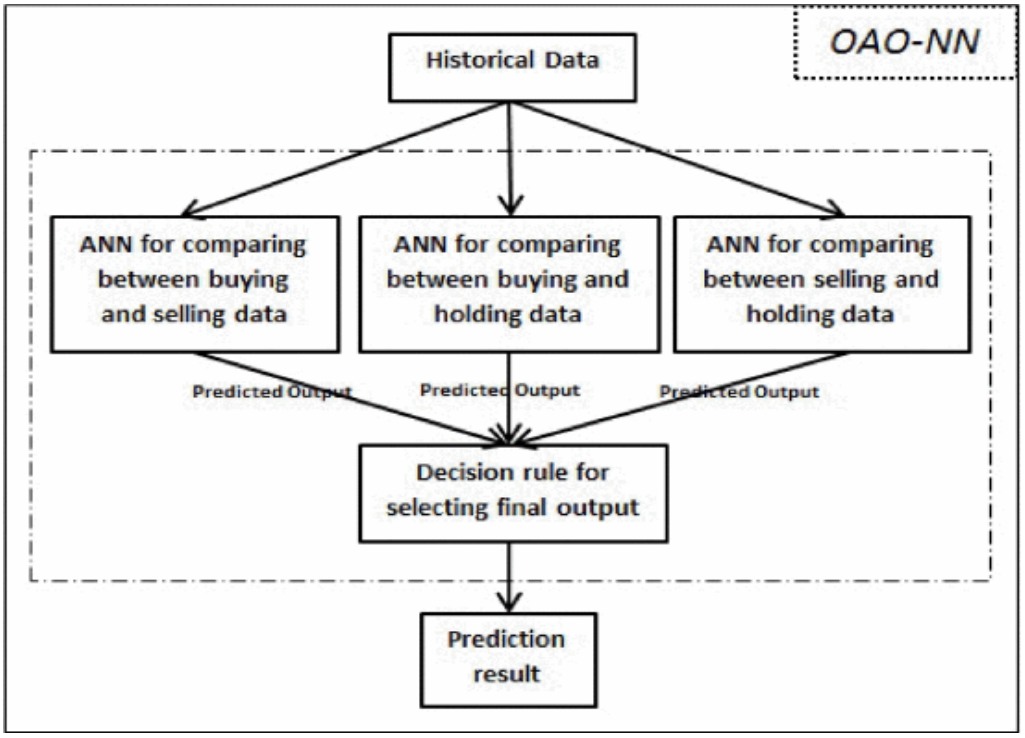


Figure 2.3: The One-Against-One Neural Network Architecture (Boonpeng and Jeatrakul, 2016)

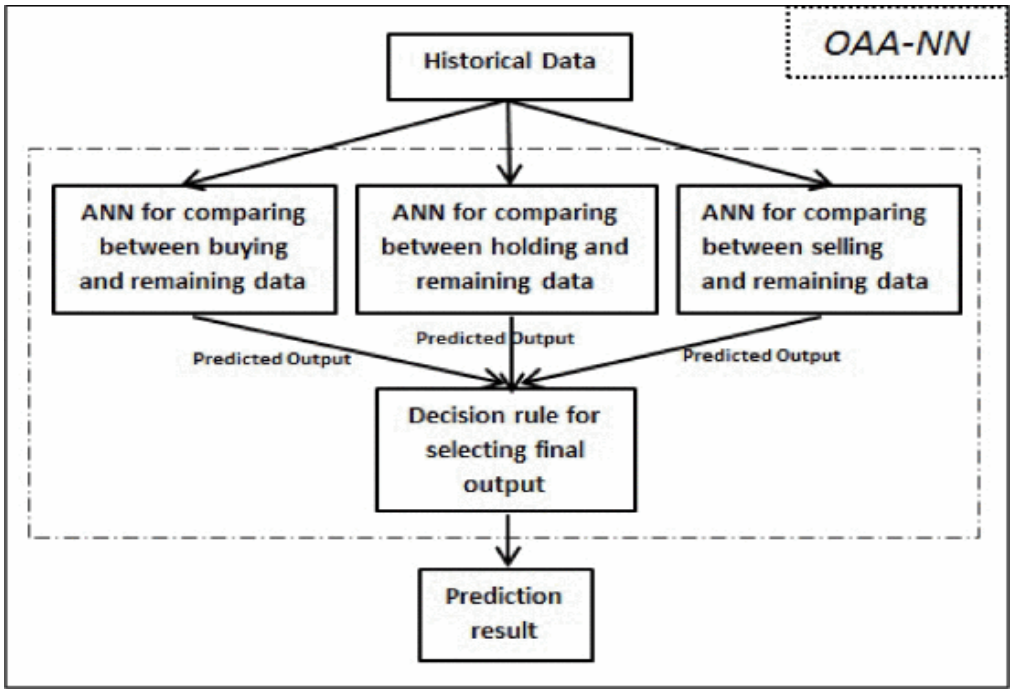


Figure 2.4: The One-Against-All Neural Network Architecture (Boonpeng and Jeatrakul, 2016)

2.5.2 Architectures and Designs

The significance of data architecture and design within the realm of data analytics cannot be overstated. These principles encompass a range of techniques, rules, models, and standards

that are pivotal in the process of gathering, sourcing, utilising, storing, and transmitting information. They play a crucial role in establishing a cohesive framework for data system interactions, which is essential for informed decision-making. A well-designed data architecture facilitates seamless connections across multiple systems while providing a creative model for data interaction. The effective application of data architecture and design principles is paramount in enabling organisations to extract valuable insights and make informed decisions based on their data.

In their 2010 study, Fasanghari and Montazer proposed a novel fuzzy expert system for equity selection. This system incorporates the recommendation guidelines employed by experts at the Tehran Stock Exchange (TSE), as illustrated in Figure 2.5. The system aims to address the inherent unpredictability of daily stock recommendations by mitigating the challenges of Multiple Criteria Decision Making (MCDM) in selecting the most suitable stocks at optimal times. It employs a graded approach to evaluate a limited number of possibilities, taking into account multiple criteria that are often imprecise and susceptible to uncertainty.

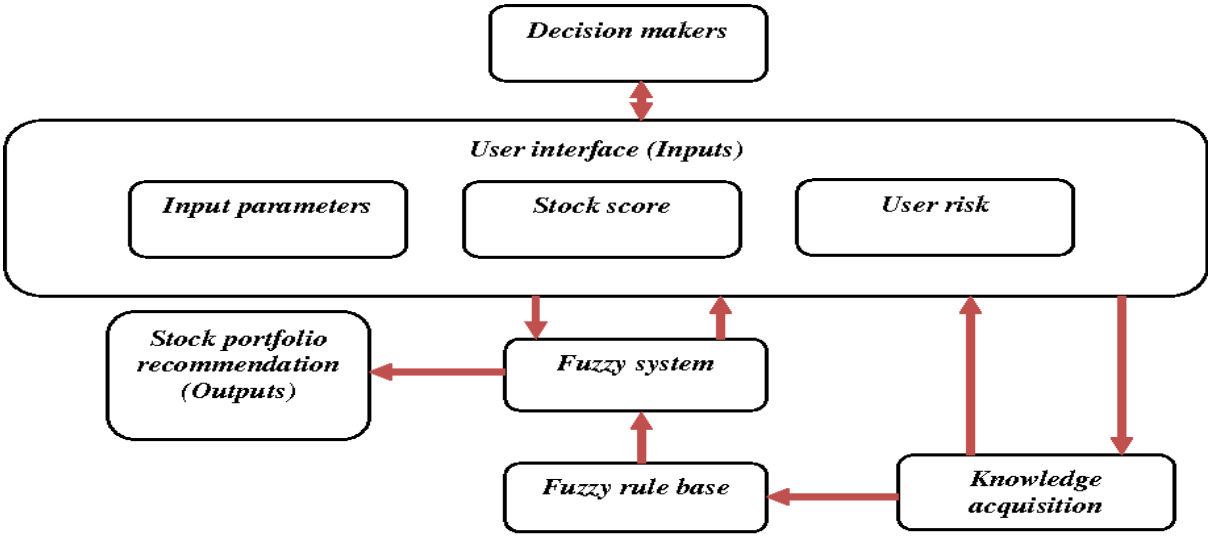


Figure 2.5: System Architecture (Fasanghari & Montazer, 2010)

Syu, Yeh, Wu, and Ho (2010) presented a self-management portfolio system that employs dynamic association mining specifically designed for the Taiwan Stock Exchange. The authors outlined the five key components of their system, as illustrated in Figure 2.6: association mining, two-stage preprocessing, self-management, adaptive closure, and trading methods.

The self-management module plays a crucial role in continuously monitoring holding positions, ensuring proactive portfolio management. Association mining serves as a resource provisioning component, aiding in the identification of valuable patterns and relationships

within the stock market data. Adaptive closure mechanisms are integrated to promote the system's sustainability, while the allocation of funds into separate units allows for effective management of trading risks.

During bear markets, the system promptly closes stock holdings to minimise drawdown risks, prioritising capital preservation. Conversely, during bullish trends, it gradually increases stock positions to capitalise on potential gains. The performance evaluation demonstrates that the proposed system consistently outperforms all baselines across various metrics, even when compared to randomly selected datasets.

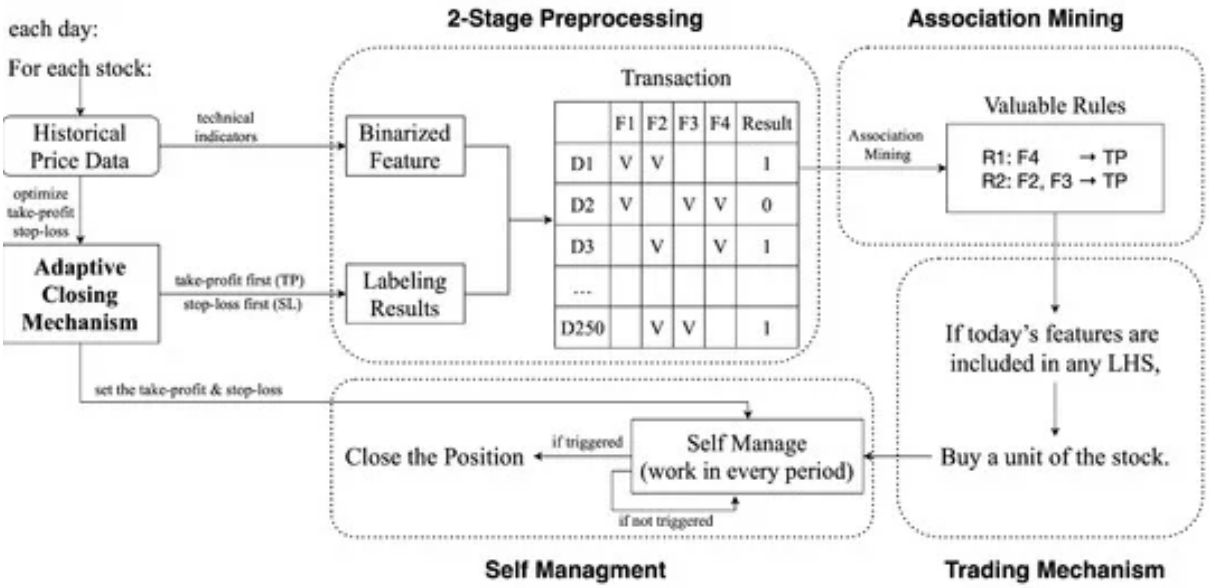


Figure 2.6: Flowchart of The Proposed System (Syu, Yeh, Wu, & Ho, 2010)

Nam and Seong (2019) introduced a novel machine learning model that incorporates causality analysis to predict changes in stock prices using financial news. Their research, titled "Financial news-based stock movement prediction utilising causality analysis of the influence on the Korean stock market," surpasses the performance of two conventional complex algorithms in experimental tests.

The study sheds light on the significant impact of news about a particular company on the stock prices of other firms operating within the same industry. Remarkably, the proposed model demonstrates the ability to predict directional changes in stock prices even in scenarios where there is an absence of financial news specifically related to the target company. Instead, the model leverages financial news pertaining to causative companies, as exemplified in Figure 2.7. This highlights the model's capacity to capture the interconnectedness and causal relationships within the stock market.

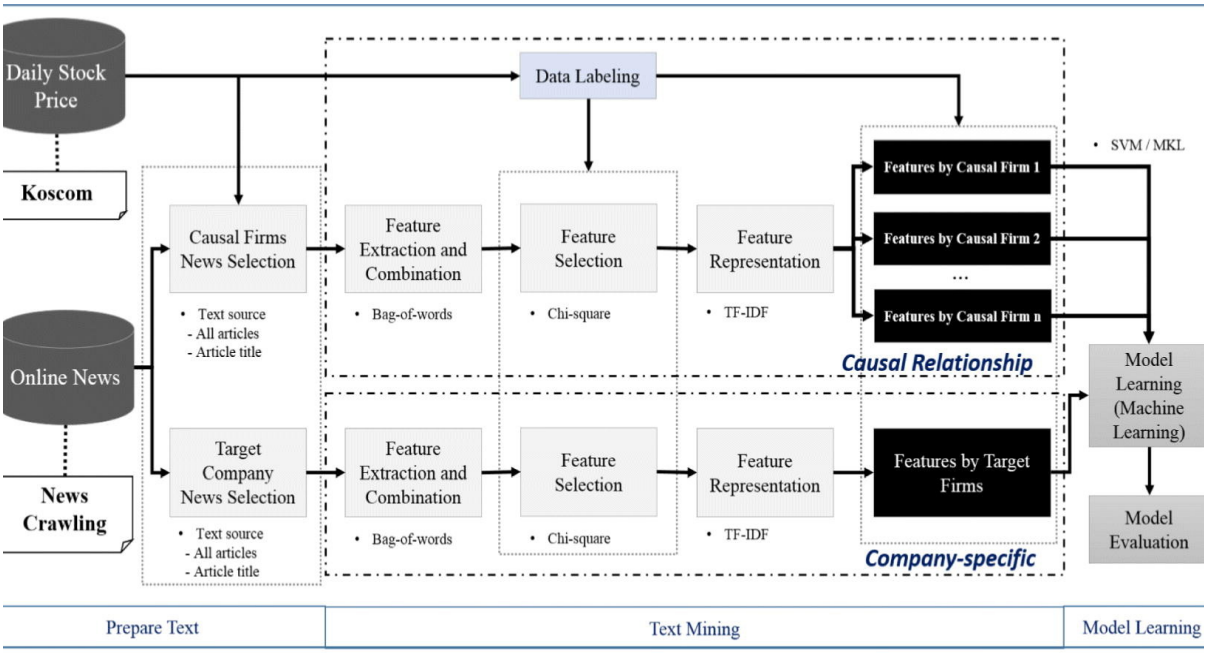


Figure 2.7: Proposed Approach (Nam & Seong, 2019)

Majgi, Balchand, and Rao (2020) conducted a study that employed machine learning techniques to predict stock market values. The researchers utilised a large-scale dataset collected from various social media platforms, including Twitter, as well as news websites. This dataset encompassed diverse textual data sources such as news articles, tweets, and other financial news materials. Natural language processing techniques were applied to analyse the textual data and extract relevant features.

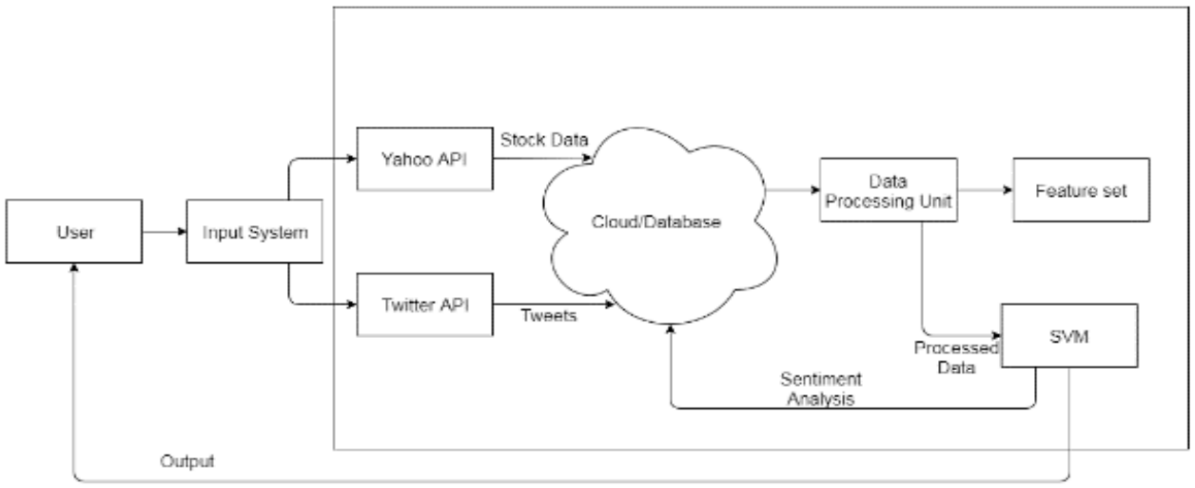


Figure 2.8: System Architecture (Majgi, Balchand and Rao, 2020)

Based on the analysed data, the researchers constructed an ensemble machine-learning model for forecasting stock market values. The model comprised four distinct modules, with one module employing Convolutional Neural Network (CNN) algorithms and the remaining three

modules utilising Long-Short Term Memory (LSTM) techniques, as illustrated in Figure 2.8. The CNN module was responsible for analysing the textual input and extracting pertinent characteristics, while the LSTM modules were tasked with modelling time series data and predicting future stock values. The study findings revealed that the proposed ensemble model outperformed other standard machine learning models in accurately predicting stock market values, exhibiting a high level of accuracy. These findings suggested that incorporating textual input from social media and news sources into machine learning models could enhance the accuracy of stock price forecasts.

Vismayaa, Pooja, Alekhya, Malavika, Nair, and Kumar (2020) devised a novel classifier-based stock trading recommendation system that leverages historical share price data and technical indicators as input features. The system's design, as depicted in Figure 2.9, exhibits a distinctive approach. To evaluate its effectiveness, the study employed stocks listed on the Bombay Stock Exchange (BSE) as the testing ground. Comparative analysis was conducted on the performance of five individual classifiers and six ensemble classifier-based decision support systems. The system's profitability for each stock was assessed using an accuracy rate and eight economic performance criteria. The findings substantiate the efficacy of the proposed strategy in delivering profitable trading recommendations.

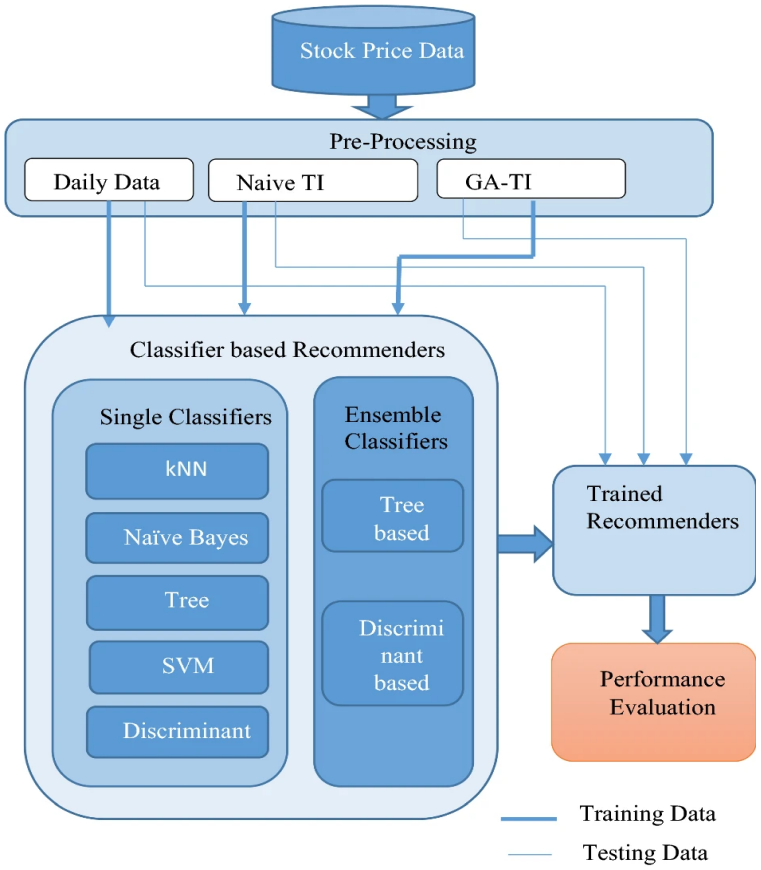


Figure 2.9: Recommender System Evaluation Process Block Diagram (Vismayaa, Pooja, Alekhya, Malavika, Nair, & Kumar, 2020)

2.5.3 Algorithms

Algorithms constitute a foundational element within both software and hardware systems, encompassing a prescribed set of instructions that systematically execute predetermined operations (Fox & Mayers, 1987). These instructions are executed in a sequential manner, forming the basis of automated processes (Forrest, 1996). In automated systems, algorithms assume a pivotal role by providing the necessary guidelines for data processing. Data processing, in turn, involves the conversion of raw data into usable and meaningful information. The chosen approach to data processing can significantly influence the speed and reliability with which queries are addressed. Hence, a thoughtful and deliberate consideration of the data processing strategy is imperative when designing an effective stock trading strategy. For instance, real-time processing should be prioritised in scenarios where availability of up-to-date information is critical for making timely and well-informed decisions (Duggal, 2022).

2.5.3.1 Sentiment Analysis

Sentiment analysis emerges as a pivotal technique for businesses seeking to gauge the emotional content conveyed within textual data. This process involves the extraction of sentiment and contextual information from text using advanced methodologies such as data mining, machine learning, and artificial intelligence techniques, as visually represented in Figure 2.10. Opinion mining, a related technique, further enhances sentiment analysis by identifying emotions in text, discerning the polarity (that is, the ratio of positive to negative sentiment), and identifying the subject and opinion holder (Al-Kabi, Gigieh, Alsmadi, Alsmadi, Gigieh, Wahsheh & Haidar, 2014). By leveraging sentiment analysis and opinion mining on social media data, companies can attain a deeper understanding of customer sentiment, evaluate their brand image, and glean invaluable insights about their customer base.

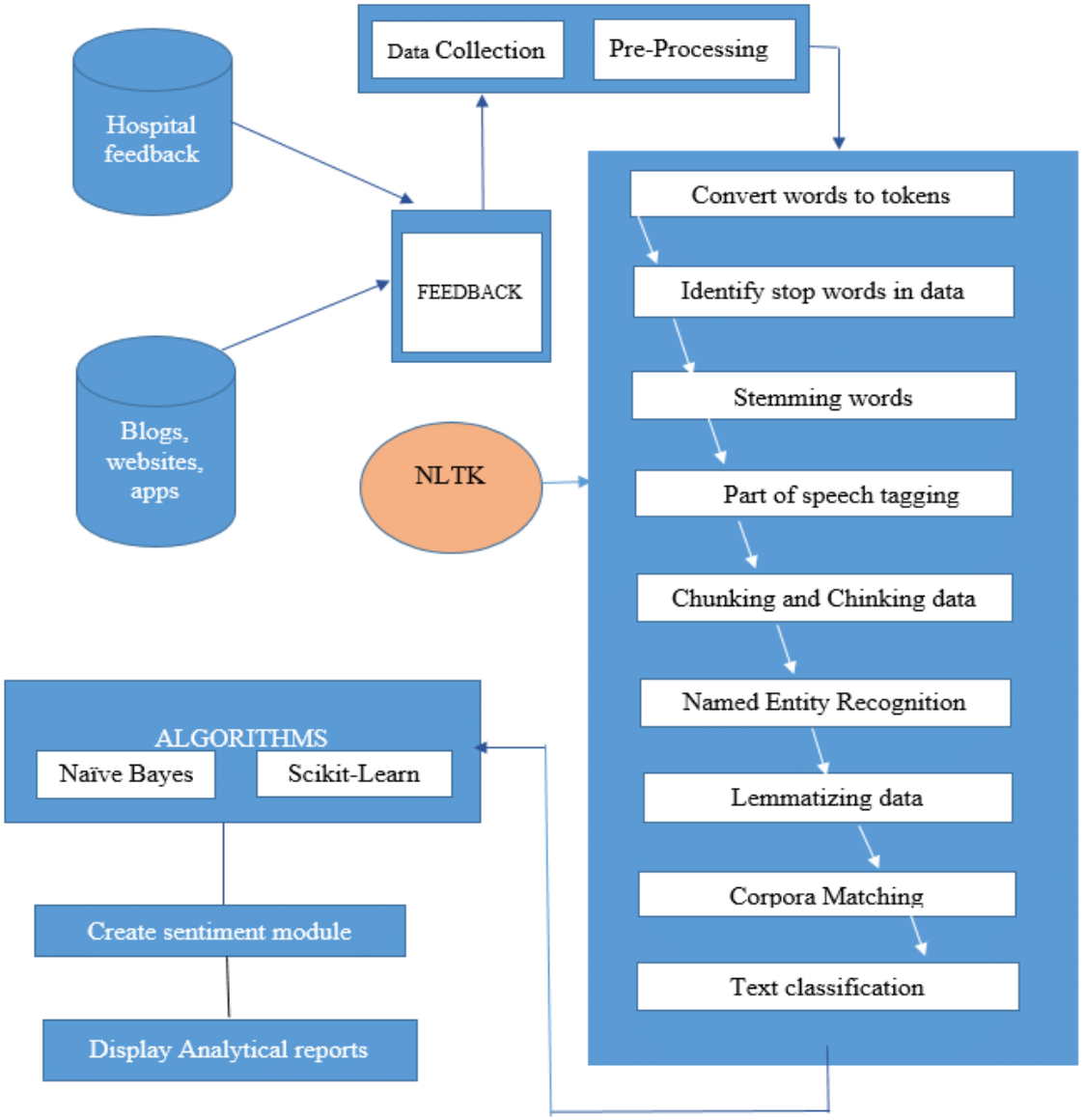


Figure 2.10: Sentiment Analysis Flowchart (Prabadevi, Reddy & Deepa, 2019)

2.5.3.2 Ensemble Models

In their study, Yang, Liu, Zhong, and Walid (2020) introduced an ensemble approach for developing a stock exchange strategy that leverages deep reinforcement algorithms to optimise capital return. The researchers explored three actor-critic-based strategies: Advantage Actor-Critic (A2C), Proximal Policy Optimisation (PPO), and Deep Deterministic Policy Gradient (DDPG).

As depicted in Figure 2.11, the ensemble technique is designed to learn and assimilate the most favourable attributes from each of the three strategies, enabling it to effectively adapt to evolving market trends. Furthermore, Figure 2.12 illustrates the notable performance of the ensemble technique in comparison to the three individual algorithms and two benchmark approaches, as assessed through risk-adjusted return, as measured by the Sharpe ratio. The

ensemble approach demonstrated superior results in terms of risk-adjusted returns, further validating its efficacy as a stock exchange strategy.

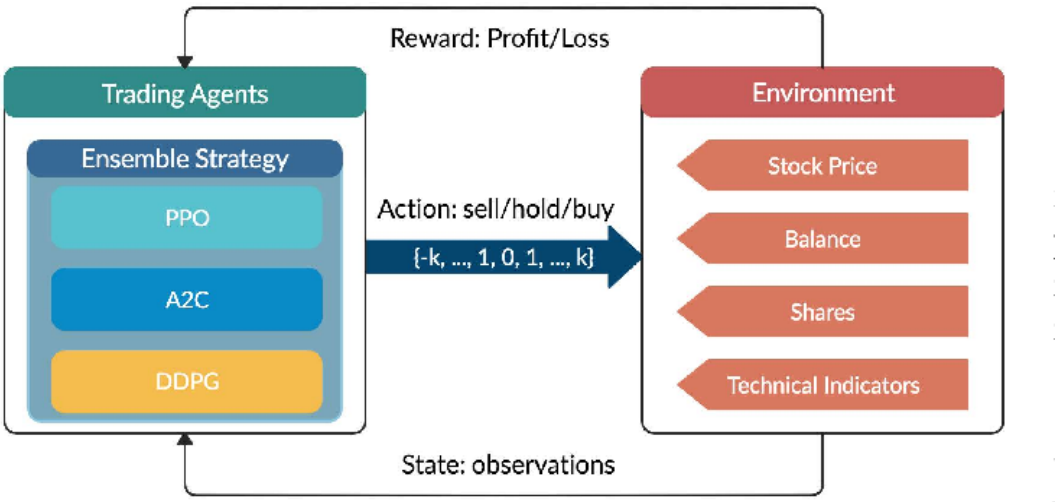


Figure 2.11: Overview of Reinforcement Learning-Based Stock Trading Strategy (Yang, Liu, Zhong & Walid, 2020)

Trading Quarter	PPO	A2C	DDPG	Picked Model
2016/01-2016/03	**0.06**	0.03	0.05	PPO
2016/04-2016/06	0.31	0.53	**0.61**	DDPG
2016/07-2016/09	-0.02	0.01	**0.05**	DDPG
2016/10-2016/12	**0.11**	0.01	0.09	PPO
2017/01-2017/03	**0.53**	0.44	0.13	PPO
2017/04-2017/06	0.29	**0.44**	0.12	A2C
2017/07-2017/09	**0.4**	0.32	0.15	PPO
2017/10-2017/12	-0.05	-0.04	**0.12**	DDPG
2018/01-2018/03	**0.71**	0.63	0.62	PPO
2018/04-2018/06	-0.08	-0.02	**-0.01**	DDPG
2018/07-2018/09	-0.17	**0.21**	-0.03	A2C
2018/10-2018/12	0.30	**0.48**	0.39	A2C
2019/01-2019/03	-0.26	-0.25	**-0.18**	DDPG
2019/04-2019/06	**0.38**	0.29	0.25	PPO
2019/07-2019/09	**0.53**	0.47	0.52	PPO
2019/10-2019/12	-0.22	**0.11**	-0.22	A2C
2020/01-2020/03	-0.36	**-0.13**	-0.22	A2C
2020/04-2020/05	-0.42	**-0.15**	-0.58	A2C

Figure 2.12: Sharpe Ratios Over Time (Yang, Liu, Zhong & Walid, 2020)

In their study, Ingle and Deshmukh (2021) introduced an ensemble deep learning model designed for stock market data prediction. This novel approach offers improved accuracy compared to previous predictive methodologies. The study employs an ensemble deep

learning framework, as depicted in Figure 2.13, which involves the identification of relevant data and the removal of irrelevant information prior to model creation. This strategy enhances the accuracy of predictions, as evidenced by the results, which indicate that the ensemble model outperforms individual deep learning models. Based on the findings, the study concludes that the proposed ensemble model effectively forecasts market trends with a high degree of accuracy. Consequently, this model serves as a valuable tool for investors and traders seeking to make informed decisions in the stock market.

