BUILDING A SOFTWARE APPLICATION TO PREDICT THE STOCK MARKET USING THE ROUGH SET THEORY

MSc Computer Science

Student Name: Oladipo Ibukunoluwa Franklin
Student Number: 22028495

Staffordshire University

A final report submitted in partial fulfillment of the requirement of Staffordshire University for the degree of Master of Science.

Abstract

There has never been a time when there was a greater need for cutting-edge forecasting technologies than there is in the ever-changing world of financial markets. This study provides an in-depth analysis and critique of a predictive software application that has been methodically developed for the purpose of forecasting the stock market. The application is supported by the cutting-edge theory known as Rough Set Theory. An in-depth investigation into the many facets of the program is presented, with an emphasis placed on its many strong points, such as its pioneering innovations, user-centric design ethos, and dedication to research-driven analysis. Nevertheless, it does not shy away from the identification of shortcomings, with significant examples being the complexities of manual data gathering and the imperatives of model openness and regulatory compliance. An array of potential, including the adoption of sophisticated prediction algorithms, the seamless incorporation of real-time data, and an emphasis on ethical issues, has been uncovered as a result of the evaluation, and these prospects combined hold the promise of increased predictive precision and adaptability. Concurrently, the report sheds light on the challenges that are looming big. These concerns range from strong competition to developing regulatory environments, data security vulnerabilities, market turbulence, and the extremely important issue of user trust. In light of these revelations, the paper outlines a course for prospective advancements in the future, propelling the application into an idealized future in which it has the ability to redefine the parameters of stock market forecasting within the realm of financial technology. This report also focuses on building a software (mobile) application that will serve as a visual representation of the predicted stock market data.

Keywords: *Financial market, Stock market, Rough Set Theory, Prediction, Mobile application.*

Acknowledgement

This thesis stands as a culmination of effort, support, and inspiration from various sources, for which I am sincerely grateful.

First and foremost, I extend my deepest appreciation to my supervisor, Abdul Rehman whose mentorship and insights were instrumental in shaping this work. His dedication to academic excellence has been a guiding light.

A debt of gratitude is owed to my colleagues and fellow researchers who provided a stimulating environment for discussions and shared their expertise, enriching my understanding of the subject matter. To my family, whose unwavering belief in my capabilities and boundless patience provided the foundation for this endeavor, I am profoundly thankful. Finally, to my friends, whose support and companionship offered moments of respite and balance amidst the demands of this pursuit, I am appreciative.

Table of Contents

Table of Figures

CHAPTER 1

1.0 Introduction

Fortunes can be amassed or lost in the blink of an eye in the complex world of financial markets, which are known for their labyrinthine complexity. The capacity to foresee market trends has transformed from a merely advantageous trait into a formidable necessity in this high-stakes environment. In order to manage the unpredictable landscape of stocks, bonds, and commodities, investors, traders, and financial institutions now rely heavily on predictive analytics. Beyond the area of profit-seeking, it is imperative to make accurate and timely predictions because they act as a critical barrier against unanticipated market volatility and economic uncertainty.

The interconnectedness of markets has offered both previously unheard-of opportunities and increased threats in today's globalized world, as the impacts of financial actions are felt well beyond national boundaries. A choice taken in one region of the world can have an immediate and profound impact on the entire global financial ecosystem. As a result, the need for reliable prediction tools has skyrocketed to the top of financial tactics. The urgency to reduce risks, grasp fleeting opportunities, and optimize investment strategies in a world where procrastination can result in large losses is what's driving this increase.

The ability of prediction algorithms to offer a route through the confusing landscape of financial markets is at the core of this need. They provide clarity amidst the confusion, enabling decision-makers to make well-informed decisions amidst the clamor of market forces. Investors can use these tools as a compass and beacon to steer clear of perilous shoals of uncertainty and toward the shores of financial success.

But much like the markets themselves, the universe of prediction tools is broad. It includes a wide range of approaches, including cutting-edge machine learning and artificial intelligence models as well as conventional financial research. Rough Set Theory, a rigorous and understandable paradigm for data analysis, has emerged as an appealing strategy amidst this diversity. It makes the promise of providing both precise predictions and the openness necessary to comprehend why and how they are made.

In this backdrop, this paper sets out on a revolutionary journey into the creation of ground-breaking software designed specifically for stock market forecasting. In an effort to give investors and financial professionals a potent tool for managing the volatile financial markets, it aims to leverage the analytical acumen of rough set theory. We will discover the essential qualities and elements that support the development of such a financial application through this investigation. Additionally, we will explore the complex methodology used to create this program, with a special emphasis on the mobile application framework created utilizing the adaptable Flutter/Dart technology.

This report will also clarify one of the most important aspects of our project: the verification of the prediction accuracy of the program. This examination intends to shed light on Rough Set Theory's usefulness as well as its unique benefits in the complex web of stock market forecasting.

1.1 Motivation

The research and creativity behind the Rough Set Theory stock market prediction program is astounding. This study is the result of tremendous labor, dedication, and unrelenting commitment to financial technology innovation. This research is driven by financial markets' constant change. In an era where technology influences investing decisions more than ever, accurate and transparent prediction tools are needed. This article showcases the creation of a unique predictive program and provides insights into Rough Set Theory and its potential to improve stock market forecasts. This effort seeks to promote financial technology inquiry, creativity, and discovery, where more prospects for development and advancement await.

1.2 Aims and Objectives

1.2.1 Aim

This research endeavor seeks to accomplish a comprehensive goal - the development of an advanced financial application rooted in Rough Set Theory for the purpose of predicting stock market movements. It endeavors to not only create a robust and user-friendly software application but also to evaluate its effectiveness. Beyond these tangible outcomes, the overarching aim is to contribute valuable insights that can enhance financial decision-making in a dynamic and intricate market environment.

1.2.2 Objectives

Identify Key Attributes and Factors for Financial Application Development

The initial goal comprises locating and carefully examining the essential characteristics and elements required for developing a financial application using rough set theory. This includes a thorough evaluation of their applicability to stock market forecasting. The knowledge gained will be used to guide the development process at later phases, ensuring that the application contains all necessary components for proper operation.

Develop a User-Friendly Financial Software Application

The design of a practical and user-friendly financial software application is the second goal, which builds on the fundamental information obtained from the first goal. This software ought to be able to accurately forecast changes in the stock market while providing a user-friendly interface. This goal entails incorporating data visualization tools that convey forecasts graphically to improve user comprehension, as well as user input options for stock preferences and time horizons.

Validate the Application's Efficacy through Comparative Analysis

Validating the application's prediction abilities is a crucial goal. To do this, historical financial data will be gathered and used to create stock market forecasts using a model based on rough set theory. The objective is to thoroughly examine the findings and evaluate the reliability and efficacy of the financial software program. This goal offers to shed light on the potential benefits and drawbacks of applying rough set theory to the challenging field of stock market forecasting.

Provide Recommendations for Future Development

The final objective is to derive actionable recommendations from the research findings. These recommendations will not only offer guidance on enhancing the application's performance and user experience but also propose potential refinements for the Rough Set Theory-based model and user interface. Moreover, the research endeavors to underscore the broader implications of its findings for financial decision-makers and investors, shedding light on the innovative avenues that technology can open in the realm of financial prediction and decision-making.

CHAPTER 2

2.1.0 Literature Review

The stock market is a volatile and complex environment that poses certain difficulties for investors and financial professionals. The ability to accurately predict stock market trends and make informed investment decisions is highly sought in today's fast-paced and competitive financial landscape. Software applications that use advances in machine learning and data analysis to predict stock market changes are attracting more and more attention.

The proposal aims to demonstrate how rough set theory can be effective in predicting stock market prices. According to Bouzayane and Saad (2020), the Rough Set Theory is chosen because it provides a powerful method for analyzing financial data uncertainties and eliminating unnecessary attributes, making it practical for making predictions in multi-attribute classification problems. By building a software program that utilizes rough set theory, this project intends to give financiers with a beneficial tool for monitoring stock market patterns and making educated investment decisions (Szul and Kokoszka 2020). The technical aspects of the application's development, such as data processing, algorithm implementation, and performance evaluation, will be discussed in depth in the subsequent sections of this proposal to demonstrate the solution's efficacy and practicality.

2.1.1 Research Background

Economic data, market mood, corporate performance, and global events all have an influence on the stock market, which is a highly complex and volatile system. According to Antosz et al., investors and financial analysts constantly strive for an advantage in stock market trend prediction and profitable investment decisions (Qi et al. 2021). However, accurately predicting the stock market's movements remains a significant challenge due to its inherent uncertainties and dynamic nature.

Traditional techniques for securities exchange examination, like principal and specialized investigation, impede catching complicated connections and examples inside vast volumes of financial information. These approaches are time-consuming and susceptible to human bias because they frequently rely on subjective interpretations and manual calculations. As a result, there is a growing interest in developing software applications that effectively predict stock market behavior by utilizing cutting-edge technologies like machine learning.

Machine learning, an artificial intelligence area, offers strong methods for evaluating and extracting patterns from big datasets. Machine learning algorithms can learn complex relationships and patterns that may not be apparent to human analysts by training models on historical stock market data and relevant financial indicators (Tabor et al. 2021). The stock market's future behavior may then be predicted using these models, providing investors with useful information.

Rough Set Theory is used in this study as a feasible technique for stock market prediction. Rough Set Theory provides a mathematical framework for dealing with uncertainty and making decisions based on weak or incomplete knowledge. (Tang et al., 2020). Using rough set theory, looking at financial data and determining the most important characteristics that influence stock market trends is now possible. This makes predictions more precise.

The study's goal is to create a software program that uses Rough Set Theory to forecast stock market fluctuations. By utilizing the rough set theory to deal with uncertainty, this application promises to give financiers a reliable and efficient tool for assessing stock market data and making educated investment decisions.

Through this examination, the potential advantages incorporate superior dynamic capacities for financial backers, diminished dependence on the abstract investigation, and expanded effectiveness in handling vast volumes of financial information (Forghani et al., 2022). This research aims to advance financial analysis and empower investors with valuable insights for successful trading strategies by comprehensively understanding the stock market and its underlying patterns.

2.1.2 Research Aim

With the use of machine learning techniques, notably the Rough Set Theory, this project intends to develop a software program that forecasts changes in the stock market.

Can the efficiency and accuracy of stock market forecasts be increased by the use of machine learning techniques, in particular the rough set theory?

2.1.3 Research Questions

1. What are the important attributes/factors to develop a financial application using rough set theory

2. How to develop financial software using rough set theory.

3. How to validate the financial software efficacy by comparing existing machine learning prediction models with rough set theory.