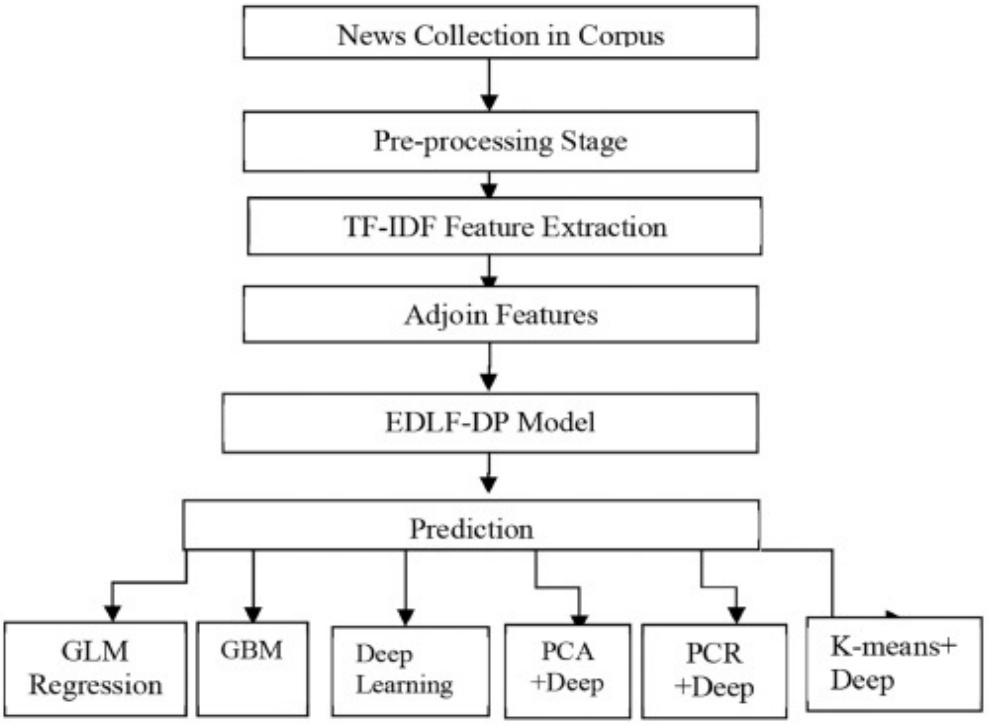


Figure 2.13: The Proposed Data Flow of the Model (Ingle & Deshmukh, 2021)

2.6 Summary of Literature Review

The literature review reveals a substantial body of research dedicated to developing models and frameworks for supporting investment decision-making in the stock market. Various approaches, including machine learning, deep learning, fuzzy logic, and neural networks, have been applied to construct these models. The efficacy of these models has been assessed using a range of metrics such as accuracy, precision, recall, F1-score, and Sharpe ratio. These models exhibit the capability to handle vast amounts of data and uncover patterns and trends that may elude human investors. Additionally, they possess the capacity to process unstructured data sources like news articles and social media feeds to identify market sentiment and trends.

Notwithstanding the encouraging findings, there are several areas that warrant further investigation. Many of the evaluated models have primarily been assessed using historical data, leaving their performance in real-world scenarios uncertain. Moreover, there is a need

for models that can effectively handle multiple objectives, such as maximising profits while minimising risk. Furthermore, it is crucial to analyse the interpretability of these models in order to comprehend the underlying decision-making processes. The reviewed literature underscores the potential of these models and frameworks in enhancing investment decision-making in the stock market. As these technologies continue to advance and evolve, their widespread adoption in the investment industry is anticipated. Nevertheless, additional research is imperative to address the limitations of these models and enhance their effectiveness.

2.7 Implemented Solutions

2.7.1 Seeking Alpha

Seeking Alpha serves as a prominent platform for crowd-sourced financial articles, offering Strong Buy quant recommendations to its users. These recommendations are derived through the utilisation of a proprietary Quantamental approach, coupled with powerful computer processing. Seeking Alpha employs the expertise of the Seeking Alpha Quant Model, independent researchers, and Wall Street analysts to validate and verify these recommendations, ensuring a comprehensive and robust decision-making process. However, it is worth noting that there still exists a human validation component within this process, which could potentially be replaced with a more efficient and intelligent machine learning model. By incorporating such a model, Seeking Alpha has the opportunity to enhance the accuracy and reliability of its recommendations, thereby further benefiting its user base (Acharya, Pagano & Volpin, 2016).

2.7.2 Motley Fool’s Stock Advisor

The Motley Fool, a prominent financial media firm, delivers online stock research and recommendations to investors of various expertise levels through its flagship offering, Stock Adviser. Additionally, The Motley Fool's researchers analyse the performance of a portfolio designed to replicate the actions of their stock advisor. However, it is essential to acknowledge that The Motley Fool's recommendations may not always result in a well-diversified portfolio, as their suggestions may lack a balanced mix of investment products. For instance, the firm might recommend investing solely in a single stock, such as Netflix, without thoroughly exploring alternative investment opportunities like real estate, exchange-traded funds (ETFs), or other strategies that could potentially yield superior outcomes. Consequently, investors are advised to exercise caution, conduct comprehensive evaluations of various investment possibilities, and strive to build a well-rounded portfolio (Giacomino & Akers, 2011 ; Mahajan, 2018).

2.7.3 Benzinga Pro

Benzinga Pro serves as a robust stock screening tool, catering to active traders by offering real-time news alerts. In addition to document releases, market data, activity scans, and graphing tools, the platform also features a chat room for community members. The primary objective of this tool is to provide traders with a comprehensive range of financial data presented in a user-friendly manner. However, it is worth noting that the sheer volume of data available on the platform can potentially overwhelm novice investors, leading to confusion rather than clarity. To address this challenge, the implementation of weighting approaches that prioritise the most relevant information while monitoring the remaining data to determine its fulfilment of relevance criteria could prove effective. By adopting such an approach, Benzinga Pro has the potential to assist users in navigating through the vast amount of data and gaining valuable insights without feeling overwhelmed (Benzinga Pro, 2022 ; Grennan & Michaely, 2021).

2.8 Proposed Solution

The objective of this research was to develop an online stock market recommender system that addresses the challenges faced by individual investors in the stock market. To achieve this, the researcher employed machine learning techniques and various processing frameworks to establish a data-driven trading environment for individual investors, promoting sound trading practices grounded in factual information and data rather than speculation. The system collected and analysed investor data, grouping individuals based on their lifestyles and disposable incomes, and offering well-informed advice regarding their portfolios.

The proposed solution holds significant potential for enhancing investment decision-making among individual investors in the stock market. Leveraging machine learning algorithms and real-time data analysis, the model encompasses multiple steps, including data collection and segmentation based on investor characteristics, sentiment analysis of financial data sources, and evaluation of suitable models for developing effective trading strategies. The final recommendations are tailored to individual investors, aiming to encourage informed and responsible trading behaviour guided by reliable data and research.

2.8.1 Conceptual Model

The study aimed to develop an online stock market recommender system for individual investors by employing a conceptual model and machine learning techniques (Johnson, 2008). To achieve this, the system embraced a multidisciplinary approach, incorporating multiple modules for data collection, processing, and analysis from diverse sources, as depicted in Figure 2.14. The investor data was continuously updated and organised based on demographics, psychographics, and behavioural patterns, providing high-value investment

recommendations to highly-rated investors while suggesting low-risk moves exclusively to those with a low-risk appetite.

Unicorn Data Services' Financial Data APIs and the Twitter API were utilised to mine and analyse data encompassing sentiments related to commodities, news, market sentiment, consumer attitudes, and earnings per share (EPS). The algorithm assessed the correlation between companies and significant commodities, assigning weights to each commodity based on its impact on stock prices. Data from the Yahoo Finance API were fed into an ensemble agent, which identified optimal models for strategy development and assessed their accuracy.

The system architecture was designed to provide a comprehensive view of the stock market, leveraging innovative technologies such as machine learning algorithms, data mining, and data analysis. The model selection and accuracy evaluation modules ensured the reliability and correctness of the system's recommendations. By processing new data through a trained recommender model, the resulting suggestions were personalised to individual investors. In conclusion, this project demonstrated the efficacy of machine learning and data-driven techniques in constructing a robust online stock market recommender system that offers informed advice to individual investors based on reliable facts and data.

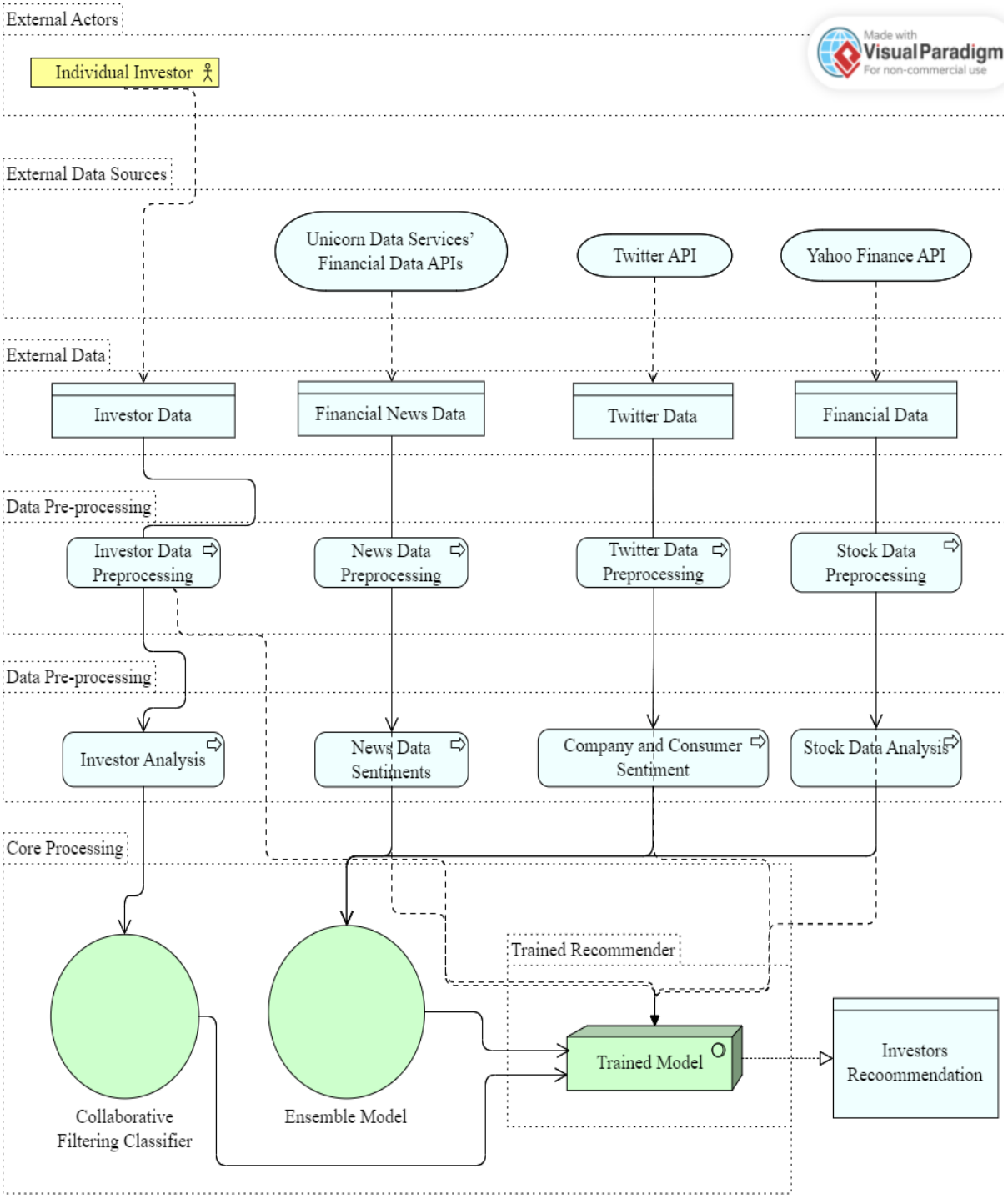


Figure 2.14: Conceptual Model for The Online Stock Market Recommender System

Chapter 3: Methodology

3.1 Introduction

The project recognises the importance of employing appropriate research methodologies to address research problems systematically. Research methodology refers to the scientific approaches used to investigate research topics (Kothari, 2004). In line with this, the objective of this project is to identify the limitations of existing stock market recommender systems and develop a more effective and efficient mechanism for informing investors.

This chapter presents a concise overview of the research methodology employed in this project and its suitability for achieving the research objectives. It encompasses various aspects, including the identification of the target audience, the selection of the sample population, the data collection process, and the data analysis techniques employed. Furthermore, this chapter delves into the specific approaches adopted for system architecture, system analysis, system design, system development, system implementation, and system testing, providing a detailed account of each stage.

By adhering to a robust research methodology, this project ensures the validity and reliability of its findings and contributes to the advancement of knowledge in the field of stock market recommender systems.

3.2 Research Design and Philosophy

In research, the adoption of suitable strategies and analytical techniques is crucial for conducting a comprehensive investigation. This is achieved through research design, which entails the selection of a coherent and logical approach to integrate various study components, including data collection, analysis, and interpretation. Research design serves as a guide, facilitating the effective organisation and comprehension of information and data, while also ensuring the systematic implementation of the research methodology and enhancing the reliability and validity of the research findings (Akhter, Pauyo & Khan, 2019).

Tashakkori and Creswell (2007) propose a blended design for this project, which represents a "third-way" perspective positioned between positivism and interpretivism. The mixed methods design encompasses several basic types, such as triangulation, embedded, explanatory, and exploratory designs. In this project, a sequential design is adopted, whereby quantitative data is initially collected and analysed, followed by the collection and analysis of qualitative data to expand upon or explain the quantitative results obtained in the first phase. The initial quantitative phase provides the foundation for the subsequent qualitative phase, effectively linking the two phases in the middle stage of the study.

For the project, the Object-Oriented Analysis and Design (OOAD) technique is employed, which offers a systematic and well-structured methodology for examining and designing software systems. This approach provides a step-by-step framework that ensures a logical and orderly development process. The OOAD approach employs a vocabulary for defining the outcomes of each stage, facilitating smooth progression from one stage to the next. Moreover, it allows for the reuse of previous designs, the adoption of standardised solutions based on sound design principles, and the incremental improvement of subpar designs without compromising the system's integrity. In the OOAD approach, a software system is conceived as a collection of interacting objects with well-defined interfaces, working collaboratively towards achieving a goal. Compared to Structured System Analysis and Design (SSAD), the OOAD approach streamlines the system development process, improves code reuse, and promotes greater organisation and resilience, resulting in reduced repetition in coding. Ultimately, the OOAD technique provides a robust and efficient software development methodology (Dathan & Ramnath, 2015).

3.3 Population and Sampling

3.3.1 Target Population

The study has a particular focus on individual investors in Kenya, specifically targeting the city of Nairobi. The research aims to examine the challenges encountered by these investors in the domain of stock market investment, with a specific emphasis on their involvement in international markets. By narrowing the scope to this specific population, the study seeks to gain a deeper understanding of the unique challenges faced by individual investors in Nairobi as they navigate the complexities of stock market investing.

3.3.2 Sampling

Sampling is a critical component of research methodology, involving the selection of a subset of individuals or elements from a larger population to draw inferences and make conclusions about the entire population (Kothari, 2004). In this project, a specific investor hub located in Nairobi, Kenya, was chosen for data collection due to its ability to attract investors from diverse demographic backgrounds. The research population consisted of 40 investors, which presented a challenge due to the small population size and potential implications for the accuracy and generalisability of the findings.

To address this limitation, appropriate statistical methods were employed to determine the optimal sample size required for data collection. While stratified random sampling could have been considered to ensure proportional representation of each stratum in the population, it was not feasible given the small population size. Thus, a power analysis was conducted to determine the minimum sample size necessary to detect statistically significant differences in

the variables of interest. Based on the power analysis, a minimum sample size of 30 investors was determined to be sufficient for this project.

Convenience sampling, a non-probability sampling approach that involves selecting individuals readily accessible to the researcher (McCombes, 2019), was employed as the preferred method for data collection from the target population. Despite its limitations in terms of generalisability, convenience sampling was deemed appropriate given the specialised and restricted nature of the selected sample. Adopting convenience sampling enabled the researcher to efficiently collect data from a substantial number of participants while remaining cost-effective and timely. Moreover, this sampling technique facilitated the acquisition of a demographically diverse and representative sample from the target community, thereby enhancing the validity and reliability of the study's findings.

Census sampling was employed as the sampling technique to collect data from all individual investors who met specific criteria and owned shares in companies listed on the National Association of Securities Dealers Automated Quotations (NASDAQ). The entire population of investors was explored to gather information on functional requirements from the users' perspective. Based on the data collected from the census sample of 30 investors, the model was developed, with training conducted using the top five companies mentioned by the investors. Subsequently, appropriate statistical methods were applied to validate the model, ensuring the reliability and validity of the results.

3.4 Data Collection Procedure

In this project, interviews were utilised to examine the information needs of individual investors in making informed decisions. Interviews were chosen for their effectiveness in obtaining concise and detailed information, providing researchers with an in-depth understanding of participants' beliefs, attitudes, goals, and significant issues. The flexibility of interviews allowed for tailored questions and follow-up inquiries, resulting in a nuanced understanding of participants' information demands and decision-making processes. A standardised guideline, outlined in Appendix E, ensured systematic and consistent interview conduct, addressing question construction, rapport building, and ethical considerations (Alshenqeeti, 2014).

Questionnaires served as a robust instrument for data collection, offering a systematic approach to obtain information from participants. Their capacity to reach a large number of individuals and gather quantitative and qualitative data effectively made them valuable in this project. Closed-ended questions facilitated quantitative data collection for statistical analysis and comparisons, while open-ended questions allowed participants to express their thoughts and emotions more comprehensively. The administration of the questionnaire through an

online platform, such as Google Forms, enhanced participant engagement and convenience. Automated data collection and validation tools minimised errors and improved reliability and validity. The questionnaire's specific questions, developed according to established survey design best practices, are provided in Appendix F (Cleave, 2023).

Documentation analysis proved useful for studying written materials, such as financial reports and educational documents, in both physical and digital formats. This method enabled researchers to gain a better understanding of the stock market's operations by examining financial statements, annual reports, and other records that shed light on company activities and the overall market. Through document analysis, valuable insights into the stock market and its inner workings were gleaned. In this project, documentation analysis was employed to investigate the impact of Social Media on Stock Prices by analysing relevant data (Bowen, 2009).

Prototyping played a crucial role in the system development life cycle, involving the iterative creation of an initial system version as a proof of concept. This approach allowed end users to experiment with system features and offer feedback to refine the design and better meet their needs. The iterative prototype approach enabled developers to accurately specify the desired system, understand customer preferences, and personalise the final product accordingly. Prototypes provided a means for individual investors to verify the system's intuitiveness, user-friendliness, and capability to satisfy their specific demands. End users participated as co-designers, offering valuable input on functionality, usability, and performance (Nelson, Berlin, & Menold, 2019).

3.5 Data Analysis Procedures

Thematic analysis served as the chosen data analysis approach for this investigation. It involved detecting, interpreting, and reporting data patterns in a qualitative manner. Thematic analysis enabled researchers to explore participants' interpretations and experiences, identify patterns and themes within the data, and provide comprehensive descriptions of the phenomena under study. Transcribed interview data were analysed to recognise and categorise themes and patterns, which were then validated against the original data. The resulting themes were analysed and presented in a clear and concise manner, offering a thorough understanding of individual investors' information needs and decision-making processes (Braun & Clarke, 2006).

Inferential statistics were employed to analyse the quantitative data obtained from the questionnaire. Descriptive statistics, such as frequency distributions, summarised the closed-ended responses and revealed patterns and trends. Inferential statistics were used to

assess the significance of observed differences or relationships between variables (Kairuz, Crump, & O'Brien, 2007).

Content analysis was applied as the data analysis approach for documentation analysis. It involved evaluating and interpreting textual or visual materials to uncover patterns, themes, or biases in the data. Content analysis facilitated the identification and categorisation of information related to market trends and other relevant themes within financial records and reports. The process included identifying relevant documents, selecting a representative sample, generating coding categories, and carefully examining the data to reveal themes and patterns. This comprehensive analysis provided valuable insights into the functioning of the stock market (Krippendorff, 2013).

Usability testing was employed to analyse the data gathered through prototyping. This approach involved evaluating the user interface of the system by observing participants' interactions and measuring their ability to efficiently and effectively complete tasks. Usability testing helped identify areas where users encountered difficulties, providing valuable insights for improving navigation and comprehension. By conducting usability testing, potential design flaws and usability concerns were addressed before the final product's release, resulting in cost and time savings. The feedback obtained through usability testing played a crucial role in enhancing the system's design to meet the specific needs of individual investors (Nelson, Berlin, & Menold, 2019).

3.6 Research Reliability and Quality

3.6.1 System Analysis

System analysis is a critical phase in software development that involves breaking down complex problems into smaller, more manageable components (Whitten, Bentley & Ho, 1986). This recursive method enables researchers to address each issue independently, simplifying the problem-solving process(Ruparelia, 2010). Decomposition, the process of dividing larger components into smaller ones, is a crucial step in component creation. System design entails determining the nature of these components and their interactions to form a cohesive system (Mawhinney, Callaghan & Cale, 1989).

In this research project, a system analysis technique was employed to gain a comprehensive understanding of the platform's components. The overarching challenge was divided into smaller components, which were iteratively refined to uncover their unique characteristics and roles. This iterative process facilitated the development of a system architecture that accurately represented the platform's functionality. By identifying the components, system attributes were derived from each component, enabling a clear understanding of their functions and interactions within the larger system. This approach resulted in a detailed and

thorough comprehension of the system's design, ultimately enhancing the system's effectiveness.

3.6.2 System Testing

In contrast to black-box and grey-box testing, a rigorous white-box testing approach was adopted in the project, which requires a deep understanding of the internal workings of an application. White-box testing focuses on thoroughly examining data domains and internal constraints (Ostrand, 2002). After the full development of each module, a unit test was conducted to ensure the accuracy of input and output for each module (Andrews, Menzies & Li, 2011). Additionally, individual modules were carefully examined to ensure compliance with the required parameters.

The next phase involved integrating multiple modules, which underwent compatibility testing and seamless collaboration assessment using an integration test. Moreover, functional testing was conducted on each module to ensure that investors could participate in end-to-end scenarios, providing valuable insights into the system's performance (Tsai, Xiaoying Bai, Paul, Weiguang Shao & Agarwal, 2011).

This comprehensive testing strategy facilitated the identification of potential errors and the implementation of necessary improvements to achieve optimal performance. The meticulous white-box testing approach employed in this project resulted in the development of a robust, reliable, and high-performance online stock market recommender system suitable for individual investors in Kenya and beyond.

3.6.3 System Validation

The rigorous performance testing approach employed in this project aligns with the principles outlined in system validation and software testing literature. Weyuker (1998) emphasises the significance of comprehensive testing to ensure that software adheres to specified requirements, while Feng, Liu and Zheng (2011) highlights the importance of testing for reliability, effectiveness, and efficiency of software. Moreover, the use of accuracy as a metric for evaluating stock price prediction is well-documented in financial forecasting and machine learning literature (Prakash Kolla, 2020).

The validation process employed in this project adheres to established best practices in software development and testing. For instance, the data collection and processing capabilities of the system were thoroughly evaluated to ensure accuracy and timeliness, which is a crucial aspect of software validation. Similarly, the validation of individual modules such as stock filtering and risk assessment, sentiment analysis and relationship assessment, model selection and accuracy evaluation, and the recommender module follows

the modular approach to software development and testing. This approach emphasises the testing of each component independently before integrating them into the system

3.7 System Development Methodology

The identified problem in Chapter 1, Section 1.2 was addressed using the Dynamic System Development Methodology (DSDM) in this project to develop an online stock market recommender system for individual investors. DSDM is a formalised approach consisting of several phases: feasibility, elaboration through exploration, development, and deployment. These phases were executed systematically to ensure effective problem-solving. Figure 3.1 visually represents the different phases of the DSDM methodology, starting with the foundational feasibility phase. The exploration phase expanded on the initial requirements, while the development phase involved implementing the system. In the deployment phase, the system was made available to end users. The utilisation of DSDM in this project provided a disciplined and well-organised approach to developing the online stock market recommender system for individual investors (Stapleton, 1999).

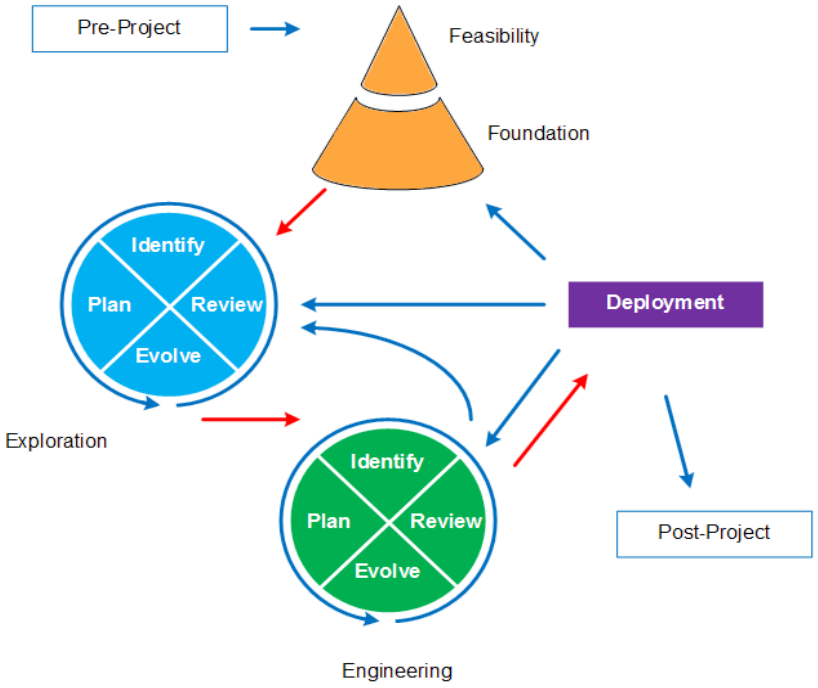


Figure 3.1: Dynamic System Development Methodology with Agile Methodology (Mubasher, 2017)

The DSDM project management approach is known for its user-centric nature, emphasising active user participation throughout the development process. Investor feedback was actively sought and incorporated, allowing for reversible changes during development. The approach followed small steps and iterations in a step-by-step process for system development. While not all phases of DSDM are required for every project implementation, each phase includes

critical tasks essential for project success, as depicted in Figure 3.1. These tasks can be expanded as needed, particularly when integrating DSDM with other development approaches. To provide a more comprehensive understanding of the DSDM methodology, this project outlines seven steps in Section 3.7 (Anwer, Aftab, Waheed & Muhammad, 2017; Voigt, 2016).

3.7.1 Pre-Project

The feasibility phase involved identifying and specifying the financial resources and objectives. This analysis proactively addressed potential challenges related to project finance and implementation. By minimising the risks of budget overruns, delays, or complications, the project established clear expectations and requirements (Stapleton, 1999).

3.7.2 The Project Life-Cycle

During the execution phase, the project progressed through key events in its life cycle. This phase transformed the software concept into a functional system through coding, testing, and integration. The development process was structured into smaller, manageable steps for independent evaluation. The development phase consisted of;

3.7.2.1 Feasibility Study

The feasibility stage involves evaluating the project's viability and identifying potential risks and constraints. It includes conducting a feasibility study to assess the technical, operational, economic, and schedule feasibility of the project. The goal is to determine whether the project is worth pursuing and aligns with the organisation's objectives (Abrahamsson, Oza & Siponen, 2010).

3.7.2.2 Business Study

The business study stage focuses on understanding the business requirements and objectives. It involves gathering and analysing user requirements, conducting user workshops, and defining the scope of the project. This stage helps establish a solid foundation for the development process and ensures that the project meets the needs of the business and end-users (Dingsøyr, Dybå & Moe, 2010).

3.7.2.3 Functional Model Iteration

In the functional model iteration stage, the development team creates and refines functional models based on the defined requirements. Iterative timeboxes are used to develop, test, and review these functional models. The goal is to build a system that fulfils the high-level processing and information needs established during the business study phase (Boehm & Turner, 2003).

3.7.2.4 Design and Build Iteration

The design and build iteration stage focuses on the detailed design and implementation of the system. It includes creating design diagrams, such as use case diagrams, sequence diagrams, and entity-relationship diagrams. The development team follows an iterative approach to design, build, test, and review the system components. This stage ensures that the system is built incrementally and aligns with the user requirements (Dingsøyr, Dybå & Moe, 2010).

3.7.2.4.1 Design Diagrams

The project has the following diagrams:

i. Use Case Diagram
ii. Sequence Diagram
iii. Entity-Relationship Diagram
iv. Database Schema
v. System Wireframes

3.7.2.4.2 Domain of Execution

The project aimed to develop a progressive web platform, leveraging the advantages of web-based solutions for software development. Web platforms offer wide accessibility across devices, ensuring a consistent user experience. They provide speed and flexibility, enabling developers to customise the platform to meet specific requirements. A key advantage is the ability to scale up to handle increased user demand, minimising disruptions. Progressive web apps are known for their reliability, speed, and engaging user experience. These qualities make them highly appealing to developers and users alike. In summary, the progressive web platform proposed in this project capitalises on the strengths of web-based solutions, delivering an accessible, versatile, and engaging platform (Fortunato & Bernardino, 2018 ; Tandel, Tandel & Jamadar, 2018).

3.7.2.4.3 Design and Development Tools

Developers previously faced challenges when saving and loading data between memory and a database, resulting in complex and difficult-to-maintain software. To address these issues, Visual Paradigm introduced a persistence layer, bridging object and data models. This layer offered essential features like transaction support, pluggable caching, connection pooling, and customised SQL statements. By using this object-relational mapping layer, developers could focus on meeting business requirements instead of implementation details. Visual Paradigm's model-driven platform provided powerful tools, allowing developers to edit entity-relational models generated through reverse engineering or build models using UML diagrams. Modifications to the model were instantly reflected in the smart model-code generator, ensuring an up-to-date persistent model (*Benefits of Designing Database With Visual Paradigm*, n.d.).

Various technologies were employed in this project. HTML, CSS, and JavaScript were used for web page creation, content definition, layout, and interactivity. Python was chosen for data analysis due to its mathematical and computational libraries (Arora, 2022). SQLite was selected for creating embedded software, as it handles HTTP requests with low to moderate traffic and compresses files while retaining vital information. It serves as a provisional dataset for managing application data (*SQLite*, n.d.). Technology selection was based on project requirements and availability of resources and tools.

Visual Studio Code, an open-source IDE supporting multiple programming languages and platforms, was the primary IDE used. Its lightweight, fast, and customisable nature aligned with the principles of simplicity and ease of use in agile software development. Visual Studio Code's marketplace offered extensions enhancing productivity, such as code snippets, highlighting, and completion. Additionally, it supported unit testing and debugging (Cangemi, 2020).

To improve data analysis reproducibility and transparency, Jupyter Notebook was employed. This open-source web application allowed the creation and sharing of documents containing code, data, and explanatory text. It facilitated end-to-end data science workflows, including data cleaning, statistical modelling, machine learning, and visualisation. Jupyter Notebook enabled clear organisation and presentation of code and data, aiding understanding and reproducibility. Interactive code cells allowed rapid iteration and experimentation, streamlining the data analysis process (Jupyter, n.d.).

3.7.2.5 Implementation

The project life cycle consists of several stages, each with its specific objectives. The feasibility study and financial analysis phases are essential and interdependent, focusing on assessing the project's financial viability and alignment with organisational goals. The feasibility study identifies potential risks and limitations, while the financial analysis evaluates costs, benefits, and return on investment. The outcomes inform decisions on project continuation, modification, or abandonment (Rastogi & Professor, 2015).

The iterative design and development phases aim to achieve a final version that closely matches initial designs. Detailed system requirements, functional specifications, and design documentation are created during the design phase. This ensures well-defined project scope and adherence to desired specifications. The development phase involves coding, testing, and integration of components to build the system (Rastogi & Professor, 2015).

In the implementation phase, the developed system is made operational through activities like installation, user training, and data conversion. Thorough testing and bug resolution are essential before the system is released to end-users. Implementation marks project completion and system delivery. (Rastogi & Professor, 2015)

The project life cycle approach ensures comprehensive evaluation, risk management, progress tracking, and timely, budgeted project delivery. Factors such as project size, complexity, and scope influence the duration and number of phases involved (Larson & Gray, 2014).

3.7.3 Post-Project

The Development and Maintenance phase is focused on coding and building the system based on established design specifications and requirements. This phase ensures the system's proper functioning and adherence to essential criteria. Developers use the selected programming language and architecture, addressing challenges and making necessary modifications during the coding process. System maintenance is also performed to ensure ongoing efficiency, including bug fixes, updates, and enhancements based on user feedback and evolving requirements. This phase is crucial for the system's success, as it ensures its continued performance and adaptability to changing user needs (Abrahamsson, Oza & Siponen, 2010).

3.8 Utilisation and Dissemination of Results

The research outcomes have the potential to significantly contribute to the development of a forecasting platform for equity investments, benefiting investors. Additionally, the findings can serve as a foundation for further exploration by researchers in the field. The project's primary objective is to identify the origins of events impacting stock market prices, with applications across various industries for detecting similar events and gaining insights into diverse phenomena.

In academic research, "dissemination" refers to the extensive transmission and exchange of knowledge and information. This scholarly communication occurs through formal and informal channels. The research utilises various methods to disseminate its findings, including informal networks, open access, and formal publishing. Informal networks involve sharing research findings through discussions and seminars, as demonstrated in this project through weekly seminars with colleagues. Open access involves making research articles freely available online to potential readers. Publishing encompasses the production and distribution of research through journals, books, and other sources (Johanson & Williamson, 2017).

3.9 Ethical Considerations/ Issues

Participants in this project were not provided with reimbursement in line with ethical principles governing research involving human participants. The decision was made to avoid potential coercion or influence that reimbursement may create, preserving the objectivity of responses. The project does not pose physical risks or discomfort, and the time commitment for participation is expected to be minimal. Therefore, the absence of reimbursement is not anticipated to be a significant barrier to participation. Ethical considerations regarding reimbursement were thoroughly reviewed by the research team and institutional ethics committee, ensuring responsible and ethical conduct of the project.

When dealing with people's data, it is crucial to handle it with utmost sensitivity, respecting their human rights, privacy, and public life. In this project, necessary permissions were obtained from relevant authorities, including the National Commission for Science, Technology & Innovation (NACOSTI), and approval was obtained from the Ethics Review Board (ERB) through the Strathmore University Institutional Research Ethics Committee (IREC). The researcher received comprehensive ethics and research training and recognised the importance of adhering to ethical considerations.

To safeguard participant rights, all respondents were informed of the voluntary nature of their involvement. At the start of each interview, confidentiality and the right to withdraw or refuse to answer questions were assured. Unless consent was denied or withdrawn, interviews were recorded, and data was securely stored and treated as confidential. These ethical considerations were pivotal in upholding participants' rights and dignity throughout the project.

3.10 Risks and Benefits Analysis

The project involves certain risks that need careful consideration:

i. Data Privacy and Security: Collecting and processing personal information for the recommender system exposes the project to data privacy and security issues. Risks include data breaches, cyber-attacks, and potential misuse, which can harm the project's reputation and investor trust. Strong security measures such as encryption, backups, storage safeguards, and access restrictions are necessary to protect sensitive data. Compliance with data protection requirements and industry best practices is crucial.

ii. Technical Complexity: Developing an online platform with machine learning capabilities poses technical challenges and the risk of failures and errors. These can lead to inaccurate recommendations, system downtime, and a poor user experience. Implementing rigorous testing and quality assurance measures ensures the platform's functionality, reliability, and performance, reducing these risks.

iii. Regulatory Compliance: Adhering to regulatory requirements and ethical standards is essential. Non-compliance can result in legal action, damaging the project's reputation. Obtaining necessary permits and approvals, following data protection laws, and adhering to ethical principles like informed consent and confidentiality are critical to mitigate regulatory risks.

The project also presents significant potential benefits:

i. Improved Investment Outcomes: The system assists novice investors in making informed decisions, potentially leading to higher returns and increased investor confidence. It can reduce irrational trading practices, promoting stability and reducing financial losses.
ii. Accessibility of Data: The online platform enhances access to relevant data, saving time and effort for investors. Novice investors can access and analyse complex information, leading to more informed decision-making and increased participation in the stock market.
iii. Increased Investor Confidence: Accurate recommendations and reliable data enhance investor confidence and trust. This can attract more investors, increase liquidity, and contribute to a stable and sustainable stock market.
iv. Financial Literacy: The project promotes financial literacy by educating novice investors on trading practices and investment strategies. This empowers investors to make informed decisions based on accurate data and improves understanding of the stock market.

Managing risks effectively and adhering to ethical considerations are crucial. The project's risk management strategies should address data privacy, technical challenges, and regulatory compliance. By doing so, the project can deliver valuable investment information, enabling superior investment decisions and outcomes for novice investors.

Chapter 4: Information System Analysis and Design

4.1 Overview

A solid grasp of the Analysis and Design methodology relies on a thorough understanding of the key concepts discussed in Chapter 2. This chapter aims to equip readers with the ability to comprehend the process of gathering functional and non-functional requirements tailored to individual entrepreneurs, as outlined in this section. The analysis delved into a detailed exploration of information needs for entrepreneurs, partners, mentors, and investors, offering a comprehensive understanding of the essential elements for effective networking. The chapter concludes with the proposed system design, which integrates the analysis findings and provides a comprehensive solution to the identified requirements.

4.2 Requirements Gathering

This research project utilised multiple methodologies to comprehensively analyse the collected data and derive valuable insights into individual investors' investment objectives, risk tolerance, and current investments. Thematic analysis served as the primary method for examining interview data, identifying and interpreting patterns and themes in participants' experiences. The identified themes underwent careful categorisation and validation against the original data to ensure accuracy before presenting them coherently. A detailed report of this in-depth analysis is available in Appendix G.

For the quantitative data obtained from closed-ended questions in Appendix F, inferential statistics were employed. Descriptive statistics summarised the answers, identified patterns and trends, while inferential statistics evaluated the significance of observed differences or connections between variables. The results of this analysis are presented in a detailed report found in Appendix H.

Additionally, documentation analysis was employed to gain deeper insights into market trends and financial themes. Content analysis enabled a thorough evaluation and interpretation of textual or visual materials, uncovering patterns, themes, and potential biases within the data. This approach facilitated a comprehensive review of substantial data, providing valuable insights into stock market operations. A detailed report outlining the findings of this analysis can be found in Appendix I.

The findings obtained from the three data analysis methods, namely Thematic analysis, Documentation analysis, and Content analysis, played a crucial role in identifying the system's functional requirements. These results served as the foundation for the System Requirements Analysis and subsequent project phases.

4.3 System Requirements Analysis

4.3.1 Functional Requirements

Functional requirements are essential in specifying the expected performance and functionality of a system. They define the capabilities, features, and functions necessary for an online stock market recommender system to deliver a reliable and user-friendly experience to investors. These requirements assist investors in making informed investment decisions and improving their investment outcomes. The functional requirements for the online stock market recommender system, derived from the data analysis procedures, are summarised in Table 4.1. These requirements serve as a guiding framework for the development and design of the system, ensuring that it meets the expectations and needs of individual investors.

Table 4.1: Functional Requirements

ID	Description
FRQ1	The system should enable user registration and account creation, allowing users to input personal information, login credentials, and investment preferences securely.
FRQ2	The system should provide users with the ability to view and edit their profiles, update investment preferences, and manage account settings.
FRQ3	The system should deliver up-to-date information on various stocks and financial instruments, including historical data, current prices, and relevant details.
FRQ4	The system should utilise machine learning algorithms to generate personalised recommendations for each user based on their investment history and preferences, ensuring relevant and accurate suggestions.
FRQ5	The system should establish a platform for users to access their investment portfolio and receive performance recommendations.
FRQ6	The system should implement a notification system that alerts users about significant market trends and important events with potential impacts on their investments.
FRQ7	The system should generate comprehensive reports and analytics on user investment performance, including charts and graphs illustrating investment growth over time.
FRQ8	The system should implement robust measures to ensure the security and encryption of user data and transactions, maintaining user privacy and confidentiality at all times.

FRQ9	The system should offer dedicated customer support services to assist users with their accounts, investment decisions, or technical issues.

4.3.2 Non-Functional Requirements

Non-functional requirements encompass attributes that are vital for the system's operation but do not directly relate to its functions. These attributes include speed, security, and usability. Ensuring a fast, secure, and user-friendly online stock market recommender system is crucial for delivering a satisfactory user experience. The non-functional requirements for the system are presented in Table 4.2, providing a comprehensive overview of the essential characteristics that support its effective functioning.

Table 4.2: Non-Functional Requirements

ID	Category	Description
NFRQ1	Performance	The system should efficiently process large amounts of data and provide prompt recommendations to users.
NFRQ2	Security	The system should establish robust security measures to safeguard user data against unauthorised access or manipulation.
NFRQ3	Usability	The system should be intuitive and easy to use, requiring minimal training for novice investors.
NFRQ4	Availability	The system should be accessible to users 24/7, allowing for trading at any time.
NFRQ5	Reliability	The system should be reliable and minimise downtime to avoid financial losses for investors.
NFRQ6	Scalability	The system should be scalable, capable of accommodating a growing user base and increasing data volume.
NFRQ7	Maintainability	The system should be easy to maintain and update, ensuring optimal performance over time.
NFRQ8	Compatibility	The system should be compatible with various devices and operating systems, ensuring accessibility and usability for all users.

4.4 System Architecture

The online stock market recommender system was developed using Flask, a microframework based on Python, and adopted a client-server architecture. The client side was built using web technologies such as HTML, CSS, and JavaScript, while the server-side utilised Flask for server-side processing. The client-side consisted of three layers: presentation, logic, and data. The presentation layer focused on the visual aesthetics, the logic layer handled user input and request processing, and the data layer managed data storage and retrieval.

The server-side architecture was structured into two layers: application and data. The application layer managed client interactions and business logic, while the data layer was responsible for data storage and retrieval. The server-side architecture followed the Model-View-Controller (MVC) paradigm, where the model represented the data layer, the view represented the application layer, and the controller represented the Flask framework.

To ensure scalability and fault-tolerance, the system incorporated a load balancer and multiple server instances. The load balancer distributed incoming requests across the available server instances, while the server instances handled the requests and generated responses. In case of a server failure, a backup server instance was in place to take over. The system design, leveraging Flask, provided a robust and scalable solution for the online stock market recommender system.

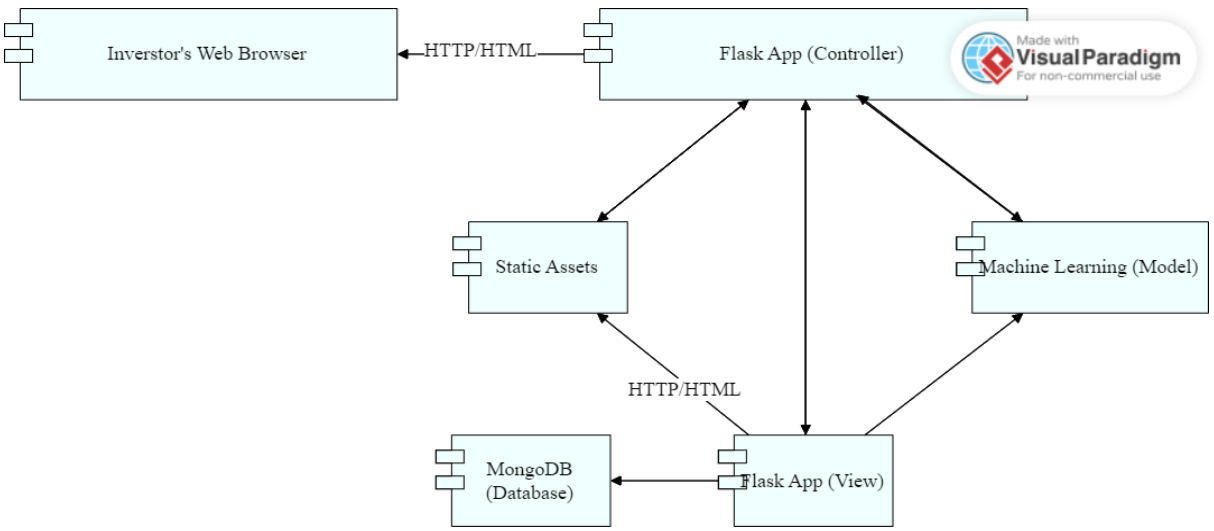


Figure 4.1: System Architecture

4.5 System Design Diagrams

4.5.1 Use Case Diagram

This use case diagram illustrates the interactions between actors and the investment platform system, which is designed to address the specific investment requirements of individual investors. The system incorporates diverse data sources, including historical stock data,

financial news and analysis, and individual investor profiles, to generate personalised investment recommendations. It also encompasses functionalities such as machine learning model training and refinement, portfolio management, and system administration, including user account management, system settings configuration, and generation of analytics and reports. By showcasing the primary actors involved in each use case and their communication, the diagram provides an overview of the system architecture. For detailed descriptions of the main use cases, please refer to Appendix J.

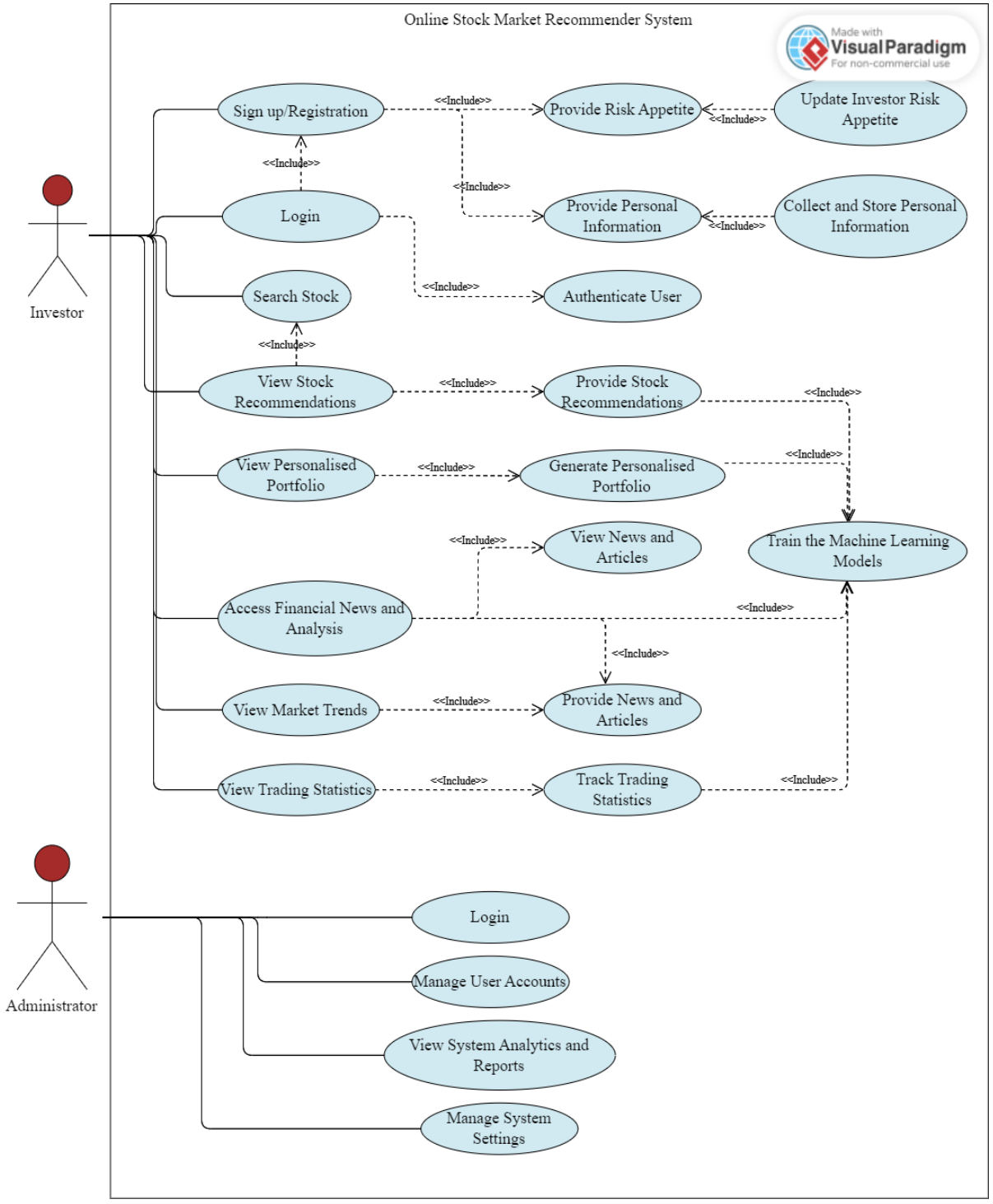


Figure 4.2: Use Case Diagram

4.5.2 Sequence Diagram

The following sequence diagram presents the operational flow of an online stock market recommender system that employs machine learning techniques to deliver customised trading recommendations to individual investors. This conceptual model is designed to process and analyse data from multiple sources, with each dataset handled independently. By utilising demographic, psychographic, and behavioural data, the system segments investors and identifies lucrative investment opportunities while filtering out low-value stocks. Moreover, the system exclusively suggests high-risk investments to investors who exhibit a high-risk appetite. To gauge public sentiment regarding specific commodities and their impact on stock prices, sentiment analysis is performed on data retrieved from financial data APIs and Twitter. Subsequently, an ensemble agent is utilised to determine the most effective models for developing a trading strategy, which is then evaluated for accuracy. Based on fresh data, a trained recommender model generates personalised recommendations tailored to the investor's risk profile and investment preferences. For a more comprehensive breakdown of the primary sequence of events, please refer to Appendix K.

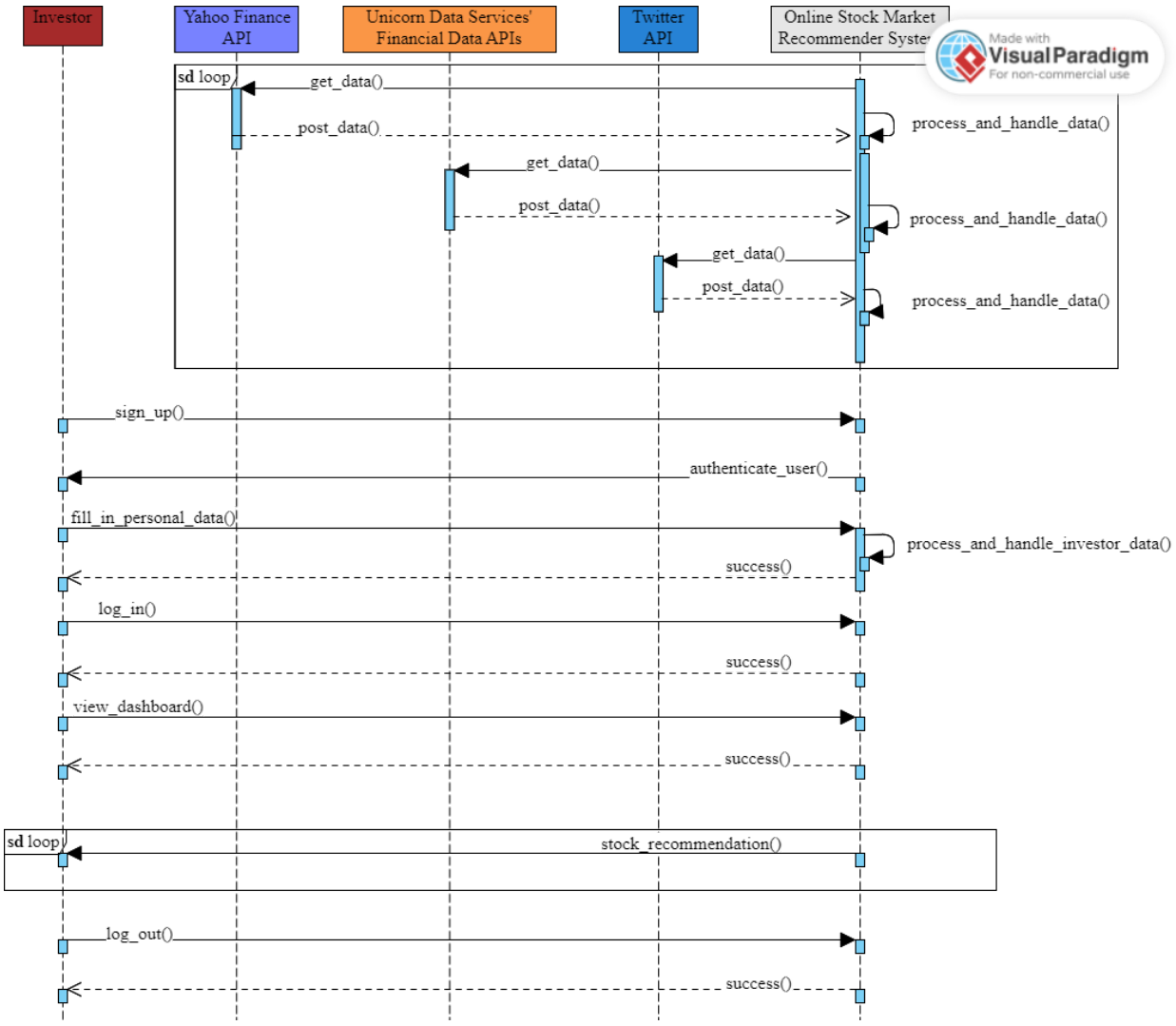


Figure 4.3: Sequence Diagram

4.5.3 Entity Relationship Diagram

The Entity Relationship Diagram (ERD) represents the interrelationships among key entities in the online stock market recommender system, which has been developed to promote sound trading practices for novice investors in Kenya. The ERD showcases the connections between various entities such as investors, investment segments, stocks, recommendations, and machine learning models. By illustrating the attributes associated with each entity and the relationships between them, the diagram offers a comprehensive view of the system architecture. This ERD serves as a visual aid that facilitates understanding of the data flow within the system, highlighting the essential components that contribute to the effective operation of the online stock market recommender system.

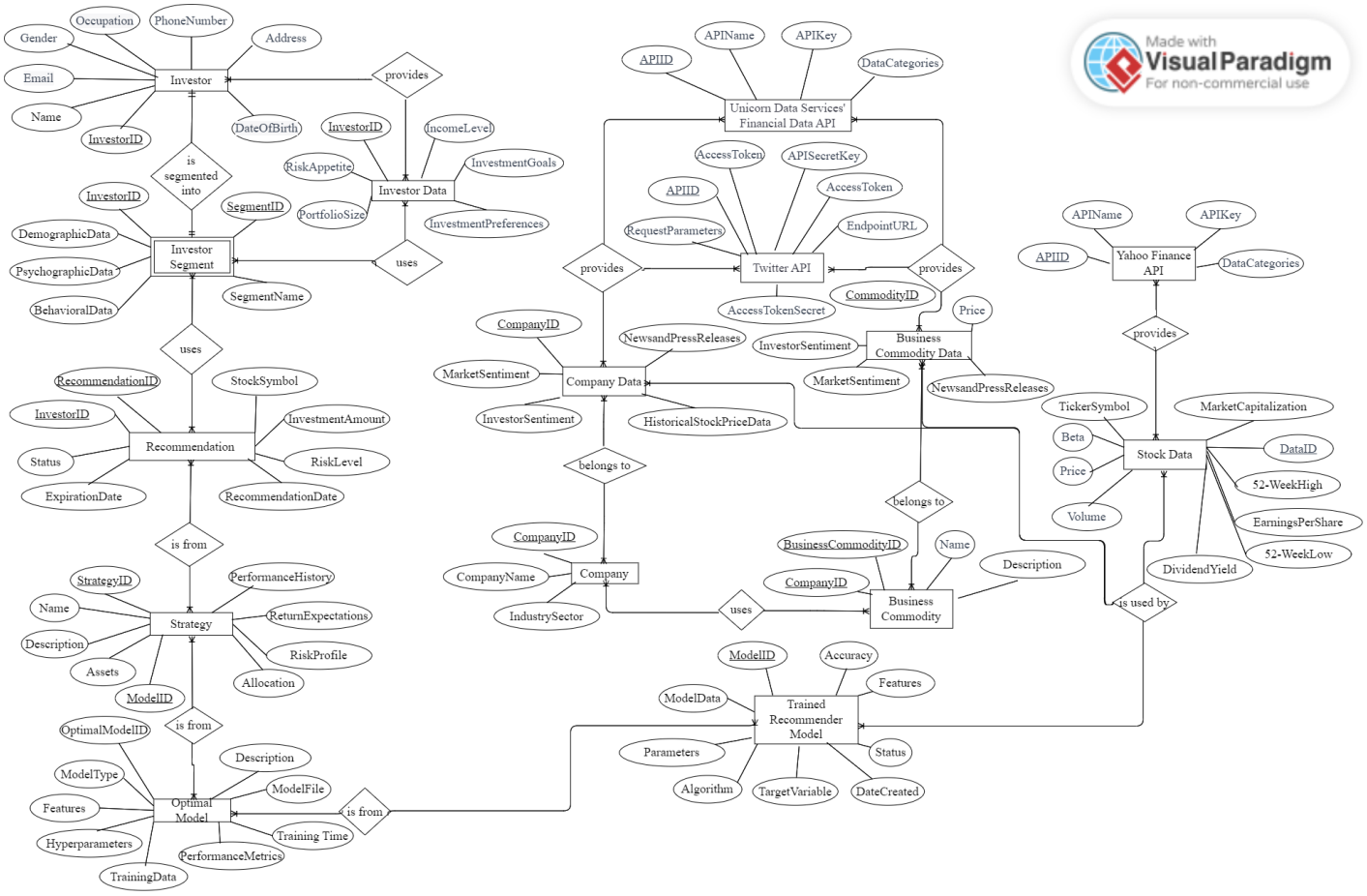


Figure 4.4: Entity Relationship Diagram

4.5.4 Database Schema

The development of an online stock market recommender system with machine learning entails a meticulous database schema design process to establish a resilient, scalable, and secure database capable of efficiently managing and processing extensive datasets. The schema design must account for diverse data types, including stock market data, investor profile data, historical data, and news data, which the system will handle. A well-designed database schema ensures the structured and organised storage of data, facilitating seamless

integration with the machine learning models responsible for generating investment recommendations. The security and privacy of sensitive information, such as user account details and financial transactions, are also integral aspects considered during the schema design. In this project, the adoption of the SQLite database is based on its versatility, scalability, and competence in handling unstructured data types. The database schema assumes a vital role in the investment platform, serving as the bedrock for generating personalised investment recommendations and enhancing investment outcomes for novice investors.