2.1.4 Objectives

1. To develop software application for stock market prediction incorporating machine learning algorithms, precisely the Rough Set Theory.

2. To evaluate the developed software's impact and practical implications on financial decision-making processes, considering its potential advantages and drawbacks in actual situations.

2.1.5 Related Work

Because accurate forecasting can result in significant financial rewards, the field of financial market prediction has attracted a lot of attention and competition. Using a variety of theories, models, and methods, researchers and analysts have spent a lot of time trying to figure out what the market's mysteries are. One arising approach that has gotten some decent forward movement is the use of unpleasant set hypothesis. In the 1980s, Zdzisaw Pawlak introduced rough set theory, which gives a framework for dealing with data that is uncertain, imprecise, or hazy. This makes it especially significant for breaking down and anticipating financial exchange patterns, which are portrayed by intricacy and vulnerability.

In the beginning, classification issues were the primary focus of the application of rough sets in both finance and economics. It was used to predict bankruptcy risks, identify failing businesses, and anticipate merger and acquisition opportunities. A deeper comprehension of the underlying mechanisms has emerged as a result of extensive research into the use of rough sets to select robust indicators for financial crises. One more basic area of use, and the essential focal point of this review, spins around creating prescient models for financial exchange developments and related exchanging techniques (Renigier-Biłozor et al. 2019). This field has seen a lot of activity because of the potential financial rewards and the complexity involved. Rough set theory has demonstrated promise in optimizing portfolio selection in a variety of market conditions, and active portfolio management is another active research field.

Methodologies have become more advanced as research in these areas has progressed. Classical rough set theory's most fundamental applications are included, as are more advanced extensions like fuzzy rough sets. Hybrid models that combine rough sets with a variety of other information discovery methods have also been investigated by researchers (Zhang et al. 2019). This evolutionary trend demonstrates the adaptability and adaptability of rough set applications when it comes to solving complex problems in financial prediction.

We will delve deeper into the literature regarding the construction of Models for predicting stock market fluctuations, financial risk management, and portfolio management using fuzzy rough sets and rough sets in the following sections (Tang et al. 2020). The investigation of rough sets in these fields has provided valuable insights that have the potential to improve decision-making tools within the field of finance. Each of these domains has its own set of challenges as well as potential rewards.

2.1.6 Trading Rules and Financial Time Series

The utilization of rough sets in financial market prediction encompasses two significant applications: financial time series prediction and the development of trading rules. This integration of data science and financial analysis arises from the recognition of the intricate nature and uncertainty inherent in financial markets, surpassing the limitations of conventional analytical techniques.

2.1.7 Time Series Forecasting

It involves using data that has already been observed to make predictions about future values. Time series data typically includes historical prices, trading volumes, and other economic indicators in the finance field. However, these datasets frequently exhibit noise, non-linear behavior, and non-stationarity, making analysis and prediction difficult.

Rough set theory provides a novel perspective on these difficulties. It empowers the catch of connections and conditions between factors in an implied way, working with successful prevailing upon loose and questionable data. In order to deal with the inherent ambiguity and haziness of financial data, rough set theory may be applied. The results of a study on using rough set theory to financial time series forecasting were favorable.. For instance, researchers used numerical attribute conversion to predict economic and stock market trends in a study that made use of rough set theory. They created time series input properties by combining delay windows, averaging, and cumulative aggregation. Using a rough set model, the study looked at a range of daily stock prices from 1990 to 1996 and used linear attribute conversion (Pal and Kar 2019). The examination laid out associations between a bunch of 10 stock costs and record values, filling in as restrictive properties, and the objective load of Applied Materials, going about as the choice variable. Classification rules were then used to represent these relationships. Initially, the decision variable was also discretized, resulting in the creation of 67 decision classes. However, this method resulted in weak, unsupported rules (Cheng et al. 2010). A system of value buckets chosen at random was used to address this, dividing the continuous value spectrum into three decision classes: expanding, stable, and diminishing. Out-of-sample data were used to assess the derived rules' robustness from August 1, 1996 to April 9, 1997. On the validation sample, the average accuracy rate of the generated rules was approximately 71%.

The authors emphasized how difficult it is to create suitable discretization algorithms, which are necessary for a variety of tools for discovering knowledge, including rough sets, especially when used with stock market time series. For the selection of attribute transformations, they suggested an iterative approach with expert input. A thorough search strategy was suggested in the absence of expert guidance.

Another article by Takahashi et al. from 2019 examined a unique application of rough set theory to financial trading, especially in the market for crude oil futures. The scientists picked this market because of its basic job in the worldwide economy and its innate unpredictability, making it an optimal proving ground for the utilization of harsh set hypothesis (Pal and Kar, 2019). Their model's primary goal was to predict future price trends and create efficient trading rules based on these predictions.

The underlying move toward their methodology was the assortment of continuous information from the unrefined petroleum fates market. A variety of market parameters typically utilized in market analysis were included in this data set. The starting and closing prices, as well as the volume of trades and the highest and lowest price points during a trading session, were all included (Takahashi et al. 2019). This data set's comprehensiveness was essential for ensuring that the subsequent application of RST was based on a solid and in-depth comprehension of market conditions.

RST was ideal for interpreting the highly volatile and unpredictable nature of futures markets due to its origins as a mathematical framework designed to manage uncertainty and imprecision within datasets. The research aims at breaking up the continuous data they had collected into distinct intervals that represented specific market states or conditions (Golan and Ziarko 1995). RST relies heavily on a process known as discretization, which enables the conversion of complex continuous data into manageable discrete states that are easier to analyze.

RST was used to further identify dependencies or relationships between the various attributes or market parameters after the data had been successfully processed and discretized. A set of rules for making decisions was then created from these relationships. These rules were basically if-then statements that connected various market states and made it clear when it would be best to buy, sell, or hold futures contracts under these particular circumstances (Takahashi et al. 2019).

On an out-of-sample data set, the model-derived trading rules were then tested. A measure of the rules' robustness and their capacity to generalize to new, unseen data are provided by the fact that this data set was distinct from the initial data used to generate the rules. Because it determines whether the rules are capable of delivering a dependable performance in real-world conditions, this validation step is of the utmost importance. Pan and others 2016) discovered that their model's trading rules outperformed conventional trading strategies in terms of profitability, demonstrating the approach's efficacy.

2.1.8 Trading Rules Generation

Rough set theory has been used to create trading rules in addition to predictions. Technical analysis, which involves identifying patterns in market prices to determine potential buying and selling opportunities, is frequently used in the development of traditional trading rules. However, due to their subjective nature and inconsistent outcomes, these conventional methods have been criticized.

Researchers have looked at the potential applications of rough set theory to forecasting stock market time series using the fundamental rough set model. As input variables, the model uses raw price data together with a range of technical indicators, such as rate of change, momentum, discrepancy, and moving averages (such as opening, closing, peak, and bottom prices). A trend indicator that is derived from daily closing prices is used to determine the decision variable.

A five-year dataset that tracked daily changes in the Johannesburg Stock Exchange All Share Index was used to assess the model's performance. A learning subset (which comprised 75% of the data) and a testing subset were separated from the overall dataset. The data were prepared using data preprocessing methods like time series conversion algorithms (Zhang et al 2019). Data discretization, reduction, and the generation of predictive (classification) rules were also carried out using software tools like ROSETTA.

The rule set's prediction accuracy is heavily influenced by discretion. The effects of equal frequency binning, Boolean reasoning, entropy, and naïve algorithms on forecast quality were studied. The results demonstrated that the equal frequency binning approach with four bins generated the greatest results, laying the groundwork for additional data processing. Researchers were able to locate a reduct, or a subset of variables, by employing the genetic algorithm feature of ROSETTA. This reduction served as the basis for the creation of a set of guidelines.

The researchers based their investigation on data sourced from the year 2011, which they employed as their learning dataset. They further selected data from the initial quarter of 2012 to function as their testing dataset. Throughout both the learning and testing periods, a linkage between the rate of net returns and the direction of stock movements was investigated, thereby emphasizing its importance as a conditional factor. The model proved impressive in its predictive accuracy when dealing with out-of-sample data, achieving a commendable success rate of 97.8%. However, it's worth noting that despite these promising results, there is potential bias due to the relatively short time span of the study and the infrequency of the observed period.

To foresee securities exchange developments inside a one-day skyline, an original procedure that consolidated choice trees and unpleasant sets was proposed by another review. Technical indicators like volume, open, high, low, and close prices were computed and used as conditional attributes. The decision attribute was defined by the price trend (Golan 1995). To work on the model's exhibition, separating strategies like the C4.5 choice tree pruning strategy were utilized. In view of the separated information, harsh set reducts and choice principles were made. Through a 10-step cross-approval strategy, the half breed framework outperformed independent unpleasant set models, feed-forward brain organizations, and innocent Bayes classifiers in order precision.

Self-putting together guides (SOM), joined with harsh sets and grouping, were recommended by one more review for monetary time series prescient examination. A market timing strategy and technical indicators that were carefully selected to represent stock trends from historical time series data were the main focus of the study. Considering the connection between's consigned bundle social events and the decision variable, another decision trademark was given out or the ongoing attribute was held. Utilizing the Chi-squared method and conventional rough set principles, discretization made the process of creating reducts and rules simpler. Additionally, self-sorting out maps were utilized to alter the solidarity values of items.

The S&P 500 index's historical time series data were compared to the model's performance using a buy-and-hold strategy. During the validation phase, the proposed system's profitability performance did not significantly improve, despite outperforming the buy-and-hold strategy during the learning phase. The system's performance varied depending on the tuning parameters. It was argued that consolidating master data when selecting the best exchange rules could enhance the presentation of the framework (Wang 2003). The proposed system outperformed the buy-and-hold strategy across a wide range of indices. In any case, the advantage of the delivered trading signals remained fragile as far as possible used.

2.1.9 Risk Management in Prediction Model

The stock market inherently entails risks and uncertainties, making it imperative to adopt effective risk management strategies (Gao et al. 2020). By leveraging the rough set theory to build predictive models, the software application aims to assist users in making informed decisions. However, it is essential to acknowledge and address the potential risks associated with relying on such models for stock market predictions.

One key risk to consider when employing the rough set theory in stock market prediction is the inherent volatility and unpredictability of financial markets. While the rough set theory can handle uncertainty and imprecision in data, it may not capture all the intricate dynamics of the market accurately. Consequently, the predictions generated by the software application may carry limitations and uncertainties, impacting their reliability.