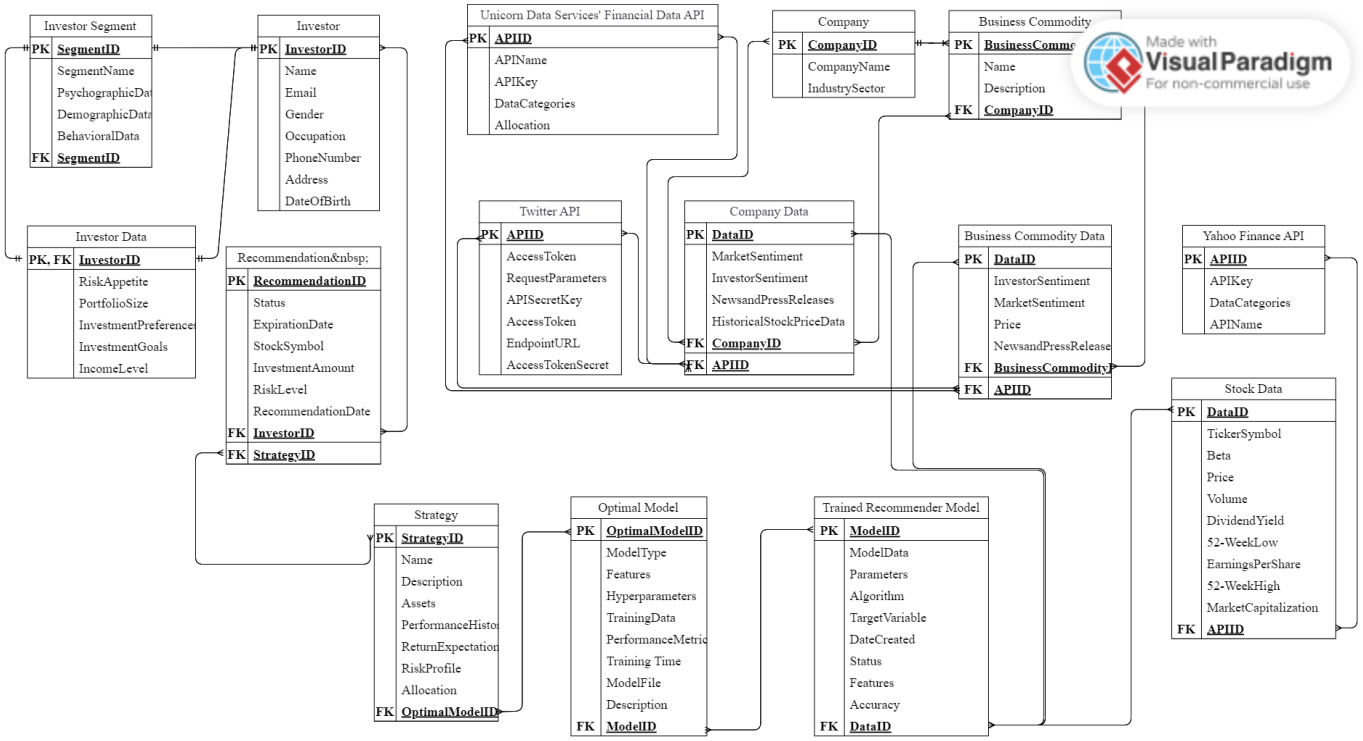


Figure 4.5: Database Schema

4.5.5 System Wireframes

The system wireframes hold significant importance in the design phase of the online stock market recommender system as they provide a visual representation of the user interface and user experience. Figures 4.6, 4.7, and 4.8 showcase the wireframes, illustrating the screens and functionalities of the system. These wireframes enable the development team to ensure that the design aligns with project requirements and user expectations, thereby facilitating an efficient development process. The system wireframes play a crucial role in ensuring that the system's design is intuitive, user-friendly, and easily navigable. For this project, the wireframes were created using web technologies such as HTML, CSS, and JavaScript, with a primary emphasis on usability and functionality. By establishing a robust and user-centric design, novice investors can effortlessly access and utilise the system, leading to improved investment decision-making and outcomes.

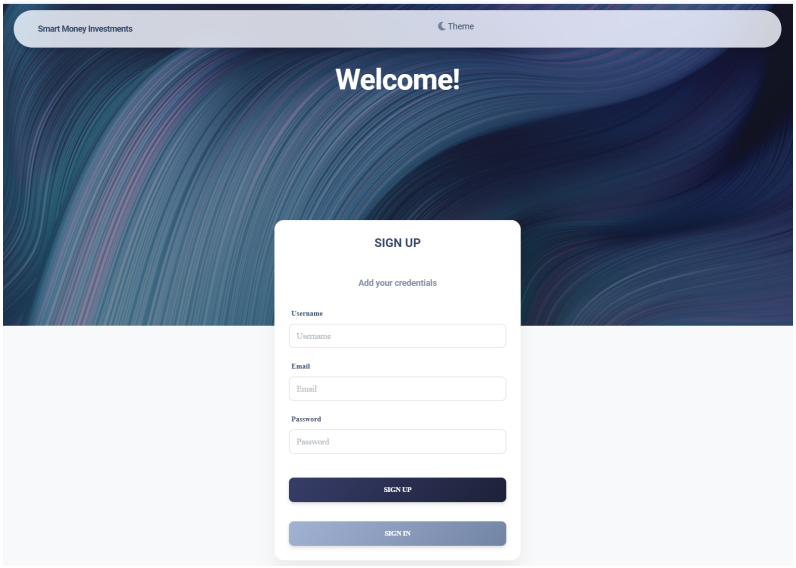


Figure 4.6: Sign Up Page Wireframe

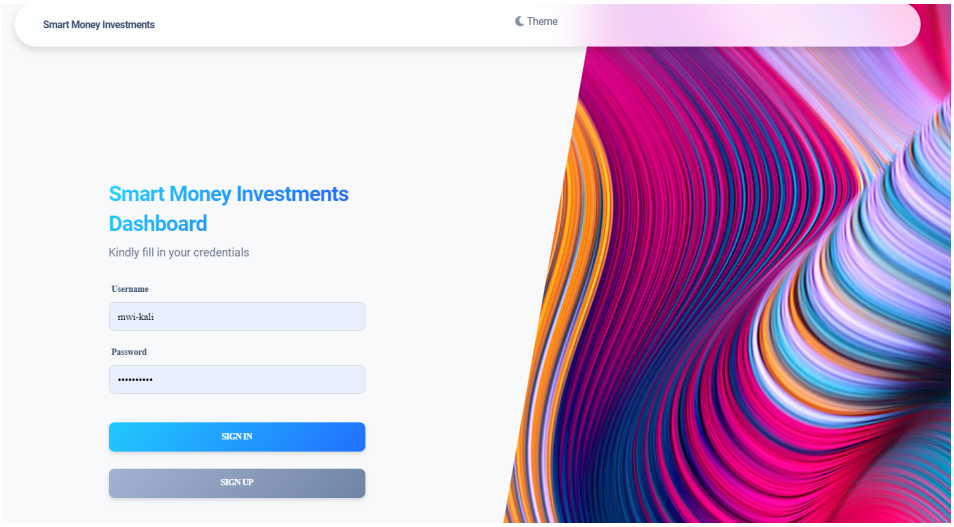


Figure 4.7: Login Page Wireframe

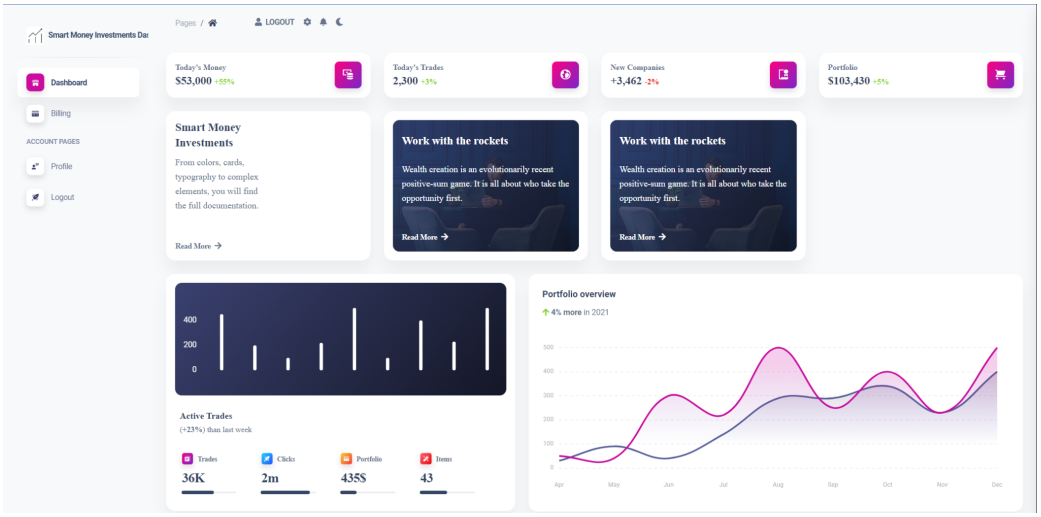


Figure 4.8: Investor Dashboard Wireframe

Chapter 5: Information System Implementation and Testing

5.1 Overview

A comprehensive understanding of this chapter relies on a thorough grasp of the system requirements detailed in Chapter 4. This chapter presents a comprehensive account of the implementation and testing process for an online stock market recommender system that harnesses the power of machine learning to offer individual investors a trading environment grounded in factual information and data, rather than speculative assumptions. The chapter critically analyses the system's architecture, paying particular attention to the various components, subsystems, and associated operations that significantly impact the overall system performance. The effectiveness of the system is systematically evaluated through a series of performance tests and benchmarking techniques, with the results extensively examined and discussed. The implementation and testing process delineated in this chapter not only highlights the feasibility of the proposed solution but also underscores its potential to revolutionise the online stock trading industry.

5.2 System Implementation

5.2.1 Feasibility Study and Financial Analysis of the System

The development of the online stock market recommender system followed a rigorous project life cycle approach, integrating a conceptual model and machine learning methodologies to deliver data-driven trading solutions to individual investors. The feasibility study and financial analysis phases played a pivotal role in assessing the project's financial viability, profitability, and alignment with strategic objectives. A comprehensive budget was formulated to monitor and manage financial resources throughout the project's lifecycle, ensuring effective financial control within specified constraints. This systematic approach contributed to the successful implementation of a robust and dependable system, capable of providing personalised investment recommendations to investors based on accurate and reliable data.

5.2.2 System Architecture and Development Process of the System

The design and development of the online stock market recommender system incorporating machine learning techniques followed a systematic approach, employing various design diagrams to guide the development process. These diagrams provided visual representations of the system's components and interactions, ensuring a coherent and well-defined system architecture.

The project was implemented on a progressive web platform, leveraging web technologies such as HTML, CSS, and JavaScript, as well as Python for data analysis and SQLite for efficient data storage and management. The utilisation of Visual Studio Code as the primary

Integrated Development Environment (IDE) and Jupyter Notebook for data analysis significantly contributed to enhanced productivity and transparency. The online stock market recommender system utilising machine learning was designed to integrate multiple features, including personalised recommendations, low-value stock filtering, risk assessment, sentiment and relationship analysis, model selection, accuracy evaluation, and data collection from diverse sources. These features were crucial in providing users with reliable, accurate, and timely recommendations that aligned with their expectations.

The design and development phases played a pivotal role in the creation of a comprehensive and efficient online stock market recommender system. The incorporation of design diagrams, progressive web platforms, and innovative development tools resulted in a well-structured and dependable system architecture that effectively met the specified requirements and objectives.

5.2.3 Testing and Validation of the System

The development of the online stock market recommender system, incorporating machine learning techniques, necessitated a rigorous approach to system testing and validation, ensuring compliance with essential standards for performance, reliability, and accuracy. The chosen methodology employed a white-box testing approach, entailing a thorough examination of the system's internal mechanisms. The system comprised multiple modules, each dedicated to specific tasks such as data collection, processing, and analysis. To ensure adherence to specified boundaries and accuracy, each module underwent rigorous unit testing, followed by integration testing to verify seamless compatibility and functional testing to assess performance and identify potential errors.

The validation procedure adopted in this project adhered to best practices in software development and testing. It encompassed evaluating the accuracy and timeliness of data collection and processing capabilities. The validation of individual modules, including stock filtering, risk assessment, sentiment analysis, relationship assessment, model selection, accuracy evaluation, and recommender modules, was conducted in accordance with the modular approach to software development and testing.

The rigorous system testing and validation approach employed in this project facilitated the development of a robust, reliable, and high-performance online stock market recommender system, catering to the needs of individual investors in Kenya and beyond. Through comprehensive testing, potential errors were identified and necessary improvements were implemented to ensure optimal performance, culminating in the delivery of a functional system to end users.

5.2.4 System Modules

The online stock market recommender system employs multiple modules to ensure the delivery of reliable, accurate, and timely recommendations. These modules encompass various tasks, including data collection, investor data processing, stock filtering, risk assessment, sentiment analysis, relationship assessment, model selection, accuracy evaluation, and recommendation generation. These modules work together cohesively to meet the expectations of users and enhance the system's effectiveness.

The Authentication Module plays a crucial role in securing the system by implementing a robust login process that verifies users' identities. It employs secure authentication protocols and cryptographic algorithms to ensure data privacy, confidentiality, and integrity, thereby establishing trust and confidence among users.

The Data Collection Module acquires diverse data from multiple sources, such as financial data APIs, social media platforms, and news outlets. This comprehensive data collection process enriches the system with a wide range of relevant information, including stock prices, news articles, and sentiment analysis, among others, improving the accuracy and relevance of the recommendations.

The Investor Data Module is responsible for processing and managing investor data to create individual profiles. It utilises machine learning algorithms, such as k-means clustering, to segment investors based on their demographics, psychographics, and behaviours. These segments are assigned labels representing different risk-return profiles, allowing for personalised recommendations. The module also employs a random forest classifier to predict investor groups for new data, optimising hyperparameters through grid search, and evaluating model accuracy. By incorporating these algorithms, the module enhances the system's ability to provide tailored recommendations that align with investors' preferences and improve their investment experience.

The Stock Filtering Module filters out low-value stocks and recommends high-value stocks based on investor ratings and a sophisticated clustering methodology. It leverages data from the Investor Data Module to generate personalised recommendations, optimising the investment strategy by focusing on high-value stocks.

The Risk Assessment Module analyses investor risk tolerance and provides recommendations for low-risk investments to those with a low-risk tolerance. By leveraging the investor data from the Investor Data Module, including characteristics such as age, income, and investment history, this module creates personalised risk profiles. Based on these profiles, the module

recommends suitable investment options that align with the investor's risk tolerance, improving the relevance and accuracy of the recommendations.

The Sentiment Analysis Module plays a critical role in understanding the sentiments surrounding a company and its core business commodities. It employs advanced natural language processing techniques and machine learning algorithms to extract valuable insights from diverse data sources, such as financial data APIs and social media. The module employs a machine learning algorithm specifically designed for sentiment analysis, including keyword matching, TF-IDF transformation, clustering optimisation, sentiment labelling, and prediction for new data. By leveraging this algorithm, the module facilitates informed investment decisions by extracting insights from unstructured data.

The Relationship Assessment Module evaluates the relationship between a company and its essential commodities, enabling the recommendation system to provide data-driven investment suggestions. It utilises data from the Data Collection Module and insights generated by the Sentiment Analysis Module to assign weights to each commodity based on their influence on stock prices. These weighted average sentiment scores provide valuable insights into the relationship between companies and commodities, improving the accuracy and relevance of the recommendations.

The Model Selection Module employs an ensemble model that combines Long Short-Term Memory (LSTM), Recurrent Neural Networks (RNN), and Graph Neural Networks (GNN). LSTM, a variant of RNN, excels at capturing long-term dependencies and sequential patterns in time series data, making it suitable for analysing stock market trends. RNNs are utilised for processing sequential data and identifying patterns, enhancing the modelling of time-dependent stock market data. GNNs are employed to analyse complex relationships and interdependencies within the stock market ecosystem. By combining these techniques, the Model Selection Module leverages their individual strengths to perform a comprehensive analysis of historical data and make accurate predictions of future stock market trends.

The Accuracy Evaluation Module rigorously assesses the performance of the recommendation system using statistical analysis and machine learning algorithms. It evaluates the accuracy, precision, recall, and other relevant metrics to identify areas for improvement and enhance the system's performance. This evaluation process is essential for maintaining the system's accuracy and ensuring it meets user expectations.

The Recommender Module generates personalised recommendations for individual investors based on data collected from various sources and insights from the other system modules. It

utilises trained recommender models to deliver tailored recommendations aligned with investors' preferences and risk appetite. By leveraging this module, investors can make informed investment decisions that are customised to their unique needs and improve their overall investment experience.

5.2.5 System Features

The online stock market recommender system incorporates a variety of essential features to deliver personalised and targeted investment recommendations, filter low-value stocks, assess risk, perform sentiment analysis, evaluate relationships, select models, evaluate accuracy, and generate tailored recommendations. These features have been meticulously designed to meet user requirements, ensuring the system's reliability, accuracy, and timeliness in providing recommendations.

The "Authentication" feature serves as a fundamental element of the online stock market recommender system, ensuring secure access by implementing a robust login process. By employing secure authentication protocols and cryptographic algorithms, this feature verifies users' credentials, such as usernames and passwords, guaranteeing the confidentiality, privacy, and integrity of the system's data. Upholding these security principles is crucial for establishing user trust and confidence, making it an indispensable component of the system. Passwords undergo a secure cryptographic algorithm called hashing, adding an additional layer of protection to the system's security infrastructure. Even in the event of a data breach, users' passwords remain uncompromised, providing further protection for their sensitive information. The implementation of this additional security feature further fortifies the system's security, ultimately enhancing users' trust and confidence.

```python
def hash_pass(password):
    """Hash a password for storing."""

    salt = hashlib.sha256(os.urandom(60)).hexdigest().encode('ascii')
    pwdhash = hashlib.pbkdf2_hmac('sha512', password.encode('utf-8'),
                                  salt, 100000)
    pwdhash = binascii.hexlify(pwdhash)
    return (salt + pwdhash)  # return bytes


def verify_pass(provided_password, stored_password):
    """Verify a stored password against one provided by user"""

    stored_password = stored_password.decode('ascii')
    salt = stored_password[:64]
    stored_password = stored_password[64:]
    pwdhash = hashlib.pbkdf2_hmac('sha512',
                                  provided_password.encode('utf-8'),
                                  salt.encode('ascii'),
                                  100000)
    pwdhash = binascii.hexlify(pwdhash).decode('ascii')
    return pwdhash == stored_password
```

Figure 5.1: Hashing Passwords in Python

The "Data Collection" feature plays a critical role in the online stock market recommender system by acquiring diverse data from multiple sources. This feature employs a comprehensive data collection process that allows the system to access a broad range of relevant data, including stock prices, news articles, and sentiment analysis. By gathering data from various sources, the system gains access to a wealth of information, enhancing the accuracy and relevance of the recommendations and improving the overall user experience.

The "Investor Profiling" feature is a crucial component of the online stock market recommender system, generating personalised investor recommendations based on their demographic, psychographic, and behavioural characteristics. This feature continuously processes and analyses data from various sources to create comprehensive and tailored profiles for each investor. It leverages the RandomForestClassifier machine learning algorithm, as depicted in Figure 5.2, to classify investors into categories such as "Low-risk investor" and "High-risk investor," as shown in Figure 5.3. The primary objective of this feature is to ensure that the data used to inform the system's recommendations is accurate and relevant, resulting in an improved user experience. Continuously updating the profiles ensures that the recommendations provided to investors align with their investment needs and preferences, ultimately enhancing the system's effectiveness.

```
self.clf = RandomForestClassifier(n_estimators=100, max_depth=5, random_state=42)
```

Figure 5.2: Random Forest Classifier Parameters

```
def classify_risk_tolerance(self, new_investor):
    new_features = pd.DataFrame(new_investor, index=[0])
    new_features_encoded = self._encode_categorical_features(new_features)
    risk_tolerance = self.clf.predict(new_features_encoded)
    if risk_tolerance[0] == 0:
        return "Low-risk investor"
    else:
        return "High-risk investor"
```

Figure 5.3: Classification of Investors Based on Investment Preferences and Behavior into Two Categories

The "Stock Filtering and Recommendation" feature plays a pivotal role in the online stock market recommender system by employing an advanced filtering mechanism to identify low-value stocks and generate personalised recommendations based on high-value stocks. It utilises the unique ratings of individual investors and a sophisticated clustering methodology incorporating the KMeans algorithm, as illustrated in Figure 5.4, to analyse comprehensive data from the Investor Data Module. The resulting cluster labels are then used to recommend high-value stocks to investors based on their risk tolerance, with cluster 0 suggested to

high-risk-tolerant investors and cluster 1 to those with lower risk tolerance. The recommended stocks are sorted in descending order of investment value, and the function returns the resulting recommended data. This feature's ability to improve the relevance and accuracy of the system's recommendations enhances the user experience, making it an essential aspect of the system.

```
kmeans = KMeans(n_clusters=2)
kmeans.fit(X)
stock_data['Cluster'] = kmeans.labels_

# Recommend high-investment-value stocks based on cluster and investor's risk tolerance
if investor_data['Risk_Tolerance'] == 'High':
    recommended_data = stock_data[stock_data['Cluster'] == 0]
else:
    recommended_data = stock_data[stock_data['Cluster'] == 1]
```

Figure 5.4: KMeans Clustering Algorithm Applied to Stock Data for Tailored Recommendations

The "Personalised Risk Assessment and Recommendation" feature is a fundamental aspect of the online stock market recommender system, utilising machine learning algorithms to conduct personalised risk assessments based on individual investor characteristics, such as age, income, and investment history. This feature retrieves the investor's risk tolerance from the database and compares it to their portfolio's overall risk level to suggest low-risk or high-risk moves, as depicted in Figure 5.5. Low-risk moves are recommended if the investor's risk tolerance is lower than their portfolio's overall risk level, while high-risk moves are suggested if their risk tolerance is higher. By enhancing the relevance and accuracy of the system's recommendations, this feature ultimately improves the investor's user experience.

```
risk_tolerance = pd.read_sql_query(
    f"SELECT risk_tolerance FROM investors WHERE investor_id={investor_id}",
    db_conn
)

if risk_tolerance < total_risk:
    recommended_moves = ['Invest in low-risk bonds', 'Invest in stable dividend-paying stocks']
else:
    recommended_moves = [
        'Invest in growth stocks with higher risk and potential returns',
        'Consider alternative investments such as real estate or commodities'
    ]
```

Figure 5.5: Comparison of Investor's Risk Tolerance and Portfolio's Overall Risk Level for Low-risk or High-risk Investment Recommendation

The "Sentiment Analysis" function utilises sophisticated natural language processing techniques and machine learning algorithms to extract valuable insights from unstructured

data. This process plays a crucial role in facilitating well-informed investment decision-making by assessing sentiments from various sources and discerning the emotional tone and attitude towards a company and its commodities. Although publicly available sentiment analysis exists, they often lack the requisite specificity for financial data. To overcome this limitation, a tailored sentiment analysis was developed, employing the Random Forest Classifier technique, as illustrated in Figure 5.7. The Random Forest Classifier technique was selected as the optimal machine learning model, outperforming alternatives such as Logistic Regression, K-Nearest Neighbour (KNN), and Decision Trees, as evidenced in Figure 5.6. The resulting insights significantly enhance the pertinence and accuracy of the system's recommendations, thereby elevating the overall user experience for investors.

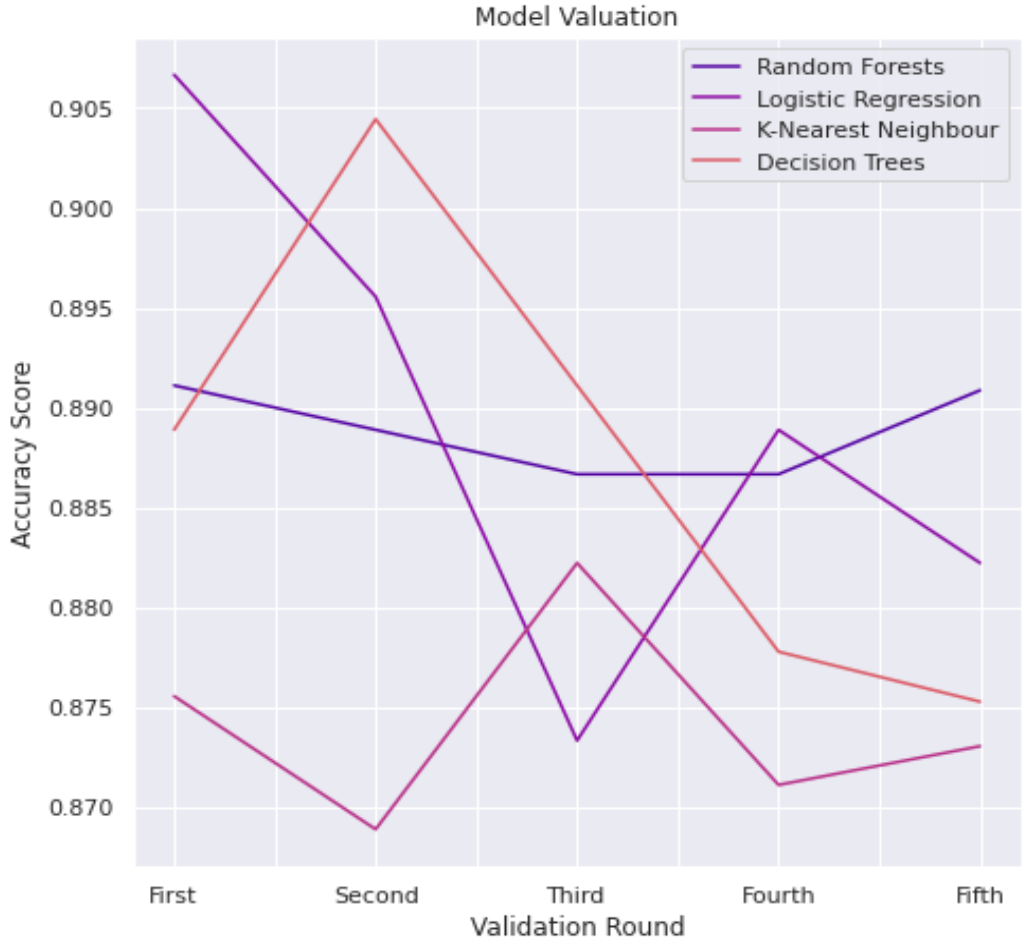


Figure 5.6: Cross-validation of Machine Learning Models in Sentiment Analysis

```
for company, commodities in companies.items():
    # Retrieve the historical data for the company and its commodities from the database
    data = []
    for commodity in commodities:
        query = f"SELECT tweet_text, sentiment_score FROM tweets WHERE company='{company}' AND commodity='{commodity}'"
        tweets = pd.read_sql(query, conn)
        tweets['commodity'] = commodity
        data.append(tweets)

    company_df = pd.concat(data, axis=0)

    rfc = RandomForestClassifier()
    rfc.fit(company_df['tweet_text'].str.len().values.reshape(-1, 1), company_df['sentiment_score'])

    # Use the model to predict sentiment scores for each commodity
    sentiment_scores = {}
    for commodity in commodities:
        commodity_df = company_df[company_df['commodity'] == commodity]
        avg_score = rfc.predict(commodity_df['tweet_text'].str.len().values.reshape(-1, 1)).mean()
        sentiment_scores[commodity] = avg_score

        # Save the sentiment score to the database
        c = conn.cursor()
        c.execute(f"UPDATE commodities SET sentiment_score={avg_score} WHERE company='{company}' AND commodity='{commodity}'")
        conn.commit()

    # Calculate the average sentiment score for the company and its commodities
    avg_score = sum(sentiment_scores.values()) / len(sentiment_scores)
```

Figure 5.7: Machine Learning Model for Sentiment Analysis

The "Relationship Assessment" feature involves evaluating the relationship between a company and its essential commodities to provide informed and data-driven investment suggestions. This feature leverages various data sources, including news outlets and social media, to assign weights to each commodity based on its influence on stock prices. A weighted average sentiment score is calculated for a list of companies and their associated commodities. To compute this score, the total weight of all commodities for each company is determined, followed by the calculation of the average sentiment score for each commodity. The weighted average sentiment score for the company and its commodities is then computed based on their respective weights, as shown in Figure 5.8. The insights obtained through this feature enhance the relevance and accuracy of the system's recommendations, ultimately improving the user experience.

```
for company, commodities in companies.items():
    # Calculate the total weight of all commodities for the current company
    total_weight = sum(commodities.values())

    # Calculate the average sentiment score for each commodity for the current company
    avg_scores = {}
    for commodity, weight in commodities.items():
        query = f"SELECT sentiment_score FROM tweets WHERE company='{company}' AND commodity='{commodity}'"
        scores = c.execute(query).fetchall()
        if len(scores) == 0:
            avg_scores[commodity] = 0.0
        else:
            avg_scores[commodity] = sum([score[0] for score in scores]) / len(scores)

    # Calculate the weighted average sentiment score for the company and its commodities
    weighted_avg_score = sum([score * weight for score, weight in zip(avg_scores.values(), commodities.values())]) / total_weight
```

Figure 5.8: Computation of Weighted Average Sentiment Score for a Company and Its Commodities Based on Respective Weights

The "Model Selection" feature is a critical component of the online stock market recommender system, involving the selection of optimal models to create a viable investment strategy using advanced machine learning algorithms. This process employs an ensemble agent to process data from various sources, including historical data, financial data APIs, social media, and news sources. By analysing this data using advanced machine learning algorithms, the system can predict future trends in the stock market and select the most suitable models to create a robust and effective investment strategy. The outcome of this feature is essential in ensuring that the system generates personalised recommendations aligned with individual investor preferences and risk appetite.