Another significant risk is the possibility of model overfitting. Overfitting occurs when a model becomes too closely aligned with the training data, compromising its ability to generalize and perform well on unseen data. To mitigate this risk, rigorous validation and testing methodologies should be employed when applying the rough set theory ((Gandhmal and Kumar 2019). Techniques such as out-of-sample validation and cross-validation can help assess and improve the model's generalization capabilities.

Data quality and availability are critical risk factors that must be addressed. The accuracy, completeness, and timeliness of the data used as input for the rough set theory model significantly influence the reliability of the predictions. Inaccurate or incomplete data can introduce biases and lead to erroneous conclusions (Gandhmal and Kumar 2019). Additionally, the availability of real-time data, crucial for generating up-to-date predictions, poses challenges in terms of data access, reliability, and potential delays. Robust data management practices and quality assurance processes are necessary to mitigate these risks.

Singh et al. 's 2018 study sought to contrast the efficacy of the rough set model with that of discriminant analysis and logit models in the context of Greek businesses. To achieve this, a greater dataset was made, containing acquiring and test tests from 80 and 36 firms, independently, tending to various current regions. The discretization of the input dataset and the selection of conditional attributes were based on expert judgment (12 out of 28 financial ratios).

The study conducted by Gandhmal and Kumar in 2019 exhibited that the application of the rough set technique outperformed traditional discriminant analysis in terms of reliability. Interestingly, the rough set model's efficacy bore a striking resemblance to that of the logit model. The research also attested to the aptness of the Valued Closeness Relation (VCR) method when it came to classifying objects that didn't exactly conform to the established rules.

A standout advantage of the rough set approach was its ability to generate succinct decision rules while also identifying a minimal collection of the most crucial attributes, also known as 'reduce'. However, it's important to bear in mind that the relevance of these induced rules was predominantly confined to the specific bank and specialized user who contributed to the selection of the decision rules and the data sample provision. That being said, the study indicated the possibility of utilizing this method with datasets from other banks, specifically ones that have a training sample filled with decision attributes, thus showcasing its potential for wider application.

Another empirical proof of the usefulness of rough set theory for predicting business failure was presented in a separate study. The author created a model for predicting company failure by utilizing data from 200 financial reports of US companies from 1986 to 1988. A learning test of 100 organizations and a testing test of one more 100 organizations made up the example (Doering et al. 2019). The creator's skill and earlier exploration educated the determination regarding eight monetary proportions for the restrictive factors in the model. For objects that didn't definitively match the produced rules, the choice class rule strength and the esteemed closeness connection were utilized in the order cycle. For the learning test, the model was accurate at 93%, while for the testing test, it was accurate at 88%. On the same dataset, this accuracy was significantly higher than the 65% accuracy of a previous recursive partitioning method that utilized the ID3 algorithm.

Moreover, a review checked the relevance of the old style unpleasant set model hypothesis in foreseeing business disappointments. The analysis used data from 291 businesses, from 1990 to 1997. By matching non-bankrupt and bankrupt businesses based on size and revenue, great care was taken to ensure that the sample was representative. The selection of conditional variables was based on theoretical support and previous research, and the discretization process involved using percentile ranking to create ten subintervals (Gao et al. 2020). 150 businesses made up the learning sample, while 141 businesses made up the testing sample. The model accomplished an exactness going somewhere in the range of 61% and 68% on the test, utilizing the VCR-based approach and producing rules in light of two reducts. The review inferred that while the unpleasant set models didn't offer a huge relative benefit over inspectors' ongoing techniques; In scenarios involving boundary region data and the search for an ideal set of explanatory variables, they demonstrated promise as an effective method for predicting business failures.

2.2.0 Rough Set Theory

The work of Wang and Leung in 1997 marked one of the early applications of rough set theory (RST) to the field of financial forecasting. The authors aimed to examine whether RST could offer a valuable new approach for predicting stock market behavior, a question prompted by the theory's established ability to handle ambiguous, uncertain, and incomplete data. The Hong Kong stock market was chosen for the study due to its inherent volatility and complexity, presenting an interesting challenge for the rough set model.

In the research under discussion, the investigators employed the principles of rough set theory to deduce decision-making guidelines from a trove of historical financial data associated with the stock market. The model harnessed an assortment of stock market indicators as attributes, which encompassed elements such as the opening and closing valuations of the stocks, the highest and lowest recorded prices in the trading session, and the aggregate number of shares transacted. The primary goal of this research was to ascertain if the closing value of a particular stock would experience an upward or downward trajectory in the subsequent trading day.

One of the key findings from Wang and Leung's work was that the rough set model did have significant potential in the domain of stock market prediction. Not only did the model show promising accuracy in its predictions, but it also had the advantage of generating transparent, understandable decision rules (Teoh et al. 2008). This transparency contrasted with more "black box" models like neural networks, where the decision-making process can be difficult to interpret. For instance, the model might generate a rule like "If today's closing price is higher than yesterday's and today's volume is higher than the average volume of the past 10 days, then tomorrow's closing price will be higher than today's." Such rules, while not guaranteeing success in every case, provide investors with a systematic basis for making decisions.

However, Wang and Leung also highlighted some limitations to the RST approach. One challenge was that the model's effectiveness could depend on the choice and number of attributes used in the analysis. Too few attributes might oversimplify the problem, while too many could overcomplicate it and make the model less interpretable. Additionally, the model was not designed to predict the magnitude of price changes, only the direction.

Chen, Li, and Yang (2020) proposed a rough set-based feature selection method in a deep learning framework for stock prediction. The central premise of the study is the use of RST for feature selection, which is integrated within a deep learning architecture. The process of feature selection is critical to eliminate noise and redundant information from the input data (Tang et a. 2020). This can improve the accuracy of the model by focusing only on the most significant features. The authors leveraged RST for this process because of its capability to handle uncertainty, vagueness, and incompleteness in data.

The deep learning component of the model uses a type of recurrent neural network (RNN) known as Long Short-Term Memory (LSTM) cells. LSTMs are known for their ability to remember past information, which is important in stock price predictions because past stock prices are often indicative of future trends (Ang and Quek 2006). This design was chosen due to the LSTM's proven proficiency in handling sequential data like time-series stock prices.

To evaluate their proposed model, the authors used a dataset that comprised daily stock price data from several major stock indices. They compared the performance of their RST-enhanced LSTM model with a standard LSTM model and other popular machine learning models, such as Support Vector Machines (SVMs) and Decision Trees (Khoza et al. 2011). The experiments demonstrated that the LSTM model with RST-based feature selection outperformed the traditional LSTM model and other models in terms of prediction accuracy. The model demonstrated superior performance in predicting both the direction and the magnitude of stock price changes.

The following table summarizes literature reviewed.

Researcher(s)	**Year**	**Title of Article**	**Proposed Idea/Objective**	**Limitation**
Renigier-Biłozor et al.	2019	Automated valuation model based on fuzzy and rough set theory for real estate market with insufficient source data	The use of rough sets for predicting financial exchange movements and related trading strategies.	Real estate markets are complex, and the valuation models require a wide range of detailed data about various attributes such as the location of the property, the size of the property, the number of rooms, age of the property, nearby amenities, and other relevant factors..
Tang et al.	2020	Evaluation of stock investments using a decision-theoretic	The construction of predictive models for stock market	Generalizability of the results

		rough set model with q-rung orthopair fuzzy information	movements, financial risk management, and portfolio management using fuzzy rough sets and rough sets.	
Pal and Kar	2019	Data discretization using fuzzistics and rule generation using rough set theory are used in time series forecasting for stock market prediction.	Rough set theory applied to financial time series forecasting using numerical attribute conversion.	Difficulty in creating suitable discretization algorithms for stock market time series.
Takahashi et al.	2019	Using generative adversarial networks to model financial time data.	Rough set theory is used in financial trading, notably in the crude oil futures market.	The robustness and generalization capabilities of the derived trading rules to new, unseen data were not thoroughly examined in the text.
Zhang et al	2019	A discussion of multi-source information fusion based on rough set theory	The application of the classical rough set model to the prediction of stock market time series.	Potential bias due to the relatively short time span of the study and the infrequency of the observed period.
Golan	1995	Rough set theory is used in stock market analysis.	The proposal of a novel procedure combining decision trees and basic sets for forecasting stock market moves.	Data used was limited both in volume and diversity compared to what is available today.

Wang	2003	The mining stock price is calculated using a fuzzy rough set technique	The suggestion of combining self-organizing maps (SOM), rough sets and clustering for financial time series predictive analysis.	The system's performance varied depending on the tuning parameters.
Gandhmal and Kumar	2019	Systematic examination and evaluation of stock market forecasting methodologies	In the context of Greek firms, a comparison of rough set models with discriminant analysis and logit models.	The relevance of the induced rules was predominantly confined to the specific bank and specialized user who contributed to the data sample.
Doering et al.	2019	Current status and potential trends in metaheuristics for rich portfolio optimization and risk management.	The application of rough set theory in predicting business failure using data from US companies.	The results might not be applicable or accurate for companies based in other countries due to various factors like different regulatory environments. This limits the external validity of the study.
Singh et al.	2023	Housing Loan Selection Criteria Based on Dominance-Based Rough Set Theory: An Indian Case	The efficiency of the rough set model in compared to discriminant analysis and logit models.	Discriminant analysis assumes normality and equal covariance matrices among groups. If these assumptions are not met, the results might be misleading.
Wang and Leung	1997	The mining stock price is calculated	Early application of rough set theory	The model's effectiveness could depend on the choice

		using a fuzzy rough set technique.	(RST) to financial forecasting, focusing on predicting stock market behavior.	and number of attributes used in the analysis. It was not designed to predict magnitude.
Chen, Li, and Yang	2020	Probabilistic fuzzy time series model approach and rough set rule induction for empirical research in stock markets.	A rough set-based feature selection method in a deep learning framework for stock prediction.	Risk of overfitting the model to that data. Overfitting means the model may predict the training data very well but fails to generalize to new, unseen data.

2.2.1 Research Gap

The studies discussed above have unveiled several research gaps. Firstly, despite the rough set model demonstrating superior predictability in comparison to discriminant analysis, there is still a need to explore alternative machine learning algorithms. The power of the rough set model and logit model was found to be comparable, prompting the question of whether other techniques like support vector machines or random forests could potentially surpass the rough set model in predicting stock market movements. Conducting further investigations into these alternative algorithms would yield valuable insights into their suitability and effectiveness within this context.

Another research gap lies in the external validity of the findings. Since the study specifically focused on Greek companies, it remains uncertain whether the results can be generalized to companies operating in different industries or geographical regions. To address this limitation, future research endeavors should replicate the study using data from companies in diverse global markets (Gandhmal and Kumar 2019). This broader scope would enable a more comprehensive assessment of the rough set model's effectiveness across various contexts. In the time series analysis for stock market prediction, there are significant research gaps that warrant further exploration. Conducting such comparative studies would provide a comprehensive understanding of the rough set model's relative strengths and weaknesses in relation to these established approaches. Moreover, the studies primarily relied on historical financial data and technical indicators as input variables. Investigating the impact of integrating these external factors into the rough set model would contribute to a more holistic understanding of its capabilities and potential improvements.

Another research gap pertains to the evaluation of the models' performance under different market conditions. The studies predominantly focused on periods of normal market behavior, neglecting the assessment of the rough set model's performance during times of market volatility or significant economic events. Analyzing the model's stability and resilience in varying market scenarios would provide valuable insights into its practical viability and effectiveness.

Lastly, the studies predominantly utilized financial data from the past few decades, omitting crucial long-term historical data that includes periods of economic downturns and recoveries. Incorporating such extended historical data would facilitate a more comprehensive evaluation of the rough set model's performance across different market cycles (Shen and Shafiq 2020). This analysis would provide insights into the model's adaptability and generalizability over extended periods, thus enhancing its practical utility.

2.3.0 Analysis of Problem/Improvement

Stock market prediction has long captivated the attention of financial researchers, investors, and technologists alike. In an era defined by data abundance and technological advancements, the quest for accurate forecasts in the ever-shifting sands of financial markets is more pertinent than ever. The financial landscape, characterized by its intricacies, unpredictable dynamics, and global interconnectedness, presents a formidable backdrop against which this problem unfolds. The very essence of financial markets, as articulated by Fama's (1970) Efficient Market Hypothesis (EMH), initially suggests an aura of impenetrable efficiency, wherein asset prices swiftly incorporate all available information. Yet, the reality, as elucidated by subsequent research from Malkiel (2003) and Lo (2004), paints a more nuanced picture, one where anomalies and exploitable patterns occasionally surface. It is this complexity, compounded by the profound financial consequences of even minor inaccuracies in predictions, that underscores the paramount importance of this research problem. In this pursuit, Rough Set Theory stands as both an intriguing mathematical framework, courtesy of Pawlak's pioneering work in 1982, and a formidable technical challenge, demanding a harmonious marriage of financial acumen and data analytical prowess. This analysis navigates through the uncharted waters of financial technology, touching upon the critical dimensions of prediction accuracy, model development, comparative analysis, user experience, and ethical considerations. The successful resolution of these intertwined challenges has the potential to usher in a transformative era in financial decision-making, armed with data-driven insights and an unwavering commitment to ethical standards.

2.3.1 Complexity of Financial Markets

The financial markets, a crucible of global economic activity, present a multifaceted landscape where traditional notions of efficiency, as outlined by Fama's (1970) Efficient Market Hypothesis (EMH), meet the realities of complexity and unpredictability. While the EMH suggests that asset prices instantaneously incorporate all available information, contemporary financial research, as articulated by Malkiel (2003) and Lo (2004), has unveiled a more intricate narrative. These scholars have meticulously dissected market anomalies and patterns that challenge the EMH's claim of market efficiency. Their research illuminates windows of opportunity for predictive modeling, where behavioral biases, shifts in market sentiment, and unexpected external shocks disrupt the presumed equilibrium of financial markets.

In addition to behavioral factors, financial markets are entangled in a web of economic, political, and global influences. Economic indicators, monetary policy decisions, geopolitical events, natural disasters, and broader global economic trends serve as a complex tapestry of variables that permeate financial markets. These factors, as highlighted by contemporary research (Zhang *et al.*, 2021), intermingle in intricate and often unforeseeable ways, amplifying the convoluted dynamics of financial markets.

The modern financial landscape is further marked by an unprecedented degree of global interconnectedness. Information and capital flow seamlessly across borders, rendering financial markets increasingly interdependent. A development in one region can cascade through global markets, a phenomenon often referred to as "contagion" (Forbes & Rigobon, 2002). In this interconnected environment, predicting precise market movements becomes an intricate puzzle, where seemingly unrelated events can trigger far-reaching consequences.

2.3.2 Model Development and Rough Set Theory

Model creation, particularly in the context of financial markets, necessitates a complex strategy that can adequately capture the nuances of market dynamics. The Efficient Market Hypothesis (EMH) and other traditional models have dominated financial research, however recent researchers (Lo, 2004; Malkiel, 2003) have questioned this assumption. Given this situation, Rough Set Theory stands out as a strong contender for modeling financial data because it can deal with uncertainty and imprecision (Pawlak, 1982).

Developed by Pawlak in 1982, rough set theory offers a mathematical framework for coping with data that may be erroneous or incomplete, a problem that regularly arises in financial markets. It makes it possible to identify decision rules and represent complex interactions. Although Rough Set Theory hasn't been used much in financial contexts, its potential is starting to be recognized (Chen et al., 2017).

Rough Set Theory can be tailored to capture financial market dynamics, according to recent studies (Kurzynski et al., 2019; Pawlak & Skowron, 2007). This model building makes use of rough set theory to find hidden patterns and linkages in financial data that may evade traditional models. Rough Set Theory is a useful technique for stock market forecasting because it can analyse massive amounts of financial data with variable degrees of precision.

However, successful model development within financial markets necessitates more than just theoretical foundations. It requires an integration of financial acumen and data analytical prowess, as demonstrated in research by Hastie et al. (2009) and Hyndman and Athanasopoulos (2018). These scholars emphasize the significance of domain knowledge in crafting predictive models that can effectively navigate the intricacies of financial markets. Therefore, marrying Rough Set Theory's mathematical capabilities with deep financial insights becomes pivotal in achieving prediction accuracy and relevance.

2.3.3 Prediction Accuracy

In the realm of stock market prediction, accuracy stands as a paramount criterion (Liu et al., 2020). The very essence of financial markets lies in their propensity to render significant financial consequences based on even minor inaccuracies in predictions. As emphasized by Granger and Ramanathan (1984), accurate predictions are indispensable for effective financial decision-making. This holds especially true in today's interconnected global markets, where financial actions and their repercussions resonate across borders (Forbes & Rigobon, 2002). Prediction models must therefore meet the challenge by offering trustworthy perceptions into market developments.

Recent studies have highlighted the value of prediction accuracy in financial modeling. Studies by Liu et al. (2020) emphasize the value of accuracy in stock price forecasts, particularly in light of the substantial risks associated with financial investments. Prediction accuracy is a crucial component of any model's effectiveness because even little changes in predictions can have significant financial repercussions.

Additionally, the development of sophisticated prediction models is made possible by the modern financial environment's growing data accessibility and computational power (Li et al., 2020; Menkveld & Yueshen, 2019). These models, especially those built on Rough Set Theory, are anticipated to give traders and investors information that are not only accurate but also useful. Models are expected to perform better than conventional methods in terms of accuracy and adaptability in the age of big data and machine learning (Nayani, Rao and Lakshmi, 2023).

CHAPTER 3

3.1 Research Methods

Onion Model of research methodology is used to methodically peel back the layers of research decisions, data collection, and analytic procedures, much like peeling back the layers of an onion reveals its inner core. This methodological technique is supported by the positivist research philosophy, which emphasizes the importance of empirical data and quantitative analysis in answering the research questions. The methodological framework underlying this study is clarified in this section as we delve into the various levels of the research onion, from the overarching research philosophy to the most in-depth data analysis methodologies. This research aims to give a thorough and rigorous inquiry into the construction and evaluation of a Rough Set Theory-based stock market prediction application by methodically designing each layer.

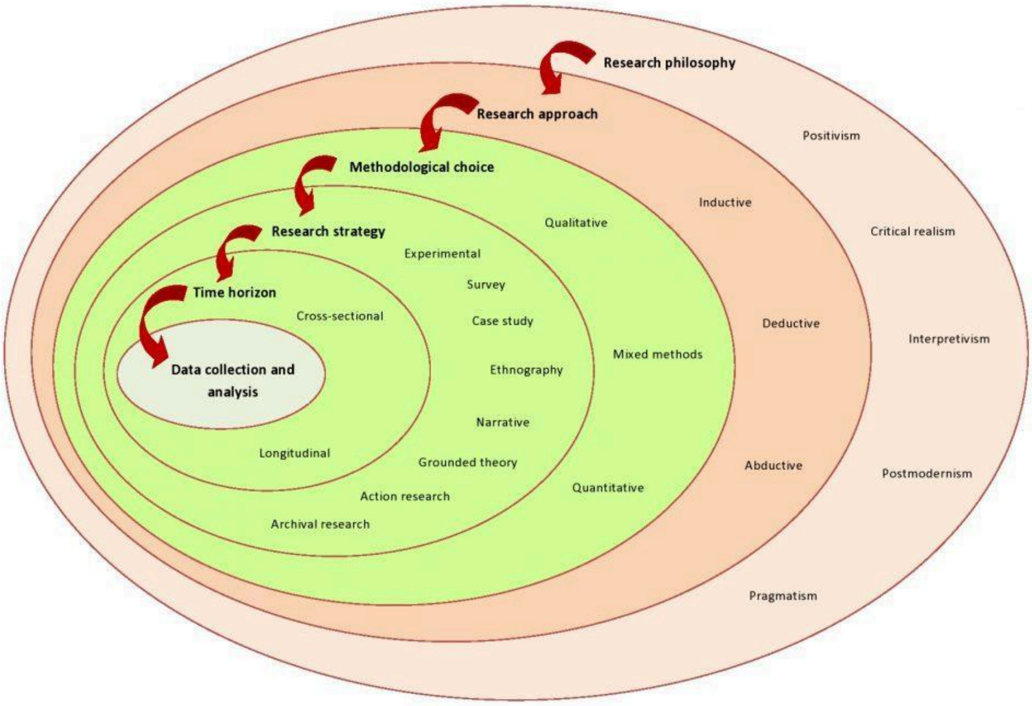


Figure 1- Research onions - (Saunders et al. 2012)

3.2 Research Philosophy

This study's methodology is based on the positivist school of thought. According to Bryman and Bell (2015), positivism is distinguished by its dedication to empirical observation, objectivity, and the utilization of numerical data. In the context of this work, designing a predictive software application for stock market forecasting and carefully assessing its accuracy are goals that are consistent with a positivist mindset.