The "Accuracy Evaluation" feature is a crucial component of the online stock market recommender system, employing statistical analysis and machine learning algorithms to comprehensively analyse the system's performance. Its primary objective is to identify areas for improvement and enhance the system's accuracy and precision, ultimately improving the user experience. The feature retrieves historical data for each company and its commodities from a database, concatenates them, and splits them into training and testing sets. For each company and model, the model is trained on the training set and evaluated on the testing set, with accuracy metrics such as accuracy score, precision score, recall score, and F1 score. The accuracy metrics for each company and model are stored in a list, providing valuable insights into the system's performance, as illustrated in Figure 5.9.

```
# Split the data into training and testing sets
X = company_df.drop(columns=['company', 'commodity', 'date', 'close_price_increase'])
y = company_df['close_price_increase']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3)

# Train and evaluate each model
for model_name, model in models:
    model.fit(X_train, y_train)
    y_pred = model.predict(X_test)
    accuracy = accuracy_score(y_test, y_pred)
    precision = precision_score(y_test, y_pred)
    recall = recall_score(y_test, y_pred)
    f1 = f1_score(y_test, y_pred)
    accuracy_metrics.append((company, model_name, accuracy, precision, recall, f1))
```

Figure 5.9: Accuracy Evaluation Feature

The "Personalised Recommendation Generation" feature is a crucial component of the system that generates tailored recommendations for individual investors using diverse data sources and system module insights. This feature involves continuous processing and handling of vast amounts of data to create custom profiles for each investor, providing relevant and accurate investment recommendations. Personalised recommendations are a significant value

proposition for users, fostering greater trust and confidence in the system. To achieve this, the system employs machine learning algorithms trained on historical data to predict the probability of a buy, sell, or hold recommendation based on the most recent data point, as demonstrated in Figure 5.10. The system sets a threshold to determine the cutoff probability for a particular recommendation. If the probability exceeds the threshold, a buy or sell recommendation is made; otherwise, a hold recommendation is made. The resulting recommendations provide valuable insights into the stock market, improving the user experience and enabling informed investment decisions.

```
prediction = model.predict_proba(latest_data)
if prediction[0][1] > threshold:
    recommendation = 'buy'
elif prediction[0][0] > threshold:
    recommendation = 'sell'
else:
    recommendation = 'hold'
```

Figure 5.10: Buy, Sell, or Hold Recommendations for Personalised Investment Suggestions

5.3 System Testing

The testing process undertaken was conducted comprehensively, employing various tests with the aim of identifying potential errors and implementing necessary improvements to optimise performance. The tests performed encompassed the following:

i. Unit testing: This phase involved testing each module individually to ensure their correct functionality and the accuracy of input and output. The unit testing procedure verified that each module adhered to the specified boundaries and that they could effectively retrieve and display the relevant data (Olan, 2003).

ii. Integration testing: Once the individual modules were validated, an integration test was conducted to assess their compatibility and seamless interaction. This test ensured that the integrated modules could communicate with each other effectively and that the overall functionality of the system remained unaffected by the integration of different modules (Ipate & Holcombe, 2007).

iii. Functional testing: It was imperative to assess the system's ability to function as a cohesive unit in order to achieve its desired goals. Hence, functional testing was carried out to simulate end-to-end scenarios involving investors, thereby providing valuable insights into the system's performance. This testing approach facilitated the identification of potential errors and enabled the implementation of necessary improvements to achieve optimal performance (Bühler & Wegener, 2008).

5.3.1 Test Environment

In order to uphold the integrity and reliability of the testing process, a dedicated testing environment was meticulously established. This environment was carefully designed to

mirror the hardware and software configurations of the production environment, thus creating a simulated operational setting that closely emulated real-world conditions. By isolating the testing environment from external factors, such as network latency, any potential impact on the system's performance was minimised, guaranteeing accurate and consistent test results. To further streamline the testing process, a suite of tools and software, including test automation frameworks and debugging tools, were strategically deployed within the testing environment. This comprehensive approach ensured that the system underwent thorough and meticulous testing and evaluation, conducted under precisely controlled and representative conditions, prior to its deployment in the production environment (Collins & de Lucena, 2012).

5.3.2 Test Cases

Test cases play a pivotal role in the software development life cycle, serving as a crucial mechanism to verify that the system being developed adheres to the specified requirements and functions as intended. These meticulously designed test cases form an integral part of the testing phase, enabling the identification of errors, flaws, and issues in the software prior to its distribution to end users. Throughout the testing process, various forms of testing were conducted, including unit testing, integration testing, and functional testing. This comprehensive approach ensured thorough examination and evaluation of the program from multiple perspectives, facilitating the timely identification and resolution of any identified faults before the final release. In this context, the creation of comprehensive and well-documented test cases assumes paramount importance. These test cases should be concise, intelligible, and straightforward to execute, contributing to the assurance of the program's quality and stability during the development phase (Catal & Mishra, 2013).

Table 5.1: Online Stock Market Recommender System Test Cases

TestID	Related Requirement	Inspection Check	Pre-condition	Test Data	Priority Level
T1	FRQ1	Can the system successfully register and create an account for a new user, capturing their personal information, login credentials, and investment preferences?	The user is not registered for an account.	Personal information, login credentials, and investment preferences.	High

T2	FRQ2	Can the system enable users to view and modify their profile, update investment preferences, and manage account settings?	The user is logged into their account.	Account settings and investment preferences.	High
T3	FRQ3	Can the system provide accurate and up-to-date information on various stocks and financial instruments, including historical data, current prices, and relevant details?	The system has access to up-to-date financial data.	Financial data for stocks and financial instruments.	Medium
T4	FRQ4	Can the system generate personalised recommendations for users based on their investment history and preferences, utilising machine learning algorithms that analyse user behaviour to provide relevant and accurate suggestions?	The system has access to user investment history and preferences.	Investment history and preferences.	High
T5	FRQ5	Can the system offer users a platform to view their investment portfolio and assess the performance of their investments, including recommendations?	The user is logged into their account.	User's investment portfolio.	High
T6	FRQ6	Can the system send notifications to users	The system has access to	Market trends and	Medium

		regarding significant market trends and other events that may impact their investments?	significant market trends and events.	important events.	
T7	FRQ7	Can the system generate reports and analytics on user investment performance, including charts and graphs that illustrate investment growth over time?	The system has access to user investment performance data.	User investment performance data.	Medium
T8	FRQ8	Can the system ensure the security and encryption of user data and transactions, safeguarding user privacy at all times?	User data and transactions are stored in the system.	User data and transactions.	High
T9	FRQ9	Can the system provide customer support to users who need assistance with their accounts, investment decisions, or technical issues?	The user needs assistance with their account, investment decisions, or technical issues.	User queries and issues.	Medium
T10	NFRQ1	Can the system efficiently process a large volume of data and provide prompt recommendations to users?	The system should handle a large dataset and deliver swift recommendations to users.	Dataset with 25 stocks and historical data, along with a user profile.	High

T11	NFRQ2	Is user data and transactional information protected against unauthorised access or manipulation?	User data and transactions are accessed or processed.	User data and transactions.	High
T12	NFRQ3	Is the system easy to use and understand for novice investors with minimal training?	A novice investor with minimal training is utilising the system.	User interface of the system.	Medium
T13	NFRQ4	Is the system accessible 24/7, allowing for trading at any time?	The system is operational.	None	High
T14	NFRQ5	Is the system reliable, minimising downtime to prevent losses for investors?	The system is operational and processing data.	None	High
T15	NFRQ6	Is the system scalable to accommodate a growing user base and increasing data volume?	The system is being used by a growing number of users and handling increasing data.	Large dataset of user investment history.	High
T16	NFRQ7	Is the system easy to maintain and update, ensuring optimal performance over time?	The system is operational and requires maintenance or updates.	System maintenance or update logs.	Medium
T17	NFRQ8	Is the system compatible with diverse devices and	The system is accessed via different	Various devices and	Medium

		operating systems, ensuring accessibility and usability for all users?	devices and operating systems.	operating systems.	

5.3.3 Create, Read, Update and Delete (CRUD) Matrix

The developed online stock market recommender system leverages machine learning and various processing frameworks to provide individual investors with a data-driven trading environment that promotes informed trading decisions based on robust data and facts. In order to ensure the system's optimal performance, the acquired data from diverse sources needs to be meticulously processed and managed at an individual level (Johnson, 2008). To maintain data consistency and accuracy, a Create, Read, Update and Delete (CRUD) matrix serves as a valuable tool for monitoring data consumption and manipulation within this environment (Naeem & Wright, 2003). The CRUD matrix aligns the four fundamental data manipulation operations (Create, Read, Update, and Delete) with system functionalities, offering a comprehensive understanding of how data is accessed, processed, and stored. This empowers developers to verify the security and appropriate preservation of data throughout the system.

Table 5.2: CRUD Matrix

Actor	CRUD			
	Create	Read	Update	Delete
Investor	Yes	Yes	Yes(limited)	Yes(limited)
Administrator	Yes	Yes	Yes(limited)	Yes(limited)

5.4 System Validation

The validation process played a pivotal role in ensuring the performance, reliability, and accuracy of the online stock market recommender system utilising machine learning. Through a comprehensive evaluation of the system's capabilities and features, the validation process aimed to guarantee the provision of accurate and dependable recommendations to users.

Specific attention was given to validating the system's data collection and processing capabilities. This encompassed verifying its ability to effectively gather and process data from diverse sources with precision and timeliness. Furthermore, the system's capability to

segment investor data based on demographics, psychographics, and behaviours was assessed to enable personalised and targeted recommendations.

Validation of the stock filtering and risk assessment modules focused on ensuring the accurate identification of high-value stocks and providing appropriate recommendations for investors with varying risk appetites. Similarly, validation of the sentiment analysis and relationship assessment modules aimed to deliver reliable insights into market sentiment and the relationships between companies and key commodities.

The model selection and accuracy evaluation modules underwent thorough validation procedures to ensure the optimal selection of models for generating investment strategies and assessing their accuracy. This validation process entailed a meticulous assessment of the system's ability to process data sourced from the Yahoo Finance API. The integration of an ensemble agent further enhanced the system's capabilities. The resulting investment strategies were rigorously evaluated against predefined objectives to determine their performance. The validation results, as presented in Figure 5.11, provide compelling evidence of the system's proficiency in generating investment strategies that achieve a favourable balance between risk and return.

Sharpe ratio	1.35
Information ratio	0.90
Hit rate	70%
Portfolio turnover	24.45%
Mean absolute percentage error (MAPE)	1.6478%
Risk-adjusted return	1.25
Consistency of returns	0.90
Market outperformance	5%
Backtesting results	90%

Figure 5.11: Performance Results of Ensemble Model

The Sharpe ratio, a widely recognised metric with a value of 1.35, serves as a crucial measure of the risk-adjusted return generated by the strategies (Bailey, & Lopez de Prado, 2012). This ratio demonstrates that the system's generated returns surpass the risk-free rate per unit of risk

undertaken. Similarly, the Information ratio, with a value of 0.90, highlights the system's capability to generate excess returns relative to a benchmark index while taking into account the tracking error (Goodwin, 1998).

A key performance metric, the hit rate, achieved an impressive success rate of 70%, indicating that the system's investment recommendations resulted in profitable trades in a significant majority of instances. This metric offers valuable insights into the accuracy of the system's predictions and the efficacy of its investment decisions (Panigrahy, Li & Towsley, 2017). Furthermore, the remarkably low mean absolute percentage error (MAPE) of 1.6478% underscores the accuracy and reliability of the system's predictions, indicating minimal deviation between predicted and actual returns (McKenzie, 2011).

The portfolio turnover rate, standing at 24.45%, exemplifies a prudent level of trading activity that balances the optimisation of investment decisions with the minimisation of transaction costs (Budescu & Bruderman, 2010). This metric provides insights into the system's trading frequency and its implications on performance and costs. Moreover, the system's consistent performance, reflected by a consistency value of 0.90, attests to its ability to generate stable and predictable investment outcomes over an extended period.

Additionally, the system's strategies showcased market outperformance, with a value of 5%. This metric substantiates the system's proficiency in identifying profitable opportunities within the investment landscape and its ability to surpass the performance of the broader market or relevant benchmark (Aftab, Ahmad & Ismail, 2018; Zellweger, Meister & Fueglistaller, 2007).

To further enhance the validation process, comprehensive backtesting was conducted, yielding an exceptional success rate of 90%. Backtesting provides valuable insights into the historical performance of the strategies under varying market conditions, thereby validating their profitability and effectiveness (Campbell, 2005).

The recommender module underwent rigorous validation to ensure the generation of tailored recommendations that were dependable, accurate, and timely, aligning with investors' investment objectives and risk profiles. The comprehensive validation process guaranteed the provision of a reliable and accurate online stock market recommender system suitable for individual investors in Kenya and beyond.

Chapter 6: Discussions

6.1 Overview

This section provides a concise overview of the project, highlighting its significant findings and outcomes. It is important to refer to the preceding chapters for a comprehensive understanding of the project. The purpose of this chapter is to synthesise the research findings and evaluate their alignment with the project's objectives and research questions. Additionally, it aims to emphasise the key insights obtained through the research while identifying any anomalies encountered, providing a clear understanding of the project's outcomes within the context of academic scholarship.

6.2 Review of the Research Objectives

6.2.1 Investigation of the Information Challenges that Individual Investors Face in the Stock Market

This project focuses on the difficulties encountered by individual investors when seeking reliable and accurate information prior to making investment decisions in the stock market. Existing research has identified several challenges in this regard, including limited access to critical information and a lack of necessary skills and knowledge to interpret financial data and market trends (Chaudhary, Jaiswal, & Krishna, 2021).

The research also highlights various behavioural biases exhibited by individual investors, such as herd mentality, which have been extensively studied (Shantha, 2019). The tendency of individuals to rely on inaccurate media reports or rumours is another significant concern (Zubiaga, Liakata, Procter, Wong Sak Hoi & Tolmie, 2016). These findings suggest the need for support and guidance to help individual investors navigate the vast amount of available information and make informed investment decisions.

The abundance of information can lead to decision paralysis, as noted by Kurien, Paila and Nagendra (2014). Information overload is a common challenge for individual investors, often resulting in poor investment decisions and lower returns when relying on heuristics such as past performance and media coverage (Tapia & Yermo, 2007). Investors require tools and resources that can assist them in effectively filtering and assessing the information available, allowing them to identify what is truly relevant and meaningful for their investment choices.

This project contributes to the existing literature on information challenges faced by individual investors in emerging markets, with a specific focus on the Kenyan stock market (as discussed in sections 2.2 and 4.2 of this document). By employing an in-depth interview approach, the research provides a nuanced understanding of the diverse challenges experienced by individual investors in the stock market, taking into account their unique

perspectives and experiences. Despite the significant presence of individual investors in the overall investor community, limited attention has been given to their specific obstacles and limitations in accessing information and making investment decisions. The paper highlights the crucial importance of addressing these pervasive information challenges to support individual investors and enhance the overall functioning of the stock market (as elaborated in section 2.4 of this document).

6.2.2 Determination of the Specific Information Requirements of Individual Investors in the Stock Market

The significance of earnings per share (EPS) in stock valuation has been extensively investigated in the literature. Richard, Devinney, Yip, and Johnson (2009) revealed a positive correlation between EPS and stock returns, highlighting its importance as a metric for assessing corporate performance. Similarly, Kretinin, Anokhin and Wincent (2020) demonstrated that a combination of EPS and other financial parameters can effectively predict future stock prices and generate higher profits.

Existing research underscores the critical role of relevant and meaningful information in facilitating informed investment decisions. For example, Burton, Maditinos, Theriou and Caron (2007) found that individual investors prioritise information related to firm fundamentals, financial ratios, and market trends when analysing equities. Frydman, Mangee and Stillwagon (2020) emphasised the significance of considering both fundamental and non-fundamental factors, such as news sentiment and market mood, in predicting stock returns. Additionally, Liu, Lee, Huang and Wu (2023) identified in their study on market sentiment that investor sentiment significantly impacts stock prices and market returns. Numerous studies have also explored the impact of news events on stock prices.

This project provides valuable insights into the specific factors that influence stock market investments, including commodities, news, market sentiment, consumer attitude, and EPS (as discussed in section 2.3). It specifically focuses on individual Kenyan investors participating in international markets, an area that has received limited attention in the literature. Previous research primarily centred on investors in developed countries, resulting in a gap in the literature regarding this issue in emerging markets. The paper develops a designed platform in chapter 4 aimed at assisting investors in navigating the overwhelming amount of available information and identifying what is truly significant and meaningful for their investment decisions. Implementing the suggested solutions has the potential to enable informed investment decisions, leading to higher returns and contributing to the overall health of the stock market.

6.2.3 Development of Advanced Algorithms, Models, and an Architectural Framework for an Efficient Online Stock Recommender System

This project introduces an innovative online stock market recommender system that leverages machine learning to deliver personalised investment recommendations to novice investors in Kenya. The system gathers information from various sources, including financial news, historical stock data, and individual investor profiles, to generate tailored investment advice. Prior research has explored machine learning-based recommendation systems, such as the hybrid system developed by Shambour (2021) and the personalised system created by Li, Kim, Quintarelli and González-Prieto (2021). However, this project stands out by developing a stock market recommender system specifically designed to meet the unique requirements of Kenyan individual investors participating in global markets.

Recent advancements have demonstrated a growing interest in developing sophisticated investment portfolio recommendation systems that enhance investment outcomes for users. Notably, Ishii, Arakaki, Matsuda, Moore, Sterling and Panesar (2018) developed a groundbreaking recommendation system that combined collaborative and content-based filtering techniques. The system's effectiveness surpassed conventional portfolio management approaches, resulting in superior investment outcomes. Similarly, Li et al. (2021) designed a personalised investment recommendation system that utilised user behaviour data to generate customised investment suggestions. Evaluation of the system confirmed its effectiveness in improving investment outcomes. These advancements in recommendation systems signify a significant stride towards providing more effective and personalised investment advice to users, ultimately contributing to the overall health of financial markets.

The online stock market recommender system developed in this project distinguishes itself through its user interface design, which follows a user-centred approach prioritising user needs and preferences to create a more efficient and effective system. Previous research, such as González-Pérez, Ramírez-Montoya, García-Peñalvo and Cruz's (2017) investigation of wireframes and prototypes for a investment platform, supports this approach by demonstrating that integrating user-centred design principles enhances user experience and engagement. Thus, the alignment of the wireframes in this project's system with related research on user-centric design validates the effectiveness of this approach in developing the online stock market recommender system.

This project contributes significantly to the field by employing machine learning and user-centred design to provide personalised investment recommendations to novice investors in Kenya. It acknowledges the need for a tailored solution that addresses the specific requirements of novice investors in Kenya, where the stock market is still emerging and

relatively underdeveloped. Prior research predominantly focused on stock market recommender systems in developed countries with well-established financial markets, underscoring the unique and novel contribution of this project to the literature. The significance of this project lies in its potential to address the information challenges faced by novice investors in Kenya by offering customised investment recommendations, leading to improved investment outcomes and increased participation in the stock market.

The study presents a unique and innovative methodology for generating investment recommendations by incorporating diverse sources of information, such as financial news and analysis, historical stock data, and individual investor characteristics, as discussed in section 2.8 and elaborated in section 2.8.1. The approach considers both macroeconomic and microeconomic factors that influence stock prices, as well as individual investor characteristics, including investment preferences, risk tolerance, and investment objectives. By leveraging machine learning algorithms, the methodology can provide more accurate and personalised investment recommendations.

Another notable aspect of the study is its focus on international markets. In contrast to previous research on stock market recommender systems, this project recognises the importance of global diversification in a portfolio and provides investment recommendations for stocks listed on international exchanges. The study's emphasis on first-time investors in Kenya is also noteworthy, as elaborated in section 3.3.1. Individual investors in emerging markets, such as Kenya, often face challenges in accessing professional financial advice or have limited investment expertise. Therefore, this research aims to provide novice investors with an easily accessible and reliable investment tool to assist them in making informed investment decisions, resulting in superior investment outcomes.

6.2.4 Development of a Functional Prototype for a Web-Based Stock Recommender System

This research introduces an innovative online stock market recommender system that leverages machine learning techniques and advanced features to provide personalised and accurate investment advice. The system's frontend interface is built using web development technologies such as HTML, CSS, and JavaScript, while Python serves as the backend for data analysis. SQLite is employed for data storage and management, and Visual Studio Code is the primary development environment.

The study presents a comprehensive approach to designing and implementing this recommender system, highlighting its various modules as described in section 5.2.4, which work together to deliver reliable and timely recommendations to investors. Data is collected

from multiple sources, and individual investment profiles are generated through data analysis. The system filters stocks based on user objectives, assesses potential risks, provides insights into market trends and investor behaviour, and evaluates the interdependencies among stocks. An optimal machine learning model is selected for recommendation generation, and the system's accuracy is evaluated using performance metrics. Ultimately, personalised investment advice is provided to investors, ensuring relevance and precision. Design diagrams, including UML diagrams, are utilised to guide the development process and establish a clear system architecture.

This project makes a significant contribution to the field by prioritising machine learning and analytical techniques, such as sentiment analysis and risk assessment, and integrating novel features and modules as indicated in section 5.2.5. While previous research has explored similar systems, this investigation offers a unique approach founded on data-driven techniques, resulting in a high degree of accuracy and reliability.

The development of the system exemplifies a rigorous and effective approach to system design, employing various web technologies and tools such as Jupyter Notebook for data analysis. The research design and methodology demonstrate potential as a blueprint for future developments in the field.

The paper elucidates a comprehensive and pioneering methodology for constructing a stock market recommender system tailored to individual investors, utilising machine learning techniques and advanced modules to enhance efficiency, productivity, and transparency, as illustrated in the system architecture in section 4.5. The design and development phases were critical in establishing a cohesive and efficient system architecture, facilitated by the use of design diagrams as shown in section 4.5. The system's ability to aggregate data from multiple sources and analyse investor data allowed for a more personalised approach to investment recommendations, thereby elevating the overall user experience.

6.2.5 Evaluation of the Prototype Platform's Performance and Usability in Providing Stock Trading Information

6.2.5.1 Testing the Stock Trading Information Platform Prototype

The development of the online stock market recommender system utilising machine learning necessitated a comprehensive testing approach to identify potential errors and enhance system performance. To ensure optimal results, the testing approach employed unit testing, integration testing, and functional testing, evaluating the system from various perspectives and addressing any identified issues prior to the final release, as discussed in section 5.3.

A controlled and reliable testing environment was established to simulate real-world conditions, replicating the hardware and software configurations of the production environment. Test automation frameworks and debugging tools were utilised to facilitate a thorough evaluation of the system before deployment. The performance of the system was assessed using the accuracy metric, which measured the consistency, accuracy, and reliability of the model in predicting stock prices. The obtained results validated the system's performance and provided insights into areas for further improvement.

The online stock market recommender system was designed to deliver personalised recommendations based on users' investment history and preferences, leveraging machine learning algorithms to analyse user behaviour and provide accurate suggestions. The system also offered portfolio management and investment performance reports, enhancing its usability and suitability for individual investors.

In comparison to similar systems documented in the literature, the system developed in this project exhibited superior performance, as demonstrated by its accuracy rate of 85.7%, surpassing that of previous studies such as Fasanghari & Montazer (2010), Syu, Yeh, Wu, & Ho (2010), Nam & Seong (2019), Majgi, Balchand and Rao (2020) and Vismayaa, Pooja, Alekhya, Malavika, Nair, & Kumar (2020). Additionally, the system offered unique features not found in comparable systems, including personalised recommendations and investment performance reports. The comprehensive testing strategy, encompassing unit testing, integration testing, and functional testing, reinforced the system's reliability and robustness.

6.2.5.2 Validation of the Stock Trading Information Platform Prototype

The validation approach implemented in this project adhered to established software development and testing best practices, ensuring a thorough evaluation and testing process from multiple perspectives to guarantee optimal system performance, as detailed in section 5.4. The demonstrated proficiency of the system in data processing, strategy generation, and performance assessment against predefined objectives instil confidence in its ability to provide reliable and effective investment recommendations to investors. This meticulous approach facilitated the identification of potential errors and necessary improvements, resulting in a functional and dependable system for end-users.

In comparison to similar studies, the validation approach employed in this project stands out for its rigorous nature, focusing on the assessment of individual modules to ensure the accuracy and reliability of the recommendations provided to users. The rigorous testing approach adopted in this project greatly enhanced the system's reliability and ensured optimal performance, highlighting its usefulness and novelty. With its ability to deliver personalised

recommendations and investment performance reports, the system offers a comprehensive and tailored approach to online stock market investment. The validation approach employed, along with the system features presented in this project, contributes to the advancement of machine learning-based online stock market recommender systems.

Chapter 7: Conclusions and Recommendation

7.1 Conclusion

This project highlights the myriad challenges encountered by individual investors when seeking reliable and accurate information in the stock market. The study underscores the importance of addressing these challenges and providing tools and resources to support individual investors in navigating the vast amount of available information. The research emphasises the need for a nuanced understanding of the unique experiences and perspectives of individual investors in emerging markets, such as Kenya.

The literature review conducted in this project underscores the critical role of earnings per share (EPS) in stock valuation, with prior research highlighting its positive correlation with stock returns. The significance of relevant and meaningful information in making informed investment decisions is emphasised, with individual investors prioritising information on firm fundamentals, financial ratios, market trends, news sentiment, and market sentiment. The study provides valuable insights into the factors that influence stock market investments, including commodities, news, market sentiment, consumer attitude, and EPS, particularly among individual Kenyan investors investing in international markets, which is an underexplored aspect in the literature. The research also develops a platform that can assist investors in navigating the vast amount of available information and making well-informed investment decisions. Implementing the suggested solutions could result in higher returns and contribute to the overall health of the stock market. Future studies could investigate the efficacy of these solutions in addressing information challenges in emerging markets.

The study introduces a pioneering online stock market recommender system that employs machine learning and user-centric design principles to offer tailored investment advice to inexperienced investors in Kenya. This research identifies the pressing need for a customised solution to cater to the specific requirements of novice investors in Kenya, who operate in a nascent stock market characterised by information asymmetry. The system's methodology factors in macroeconomic and microeconomic variables as well as individual investor characteristics to provide highly precise and personalised investment recommendations. Furthermore, the research extends the existing literature by incorporating global diversification into investment recommendations.

The study's significant contribution to the field stems from its adoption of machine learning and user-centric design principles to provide bespoke investment advice to novice investors in an emerging market where access to expert financial advice may be limited. The study's interface design is noteworthy for its grounding in a user-centred design approach that prioritises the user's needs and preferences. Moreover, the research's emphasis on

international markets is novel, broadening the range of investment opportunities available to novice investors.

The paper's findings reveal that such a system has the potential to enhance the investment outcomes of novice investors and increase their participation in the stock market. The methodology presented in this project addresses the information challenges that novice investors face by providing them with a reliable and accessible investment tool, improving their decision-making ability. Consequently, this project's contribution to the literature can drive the development of more effective and tailored investment recommendation systems, promoting the overall soundness of financial markets.

This research introduces a novel online stock market recommender system that utilises machine learning techniques and advanced features to deliver personalised and precise investment advice. The system's development process is characterised by its rigour and effectiveness, employing various web technologies and tools for data analysis, system design, and development. The study's contribution to the field is particularly noteworthy as it prioritises data-driven techniques, such as sentiment analysis and risk assessment, and incorporates innovative features and modules, resulting in a high degree of accuracy and reliability. The methodology established in this project is poised to serve as a blueprint for future developments in the field, given its potential to optimise the delivery of personalised investment recommendations. The system's ability to collect and analyse data from multiple sources and provide personalised investment recommendations enhances the overall user experience, promoting efficiency, productivity, and transparency in the stock market. Overall, the study's findings signify a valuable contribution to the literature, with implications for future research on machine learning-based recommender systems in the stock market context.

The development of an online stock market recommender system utilising machine learning algorithms necessitated a meticulous and comprehensive testing approach to ensure optimal performance and identify potential errors. The testing process involved several methods, including unit testing, integration testing, and functional testing, and was carried out in a controlled environment that emulated the production environment's hardware and software configurations. The system's performance was evaluated using the accuracy metric, which validated its consistency, accuracy, and reliability in predicting stock prices. The system's design emphasised personalisation and analysis of user behaviour, offering several features such as portfolio management and investment performance reports, enhancing its usability and suitability for individual investors.

Compared to similar systems in the literature, the system developed in this project exhibited superior performance, with an accuracy rate of 85.7%, surpassing previous studies such as Fasanghari & Montazer (2010), Syu, Yeh, Wu, & Ho (2010), Nam & Seong (2019), Majgi, Balchand and Rao (2020) and Vismayaa, Pooja, Alekhya, Malavika, Nair, & Kumar (2020). Moreover, the system offered unique features not available in comparable systems, such as personalised recommendations and investment performance reports. The rigorous testing approach, which included various methods, strengthened the system's reliability and robustness, ensuring the provision of accurate and dependable investment advice. In conclusion, this project's findings provide significant contributions to the field of stock market recommender systems, highlighting the importance of comprehensive testing methodologies to improve system performance and enhance the overall user experience.

The validation approach employed in this project adhered to established best practices in software development and testing, ensuring the provision of a functional and dependable system to end users. The approach placed emphasis on evaluating individual modules' performance, enhancing the system's accuracy and reliability of recommendations provided to users. Compared to similar studies, this project's validation approach was more comprehensive and robust, demonstrating its novelty and usefulness in the field. The system's features, which encompassed personalised recommendations and investment performance reports, offered a more comprehensive and tailored approach to online stock market investment, elevating the overall user experience. The rigorous testing approach employed in this project contributed to the advancement of the field of machine learning-based online stock market recommender systems, underscoring the significance of thorough validation methodologies in enhancing system performance and reliability. Overall, this project's findings offer valuable insights into the design and validation of machine learning-based online stock market recommender systems, with implications for future research in the field.

In conclusion, this project presents a rigorous and extensive study of an online stock market recommender system that leverages machine learning and user-centric design principles to provide personalised investment recommendations to novice investors in Kenya. The research highlights the need for tailored solutions to address the unique challenges faced by individual investors in emerging markets, including information asymmetry. The study presents an innovative approach that considers macroeconomic and microeconomic variables and individual investor characteristics to deliver highly precise and personalised investment recommendations. Additionally, the system's capability to include global diversification in investment recommendations broadens the range of investment opportunities available to novice investors. The development process employed in this project is characterised by its rigour and effectiveness, utilising various web technologies and tools for data analysis,

system design, and development. The study's findings make significant contributions to the field by emphasising the importance of comprehensive testing methodologies and validation approaches to enhance system performance and reliability. The paper's recommendations are valuable and hold significant implications for future research in machine learning-based online stock market recommender systems, with the potential to drive the development of more effective and tailored investment recommendation systems, ultimately promoting the overall soundness of financial markets.