According to Remenyi et al. (1998), positivism places a strong emphasis on the methodical gathering of empirical evidence to draw conclusions and test theories. It is ideal for this research since it makes it possible to quantitatively analyze historical stock market data and determine the effectiveness of the predictive model in an unbiased manner. This study, which adheres to a positivist ideology, attempts to advance empirical knowledge of the use of Rough Set Theory in the context of stock market forecasting.

The positivist school of thought emphasizes the significance of objectivity in research and encourages scientists to avoid subjectivity and prejudice in the gathering and evaluation of data (Saunders et al., 2018). When assessing the precision of prediction models in the fast-paced, high-stakes environment of financial markets, this dedication to objectivity is essential. To assure the validity and dependability of its findings, the research adopts a structured and systematic approach that is informed by positivist ideas.

3.3 Research Approach

This study uses a deductive research approach, which is consistent with positivism's general research philosophy. Testing hypotheses that are based on accepted ideas and concepts defines the deductive approach (Bryman & Bell, 2015). It starts with theoretical claims and looks for empirical data to support or contradict them.

The decision to use a deductive approach is appropriate given the goal of the study, which is to create and assess a predictive software program for stock market forecasting. This research aims to systematically evaluate the performance of the Rough Set Theory-based model by constructing hypotheses based on established principles of Rough Set Theory and existing machine learning prediction models.

A organized, hypothesis-driven inquiry into the research issues is also made possible by the deductive approach, which also offers a rational way to assess the efficacy and performance of the predictive model. It makes it easier to test out specific hypotheses about how well the application predicts stock market trends.

The deductive method emphasizes the significance of meticulous data analysis and the impartial assessment of the model's effectiveness. By using this strategy, the study makes sure that the assessment of the predictive model remains supported by empirical data and adds to the body of knowledge on stock market forecasting.

3.4 Research Strategy

The development and assessment of a stock market forecasting software program based on rough set theory are investigated in this work using a quantitative research methodology. The emphasis on the gathering and analysis of numerical data to test hypotheses and develop statistical insights is what distinguishes a quantitative research method (Bryman & Bell, 2015).

For the purposes of this study, which include evaluating the predictive model's accuracy and contrasting it with other machine learning prediction models, quantitative research is well suited. The research's objective is to offer exact and quantified insights into the Rough Set Theory-based model's forecasting powers by using quantitative data analysis approaches.

The quantitative research technique also fits with the requirement for systematic, objective data gathering and analysis, which is essential in the field of financial markets where precision and accuracy are key. It allows for the objective assessment of the model's performance through the analysis of numerical data, such as historical stock prices, trading volumes, and performance metrics.

The quantitative research approach additionally makes it easier to apply statistical tests and regression modeling to evaluate the model's predicted efficacy. It makes it possible to produce empirical data that can be used to support claims about how well Rough Set Theory predicts the stock market.

3.5.0 Research Choices

This study makes particular research decisions relating to the type of data gathered and the temporal span of analysis within the broader framework of the quantitative research strategy. The selected research options are:

Data Collection

The majority of the data came from reputable data collection sites like Kaggle, which offer accurate historical stock prices, trade volumes, and economic indicators. The prediction model was assessed using this data as the foundation. To measure engagement and usability, user interaction data from the mobile app, including stock selections, user preferences, and query data, was also gathered.

Cross-Sectional Data Collection

The gathering of cross-sectional data is one of the research options. Cross-sectional data provide a moment in time snapshot of a certain phenomenon (Tran and Huh, 2022). Cross-sectional data collection in the context of this study entails obtaining information at particular time intervals on stock market factors, such as stock prices, trading volumes, and economic indicators. These facts provide information about the stock market's condition at various points in time.

It is possible to examine the state of the stock market and the model's level of predictability at specific points in time by using cross-sectional data. It makes it easier to evaluate how well the model captures short-term changes in stock prices and market dynamics.

Data Analysis: Quantitative

Data analysis in this study involves a quantitative approach, including descriptive statistics to summarize historical stock data in a very human-readable form. Specialized software such Python is employed for robust analysis. This comprehensive approach aims to provide empirical evidence regarding the Rough Set Theory-based predictive software's effectiveness in forecasting stock market trends.

CHAPTER 4

4.0 Design of Artefact

Our exploration now takes us deep into the intricacies of the predictive software application, where we uncover the architectural decisions, coding methodologies, and user interface intricacies that collectively shape its form and functionality. In this chapter, we navigate the technical terrain of this innovative tool, dissecting the core components that underlie its predictive capabilities. We shine a spotlight on the careful fusion of technology and user-centric design principles that define the software's essence, creating a holistic view of how it transforms abstract concepts into a tangible, user-friendly application.

The design of the artefact was in two phases: first is the designing of the mobile app protype so we can see how the final application would look like. The second phase is the outlining of the steps that would be carried out to make the prediction using the Rough Set Theory.

The first step of the design of the artefact includes the design of the user-friendly mobile application. This will give us an overview of how the final application would look like. The design of the app has discussed earlier would be built using the cross-platform technology: Flutter/Dart.

We would outline the packages or libraries that would be used in building this below.

1. Flutter_lints: This package allows us to adhere to strict coding practices while giving us the flexibility we need to build a complex application.
2. Fl_chart: This package enables us to create stunning graph-like charts that are easy to understand and can be used by anyone.

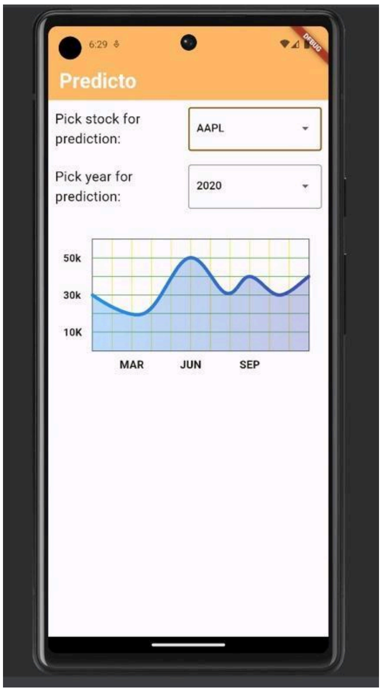


Figure 2 - User-interface of the mobile application

Hardcoded data has been used initially for the sake of quickly bootstrapping the design so a prototype of the final application can be built. These hardcoded data include a subset of the stocks in the market like: AAPL, TSLA, BMY, GOOGL, etc.

Arbitrary years for view the prices of the stocks as well as the predictions were used: 2020, 2023, 2024, 2028, etc.

4.1 Architectural Design

4.1.1 Client-Side Components

On the client side, the application has been painstakingly developed using the Flutter framework, which has highly developed features that are compatible with a variety of different platforms. Dart, the programming language that was ultimately chosen, is an absolute requirement for the application if the developers want to ensure that it is both responsive and quick. A just-in-time (JIT) compiler is one of the features that distinguishes Dart from other programming languages. This is one of the things that makes Dart apart. During the process of software development, the use of this kind of compiler enables rapid iterations and code updates in real time. The deft use of a large number of Flutter widgets speeds up the process of building a user-friendly interface while also reducing the amount of work involved in the process. The FlatButton and DropDownButton and AutoComplete components stand out as being very effective for improving the user experience by simplifying the processes of stock selection and interaction. Of all of them, these two components stand out as being exceptionally beneficial.

4.2 Technical Specifications

4.2.1 Dart Programming Language

The application was built with the help of the programming language known as Dart, which is renowned for its focus on practicality, its reliable typing system, and its cutting-edge object-oriented structure. Dart offers a robust foundation that makes it possible to write code that is both readable and easy to maintain. The just-in-time (JIT) compiler that this platform use allows developers the ability to edit code in real time while yet maintaining the integrity of the codebase across the various rounds of development. The comprehensive type system that Dart use helps to prevent errors, which in turn increases the dependability and efficiency of the program as a whole.

4.2.2 fl_chart Library

The fl_chart library is a complex and feature-rich tool that was built for the purpose of creating dynamic and interactive charts. It is prominently featured in the user interface of the application. This library provides a full range of charting tools, which makes it possible to see stock market forecasts in an organized and straightforward manner. The adaptability and responsiveness of fl_chart make it possible for a great degree of customisation to be applied to chart design. This ensures that users receive an interesting and informative representation of data in visual form.

4.2.3 Data Processing Capabilities

The application's data processing skills, in particular those pertaining to the interpretation of prediction data that is saved in JSON format, are essential to the basic functionality of the application. The powerful data processing capabilities of Dart are leveraged in order to handle this work in an effective manner. The language's built-in data manipulation facilities and capabilities for parsing JSON ensure that stock market forecasts are presented in a manner that is both accurate and efficient. This not only improves the application's performance but also adds to its general reliability, which is a significant benefit.

4.2.4 Widget-Based User Interface

The user interface (UI) of the program has been painstakingly constructed utilizing the widget-based design approach of Flutter, which is well-known for its adaptability and user-friendliness. The user interface (UI) is enhanced with a wide variety of interactive and user-friendly Flutter widgets, including but not limited to DropDownButton and FlatButton components. These widgets are used in a strategic manner to achieve the desired effect. The user experience is improved as a result of the implementation of these widgets, and the development process is sped up as a result of the provision of pre-built user interface components that comply with contemporary design standards. The user interface of the program is dynamic, responsive, and designed to provide users with an experience that is engaging and easy to understand.

4.2.5 Stock Selection

The user interface includes a mechanism for selecting stocks that is powered by Flutter widgets, specifically the AutoComplete components. These widgets guarantee a smooth and fast selection process and make it possible for users to choose from a comprehensive list of financial instruments in an easy and uncomplicated manner. Behind the scenes, the data retrieval process is tuned to reduce latency as much as possible, which ensures a brisk and effective stock selection experience for the user.

4.2.6 Year Selection for Prediction

Users are empowered to specify the precise target year for which stock market predictions are sought. This pivotal feature seamlessly integrates into the UI, offering users a granular level of control. Flutter's interactive widgets facilitate user input, enabling swift and precise specification of the desired prediction timeframe.

4.2.7 Interactive Graphical Representation

At the core of the UI lies an interactive graphical representation of stock trends, harnessed through the robust fl_chart library. This library's versatility shines in creating dynamic and visually captivating charts. The line graph elegantly showcases historical and predicted stock prices, permitting users to delve deep into temporal trends. The interactive capabilities of the graph extend to touch gestures, providing users with the freedom to zoom, pan, and scrutinize specific data points. This immersive graph empowers users with a comprehensive view of stock behavior, reinforcing their decision-making process.

The next phase of the artefact design is outlining the steps that must be followed to generate the prediction on the stock market dataset.

The dataset chosen for this is ‘sp500’ which includes 500 companies stock data. It includes various companies like Apple, Amazon etc. First the dataset is loaded and converted into separate datasets for each company just for the easiness and liberty for the ‘Rough Set Theory’ to predict the future data and not to add much load on the memory. Below are the main steps taken for the rough set theory prediction:

1. **Data Preprocessing**:

   Data preprocessing serves as the fundamental basis for rough set theory. This guarantees the integrity of the data by ensuring its consistency and absence of missing values. The presence of complete data is of utmost importance in rough set theory, as it guarantees the precise construction of the discernibility matrix, which is a basic notion in this field. By imputing missing values and addressing infinite values, we are maintaining the data integrity of our dataset in preparation for subsequent analysis.

2. **Generating Target Dates**:

   This establishes the temporal range within which our forecasts are made. The granularity of our decision system is established, guaranteeing that our analysis is congruent with time intervals of significance.

3. **Upper and Lower Limit Calculation**:

The concepts of lower and upper approximations play a crucial role in rough set theory. The through the process of establishing the upper and lower boundaries effectively approximates these ideas. The boundary region, denoting the disparity between the upper and lower approximations, signifies the inherent uncertainty within our dataset. Through the process of calculating these limitations, we are effectively quantifying the degree of "roughness" or inherent uncertainty present within our dataset.

4. **Predictions Generation**:

The phase of decision rule creation in rough set theory. Probabilistic decision rules are being formulated by utilising the boundary regions and approximations. The rules, which are generated from the intrinsic patterns observed in the data, are subsequently employed to forecast future data points, so demonstrating the predictive efficacy of rough set theory.

5. **Sample Data Processing and Predictions**:

This segment provides a practical illustration of the analysis based on rough sets. The given scenario serves as a representative model of the complete procedure, wherein it is implemented on a subset of data, enabling us to observe the profound impact of rough set theory in practical application.

6. **Final Predictions and Output**:

The ultimate outcome of our investigation is the production of practical and implementable ideas. The predictions, which are based on the decision rules and approximations, serve as evidence of the effectiveness of rough set theory. Subsequently, these data are systematically arranged, so facilitating subsequent examination, visualisation, or decision-making processes.

4.3 Implementation testing & validation of Artefact

The journey from concept to reality is navigated through the phases of implementation, testing, and validation. In this pivotal stage, the predictive software application takes shape, driven by the technical acumen of development, the scrutiny of testing, and the confirmation of validation. Implementation sees the application's architecture, user interface, and data handling capabilities come to life. The rigorous testing phase verifies the application's reliability, performance, and adherence to design principles. Validation, on the other hand, stands as the litmus test for the predictive model, measuring its accuracy against historical data and benchmarking it against established machine learning models. Through an iterative process of refinement, this triad of phases ensures the application's readiness to empower users with accurate stock market forecasts while offering a seamless and engaging user experience.

4.3.1 Implementation

The implementation phase of our predictive software application is a meticulous process that bridges the conceptual design with real-world functionality. Leveraging the versatile Flutter framework in conjunction with the Dart programming language, the application was brought to life.

Technical Components

The technical components of the application encompass a multitude of features that have been diligently crafted. Stock selection is facilitated through the utilization of Flutter's DropDownButton widgets, affording users a seamless and intuitive means to choose from a diverse spectrum of financial instruments. Year specification, a critical facet for accurate predictions, is seaml essly integrated into the user interface (UI), empowering users to precisely define the temporal scope of their forecasts. At the core of the UI lies an interactive and dynamic graphical representation of stock trends. This functionality is realized through the integration of the fl_chart library, known for its versatility and responsiveness in creating visually engaging charts. Complementing this graphical element, the application furnishes real-time predictive indicators. These indicators offer concise insights into whether the anticipated stock prices are poised for an upswing or decline, augmenting the user experience and decision-making process.

Data Integration

The current data integration method involves manual insertion through JSON files. Historical stock market data is structured and processed to support the predictive modeling engine. While this manual data insertion is effective for the current iteration, the architecture remains future-ready for automated data collection mechanisms, ensuring adaptability and scalability as data requirements evolve.

The implementation of the stock market prediction using the Python programming language is outlined below. This implementation is what generates the final CSV file which was converted to a JSON file to be included in the mobile application.

1. We add the dataset into the work environment.

```
[ ] sp500_data = pd.read_csv('dataset/sp500.csv')

    sp500_data.head()

    C:\Users\muham\AppData\Local\Temp\ipykernel_18096\2360185141.py:1: DtypeWarning: Columns (1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,4
      sp500_data = pd.read_csv('dataset/sp500.csv')
```

	Date	Close	Close.1	Close.2	Close.3	Close.4	Close.5	Close.6	Close.7	Close.8	...	Volume.493	Volume.494	Volume.495	Volume.496	Volume.497	Volume.498	Volume.499	Volume.500	Volume.5
0	NaN	A	AAL	AAP	AAPL	ABBV	ABC	ABT	ACGL	ACN	...	WYNN	XEL	XOM	XRAY	XYL	YUM	ZBH	ZBRA	ZIC
1	04/01/2010	20.22970963	4.49687624	36.70385742	6.496294022	NaN	21.76925087	19.33639908	7.994443893	32.70815277	...	4741400	2670400	27809100	1051400	NaN	2962274	805872	168800	39746
2	05/01/2010	20.00995445	5.00595665	36.4856987	6.507524967	NaN	21.61393547	19.18017197	7.967778206	32.91030502	...	5644300	4321400	30174700	763400	NaN	3298757	1769643	168800	56055
3	06/01/2010	19.93887138	4.798553467	36.80384827	6.404016018	NaN	21.40957069	19.28668976	7.93333292	33.26016235	...	2738800	2164500	35044700	1595100	NaN	4178981	1315619	385300	126152
4	07/01/2010	19.91301727	4.939965725	36.79476929	6.392177105	NaN	21.06622505	19.44647408	7.886666775	33.22906113	...	2388500	3041700	27192100	1096100	NaN	2452472	1734005	183600	247168

5 rows × 2516 columns

Figure 3 – adding dataset to our work environment

2. When that has been done we make seperate datasets for the columns becuase they represent each companies and they are best in ordered form.

```
[ ] ticker_symbols = sp500_data.iloc[0, 1::2].values

    output_directory = "companies_data"
    os.makedirs(output_directory, exist_ok=True)
```

```
[ ] for idx, ticker in enumerate(ticker_symbols):
        close_col = 'Close' if idx == 0 else f'Close.{idx}'
        volume_col = 'Volume' if idx == 0 else f'Volume.{idx}'

        if close_col in sp500_data.columns and volume_col in sp500_data.columns:
            company_data = sp500_data[['Date', close_col, volume_col]].copy()
            company_data.columns = ['Date', 'Close', 'Volume']
            company_data = company_data.iloc[1:].reset_index(drop=True)

            company_data.to_csv(os.path.join(output_directory, f'{ticker}.csv'), index=False)

    output_directory

    'companies_data'
```

Figure 4 - separating dataset

3. We plot the Apple stock dataset to see that everything is alright.

```
[ ] aapl_data = pd.read_csv(os.path.join(output_directory, 'AAPL.csv'))
    aapl_data['Date'] = pd.to_datetime(aapl_data['Date'])

    C:\Users\muham\AppData\Local\Temp\ipykernel_18096\1179616849.py:2: UserWarning: Parsing dates in DD/MM/YYYY format when dayfirst=False (the default) was specified. This may lead to
      aapl_data['Date'] = pd.to_datetime(aapl_data['Date'])

[ ] aapl_data.head()
```

	Date	Close	Volume
0	2010-04-01	28.915880	1876000
1	2010-05-01	28.994598	2186900
2	2010-06-01	28.819681	1147200
3	2010-07-01	28.933376	1272400
4	2010-08-01	29.773031	2068200

Figure 5 - Checking that dataset is outputing expected data

4. Next, we iterate over all the data in the dataset to see their basic statistics like Mean, Median, Average close, etc

```python
for file in all_files:
    if file.endswith('.csv'):
        filepath = os.path.join(output_directory, file)
        company_data = pd.read_csv(filepath)
        ticker = file[:-4]
        close_stats = company_data['Close'].describe()
        volume_stats = company_data['Volume'].describe()

        stats = {
            'Company': ticker,
            'Average_Close': close_stats['mean'],
            'Median_Close': close_stats['50%'],
            'Min_Close': close_stats['min'],
            'Max_Close': close_stats['max'],
            'Average_Volume': volume_stats['mean'],
            'Median_Volume': volume_stats['50%'],
            'Min_Volume': volume_stats['min'],
            'Max_Volume': volume_stats['max']
        }

        all_stats.append(stats)
```

```python
summary_df = pd.DataFrame(all_stats)
```

```python
summary_df.head()
```

	Company	Average_Close	Median_Close	Min_Close	Max_Close	Average_Volume	Median_Volume	Min_Volume	Max_Volume
0	A	62.827543	43.808224	17.431160	177.025680	2.971345e+06	2310100.0	271900.0	25368667.0
1	AAL	67.310050	54.425762	29.847603	146.694687	1.018918e+06	884700.0	99900.0	9764800.0
2	AAP	24.860057	20.910000	3.770965	56.988728	1.776176e+07	8997900.0	1158400.0	428617100.0
3	AAPL	93.672539	80.877342	26.493113	215.566635	9.298716e+05	794700.0	151900.0	19062300.0
4	ABBV	120.309316	130.224701	35.594917	230.368683	1.084406e+06	894500.0	130700.0	21832300.0

Figure 6 - Print out statistics on data

5. Next, we take the Apple stock data and perform the Rough Set Theory operation on it. This will be the test data, our internal validation.

```python
apple_data = pd.read_csv(os.path.join('companies_data', "AAPL.csv"))
apple_data.head()
```

	Date	Close	Volume
0	04/01/2010	28.915880	1876000
1	05/01/2010	28.994598	2186900
2	06/01/2010	28.819681	1147200
3	07/01/2010	28.933376	1272400
4	08/01/2010	29.773031	2068200

```python
n_bins = 10
discretizer = KBinsDiscretizer(n_bins=n_bins, encode='ordinal', strategy='quantile')
apple_data[['Close', 'Volume']] = discretizer.fit_transform(apple_data[['Close', 'Volume']])

apple_data.head()
```