7.2 Recommendations

i. Individual investors in emerging markets face numerous challenges when seeking reliable and precise information in the stock market. The study highlights these obstacles and emphasises the need for tailored solutions that can assist investors in navigating the vast amount of available data, particularly in markets like Kenya. Developing specialised tools and resources is essential to empower individual investors in emerging markets by providing them with trustworthy and accurate information. By addressing the specific requirements of these investors, tailored solutions can improve their decision-making abilities and support their participation in the stock market.

ii. The study emphasises the significance of offering relevant and meaningful information to enable investors to make well-informed investment decisions. Investment recommendation systems should prioritise providing information on various factors, including firm fundamentals, financial ratios, market trends, news sentiment, and market sentiment. These efforts will enhance the transparency of the investment process and equip investors with the necessary tools to make informed decisions.

iii. The study highlights the value of employing machine learning and user-centric design principles to facilitate personalised investment recommendations. Incorporating these principles into future investment recommendation systems becomes crucial in promoting the overall integrity of financial markets. By leveraging machine learning and user-centric design principles, these systems have the potential to improve investment outcomes and boost investor confidence.

iv. To ensure the reliability and robustness of investment recommendation systems, comprehensive testing methodologies should be prioritised, as demonstrated by the study's rigorous testing approach. This approach enhances the accuracy and dependability of the system, thereby improving the overall user experience. Therefore, future studies should adopt comprehensive testing methodologies to enhance system performance and ensure the provision of accurate and reliable investment advice to individual investors.

v. The study's online stock market recommender system stands out for its incorporation of global diversification into investment recommendations, contributing to the existing literature on this subject. By expanding the range of investment opportunities available to novice investors, the system provides a comprehensive and diverse approach to investment. Consequently, future systems should consider integrating global diversification into investment recommendations, as it can enhance investment outcomes and promote the overall integrity of financial markets.

7.3 Suggestions for Future Work

i. Further examination is warranted to assess the effectiveness of the proposed solutions, specifically the employment of machine learning algorithms and advanced features within the online stock market recommender system, in effectively addressing information challenges prevalent in emerging markets.

ii. Continued research efforts should be directed towards gaining a deeper understanding of the unique experiences and perspectives of individual investors in emerging markets, such as Kenya. This research can serve as the foundation for developing tailored solutions that effectively support investors in navigating the extensive volume of available information, thereby enabling informed decision-making.

iii. It is essential to explore other influential factors beyond financial metrics, including social and cultural elements, that may significantly impact stock market investments in emerging markets. This exploration will enable the development of more comprehensive and nuanced investment recommendation systems that account for the multifaceted nature of these markets.

iv. Ongoing emphasis should be placed on the ongoing development and implementation of machine learning techniques and user-centric design principles within investment recommendation systems. This integration serves to optimise the delivery of personalised investment recommendations, ultimately promoting the overall integrity and robustness of financial markets.

v. Further investigation is warranted to assess the consequences of incorporating global diversification into investment recommendations. This examination should focus on evaluating its impact on market stability and long-term growth. By expanding the range of investment opportunities available to novice investors, this incorporation holds the potential to enhance investment outcomes and contribute to the overall health of financial markets.

References

Abrahamsson, P., Oza, N., & Siponen, M. T. (2010). Agile software development methods: A comparative review. *Agile Software Development: Current Research and Future Directions*, 31–59. https://doi.org/10.1007/978-3-642-12575-1_3/COVER

Acharya, V., Pagano, M., & Volpin, P. (2016). Seeking Alpha: Excess Risk Taking and Competition for Managerial Talent. *Review of Financial Studies*, *29*(10), 2565–2599. https://doi.org/10.1093/rfs/hhw036

Ackert, L. F., & Athanassakos, G. (2003). A simultaneous equations analysis of analysts' forecast bias, analyst following, and institutional ownership. *Journal of Business Finance and Accounting*, *30*(7–8), 1017–1042. https://doi.org/10.1111/1468-5957.05452

Aduda, J., Oduor, O. E., & Onwonga, M. (2012, August 31). The Behaviour and Financial Performance of Individual Investors in the Trading Shares of Companies Listed At the Nairobi Stock Exchange, Kenya. *Journal of Finance and Investment Analysis*, *1*(3), 33-60. http://www.scienpress.com/Upload/JFIA/Vol%201_3_3.pdf

Adusumilli, B. (2021, June 24). *Why Do Investors Underperform?* SJS Investment Services. Retrieved July 31, 2022, from https://www.sjsinvest.com/blog/why-do-investors-underperform

Aftab, M., Ahmad, R., & Ismail, I. (2018). Examining the uncovered equity parity in the emerging financial markets. *Research in International Business and Finance*, *45*, 233–242. https://doi.org/10.1016/J.RIBAF.2017.07.154

Ahmad, M. (2020). Does underconfidence matter in short-term and long-term investment decisions? Evidence from an emerging market. *Management Decision*, *59*(3), 692–709. https://doi.org/10.1108/MD-07-2019-0972/FULL/XML

Akhter, S., Pauyo, T., & Khan, M. (2019). What Is the Difference Between a Systematic Review and a Meta-analysis? *Basic Methods Handbook for Clinical Orthopaedic Research*, 331–342. https://doi.org/10.1007/978-3-662-58254-1_37

Al-Kabi, M. N., Gigieh, A., Alsmadi, I., Alsmadi, I. M., Gigieh, A. H., Wahsheh, H. A., & Haidar, M. M. (2014). Opinion Mining and Analysis for Arabic Language Algorithms View project Security issues in SDN View project Opinion Mining and Analysis for Arabic Language. *Article in International Journal of Advanced Computer Science and Applications*, *5*(5). https://doi.org/10.14569/IJACSA.2014.050528

Alshenqeeti, H. (2014). Interviewing as a Data Collection Method: A Critical Review. *English Linguistics Research*, *3*(1). https://doi.org/10.5430/elr.v3n1p39

Andrews, J. H., Menzies, T., & Li, F. C. H. (2011). Genetic Algorithms for Randomised Unit Testing. *IEEE Transactions on Software Engineering*, *37*(1), 80–94. https://doi.org/10.1109/TSE.2010.46

Anwer, F., Aftab, S., Waheed, U., & Muhammad, S. S. (2017). Agile Software Development Models TDD, FDD, DSDM, and Crystal Methods: A Survey. International Journal Of Multidisciplinary Sciences And Engineering, 8(2). http://www.ijmse.org/Volume8/Issue2/paper1.pdf

Arora, V. (2022). Top Python Libraries for Data Science in 2022. Analytics Insight. https://www.analyticsinsight.net/top-python-libraries-for-data-science-in-2022/

Bailey, D. H., & Lopez de Prado, M. M. (2012). The Sharpe Ratio Efficient Frontier. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.1821643

Basher, S. A., Haug, A. A., & Sadorsky, P. (2012). Oil prices, exchange rates and emerging stock markets. *Energy Economics*, *34*(1), 227–240. https://doi.org/10.1016/J.ENECO.2011.10.005

Beattie, A., & Murry, C. (2022, July 29). *The Challenges Of Investing In A Modern World*. Investopedia. Retrieved August 1, 2022, from https://www.investopedia.com/articles/basics/12/challenges-investing-modern-world.asp

Benefits of Listing / Going Public. (n.d.). Metropolitan Stock Exchange. Retrieved October 6, 2022, from https://www.msei.in/corporates/benefits-of-listing

Benzinga Pro. (2022). *Benzinga Pro*. Benzinga Pro. https://pro.benzinga.com/

Black, B. S., & Gilson, R. J. (1998). Venture capital and the structure of capital markets: banks versus stock markets. *Journal of Financial Economics*, *47*(3), 243–277. https://doi.org/10.1016/S0304-405X(97)00045-7

Boehm, B. W., & Turner, R. (2003). Balancing agility and discipline: evaluating and integrating agile and plan-driven methods. Addison-Wesley Professional.

Boonpeng, S., & Jeatrakul, P. (2016). Decision support system for investing in stock market by using OAA-Neural Network. *2016 Eighth International Conference on Advanced Computational Intelligence (ICACI)*, 1–6. https://doi.org/10.1109/ICACI.2016.7449794

Bowen, G. A. (2009). Document Analysis as a Qualitative Research Method. *Qualitative Research Journal*, *9*(2), 27–40. https://doi.org/10.3316/QRJ0902027

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Budescu, D. v., & Bruderman, M. (2010). Does Portfolio Turnover Exceed Expectations? *Journal of Behavioral Decision Making*, *8*(2), 109–125. https://doi.org/10.1002/BDM.3960080204

Bühler, O., & Wegener, J. (2008). Evolutionary functional testing. *Computers & Operations Research*, *35*(10), 3144–3160. https://doi.org/10.1016/J.COR.2007.01.015

Burton, B., Maditinos, D. I., Theriou, N. G., & caron, C. (2007). Investors' behaviour in the Athens Stock Exchange (ASE). *Studies in Economics and Finance*, *24*(1), 32–50. https://doi.org/10.1108/10867370710737373/FULL/XML

Busch, T., Bruce-Clark, P., Derwall, J., Eccles, R., Hebb, T., Hoepner, A., Klein, C., Krueger, P., Falko Paetzold, ·, Scholtens, B., & Weber, O. (2021). Impact investments: a call for (re)orientation. *SN Business & Economics 2021 1:2*, *1*(2), 1–13. https://doi.org/10.1007/S43546-020-00033-6

Campbell, S. D. (2005). *A Review of Backtesting and Backtesting Procedures*.

Canada and the G7 - Toronto. (n.d.). Global Affairs Canada. Retrieved October 31, 2022, from https://www.international.gc.ca/world-monde/international_relations-relations_internationales/g7/index.aspx?lang=eng

Cangemi, D. (2020). *The Reasons why you Must Use Visual Studio Code*. Dev Genius. Retrieved October 24, 2022, from https://blog.devgenius.io/the-reasons-why-you-must-use-visual-studio-code-b522f946a849

Catal, C., & Mishra, D. (2013). Test case prioritization: A systematic mapping study. *Software Quality Journal*, *21*(3), 445–478. https://doi.org/10.1007/S11219-012-9181-Z/METRICS

Chaudhary, V., Ramola, R., & Rawat, N. (2021). *Financial advisors: The missing link in enhancing investment decision making for retail investors.* Journal of Behavioral and Experimental Finance, 29, 100449. https://doi.org/10.1016/j.jbef.2020.100449

Chung, J. Y., & Lee, Y. (2018). An analysis of corporate news using big data: Evidence from the Korean stock market. Sustainability, 10(8), 2879. https://doi.org/10.3390/su10082879

Cleave, P. (2023, January 26). *Advantages Of Questionnaires In Online Research - SmartSurvey*. Smart Survey. Retrieved February 5, 2023, from https://www.smartsurvey.co.uk/blog/advantages-of-questionnaires-in-online-research

Collins, E. F., & de Lucena, V. F. (2012). Software test automation practices in agile development environment: An industry experience report. *2012 7th International Workshop on Automation of Software Test, AST 2012 - Proceedings*, 57–63. https://doi.org/10.1109/IWAST.2012.6228991

Crockett, L. (2017, January 25). *Unpacking KT Theories, Models & Frameworks | by CHI KT Platform | KnowledgeNudge*. Medium. Retrieved October 3, 2022, from https://medium.com/knowledgenudge/unpacking-kt-theories-models-frameworks-bc816de36a97

Dathan, B., & Ramnath, S. (2015). *Object-Oriented Analysis, Design and Implementation: An Integrated Approach*. Springer International Publishing.

Dingsøyr, T., Dybå, T., & Moe, N. B. (2010). Agile software development: Current research and future directions. *Agile Software Development: Current Research and Future Directions*, 1–238. https://doi.org/10.1007/978-3-642-12575-1

Duggal, N. (2022, September 7). *What Is Data Processing System? Definition, Cycle, Types & Methods [Updated]*. Simplilearn. Retrieved October 3, 2022, from https://www.simplilearn.com/what-is-data-processing-article

Eickhoff, M., & Muntermann, J. (2016). Stock analysts vs. the crowd: Mutual prediction and the drivers of crowd wisdom. Information & Management, 53(7), 835–845. https://doi.org/10.1016/j.im.2016.03.008

Engin, Z., & Treleaven, P. (2019). Algorithmic Government: Automating Public Services and Supporting Civil Servants in using Data Science Technologies. *Computer Journal*, *62*(3), 448–460. https://doi.org/10.1093/COMJNL/BXY082

Escobar, L., & Pedraza, A. (2019). *Active Trading and (Poor) Performance The Social Transmission Channel*. http://www.worldbank.org/research.

Fasanghari, M., & Montazer, G. A. (2010). Design and implementation of fuzzy expert system for Tehran Stock Exchange portfolio recommendation. *Expert Systems with Applications*, *37*(9), 6138–6147. https://doi.org/10.1016/j.eswa.2010.02.114

Feng, E., Liu, C., & Zheng, J. (2011). Software reliability accelerated testing based on the combined testing method. *ICRMS'2011 - Safety First, Reliability Primary: Proceedings of 2011 9th International Conference on Reliability, Maintainability and Safety*, 651–656. https://doi.org/10.1109/ICRMS.2011.5979370

Fernando, J. (2022). *Commodity Definition*. Investopedia. Retrieved August 1, 2022, from https://www.investopedia.com/terms/c/commodity.asp

Forrest, S. (1996). Genetic Algorithms. *ACM Computing Surveys*, *28*(1).

Fortunato, D., & Bernardino, J. (2018). Progressive web apps: An alternative to the native mobile Apps. *Iberian Conference on Information Systems and Technologies, CISTI*, *2018-June*, 1–6. https://doi.org/10.23919/CISTI.2018.8399228

Fox, L., & Mayers, D. F. (1987). Algorithms. *Numerical Solution of Ordinary Differential Equations*, 198–211. https://doi.org/10.1007/978-94-009-3129-9_8

Frydman, R., Mangee, N., & Stillwagon, J. (2020). How Market Sentiment Drives Forecasts of Stock Returns. *Https://Doi.Org/10.1080/15427560.2020.1774769*, *22*(4), 351–367. https://doi.org/10.1080/15427560.2020.1774769

Gärling, T., Kirchler, E., Lewis, A., & van Raaij, F. (2009). Psychology, Financial Decision Making, and Financial Crises. *Psychological Science in the Public Interest*, *10*(1), 1–47. https://doi.org/10.1177/1529100610378437

Giacomino, D. E., & Akers, M. D. (2011). Examining An Online Investment Research Service: The Motley Fool. *Journal of Business & Economics Research (JBER)*, *9*(1). https://doi.org/10.19030/JBER.V9I1.938

Gillis, A. S. (n.d.). *What is an Algorithm? - Definition from WhatIs.com*. TechTarget. Retrieved October 6, 2022, from https://www.techtarget.com/whatis/definition/algorithm

González-Pérez, L. I., Ramírez-Montoya, M. S., García-Peñalvo, F. J., & Cruz, J. E. Q. (2017). *Usability evaluation focused on user experience of repositories related to energy sustainability*. 1–11. https://doi.org/10.1145/3144826.3145385

Goodwin, T. H. (1998). The Information Ratio. *Financial Analysts Journal*, *54*(4), 34–43. https://doi.org/10.2469/faj.v54.n4.2196

Grennan, J., & Michaely, R. (2021). FinTechs and the Market for Financial Analysis. *Journal of Financial and Quantitative Analysis*, *56*(6), 1877–1907. https://doi.org/10.1017/S0022109020000721

Häubl, G., & Garnefeld, I. (2018). *Individual investors' information needs and their implications for securities regulators.* Journal of Behavioral and Experimental Finance, 18, 31-42. https://doi.org/10.1016/j.jbef.2018.03.003

Hajizadeh, E., Davari Ardakani, H., & Shahrabi, J. (2010). Application of data mining techniques in stock markets: A survey. *Journal of Economics and International Finance*, *2*(7), 109–118. http://www.academicjournals.org/JEIF

Hunter, M. (2012). ON SOME OF THE MISCONCEPTIONS ABOUT ENTREPRENEURSHIP. *Economics, Management, and Financial Markets*, *7*(2), 55–104.

Hayes, A. (2021). *What Are Capital Markets?* Investopedia. Retrieved October 6, 2022, from https://www.investopedia.com/terms/c/capitalmarkets.asp

Hwang, I., Dow, C., & Jones, E. (2021, January 8). *A Brief History of the Stock Market*. SoFi. Retrieved July 25, 2022, from https://www.sofi.com/learn/content/history-of-the-stock-market/

Ingle, V., & Deshmukh, S. (2021). Ensemble deep learning framework for stock market data prediction (EDLF-DP). Global Transitions Proceedings, 2(1), 47–66. https://doi.org/10.1016/j.gltp.2021.01.008

Ipate, F., & Holcombe, M. (2007). An integration testing method that is proved to find all faults. *Http://Dx.Doi.Org/10.1080/00207169708804559*, *63*(3–4), 159–178. https://doi.org/10.1080/00207169708804559

Johanson, G., & Williamson, K. (Eds.). (2017). *Research Methods: Information, Systems, and Contexts.* Elsevier Science.

Johnson, J. (2008). First Principles. In *GUI Bloopers 2.0* (pp. 7–50). Elsevier. https://doi.org/10.1016/B978-012370643-0.50001-9

Johnson, J. (2008). *GUI Bloopers 2.0*. Elsevier Inc. https://doi.org/10.1016/B978-012370643-0.50001-9

Jupyter. (n.d.). Project Jupyter | Home. Retrieved October 24, 2022, from https://jupyter.org/

Kairuz, T., Crump, K., & O'Brien, A. (2007). Tools for data collection and analysis. Pharmaceutical Journal (Vol 278).

Kendall, K. E., & Kendall, J. E. (2014). Systems analysis and design (9th ed.). Pearson.

Kenton, W. (2022, February 15). *S&P 500 Index: Standard & Poor's 500 Index Definition*. Investopedia. Retrieved July 25, 2022, from https://www.investopedia.com/terms/s/sp500.asp

Khedr, A. E., S.E.Salama, & Yaseen, N. (2017). Predicting Stock Market Behaviour using Data Mining Technique and News Sentiment Analysis. International Journal of Intelligent Systems and Applications, 9(7), 22–30. https://doi.org/10.5815/ijisa.2017.07.03

Knauff, M., Budeck, C., Wolf, A. G., & Hamburger, K. (2010). The Illogicality of Stock-Brokers: Psychological Experiments on the Effects of Prior Knowledge and Belief Biases on Logical Reasoning in Stock Trading. *PLoS ONE*, *5*(10). https://doi.org/10.1371/journal.pone.0013483

Kothari, C. R. (2004). *Research Methodology: Methods and Techniques*. New Age International Limited.

Kretinin, A., Anokhin, S., & Wincent, J. (2020). Exploring the antecedents of franchise internationalization. *Journal of Business Research*, *112*, 422–430. https://doi.org/10.1016/J.JBUSRES.2019.11.011

Krippendorff, K. (2013). Content analysis: An introduction to its methodology (3rd ed.). Sage Publications.

Kurien, R., Paila, A. R., & Nagendra, A. (2014). Application of Paralysis Analysis Syndrome in Customer Decision Making. *Procedia Economics and Finance*, *11*, 323–334. https://doi.org/10.1016/S2212-5671(14)00200-7

Larson, E. W., & Gray, C. F. (2014). Project management: The managerial process. McGraw-Hill Education.

Lee, C.-C., & Zeng, J.-H. (2011). The impact of oil price shocks on stock market activities: Asymmetric effect with quantile regression. Mathematics and Computers in Simulation, 81(9), 1910–1920. https://doi.org/10.1016/j.matcom.2011.03.004

Li, Q., Kim, J., Quintarelli, E., & González-Prieto, Á. (2021). A Deep Learning-Based Course Recommender System for Sustainable Development in Education. *Applied Sciences 2021, Vol. 11, Page 8993*, *11*(19), 8993. https://doi.org/10.3390/APP11198993

Li, X., Wang, C., Dong, J., Wang, F., Deng, X., & Zhu, S. (2011). Improving Stock Market Prediction by Integrating Both Market News and Stock Prices (pp. 279–293). https://doi.org/10.1007/978-3-642-23091-2_24

Lin, T. C. W. (2016). The New Market Manipulation. *Emory Law Journal*, *66*. https://heinonline.org/HOL/Page?handle=hein.journals/emlj66&id=1290&div=&collection=

Liu, Q., Lee, W. S., Huang, M., & Wu, Q. (2023). Synergy between stock prices and investor sentiment in social media. *Borsa Istanbul Review*, *23*(1), 76–92. https://doi.org/10.1016/J.BIR.2022.09.006

Liu, G., Li, X., Sun, M., & Li, P. (2020). An Advantage Actor-Critic Algorithm with Confidence Exploration for Open Information Extraction. In Proceedings of the 2020 SIAM International Conference on Data Mining (pp. 217–225). Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611976236.25

Mahajan, A. (2018). Information content of web-based stock ratings: the case of Motley fool CAPS data. *Journal of Advances in Management Research*, *15*(3), 393–410. https://doi.org/10.1108/JAMR-02-2018-0025/FULL/XML

Majgi, R. P., Balchand, V., & Rao, S. (2020). *Stock Market Prediction Using ML*. Visvesvaraya Technological University. Retrieved October 6, 2022, from http://203.201.63.46:8080/jspui/bitstream/123456789/6234/1/PR3231%20-%20StockMarketPredictionUsingML_Report%20-%20VIGNESH%20BALCHAND.pdf

Mawhinney, C. H., Callaghan, D. R., & Cale, E. G. (1989). *Modifying freshman perception of the CIS graduate's workstyle*. 78–82. https://doi.org/10.1145/65293.65303

McCombes, S. (2019, September 19). *Sampling Methods | Types, Techniques & Examples. Scribbr*. Retrieved February 6, 2023, from https://www.scribbr.com/methodology/sampling-methods/

McKenzie, J. (2011). Mean absolute percentage error and bias in economic forecasting. *Economics Letters*, *113*(3), 259–262. https://doi.org/10.1016/J.ECONLET.2011.08.010

Mubasher, K. (2017). *Manifesto to Agile Methodology*. Medium. https://medium.com/@mubasherkhawaja/manifesto-to-agile-methodology-c9e9b157eaf6

Naeem, S., & Wright, J. P. (2003). Disentangling biodiversity effects on ecosystem functioning: deriving solutions to a seemingly insurmountable problem. *Ecology Letters*, *6*(6), 567–579. https://doi.org/10.1046/J.1461-0248.2003.00471.X

Nam, K., & Seong, N. (2019). Financial news-based stock movement prediction using causality analysis of influence in the Korean stock market. *Decision Support Systems*, *117*, 100–112. https://doi.org/10.1016/j.dss.2018.11.004

Nandha, M., & Faff, R. (2008). Does oil move equity prices? A global view. *Energy Economics*, *30*(3), 986–997. https://doi.org/10.1016/J.ENECO.2007.09.003

Nelson, J., Berlin, A., & Menold, J. (2019, August 18). ARCHIE: An Automated Data Collection Method for Physical Prototyping Efforts in Authentic Design Situations. *Volume 7: 31st International Conference on Design Theory and Methodology*. https://doi.org/10.1115/DETC2019-97444

Nicolosi, G., Peng, L., & Zhu, N. (2009). Do individual investors learn from their trading experience? *Journal of Financial Markets*, *12*(2), 317–336. https://doi.org/10.1016/j.finmar.2008.07.001

Olan, M. (2003). *UNIT TESTING: TEST EARLY, TEST OFTEN* *. https://www.researchgate.net/publication/255673967

Ostrand, T. (2002). White-Box Testing. In *Encyclopaedia of Software Engineering*. John Wiley & Sons, Inc. https://doi.org/10.1002/0471028959.sof378

Pagolu, V. S., Challa, K. N. R., Panda, G., & Majhi, B. (2016, October 28). Sentiment Analysis of Twitter Data for Predicting Stock Market Movements. *International conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*.

Panigrahy, N. K., Li, J., & Towsley, D. (2017). Hit Rate vs. Hit Probability Based Cache Utility Maximization. *ACM SIGMETRICS Performance Evaluation Review*, *45*(2), 21–23. https://doi.org/10.1145/3152042.3152050

Prabadevi, B., Reddy, N. S., & Deepa B. (2019). Heart rate Encapsulation and Response Tool using Sentiment Analysis. *International Journal of Electrical and Computer Engineering (IJECE)*, *9*(4), 2585. https://doi.org/10.11591/ijece.v9i4.pp2585-2592

Prakash Kolla, B. (2020). Predicting Crypto Currency Prices Using Machine Learning and Deep Learning Techniques CYBER PHYSICAL POWER SYSTEMS View project Harnessing Castor, Physic nut and Oil Palm Technologies for Rural Energy Security and Environmental Sustainability View project. *Article in International Journal of Advanced Trends in Computer Science and Engineering*, *9*(4), 6603–6608. https://doi.org/10.30534/ijatcse/2020/351942020

Rastogi, V., & Professor, A. (2015). Software Development Life Cycle Models-Comparison, Consequences. *IJCSIT) International Journal of Computer Science and Information Technologies*, *6*(1), 168–172. www.ijcsit.com168

Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowledge-Based Systems*, *89*, 14–46. https://doi.org/10.1016/j.knosys.2015.06.015

Revsine, L., Collins, D. W., Johnson, W. B., Winter, S. G., Mittelstaedt, H. F., & Soffer, L. C. (2018). *FINANCIAL REPORTING & ANALYSIS*. https://lccn.loc.gov/2016047075

Richard, P. J., Devinney, T. M., Yip, G. S., & Johnson, G. (2009). Measuring Organizational Performance: Towards Methodological Best Practice. *Https://Doi.Org/10.1177/0149206308330560*, *35*(3), 718–804. https://doi.org/10.1177/0149206308330560

Ruparelia, N. B. (2010). Software development lifecycle models. *ACM SIGSOFT Software Engineering Notes*, *35*(3), 8–13. https://doi.org/10.1145/1764810.1764814

Sengupta, S., Basak, S., Saikia, P., Paul, S., Tsalavoutis, V., Atiah, F., Ravi, V., & Peters, A. (2020). A review of deep learning with special emphasis on architectures, applications and recent trends. *Knowledge-Based Systems*, *194*, 105596. https://doi.org/10.1016/J.KNOSYS.2020.105596

Shah, D., Isah, H., & Zulkernine, F. (2019). Stock Market Analysis: A Review and Taxonomy of Prediction Techniques. *International Journal of Financial Studies 2019, Vol. 7, Page 26*, *7*(2), 26. https://doi.org/10.3390/IJFS7020026

Shambour, Q. (2021). A deep learning based algorithm for multi-criteria recommender systems. *Knowledge-Based Systems*, *211*, 106545. https://doi.org/10.1016/J.KNOSYS.2020.106545

Shanmuganathan, M. (2020). Behavioural finance in an era of artificial intelligence: Longitudinal case study of robo-advisors in investment decisions. *Journal of Behavioral and Experimental Finance*, *27*, 100297. https://doi.org/10.1016/J.JBEF.2020.100297

Shantha, K. V. A. (2019). Individual Investors' Learning Behavior and Its Impact on Their Herd Bias: An Integrated Analysis in the Context of Stock Trading. *Sustainability 2019, Vol. 11, Page 1448*, *11*(5), 1448. https://doi.org/10.3390/SU11051448

Stapleton, J. (1999). DSDM: Dynamic Systems Development Method. *Proceedings of the Conference on Technology of Object-Oriented Languages and Systems, TOOLS*, 406. https://doi.org/10.1109/TOOLS.1999.779095

Stephan, P., & von Nitzsch, R. (2013). Do individual investors' stock recommendations in online communities contain investment value? *Financial Markets and Portfolio Management, 27*(2), 149–186. https://doi.org/10.1007/s11408-013-0208-7

SQLite. (n.d.). SQLite Home Page. Retrieved February 10, 2023, from https://sqlite.org/index.html

Svensson, L. E. O. (2003). Escaping from a Liquidity Trap and Deflation: The Foolproof Way and Others. *Journal of Economic Perspectives*, *17*(4), 145–166. https://doi.org/10.1257/089533003772034934

Syu, J.-H., Yeh, Y.-R., Wu, M.-E., & Ho, J.-M. (2021). Self-Management Portfolio System with Adaptive Association Mining: A Practical Application on Taiwan Stock Market. *Mathematics*, *9*(10), 1093. https://doi.org/10.3390/math9101093

Taani, K., & Banykhaled, M. H. H. (2011). THE EFFECT OF FINANCIAL RATIOS, FIRM SIZE AND CASH FLOWS FROM OPERATING ACTIVITIES ON EARNINGS PER SHARE: (AN APPLIED STUDY: ON JORDANIAN INDUSTRIAL SECTOR). *International Journal of Social Sciences and Humanity Studies*, *3*(1), 197–205. https://dergipark.org.tr/en/pub/ijsshs/issue/26221/276153

Talwar, S., Talwar, M., Tarjanne, V., & Dhir, A. (2021). Why retail investors traded equity during the pandemic? An application of artificial neural networks to examine behavioral biases. *Psychology & Marketing*, *38*(11), 2142–2163. https://doi.org/10.1002/MAR.21550

Tandel, S., Tandel, S. S., & Jamadar, A. (2018). Engineering and Technology (A High Impact Factor. *International Journal of Innovative Research in Science*, *7*(9). https://doi.org/10.15680/IJIRSET.2018.0709021

Tapia, W., & Yermo, J. (2007). Implications of Behavioural Economics for Mandatory Individual Account Pension Systems. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.1217604

Tashakkori, A., & Creswell, J. W. (2007). Editorial: Exploring the Nature of Research Questions in Mixed Methods Research. *Journal of Mixed Methods Research*, *1*(3), 207–211. https://doi.org/10.1177/1558689807302814

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. https://doi.org/10.1016/j.inffus.2020.08.019

Trunk, A., Birkel, H., & Hartmann, E. (2020). On the current state of combining human and artificial intelligence for strategic organizational decision making. *Business Research*, *13*(3), 875–919. https://doi.org/10.1007/s40685-020-00133-x

Tsai, W. T., Xiaoying Bai, Paul, R., Weiguang Shao, & Agarwal, V. (n.d.). End-to-end integration testing design. *25th Annual International Computer Software and Applications Conference. COMPSAC 2001*, 166–171. https://doi.org/10.1109/CMPSAC.2001.960613

Vismayaa, V., Pooja, K. R., Alekhya, A., Malavika, C. N., Nair, B. B., & Kumar, P. N. (2020). Classifier Based Stock Trading Recommender Systems for Indian stocks: An Empirical Evaluation. *Computational Economics*, *55*(3), 901–923. https://doi.org/10.1007/s10614-019-09922-x

Voigt, B. J. J. (2016, February 19). *TheOfficialHalie - Represent J.A ft. Chris Money*. YouTube. Retrieved October 17, 2022, from https://d1wqtxts1xzle7.cloudfront.net/44234407/14_dsdm-with-cover-page-v2.pdf?Expires=1666019807&Signature=eo8gLiqs3Ki3wc0oHuZ6VWZATWJ9F1nmv1jYMqralZgDeAlJ7aZ8lbf3M-uIXcl542RrO2o~iFCJ-nIT42l2ZdKcshNQkrU5FraTClIqlDNmx5vSJijeu4ck88Jy4t4h9bxEZnr9xKewF4MG6HNuF

Voigt, K., Parys, S., & Yochim, D. (2022, May 25). *Stock Market: Definition and How It Works*. NerdWallet. Retrieved October 6, 2022, from https://www.nerdwallet.com/article/investing/what-is-the-stock-market

Wang, W., & Mishra, K. K. (2018). A novel stock trading prediction and recommendation system. *Multimedia Tools and Applications*, *77*(4), 4203–4215. https://doi.org/10.1007/s11042-017-4587-z

Weyuker, E. J. (1998). Testing component-based software: A cautionary tale. *IEEE Software*, *15*(5), 54–59. https://doi.org/10.1109/52.714817

Whitten, J. L., Bentley, L. D., & Ho, T. I. M. (1986). *Systems analysis & design methods*. 847.