	Date	Close	Volume
0	04/01/2010	0.0	9.0
1	05/01/2010	0.0	9.0
2	06/01/2010	0.0	7.0
3	07/01/2010	0.0	8.0
4	08/01/2010	0.0	9.0

```python
def create_lagged_features(df, lag_count=3):
    df_lagged = df.copy()
    for i in range(1, lag_count + 1):
        df_lagged[f"Close_Lag_{i}"] = df_lagged["Close"].shift(i)
        df_lagged[f"Volume_Lag_{i}"] = df_lagged["Volume"].shift(i)
    return df_lagged.dropna().reset_index(drop=True)
```

```python
lag_count = 3
apple_data_lagged = create_lagged_features(apple_data, lag_count)
apple_data_lagged.head()
```

	Date	Close	Volume	Close_Lag_1	Volume_Lag_1	Close_Lag_2	Volume_Lag_2	Close_Lag_3	Volume_Lag_3
0	07/01/2010	0.0	8.0	0.0	7.0	0.0	9.0	0.0	9.0
1	08/01/2010	0.0	9.0	0.0	8.0	0.0	7.0	0.0	9.0
2	11/01/2010	0.0	9.0	0.0	9.0	0.0	8.0	0.0	7.0
3	12/01/2010	0.0	8.0	0.0	9.0	0.0	9.0	0.0	8.0
4	13/01/2010	0.0	8.0	0.0	8.0	0.0	9.0	0.0	9.0

Figure 7 - Performing Rough Set Theory operation on test data

6. We then generate the rules which would be used in the prediction. This is an important step in utilizing the Rough Set Theory.

```
[ ] def get_rules(data, n_bins):
        limits = pd.DataFrame(0, index=range(n_bins), columns=range(n_bins))

        for i in range(1, len(data)):
            prev_bin = int(data.iloc[i-1])
            curr_bin = int(data.iloc[i])
            limits.at[prev_bin, curr_bin] += 1


        limits = limits.div(limits.sum(axis=1), axis=0)
        return limits.fillna(0)


    limits = get_rules(apple_data['Close'], n_bins)

    limits
```

	0	1	2	3	4	5	6	7	8	9
0	0.964706	0.035294	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	0.032353	0.950000	0.017647	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.014706	0.905882	0.079412	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.076471	0.894118	0.026471	0.002941	0.000000	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.000000	0.023599	0.941003	0.035398	0.000000	0.000000	0.000000	0.000000
5	0.000000	0.000000	0.000000	0.002941	0.032353	0.908824	0.055882	0.000000	0.000000	0.000000
6	0.000000	0.000000	0.000000	0.000000	0.000000	0.052941	0.902941	0.044118	0.000000	0.000000
7	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.041176	0.935294	0.023529	0.000000
8	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.020588	0.876471	0.102941
9	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.100295	0.899705

Figure 8 - Generating Rough Set Rules

7. Finally, the make the prediction.

```
[ ] def predict_next_bin(current_bin, limits):
        return np.random.choice(limits.columns, p=limits.loc[current_bin].values)

    def generate_predictions(initial_bin, limits, n_days):
        predictions = [initial_bin]
        for _ in range(n_days):
            next_bin = predict_next_bin(predictions[-1], limits)
            predictions.append(next_bin)
        return predictions


    n_days = 252 * 5


    initial_bin = int(apple_data['Close'].iloc[-1])
    predicted_bins = generate_predictions(initial_bin, limits, n_days)

    predicted_bins[:10]

    [9, 9, 9, 9, 8, 8, 9, 9, 9, 9]
```

```
[ ] bin_edges = discretizer.bin_edges_[0]
    predicted_ranges = [(bin_edges[int(bin)], bin_edges[int(bin) + 1]) for bin in predicted_bins[:10]]

    predicted_ranges

    [(170.7268402, 215.5666351),
     (170.7268402, 215.5666351),
     (170.7268402, 215.5666351),
     (170.7268402, 215.5666351),
     (132.30180666, 170.7268402),
     (132.30180666, 170.7268402),
     (170.7268402, 215.5666351),
     (170.7268402, 215.5666351),
     (170.7268402, 215.5666351),
     (170.7268402, 215.5666351)]
```

Figure 9 - Making the prediction

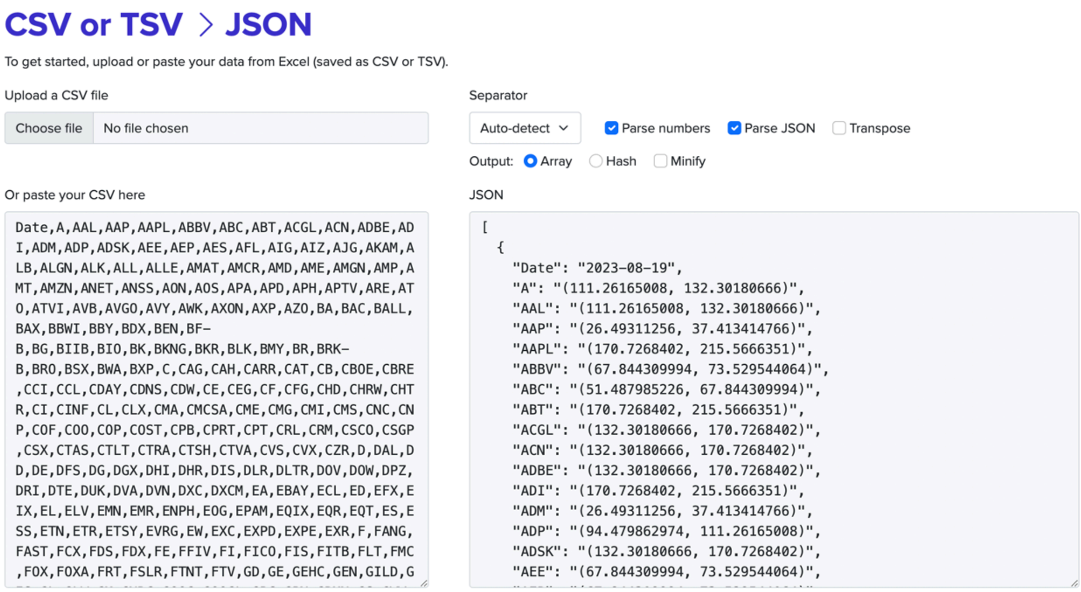
The prediction gives us a lower and upper band within which our prediction lies.

8. We now perform the same Rough Set prediction operation on our whole dataset which represents our real dataset and outputting the result into a predicted_forecast.csv file.

```
[ ]
df_all_predictions_final = pd.concat([process_company_data_final(os.path.join('companies_data', file), target_dates) for file in data_files])

[ ]
df_pivot = df_all_predictions_final.pivot(index='Date', columns='Company', values='Predicted_Price_Range')
df_pivot = df_pivot.reset_index()
output_pivot_file_path = "predicted_forecast.csv"
df_pivot.to_csv(output_pivot_file_path, index=False)

[ ] output_pivot_file_path

'predicted_forecast.csv'

[ ]
```

Figure 10 - Performing prediction on our entire dataset.

This predicted_forecast.csv file was then converted to a JSON object which was included in the application to allow users to rapidly view the predictions for any stock of their choosing.



4.3.2 Testing

The testing phase within our development lifecycle is an exhaustive and meticulous process designed to ascertain the reliability, functionality, and performance of our predictive software application. This multifaceted approach encompasses various testing methodologies to ensure a robust and dependable end product.

Unit Testing

The testing process commences with unit testing, which involves a meticulous analysis of individual code components. Every function, method, and class undergoes a series of test cases to verify its precision and functionality. Unit tests play a crucial role in ensuring the quality of software applications by verifying that the essential components align with their intended design criteria.

```dart
void main() {

  TestWidgetsFlutterBinding.ensureInitialized();

  test('Test that the predicted json file from assets IS NOT EMPTY', () async{

    List<PredictedStocks> assets = await loadJsonFromAssets();

    expect(assets.length, isNot(0));

});

}
```

Figure 11 - This verifies that the JSON file is not empty

Integration Testing

In addition to unit testing, integration testing examines the cohesive interaction across distinct modules and components inside the program. This stage evaluates the exchange of data, communication, and synchronization between the user interface, data processing layers. The identification and resolution of any bottlenecks, conflicts, or inconsistencies that may arise from the integration of components play a crucial role in this process.

Performance Testing

Performance testing is a rigorous examination of the application's behavior under varying conditions. It assesses response times, resource utilization, and system stability across different usage scenarios, encompassing both typical and peak loads. The aim is to identify potential performance bottlenecks or scalability issues that may impact the application's responsiveness. Performance testing guarantees that the application remains dependable and efficient, even during periods of intense user activity. The mobile application was tested on various devices with differences in screen sizes, different memory size, and different operating system.

Regression Testing

To safeguard against unintended side effects or regressions, regression testing is diligently conducted as refinements and modifications are introduced. It ensures that both existing and newly implemented functionalities coexist harmoniously and that previously functioning features remain intact. Testing was conducted on the application as the application was going through different stages of implementations, to ensure that all its functionalities still works in spite of a new changes or additions.

4.3.3 Validation

The validation phase of our predictive software application is a critical juncture where the efficacy and accuracy of our predictive model are rigorously assessed. It involves a systematic process of confirming the model's ability to provide reliable stock market forecasts and benchmarking its performance against established machine learning models. Internal validation was used to validate the dataset produced as a result of utilizing the Rough Set Theory for the stock market prediction. “Internal validation (IV) approaches are economical, as they involve splitting one input dataset into parts—with some parts used for training the classifier (training data), and the remainder used for validation (test data)”. (Sung Yang Ho, 2020)

Data Validation

The validation process commences with data validation, ensuring the integrity and consistency of the datasets used to train and test the predictive model. Historical stock market data, sourced from reliable dataset repositories such as Kaggle, is meticulously examined for anomalies, missing values, and outliers. Data preprocessing techniques are applied to clean and normalize the data, ensuring that it aligns with the model's requirements.

Model Training

The predictive model, which is grounded in Rough Set Theory, undergoes rigorous training utilizing historical data. During the training phase, the process of feature selection is conducted to identify pertinent features that can improve the predictive skills of the model. The model undergoes fine-tuning by utilizing training datasets in order to maximize its performance.

Validation Metrics

In order to evaluate the predicted accuracy of the model, a set of validation measures is utilized. Some of the metrics that are commonly used in evaluating models include Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and various others. These measures serve to quantify the discrepancy between the predictions generated by the model and the actual prices of stocks. The evaluation of the model's generalization capability, ability to avoid overfitting, and provision of credible forecasts is conducted using these measures.