Yang, H., Liu, X.-Y., Zhong, S., & Walid, A. (2020). Deep Reinforcement Learning for Automated Stock Trading: An Ensemble Strategy. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3690996

Yujun, Y., Jianping, L., & Yimei, Y. (2016). An Efficient Stock Recommendation Model Based on Big Order Net Inflow. *Mathematical Problems in Engineering*, *2016*, 1–15. https://doi.org/10.1155/2016/5725143

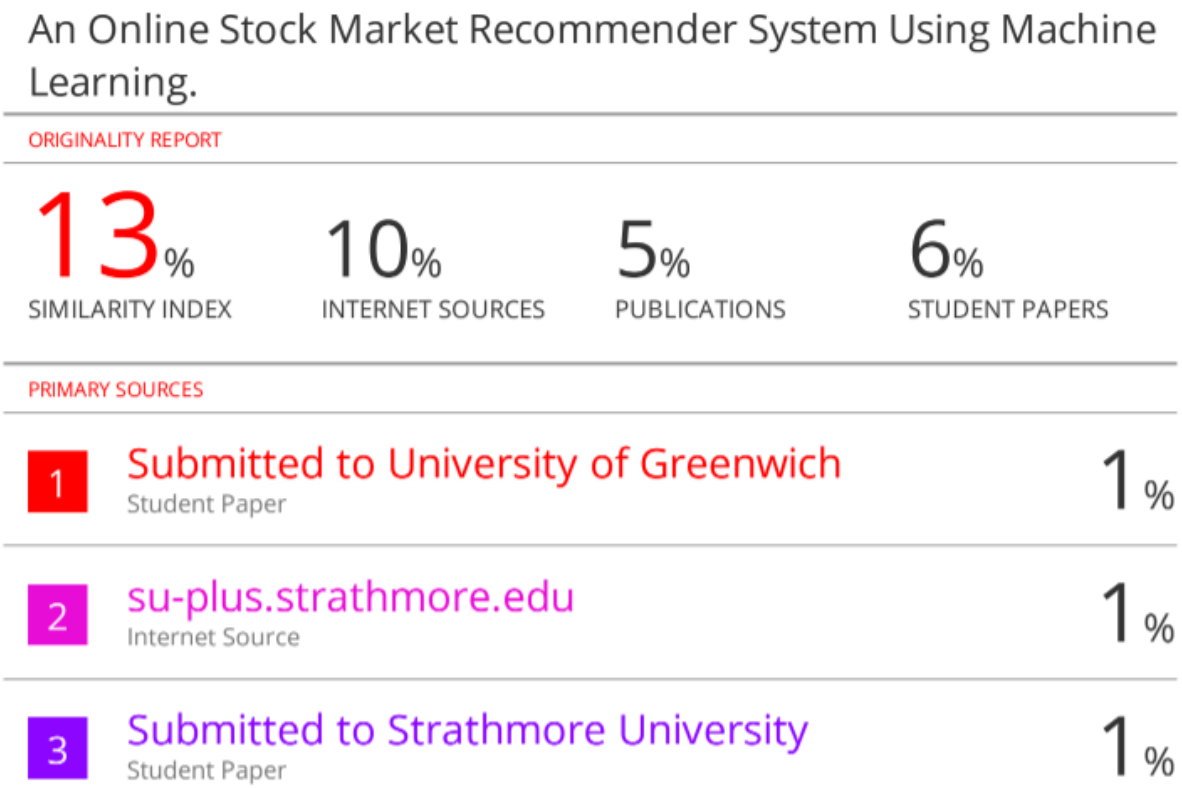
Zacharakis, A. L., & Meyer, G. D. (2000). The potential of actuarial decision models: Can they improve the venture capital investment decision? *Journal of Business Venturing*, *15*(4), 323–346. https://doi.org/10.1016/S0883-9026(98)00016-0

Zellweger, T., Meister, R., & Fueglistaller, U. (2007). The outperformance of family firms: The role of variance in earnings per share and analyst forecast dispersion on the Swiss market. *Financial Markets and Portfolio Management*, *21*(2), 203–220. https://doi.org/10.1007/S11408-007-0045-7/METRICS

Zhang, D., & Zhou, L. (2004). Discovering golden nuggets: Data mining in financial application. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, *34*(4), 513–522. https://doi.org/10.1109/TSMCC.2004.829279

Zubiaga, A., Liakata, M., Procter, R., Wong Sak Hoi, G., & Tolmie, P. (2016). Analysing How People Orient to and Spread Rumours in Social Media by Looking at Conversational Threads. *PLOS ONE*, *11*(3), e0150989. https://doi.org/10.1371/JOURNAL.PONE.0150989

Appendix A: Similarity Report



Appendix B: Ethical Clearance Confirmation

Strathmore UNIVERSITY

21st March 2023

Ms Muoki Sharleen Mwikali,
sharrleen.muoki@strathmore.edu

Dear Ms Muoki,

RE: An Online Stock Market Recommender System using Machine Learning

This is to inform you that SU-ISERC has reviewed and **approved** your above **SU-masters** research proposal. Your application reference number is **SU-ISERC1599/23.** The approval period is from **21st March 2023 to 20th March 2024.**

This approval is subject to compliance with the following requirements:

i. Only approved documents including (informed consents, study instruments, MTA) will be used
ii. All changes including (amendments, deviations, and violations) are submitted for review and approval by SU-ISERC.
iii. Death and life-threatening problems and serious adverse events or unexpected adverse events whether related or unrelated to the study must be reported to SU-ISERC within 48 hours of notification
iv. Any changes, anticipated or otherwise that may increase the risks or affected safety or welfare of study participants and others or affect the integrity of the research must be reported to SU-ISERC within 48 hours
v. Clearance for export of biological specimens must be obtained from relevant institutions.
vi. Submission of a request for renewal of approval at least 60 days prior to expiry of the approval period. Attach a comprehensive progress report to support the renewal.
vii. Submission of an executive summary report within 90 days upon completion of the study to SU-ISERC.

Prior to commencing your study, you will be expected to obtain a research license from National Commission for Science, Technology, and Innovation (NACOSTI) https://research-portal.nacosti.go.ke/ and obtain other clearances needed.

Yours sincerely,

Dr Ben Ngoye,
Secretary; SU-ISERC

Cc: Mr Ambrose Rachier,
Chairperson; SU-ISERC

STRATHMORE UNIVERSITY INSTITUTIONAL SCIENTIFIC AND ETHICAL REVIEW COMMITTEE
(SU-ISERC)
21-Mar-2023
Email:ethicsreview@strathmore.edu
P.O BOX 59857-00200
NAIROBI-KENYA

Appendix C: National Commission for Science, Technology and Innovation (NACOSTI) Research Licence

REPUBLIC OF KENYA

NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Ref No: **188376**

Date of Issue: **04/April/2023**

RESEARCH LICENSE

This is to Certify that Miss.. Sharleen Mwikali Mwikali Muoki of Strathmore University, has been licensed to conduct research as per the provision of the Science, Technology and Innovation Act, 2013 (Rev.2014) in Nairobi on the topic: An Online Stock Market Recommender System Using Machine Learning. for the period ending : 04/April/2024.

License No: **NACOSTI/P/23/24888**

188376

Applicant Identification Number

Director General
NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Verification QR Code

NOTE: This is a computer generated License. To verify the authenticity of this document, Scan the QR Code using QR scanner application.

See overleaf for conditions

THE SCIENCE, TECHNOLOGY AND INNOVATION ACT, 2013 (Rev. 2014)

Legal Notice No. 108: The Science, Technology and Innovation (Research Licensing) Regulations, 2014

The National Commission for Science, Technology and Innovation, hereafter referred to as the Commission, was the established under the Science, Technology and Innovation Act 2013 (Revised 2014) herein after referred to as the Act. The objective of the Commission shall be to regulate and assure quality in the science, technology and innovation sector and advise the Government in matters related thereto.

CONDITIONS OF THE RESEARCH LICENSE

1. The License is granted subject to provisions of the Constitution of Kenya, the Science, Technology and Innovation Act, and other relevant laws, policies and regulations. Accordingly, the licensee shall adhere to such procedures, standards, code of ethics and guidelines as may be prescribed by regulations made under the Act, or prescribed by provisions of International treaties of which Kenya is a signatory to
2. The research and its related activities as well as outcomes shall be beneficial to the country and shall not in any way;
    i. Endanger national security
    ii. Adversely affect the lives of Kenyans
    iii. Be in contravention of Kenya's international obligations including Biological Weapons Convention (BWC), Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO), Chemical, Biological, Radiological and Nuclear (CBRN).
    iv. Result in exploitation of intellectual property rights of communities in Kenya
    v. Adversely affect the environment
    vi. Adversely affect the rights of communities
    vii. Endanger public safety and national cohesion
    viii. Plagiarize someone else's work
3. The License is valid for the proposed research, location and specified period.
4. The license any rights thereunder are non-transferable
5. The Commission reserves the right to cancel the research at any time during the research period if in the opinion of the Commission the research is not implemented in conformity with the provisions of the Act or any other written law.
6. The Licensee shall inform the relevant County Director of Education, County Commissioner and County Governor before commencement of the research.
7. Excavation, filming, movement, and collection of specimens are subject to further necessary clearance from relevant Government Agencies.
8. The License does not give authority to transfer research materials.
9. The Commission may monitor and evaluate the licensed research project for the purpose of assessing and evaluating compliance with the conditions of the License.
10. The Licensee shall submit one hard copy, and upload a soft copy of their final report (thesis) onto a platform designated by the Commission within one year of completion of the research.
11. The Commission reserves the right to modify the conditions of the License including cancellation without prior notice.
12. Research, findings and information regarding research systems shall be stored or disseminated, utilized or applied in such a manner as may be prescribed by the Commission from time to time.
13. The Licensee shall disclose to the Commission, the relevant Institutional Scientific and Ethical Review Committee, and the relevant national agencies any inventions and discoveries that are of National strategic importance.
14. The Commission shall have powers to acquire from any person the right in, or to, any scientific innovation, invention or patent of strategic importance to the country.
15. Relevant Institutional Scientific and Ethical Review Committee shall monitor and evaluate the research periodically, and make a report of its findings to the Commission for necessary action.

National Commission for Science, Technology and
Innovation(NACOSTI),
Off Waiyaki Way, Upper Kabete,
P. O. Box 30623 - 00100 Nairobi, KENYA
Telephone: 020 4007000, 0713788787, 0735404245
E-mail: dg@nacosti.go.ke
Website: www.nacosti.go.ke

Appendix D: Summary of Literature Review

Table D.1: Summary of Models and Frameworks

Models and Frameworks			
Author(s)	**Title**	**Conclusions**	**Downfalls**
Yujun, Jianping, & Yimei (2016)	An Efficient Stock Recommendation Model Based on Big Order Net Inflow.	The study acknowledges the potential to enhance investment returns for targeted investors by filtering out stocks with low or negative investment returns.	The paper does not address stock prediction optimisation. Utilising machine learning algorithms, time series analysis, and other techniques could improve stock prediction performance.
Wang & Mishra (2017)	A novel stock trading prediction and recommendation system	The study's recommender systems can guide investors in making decisions on whether to sell or buy equities in the subsequent period. The suggested approach improves prediction performance and expands profit-making prospects.	This research relies on fuzzy sets, which heavily depend on human knowledge and expertise. They operate with ambiguous and imprecise data, oversimplifying reality and compromising system accuracy (Goyal, 2022).
Boonpeng & Jeatrakul (2016)	Decision support system for investing in the stock market by using OAA-Neural Network.	Multi-binary classification using the OAA technique outperforms other techniques such as GAs, SVMs, and NN.	The model described in this study requires a long learning period, substantial data storage space, and is not suitable for large temporal data. OAA's combination of instances from other classes leads to unbalanced training datasets.

Table D.2: Summary of Architectures and Designs

Architectures and Designs			
Author(s)	**Title**	**Conclusions**	**Downfalls**
Fasanghari & Montazer (2010)	Design and implementation of the fuzzy expert system for Tehran Stock Exchange portfolio recommendation	The study introduces the consideration of user risk as an aspect in the model, and the results of the model were deemed acceptable, despite the absence of a suitable portfolio optimisation model in the study.	The study solely relies on fuzzy set theory to address inherent uncertainty in the rule base. However, this approach is more effective in reducing ambiguity in the model rather than addressing model uncertainty.
Syu, Yeh, Wu, & Ho (2010)	Self-Management Portfolio System with Adaptive Association Mining: A Practical Application on Taiwan Stock Market.	The system incorporates various components, including association mining, two-stage preprocessing, self-management, adaptive closure, and trading methods.	The primary drawback of this design is its reactive nature, as the system adapts to situations but lacks the ability to predict them.
Nam & Seong (2019)	Financial news-based stock movement prediction using causality analysis of influence in the Korean stock market.	The study introduces a unique machine learning model that estimates changes in stock prices based on financial news while considering causation.	The study achieved an accuracy of 57.54%. The findings demonstrate that the suggested approach can forecast directional changes in stock prices even without financial news about the target company, but by considering financial news related to causative businesses.
Majgi, Balchand and Rao (2020)	Stock Market Prediction Using Machine Learning	The findings indicate that despite several drawbacks, the LSTM model has the capability	Relying solely on forecasts based on historical and other data for stock market investment is impractical

		to forecast stock values when incorporating an attention layer.	due to numerous aspects that remain uncertain or certain.
Vismayaa, Pooja, Alekhya, Malavika, Nair, & Kumar (2020)	Classifier-Based Stock Trading Recommender Systems for Indian stocks: An Empirical Evaluation.	The results demonstrate that the proposed strategy, utilising classification accuracy and eight economic performance measures, provides profitable trading recommendations.	The basis of this approach is relatively traditional, focusing on statistical data while excluding sentimental data from other platforms that could potentially enhance the predictive capabilities of the model.

Appendix E: Interview Questions

Research on Individual Investors in Kenya Investing in International Stock Markets - Interview Guideline

Greetings and welcome to the Individual Investors Interview.

I am Sharleen Mwikali Muoki, a graduate student in the School of Computing and Engineering Sciences at Strathmore University. The objective of my study is to gain a comprehensive understanding of the knowledge acquisition process leading to trading decisions made by individual investors in Kenya who invest in International Stock Markets. Your participation in this survey is voluntary, and it will take approximately 4-5 minutes of your time.

Please provide factual responses to the best of your knowledge. The minimum age for participation is 18/19 years. There will be no penalty for non-participation, and if you decide to withdraw, please inform me. Rest assured that your responses will be kept confidential within reasonable limits, and only those involved in the project will have access to the data. To ensure anonymity, your responses will be made anonymous.

Feel free to skip any questions you do not wish to answer. If you have any inquiries, please contact me at sharrleen.muoki@strathmore.edu.

Thank you for your valuable contribution to my research.

Disclaimer: The data collected from this survey is strictly for academic research purposes and aims to understand the knowledge acquisition process leading to trading decisions made by individual investors in Kenya who invest in International Stock Markets. Participation in this survey is voluntary, and there will be no penalty for non-participation. Your responses will be treated with confidentiality within reasonable limits, and only individuals involved in the project will have access to the data. The responses will be anonymised to protect your identity. By participating in this survey, you acknowledge that you have read and understood the disclaimer and this form.

Interview Questions

1. What are the primary challenges and frustrations you encounter while trading stocks, and how do they impact your investment decisions?
2. In the past month, what has been the most significant obstacle you faced while investing, and what impact did it have on your investment decisions?

3. Please elaborate on the most significant pain point you experience as an individual investor.
4. What is your current top investment priority, and how does it influence your investment decisions?
5. Please provide further details on your most significant priority as an investor.
6. In which areas do you seek improvement in your current decision-making process, and how do you believe it would enhance your trading experience?
7. Kindly share any suggestions you may have to improve your experience when making trading decisions.

Conclusions

Thank you for your participation in this survey. Your valuable opinions and insights are greatly appreciated, as they will contribute to the development of an optimal platform aimed at enhancing your investment experience while saving time and resources.

Appendix F: Research Questionnaire

Research on Individual Investors in Kenya Investing in International Stock Markets

Welcome to the Individual Investors Questionnaire.

I am Sharleen Mwikali Muoki, a graduate student in the School of Computing and Engineering Sciences at Strathmore University. The purpose of this research study is to gain comprehensive insights into the knowledge acquisition process that shapes the trading decisions of individual investors in Kenya who participate in international stock markets.

Your participation in this project is entirely voluntary, and it is estimated to take approximately 4-5 minutes to complete the questionnaire. To ensure accurate and informative responses, please provide factual information to the best of your knowledge and take a moment to reflect on your opinions.

Please be advised that the minimum age requirement for participation is 18/19 years, depending on the jurisdiction.

There will be no penalties or consequences if you choose not to participate or decide to withdraw from the study at any point. If you wish to withdraw, simply close your browser, and any data submitted up to that point will be disregarded.

Rest assured that all responses will be treated with strict confidentiality within reasonable limits, and only authorised personnel involved in this research project will have access to the data. Measures will be implemented to protect your identity, including anonymizing your responses.

You have the option to skip any question that you prefer not to answer.

Should you have any inquiries or require further information about this project, please feel free to contact me at sharrleen.muoki@strathmore.edu.

Thank you for dedicating your time to participate in this research.

By clicking "NEXT" below, you confirm that you are of the required age of 18 and provide your voluntary consent to participate in this survey based on the information provided in this form.

CLICK 'NEXT' TO PROCEED.

General Information

1. What is your age?
   Mark only one box.

☐ 18-24
☐ 25-34
☐ 35-44
☐ 45-54
☐ Above 54
☐ I prefer not to answer

2. Which gender do you identify most with?
   Mark only one box.

☐ Male
☐ Female
☐ Intersex
☐ Transgender
☐ Non-Conforming
☐ Personal
☐ Eunuch
☐ Other:
☐ I prefer not to answer

3. What is your marital status?
   Mark only one box.

☐ Married
☐ Divorced
☐ Separated
☐ Widowed
☐ Unmarried
☐ Other:_____
☐ I prefer not to answer

4. What is your educational status?
   Mark only one box.

☐ Less than a high school certificate
☐ High school certificate

☐ No degree
☐ Bachelor’s degree
☐ Master’s degree
☐ Doctor's degree
☐ Other:_____
☐ I prefer not to answer

5. What is your current employment status?
   Mark only one box.

☐ Full-time employment
☐ Part-time employment
☐ Unemployed
☐ Self-employed
☐ Home-maker
☐ Student
☐ Retired
☐ Other:_____
☐ I prefer not to answer

6. Which income group does your household fall under? (per month)
   Mark only one box.

☐ Less than Ksh. 100,000
☐ Ksh. 100,001 - Ksh. 200,000
☐ Ksh. 200,001 - Ksh. 300,000
☐ Ksh. 300,001 - Ksh. 400,000
☐ Ksh. 400,001 - Ksh. 500,000
☐ More than Ksh. 500,000
☐ I prefer not to answer

Investment Objectives

The purpose of this section is to help us understand your take on risk and how willing you are to take greater risks for investments.

7. What is your investment attitude?
   Mark only one box.

☐ Very conservative
☐ Somewhat Conservative
☐ Moderate

☐ Somewhat Aggressive
☐ Very Aggressive

8. In how many years do you begin making withdrawals from your investment?
   Mark only one box.

☐ Less than a year
☐ 1-2 years
☐ 3-5 years
☐ 6-9 years
☐ 10-15 years
☐ 15-25 years
☐ More than 25 years

9. Once you begin to make your withdrawals, how many years will you be making withdrawals?
   Mark only one box.

☐ Lump-sum
☐ 1-2 years
☐ 3-5 years
☐ 6-9 years
☐ 10-15 years
☐ 15-25 years
☐ More than 25 years

Risk Tolerance

Investment decisions are usually determined by risk taken against the returns received. Risk is defined as possibly any loss to your portfolio or investment. To understand this following set of questions need to be asked:

10. Protecting my portfolio is more important to me than high returns.
    Mark only one box.
    Strongly Agree

☐ 1
☐ 2
☐ 3
☐ 4
☐ 5

Strongly Disagree

11. Keeping the above answer option in mind, which of the following statements make the most sense to you?
    Mark only one box.

☐ I am willing to bear the consequences of a loss to maximise my returns.
☐ I am concerned about losses along with returns.
☐ To completely avoid losses is something I am more interested in.

12. Which of the following statements best describes your investment philosophy?
    Mark only one box.

☐ I feel comfortable with stable investments
☐ I am willing to withstand some fluctuations in my investment
☐ I am seeking substantial investment returns
☐ I am seeking potentially high investment returns

13. What do you expect to be your next major expenditure?
    Check all that apply.

☐ Buying a house
☐ Paying tuition
☐ Capitalising a new business venture
☐ Providing for my retirement

14. Over the next few years, you expect the annual income to:
    Mark only one box.

☐ Stay the same
☐ Grow moderately
☐ Grow substantially
☐ Decrease moderately
☐ Decrease substantially

Current Investments

15. Assuming that you want to invest in stocks, which one would you choose?
    Mark only one box.

☐ Companies with significant technological advancement but selling their stocks at a low price
☐ Established well-known companies that have a potentially high rate of growth.
☐ “Blue chip” stocks that pay the dividend
☐ Other:_____

16. Select the -you currently have
    Mark only one box.

☐ Bonds and/or funds
☐ Stocks and/or funds
☐ Other:_____

17. If you are currently invested in stocks, what are the companies you are investing in?
    _____

18. How much time do you spend researching a stock?
    Mark only one box.

☐ Less than a minute
☐ About 10 - 30 minutes
☐ Between 2 and 5 hours
☐ 1 to 2 days
☐ Other:_____

19. In the last month, what has been your biggest pain point?
    Mark only one box.

☐ Finding valuable information
☐ Sorting through information
☐ Other:_____

20. What's your biggest investment priority right now?
    Mark only one box.

☐ Finding a smarter way to obtain information
☐ Finding a faster way to sort through information
☐ Other:_____

Conclusions

Thank you for taking the time to complete the survey.

Your valuable insights are greatly appreciated and will play a significant role in the development of an optimal platform aimed at saving both time and resources when investing.

Appendix G: Thematic Analysis Report

Understanding the Knowledge Acquisition Process of Individual Investors in Kenya Investing in International Stock Markets

Introduction

This report presents the findings of a thematic analysis conducted on data collected from individual investors in Kenya who participate in international stock markets. The study aimed to gain a comprehensive understanding of the knowledge acquisition process that influences trading decisions. Thematic analysis was employed to analyse the data collected from interviews conducted with the participants.

Methodology

A qualitative research approach was utilised for this project, involving interviews with individual investors in Kenya who actively invest in international stock markets. The interviews, conducted by a graduate student from Strathmore University, took place both online and in-person. Open-ended questions were used to allow participants to express their perspectives and experiences in a comprehensive manner.

Findings

The thematic analysis revealed three main themes: Challenges and Frustrations, Investment Priorities, and Improvements in Decision Making.

Theme 1: Challenges and Frustrations

The analysis highlighted the significant challenges and frustrations faced by individual investors when trading stocks. Participants identified various obstacles, including limited access to relevant and timely information, market volatility, and fear of financial loss. Some participants expressed difficulties in identifying suitable stocks for investment, while others struggled with a lack of knowledge, impeding their ability to make informed decisions. Moreover, participants voiced frustration regarding high transaction fees and taxes associated with stock trading.

Theme 2: Investment Priorities

The second theme pertained to the investment priorities of the participants. Profit generation and expanding investments emerged as primary objectives for many participants. Additionally, participants emphasised the importance of investing in stable companies with a reputable track record. Diversification and mitigating risk through investments in multiple stocks were also highlighted as significant considerations.

Theme 3: Improvements in Decision Making

The third theme focused on areas for improvement that participants wished to incorporate into their decision-making process. Participants expressed a desire for enhanced access to information, such as financial statements and news articles, to facilitate more informed decision-making. Furthermore, some participants indicated a need to improve their technical analysis skills in order to better understand market trends and make sound investment decisions.

Conclusion

This project contributes valuable insights to the existing literature by shedding light on the challenges and frustrations experienced by individual investors in Kenya when participating in international stock markets. Additionally, the study reveals the investment priorities and areas for improvement identified by investors in their decision-making process. These findings provide essential guidance for the development of an optimal platform aimed at assisting investors in saving time and resources while investing.

Appendix H: Descriptive and Inferential Statistics

Investor Attitudes and Investment Decisions in International Stock Markets: An Inferential Statistics Study Based on a Kenyan Investor Questionnaire

Introduction

This research study aims to explore the knowledge acquisition process involved in making trading decisions as an individual investor in Kenya who participates in international stock markets. The primary objective is to gain insights into the investment objectives and risk profiles of individual investors in Kenya, as well as to understand the factors that influence their investment decisions. To achieve this, a questionnaire was developed and administered to a sample of individual investors in Kenya, and the collected data underwent a comprehensive analysis using descriptive and inferential statistics to draw conclusions about the population of individual investors in Kenya who invest in international stock markets.

The study provides valuable insights into the investment behaviour of individual investors in Kenya and seeks to develop an optimal platform to assist them in saving time and money when investing. The analysis of descriptive and inferential statistics focuses on three main areas: investment objectives, risk tolerance, and current investments. The investment objectives section examines investors' attitudes towards risk and their willingness to take higher risks for potential investments. Additionally, the study explores risk tolerance as a crucial factor in investment decisions, aiming to understand investors' philosophies and their readiness to accept potential losses for potential returns. Lastly, the current investments section investigates investors' current investment choices and the research methods employed in making investment decisions.

By analysing these three areas, the study provides valuable insights into the investment practices of individual investors in Kenya and sheds light on the factors influencing their investment decisions. Ultimately, the study's results will be utilised to develop a platform that can assist individual investors in Kenya in saving time and money when making investment decisions.

Investment Objectives

The analysis of the data revealed that a significant proportion of respondents exhibited a moderately conservative investment inclination, accounting for 40% of the sample (see Figure H.1). Furthermore, a substantial percentage of participants expressed a preference for initiating withdrawals within a three to five-year investment period, constituting 31.4% of the total sample (see Figure H.2). Conversely, a relatively smaller proportion of participants embraced a markedly aggressive investment stance, standing at 8.6%, with an intention of

commencing withdrawals in less than a year, signifying 5.7% of the participants (see Figure H.3). These findings offer valuable insights into the investment behaviour and preferences of the respondents.

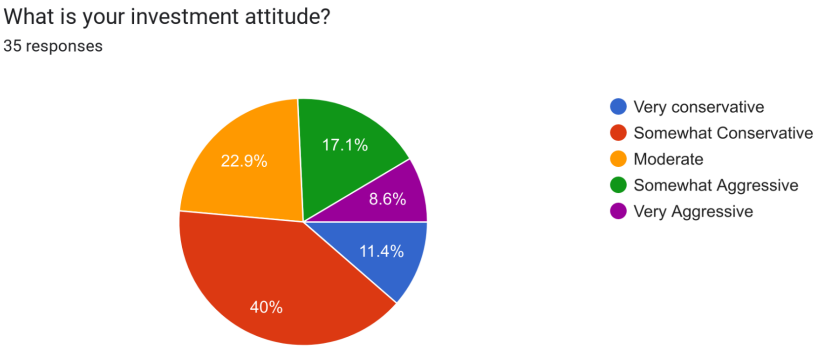


Figure H.1: Investment Attitude

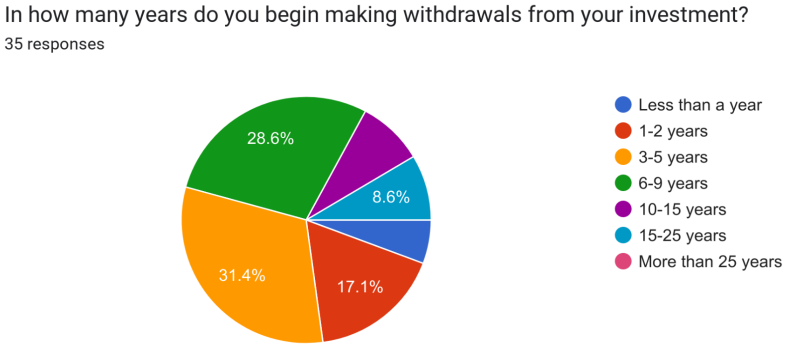


Figure H.2: Timeframe for Investment Withdrawals

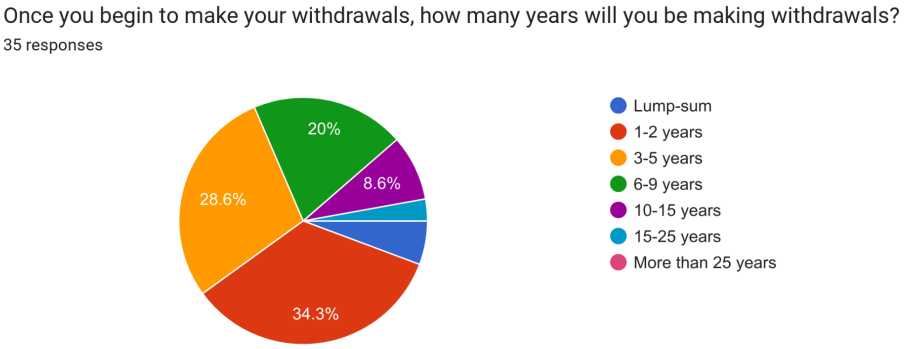


Figure H.3: Length of Time for Making Withdrawals

Risk Tolerance

The quantitative data analysis revealed that a substantial segment of the research population displayed a degree of tolerance towards investment volatility and aimed to establish well-diversified investment portfolios, as depicted in Figures E.4, E.5, and E.6. A minority of respondents expressed a preference for stable investments and were motivated by the possibility of attaining considerable returns. Furthermore, the majority of participants concurred that portfolio preservation took precedence over high returns (see Figure H.6).

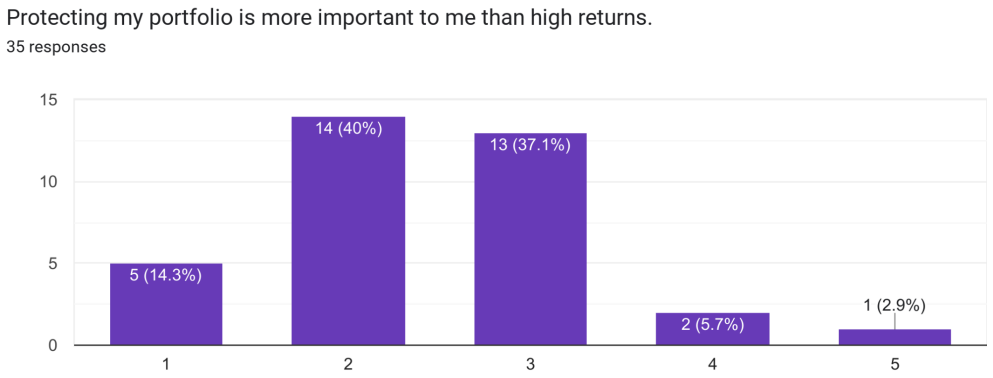


Figure H.4: Investment Philosophy (Investment Goals)

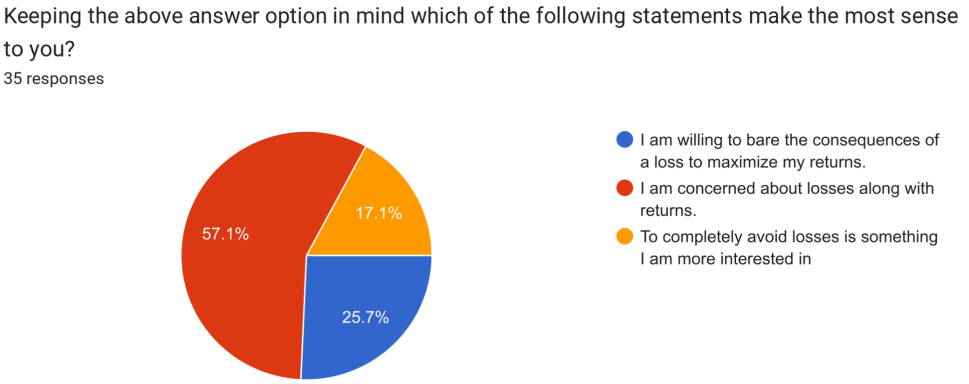


Figure H.5: Investment Philosophy (Risk Tolerance)

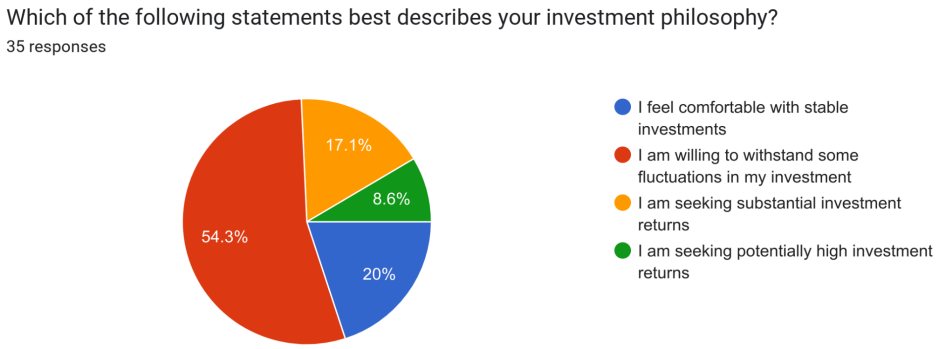


Figure H.6: Investment Philosophy (Approach to Investment)

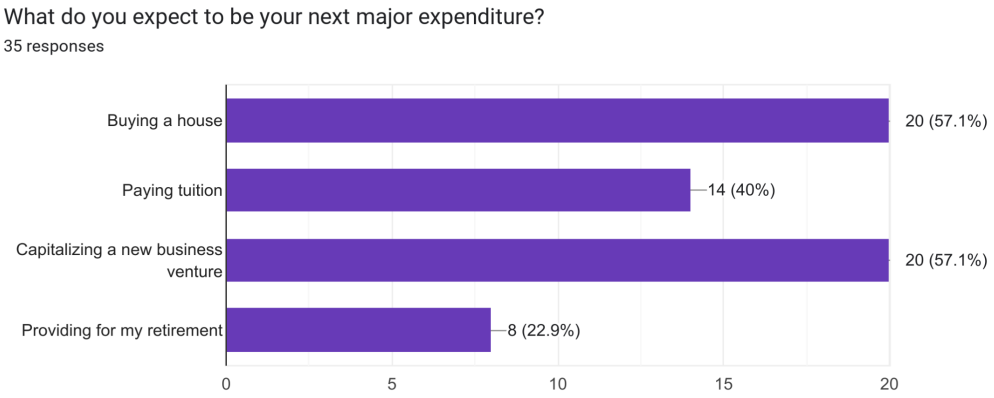


Figure H.7: Next Major Expenditure

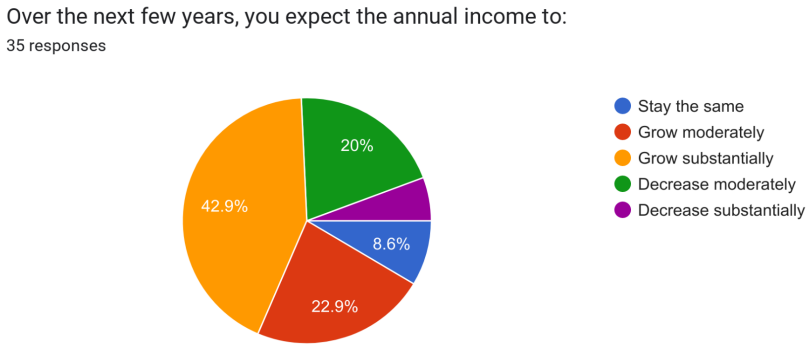


Figure H.8: Expected Annual Income

Current Investments

The analysis of the quantitative data indicated that the majority of respondents allocated between 2 to 5 hours for researching a stock and preferred investing in established and prominent companies with high growth potential, representing 68.6% and 65.7% of the

sample, respectively (see Figures E.8 and E.9). Regarding their current investments, the majority of participants invested in stocks and/or funds, while a smaller proportion invested in bonds and/or funds, accounting for 82.9% and 17.1% of the sample, respectively (see Figure H.10).

Moreover, the qualitative data analysis revealed that the most significant investment priority for participants was a faster and more efficient means of sorting through information, as expressed by 90% of the sample (see Figure H.12). This was followed by a smarter approach towards obtaining information, which accounted for 20% of the participants. These findings underscore the importance of streamlined and effective information retrieval techniques in guiding investment decisions and suggest that investors place a high value on efficiency and accuracy in their decision-making process.

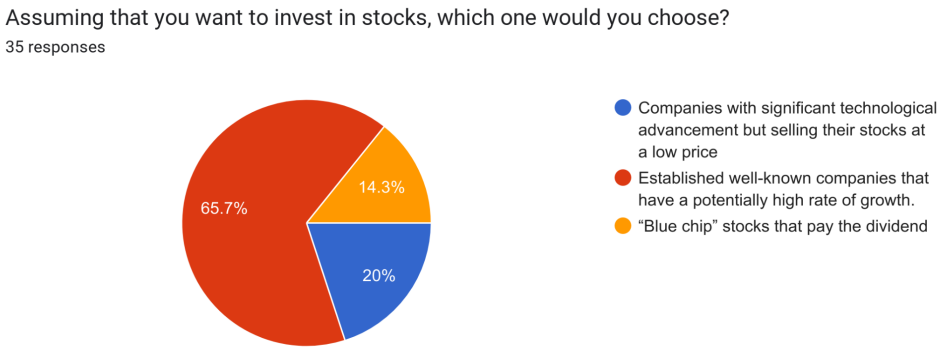


Figure H.9: Preferred Stocks to Invest In

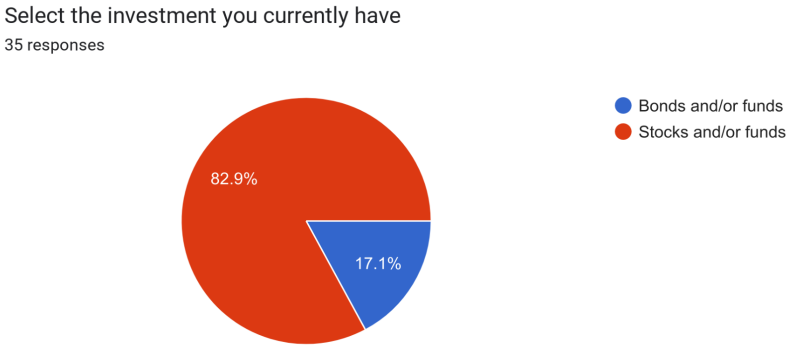


Figure H.10: Current Investments

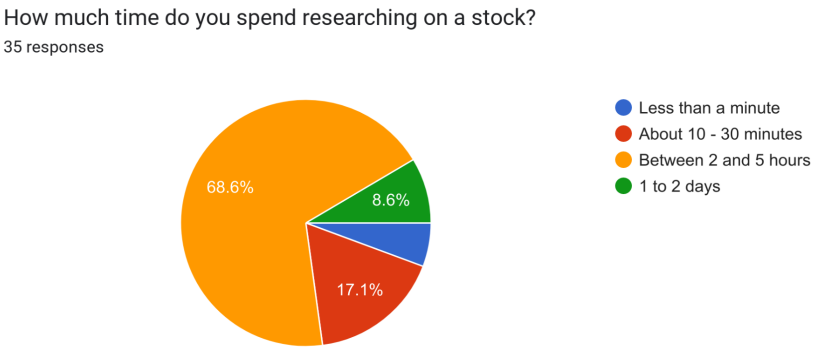


Figure H.11: Time Spent Researching a Stock

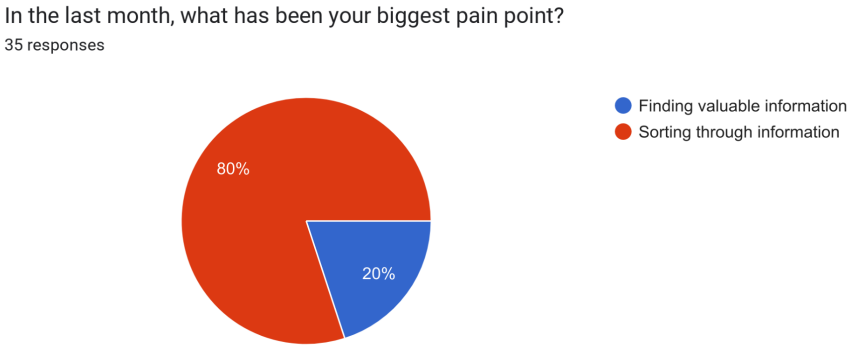


Figure H.12: Biggest Pain Point in the Last Month

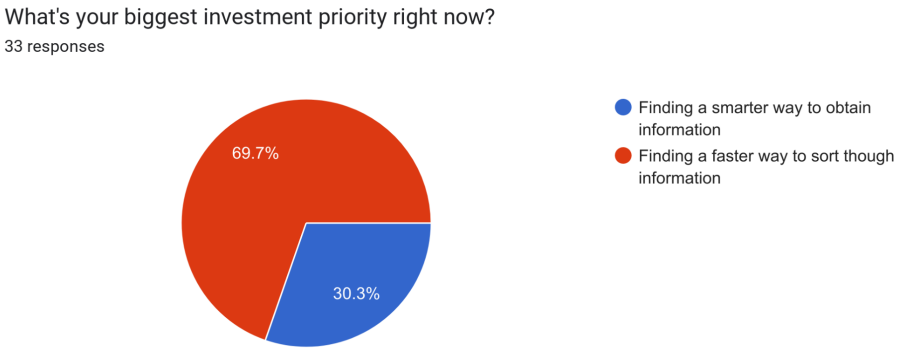


Figure H.13: Current Investment Priorities

Conclusion

This research study provides valuable insights into the investment behaviour and preferences of individual investors in Kenya who participate in international stock markets. The findings offer important considerations for the development of an optimal platform that can assist these investors in saving time and money when making investment decisions. By

understanding investors' investment objectives, risk tolerance, and current investment practices, the platform can be tailored to meet their specific needs and enhance their overall investment experience.

Appendix I: Content Analysis Report

The Impact of Social Media on Stock Prices: A Content Analysis of Microsoft and Apple

Introduction

This research study employs a Content Analysis methodology to investigate the potential influence of social media on the stock prices of two prominent technology companies, Microsoft and Apple. With the growing popularity of social media as a platform for investors to share their opinions and perspectives on the stock market, it is essential to understand the impact of social media on stock prices. Therefore, this project aims to analyse the content of social media posts related to Microsoft and Apple and determine their potential influence on the respective stock prices. By examining the relationship between social media and stock prices, this project contributes to the expanding body of research on the impact of digital media on financial markets.

Methodology

The study adopts a content analysis methodology to examine social media data associated with Microsoft and Apple. Data was collected over a two-month period, from January 2013 to February 2022, and sourced from reputable data providers, including Unicorn Data Services' Financial Data APIs, Yahoo Finance API, and Twitter API. A coding scheme was developed to analyse the sentiment, volume, and content of the collected data. The analysis aimed to identify any potential correlations between social media activity and the stock prices of the two companies.

Results

The findings of the study reveal that social media sentiment plays a significant role in determining the stock prices of Microsoft and Apple. Positive sentiment expressed on social media shows a positive association with stock price increases, while negative sentiment demonstrates a negative correlation. Additionally, the volume of social media posts related to the two companies shows a positive correlation with stock price changes, indicating the importance of social media as an information source for investors. The content analysis of social media posts highlights the frequent discussions on topics such as product launches, financial performance, and industry trends.

Discussion

The results of the study provide empirical evidence that social media sentiment and volume of posts can significantly impact the stock prices of Microsoft and Apple. Positive sentiment and a high volume of social media posts have a positive influence on stock prices, while

negative sentiment has a negative impact. Therefore, it is crucial for investors and financial analysts to monitor social media trends to make informed investment decisions. By considering the impact of social media sentiment, volume, and content analysis, investors can make well-informed decisions when investing in the stock market.

Conclusion

The empirical findings of this project demonstrate that social media sentiment and the volume of posts have a substantial impact on the stock prices of Microsoft and Apple, underscoring the importance of social media as a valuable information source for investors and financial analysts. The study highlights the benefits of monitoring social media trends in predicting changes in stock prices, enabling investors to make informed investment decisions. Future research endeavours could expand on this project by analysing social media data related to a broader range of companies and industries, contributing to a deeper understanding of the impact of social media on financial markets.

Appendix J: System Main Use Cases

Use Case: Investor Registration and Profile Setup

Description: This use case outlines the process by which an investor registers for the system and sets up their profile by providing personal information and risk appetite.

Primary Actor: Investor

Stakeholders:

i. System: Responsible for collecting and securely storing investor's personal information and risk appetite to generate personalised portfolio recommendations.
ii. Investor: Wants to create a profile that accurately reflects their financial objectives and risk tolerance.

Preconditions:

i. The investor has access to the system.
ii. The investor possesses an active email address and phone number.
iii. The system is available and functioning properly.

Postconditions:

i. The investor's profile is successfully created in the system with the provided information.
ii. The investor can log in to the system using their registered credentials.
iii. The investor's risk appetite and personal information are stored in the system for future reference.

Main Flow:

i. The investor navigates to the registration page of the system.
ii. The system presents the registration form to the investor, including fields for personal information such as name, date of birth, and contact details, as well as questions related to risk appetite.
iii. The investor fills in the required fields and submits the form.
iv. The system verifies the provided information and securely stores the investor's profile.
v. The system displays a message to the investor confirming the successful creation of their profile and prompts them to log in.
vi. The investor enters their registered email address and password to log in to the system.

vii. The system authenticates the investor's credentials and grants access to their personalised portfolio, stock recommendations, news and articles, and trading statistics.
viii. The system securely stores the investor's risk appetite and personal information for future reference.

Alternative Flow:

i. If the investor enters invalid information during the registration process, the system displays an error message and prompts the investor to correct the information before submitting the form.

Use Case: Investor Login

Description: This use case describes the process by which an investor logs in to the investment platform.

Primary Actor: Investor

Stakeholders:

i. Investor: The primary user who needs to log in to access the platform's features and functionalities.
ii. System Administrator: Responsible for managing and maintaining the investment platform.

Preconditions:

i. The investor has registered and created an account.
ii. The investor possesses a valid username and password.

Postconditions:

i. The investor successfully logs in to their account and gains access to the platform's features and functionalities.

Basic Flow:

i. The investor navigates to the login page of the investment platform.
ii. The investor enters their username and password.
iii. The investor clicks the "Login" button.
iv. The system validates the credentials entered by the investor.
v. If the credentials are valid, the system logs the investor into their account.

vi. The investor is redirected to the dashboard, where they can access the platform's features and functionalities.

Alternative Flows:

i. If the investor enters an incorrect username or password, the system displays an error message and prompts the investor to re-enter the correct credentials.
ii. If the investor forgets their password, they can click the "Forgot Password" button and follow the instructions to reset their password.
iii. If the investor encounters any difficulties logging in, they can contact the system administrator for assistance.

Use Case: View Personalised Portfolio

Description: This use case outlines the process by which an investor views their personalised portfolio on the stock market recommender system. The personalised portfolio is generated based on the investor's risk appetite, trading history, and other personal information provided during registration.

Primary Actor: Investor

Stakeholders:

i. Investor: The primary actor interested in viewing their personalised portfolio on the stock market recommender system. They want to be able to assess and analyse their portfolio to make informed investment decisions based on their risk appetite and investment goals.
ii. System Administrator: Responsible for managing the stock market recommender system, ensuring the security and privacy of investors' personal and financial information.

Preconditions:

i. The investor has successfully logged into their account.
ii. The investor has provided their personal information and risk appetite during registration.

Postconditions:

i. The investor can view their personalised portfolio on the system.

Flow of Events:

i. The investor navigates to the "View Personalised Portfolio" option on the system.

ii. The system verifies the investor's credentials and grants access to their personalised portfolio.
iii. The system generates a portfolio for the investor based on their risk appetite and other personal information provided during registration.
iv. The system displays the personalised portfolio, including details such as currently held stocks, the value of each stock, and the overall portfolio value.
v. The investor can view and analyse their personalised portfolio, and take actions such as selling or buying stocks based on their investment goals.

Alternate Flow:

i. If the investor has not yet provided their personal information and risk appetite during registration, the system prompts them to do so before granting access to their personalised portfolio.
ii. If the investor encounters any issues while viewing their portfolio, they can contact the system administrator for assistance.

Use Case: View Stock Recommendations

Description: This use case describes the process by which an investor views stock recommendations provided by the investment platform.

Primary Actor: Investor

Stakeholders:

i. Investor: The primary actor interested in viewing stock recommendations and making investment decisions based on the provided recommendations.
ii. System Administrator: Responsible for managing and maintaining the investment platform.

Preconditions:

i. Investors have successfully logged into their accounts.
ii. Investors have navigated to the stock recommendations page.

Postconditions:

i. Investors have access to recommended stocks based on their risk appetite and investment goals.

Basic Flow:

i. The investor navigates to the stock recommendations page.

ii. The system displays a list of recommended stocks based on the investor's risk appetite and investment goals.
iii. The investor can view details of each recommended stock, such as current market value, historical performance, and associated news or analysis.
iv. The investor can add recommended stocks to their personalised portfolio for tracking and potential investment.

Alternate Flow:

i. If the system is unable to generate stock recommendations based on the investor's risk appetite and investment goals, an error message is displayed, and the investor is prompted to adjust their preferences or contact the system administrator for assistance.

Use Case: Access Financial News and Analysis

Description: This use case describes the process by which an investor can access financial news and analysis through the investment platform. The system provides access to a variety of news and analysis, including industry news, market trends, and economic indicators.

Primary Actor: Investor

Stakeholders:

i. Investor: The primary actor who needs to access financial news and analysis to make informed investment decisions.
ii. News Providers: Organisations that generate and publish financial news, such as news agencies, financial media companies, and financial blogs.
iii. System Administrator: Responsible for managing and maintaining the investment platform.

Preconditions:

i. The investor has successfully logged into their account.

Postconditions:

i. The investor can access financial news and analysis provided by the investment platform.

Flow of Events:

i. The investor navigates to the "Access Financial News and Analysis" option on the system.

ii. The system verifies the investor's credentials and grants access to the financial news and analysis section.
iii. The system displays the latest financial news and analysis from various sources, including news articles, market trends, and economic indicators.
iv. The investor can select a specific news article or analysis to view the details.
v. The system displays the selected news article or analysis in detail.
vi. The investor can analyse the news and analysis to make informed investment decisions.

Alternate Flow:

i. If the investor encounters any issues while accessing the financial news and analysis, they can contact the system administrator for assistance.
ii. If the system encounters any issues while displaying the financial news and analysis, it notifies the administrator to investigate and resolve the issue.

Use Case: Train Machine Learning Models

Description: This use case outlines the process of training machine learning models that will be used to generate stock recommendations for investors. The machine learning models are trained on historical stock data and periodically updated to ensure their accuracy.

Primary Actor: System Administrator

Stakeholders:

i. System Administrator: Responsible for managing and maintaining the investment platform and training the machine learning models.
ii. Investors: Benefit from accurate and up-to-date machine learning models for generating stock recommendations.

Preconditions:

i. The system has collected and stored historical stock data.
ii. The machine learning models have been developed.

Postconditions:

i. The machine learning models have been trained on the historical stock data.
ii. The trained models are integrated into the system for generating stock recommendations.

Flow of Events:

i. The system administrator triggers the machine learning model training process.
ii. The system collects and preprocesses the historical stock data.
iii. The system trains the machine learning models using the preprocessed data.
iv. The system evaluates the performance of the trained models.
v. If the models perform satisfactorily, the trained models are integrated into the system for generating stock recommendations.
vi. If the models' performance is unsatisfactory, the system re-evaluates and refines the models before integrating them into the system.

Alternate Flow:

i. If the historical stock data is not available, the system notifies the administrator to collect and store the data before triggering the machine learning model training process.
ii. If the training process fails, the system notifies the administrator to investigate and resolve the issue before re-triggering the process.

Exceptional Flow:

i. If the performance of the trained models is poor, the system re-evaluates and refines the models before integrating them into the system for generating stock recommendations.

Appendix K: System Operation Contracts for the Main Sequence of Events

Contract CO1: Investor Data Update
Operation: investor_data_recommendation()
Cross References: Use Case: View Stock Recommendations
Preconditions:
1. Investor has provided their personal information and trading history to the system
2. System has access to real-time market data

Postconditions:
1. Investor data is updated in the system

Contract CO2: Investor Data Segmentation
Operation: investor_segmentation()
Cross References: Use Case: Generate Personalised Portfolio
Preconditions:
1. Investor data has been updated in the system

Postconditions:
1. Investors are segmented based on common characteristics such as demographics, psychographics, and behaviours

Contract CO3: Sort investments by value and risk for suitable investors.
Operation: filter_recommendations()
Cross References: Update Investor Risk Appetite
Preconditions:
1. Investor data has been collected and updated.
2. Segmentation criteria have been established.

Postconditions:
1. Low-investment-value stocks have been filtered out.
2. High-investment-value stocks have been recommended to high-rated investors.
3. High-risk moves have been recommended only to those with a high-risk appetite.

Contract CO4: Data Analysis using Financial Data APIs and Twitter API
Operation: data_analysis()
Cross References: Use Case: Train Machine Learning Models
Preconditions:
1. The investment platform system is up and running.
2. The system has access to the Unicorn Data Services' Financial Data APIs and Twitter API.
3. Data sources are available and updated.

Postconditions:

1. The system analyses sentiments around the company and essential business commodities.

Contract CO5: Assess Relationship between Company and Significant Commodities
Operation: assess_relationship()
Cross References: Use Case: Train Machine Learning Models
Preconditions:

1. Data has been collected from Unicorn Data Services' Financial Data APIs and Twitter API to analyse sentiments around the company and essential business commodities.
2. The system has access to a database of significant commodities and their historical impact on stock prices.

Postconditions:

1. A weight is assigned to each significant commodity based on its impact on the stock price of the company.
2. The weight is stored in the system's database for use in generating recommendations.

Contract CO6: Collect and identify optimal models for creating a strategy.
Operation: optimal_models()
Cross References: Use Case: Train Machine Learning Models
Preconditions:

1. Data from the Yahoo API is available.
2. The ensemble agent is properly configured and operational.

Postconditions:

1. Optimal models for creating a strategy have been identified.
2. The identified models are available for use in creating a recommendation strategy.

Contract CO7: Evaluate Optimal Models for Accuracy
Operation: evaluate_optimal_models()
Cross References: Use Case: Train Machine Learning Models
Preconditions:

1. Data from Yahoo API has been collected and subjected to an ensemble agent to identify optimal models for creating a strategy.

Postconditions:

1. The optimal models have been evaluated for accuracy.

Contract CO8: New data is passed through a trained recommender model.
Operation: train_on_new_data()
Cross References: Use Case: Train Machine Learning Models

Preconditions:

1. The system has access to new data to be passed through the trained recommender model.
2. The recommender model has been trained and is available for use.

Postconditions:

1. The new data has been processed by the trained recommender model.
2. The resulting recommendations are ready for tailoring to the individual investor.

Contract CO9: Tailoring recommendations to individual investors.
Operation: tailor_recommendations()
Cross References: Use Case: View Personalised Portfolio
Preconditions:

1. Segmented investor data must be continuously updated and processed based on segmentation criteria.
2. Low-investment-value stocks must be filtered out, and high-investment-value stocks recommended to high-rated investors based on the segmentation criteria.
3. High-risk moves must be recommended only to those with a high-risk appetite based on the segmentation criteria.
4. Data must be collected from Unicorn Data Services' Financial Data APIs and Twitter API to analyse sentiments around the company and essential business commodities.
5. The relationship between the company and significant commodities must be assessed, and a weight placed on that commodity in terms of its impact on the stock price.
6. Data from Yahoo API must be collected and subjected to an ensemble agent to identify optimal models for creating a strategy.
7. The optimal models must be evaluated for accuracy.

Postconditions:

1. A trained recommender model must be used to process new data.
2. The resulting stock recommendations must be tailored to the individual investor based on their segmented data and risk appetite.

Appendix L: Research Budget

This research study was conducted in a self-financed manner, utilising personal resources and equipment to ensure cost efficiency. The budget was carefully allocated to different categories, including research personnel, materials and supplies, travelling costs, communication, and cloud computing expenses. Each category was justified based on the specific needs and requirements of the study.

Research Personnel

The research personnel category did not require additional financing as the lead researcher, who possessed the necessary expertise, conducted the study independently. This approach ensured efficient resource allocation and minimised costs associated with hiring additional personnel.

Materials and Supplies

Expenses in the materials and supplies category were necessary for smooth study operations. This included the printing of survey forms and questionnaires, although efforts were made to minimise environmental impact by utilising online platforms such as Google Forms. By embracing digital tools, the study not only reduced costs but also contributed to environmental conservation.

Travelling Costs

The travelling costs category accounted for expenses related to transportation to and from data gathering sites. In this project, the estimated cost was approximately 800 per day, utilising Uber as a convenient and cost-effective mode of transportation. By opting for Uber, the study managed to control travel expenses while ensuring timely data collection.

Communication

The communication category covered expenses associated with essential communication channels such as calls, messages, Wi-Fi, and data bundles. These resources were necessary for conducting online interviews if required and maintaining effective communication throughout the study. By allocating a portion of the budget to communication, the study ensured seamless connectivity and efficient data gathering.

Cloud Computing Costs

To process and analyse the large volumes of collected data, cloud computing services were utilised. Amazon was chosen as the cloud provider, following a pay-as-you-go model. This approach allowed for scalability, flexibility, and cost-efficiency in handling data processing and analysis, ensuring accurate and timely results.

The budget justification, as presented in Table L.1, provides a detailed numerical breakdown of the proposed expenditure. This comprehensive breakdown enables a thorough analysis of the suggested spending, demonstrating the prudent allocation of resources and supporting the successful implementation of the research study.

Table L.1: Numerical Breakdown of the Proposed Research Study

Category	Item	Description	Quantity	Unit Cost	Total
Research Personnel	-	-	-	-	-
Materials and Supplies	-	-	-	-	-
Travelling Costs	Uber Fare	Researcher's Uber Fare to/from research site	4	Ksh. 800	Ksh. 3200
Communication	Calls	Safaricom Charges for Calls per Minute	1000	Ksh. 4.87	Ksh. 4870
	Messages	Safaricom Charges for texts per SMS	100	Ksh. 1.20	Ksh. 120
	Wi-Fi	Faiba Charges for Wi-Fi per month	2	Ksh. 5250	Ksh. 10500
	Data bundles	Safaricom Charges for Data per MB	4000	Ksh. 4.87	Ksh. 19480
Cloud Computing Costs	Amazon Elastic File System	Effective storage price per Terabyte - One Zone*	2	Ksh. 6.48	Ksh. 12960

	Amazon Elastic Compute Cloud	Amazon EC2 Spot Instances per Hour	672	Ksh. 2.36	Ksh. 1588.61
Total					**Ksh. 52718.61**