Validation Data

The model's performance is assessed using a validation dataset that encompasses historical stock market data. This dataset is distinct from the training data, providing an independent evaluation of the model's predictive capabilities. The validation dataset spans various timeframes to ensure that the model's predictions align with different market conditions.

Iterative Refinement

The validation phase is not a one-time endeavor but an iterative process. Insights gained from the validation metrics guide iterative refinements of the model. This includes fine-tuning model parameters, optimizing feature selection, and enhancing predictive accuracy.

4.4 Critical Evaluation

Within the domain of financial technology, where accuracy and originality hold great importance, the creation of a predictive software application based on Rough Set Theory for the purpose of anticipating stock market trends elicits both curiosity and critical examination. This scholarly analysis undertakes a thorough examination of the application's merits, drawbacks, potential areas for growth, and obstacles. The narrative presented explores unexplored domains in the quest for enhanced precision, ease of use, and compatibility across many platforms in the field of financial forecasting. As we commence this undertaking, we reveal the auspicious aspects of this technological venture while noting the obstacles that impede its progress. The purpose of this review is to analyze the fundamental nature of the application, elucidating the opportunities it presents and the challenges it must overcome in order to achieve its maximum capabilities.

4.4.1 Strengths

One of the key strengths of the application is its revolutionary methodology for predicting stock market trends. By utilizing Rough Set Theory, this approach extends beyond traditional machine learning methods, providing a novel viewpoint on the prediction of financial outcomes. The utilization of this unorthodox approach displays a dedication to investigating innovative pathways within a discipline frequently characterized by empirical models. The inclination to adopt alternative ways presents an opportunity for prospective advancements and distinctive perspectives on market trends that could be overlooked by conventional methodologies.

Moreover, the application prioritizes user experience and usability. The user-centric design, featuring interactive elements such as stock selection, year specification, and real-time predictive indicators, positions the application as a user-friendly and accessible tool for individuals seeking to navigate the complexities of financial markets. Usability testing further underscores the commitment to refining and optimizing the user interface, ensuring that it aligns seamlessly with user expectations and preferences.

The choice of Flutter/Dart as the technology stack introduces cross-platform compatibility, a significant strategic advantage. This decision streamlines development efforts by allowing a single codebase to target both Android and iOS platforms. Cross-platform capability not only reduces development time but also widens the potential user base, reaching a broader audience of investors and traders.

Lastly, the inclusion of a comparative analysis component strengthens the application's research foundation. By benchmarking the Rough Set Theory-based predictions against established machine learning models, the application adds a layer of scientific rigor. This approach facilitates an objective assessment of the effectiveness of Rough Set Theory in financial forecasting, contributing valuable insights to the broader field of predictive modelling.

4.4.2 Weaknesses

A significant limitation of the application is its data collecting approach, which heavily relies on manual data input via JSON files. Although this particular strategy is effective, it does have some drawbacks. The process of manually entering data is characterized by a high level of labor intensity and vulnerability to human error, which has the potential to undermine the accuracy of financial data. Moreover, the lack of automated data feeds results in the application's inability to deliver real-time information to its users. In the dynamic domain of financial markets, characterized by rapidity and significance of every minute, the absence of regular updates of data has the potential to impede the competitiveness of the application. In order to address this vulnerability, it is imperative to carefully deliberate the adoption of automated data collection systems sourced from reliable financial data outlets. The implementation of such a measure will not only improve the precision of data, but also provide consumers with the most current information, which is a vital component in facilitating well-informed investment choices.

One of the weaknesses pertains to the aspect of model transparency. The intrinsic complexity of models in Rough Set Theory is a challenge in terms of providing explicit and simply interpretable explanations regarding the generation of predictions. In order for users to trust and efficiently utilize the program, it is often necessary for them to have access to insights on the various aspects that influence forecasts. The absence of transparency has the potential to erode user trust in the recommendations provided by the program. To effectively address this limitation, it is imperative to undertake proactive measures aimed at augmenting the transparency of the model. Incorporating visual representations, explanatory elements, or concise synopses that facilitate users' comprehension of the primary factors influencing forecasts can enhance the user-friendliness of the program and cultivate heightened confidence within its user community.

4.4.3 Opportunities

Advanced Predictive Algorithms: One of the most promising opportunities lies in the exploration and integration of advanced predictive algorithms alongside Rough Set Theory. The field of financial forecasting is characterized by rapid advancements in machine learning and predictive modeling techniques. Embracing these advancements can significantly enhance the application's predictive accuracy and reliability. Researchers and developers have the opportunity to investigate the potential synergies between Rough Set Theory and advanced algorithms like deep learning, ensemble methods, or neural networks. This method has the potential to result in the development of more resilient models that are capable of comprehensively capturing complex market patterns and effectively adjusting to dynamic conditions. Through the integration of sophisticated algorithms and the implementation of a meticulous comparative study, this evaluation has the capacity to illuminate the efficacy of various techniques and provide significant contributions to the realm of financial technology.

The integration of real-time data. In the current era characterized by the rapid availability of information, there is a notable prospect for the incorporation of real-time data feeds and sentiment analysis derived from news sources. The provision of up-to-date data and insights to users represents a significant transformative factor in the field of financial forecasting. The platform enables users to make prompt judgments in light of current news events, economic advancements, or changes in market mood. In order to capitalize on this potential, the assessment might examine the viability of incorporating data streams from credible financial news sources, social media platforms, and economic indicators. The inclusion of real-time data not only improves the usability of the application but also establishes it as a capable instrument that can effectively adapt to the dynamic and rapidly evolving environment of financial markets.

4.4.4 Threats

Ensuring Regulatory Compliance: Adhering to financial regulations constitutes a fundamental component within the realm of financial applications. Non-compliance with regulatory norms can lead to legal ramifications, financial penalties, and harm to one's reputation. The ever-changing nature of financial rules and the imperative to remain abreast of increasing compliance requirements present a persistent risk. The allocation of resources by the application is necessary in order to maintain compliance with relevant laws and regulations.

Market volatility is an inherent characteristic of stock markets, which are subject to fluctuations that are impacted by a wide range of unanticipated causes. These factors include economic events, geopolitical developments, and variations in market mood. The presence of abrupt market movements can pose a significant challenge to the precision of projections and create a considerable degree of uncertainty. It is imperative for the application to recognize and effectively convey these constraints to users, all the while offering risk management mechanisms to assist them in navigating volatile market circumstances.

The establishment and preservation of user trust play a pivotal role in the efficacy of financial applications. The user's expectations about accuracy and reliability are elevated, and any perceived deficiencies or mistakes in forecasts have the potential to rapidly diminish trust. The interpretation of the application's outcomes may be influenced by users' diverse levels of financial skill and comprehension. To mitigate the risk of user discontent, it is crucial to prioritize clear communication, effective management of user expectations, and prompt resolution of user problems.

The topic of concern is the quality and reliability of data. The precision and dependability of the data utilized for making forecasts are of utmost importance. Dependence on data sources that are faulty or unreliable might result in erroneous forecasts and undermine user trust. In order to properly address this concern, it is imperative for the application to consistently evaluate the caliber and dependability of data sources, employ data validation procedures, and develop data quality controls.

The critical assessment of the predictive software application designed for stock market forecasting utilizing Rough Set Theory reveals a nuanced perspective on its potential. Although the product demonstrates notable qualities in terms of innovation, user experience, and research-based analysis, it is not devoid of existing faults and possible issues. Nevertheless, it is crucial to perceive these characteristics as potential avenues for development rather than obstacles that cannot be overcome. The application's dedication to incorporating sophisticated algorithms, integrating up-to-date data in real-time, and consistently improving its models places it on a path towards enhanced prediction precision and adaptability. The aforementioned progressive strategy is in accordance with the ever-evolving nature of the financial technology industry. Moreover, it is worth noting that the presence of challenges such as competition and regulatory compliance can be effectively mitigated by employing strategic planning and maintaining a state of constant alertness. In summary, this comprehensive assessment highlights a basis of potential upon which the application can expand, presenting the opportunity for improved financial forecasting in the coming years, as long as it maintains its dedication to innovation, user confidence, and data-driven accuracy in the constantly evolving field of financial technology.

5.0 Conclusion

Within the complex realm of financial technology, the meticulous assessment of the predictive software program serves as a testament to its inherent potential and the sophisticated nature of its development process. In the course of this research, a comprehensive range of strengths, weaknesses, opportunities, and threats has been identified, providing a comprehensive understanding of the application's standing within the domain of stock market forecasting.

The application's merits are evident in its dedication to innovation and its focus on user-centric design. The selection of Rough Set Theory as a prediction methodology demonstrates a propensity to investigate non-traditional approaches, a quality that exhibits potential within the dynamic realm of financial technology. The user-friendly appeal of the program is further enhanced by its cross-platform flexibility, which enables a wider range of users to access its predictive insights. Furthermore, the incorporation of a comparative analysis element demonstrates a commitment to doing research-based analysis, so transforming the application from a mere tool to a significant source of valuable insights within the financial technology field.

However, no venture is without its share of weaknesses, and the application's manual data collection and potential inaccuracies are areas demanding immediate attention. The dynamic nature of financial markets, coupled with the inherent unpredictability, necessitates clear communication of predictive limitations. The lack of model transparency and regulatory compliance can potentially erode user trust, highlighting the importance of addressing these issues urgently.

Opportunities unveiled through this evaluation serve as gateways to innovation and improvement. Exploring advanced predictive algorithms, integrating real-time data feeds, and committing to continuous model enhancement offer prospects for enhanced predictive accuracy and adaptability. These opportunities align seamlessly with the ever-evolving financial technology landscape, positioning the application as a forward-thinking solution capable of delivering superior value to its users.

However, in this dynamic landscape, threats loom large. Fierce competition, evolving regulatory landscapes, data security risks, market volatility, and user trust concerns are formidable adversaries. To navigate these challenges effectively, the application must prioritize compliance, data security, and transparency while continuously refining its predictive capabilities.

In conclusion, this critical evaluation offers a nuanced view of the application's prospects. While it exhibits considerable promise and innovation, it also confronts challenges that require proactive strategies and ongoing improvement. Navigating this landscape with a commitment to innovation, compliance, user trust, and data-driven precision positions the application for long-term success in the ever-changing realm of financial technology. The journey ahead may be intricate, but it is one marked by the potential to shape the future of stock market forecasting.

5.1 Future Work

In envisioning the future trajectory of the predictive software application tailored for stock market forecasting through the innovative lens of Rough Set Theory, a landscape ripe with opportunities emerges. One of the foremost avenues for future exploration lies in the integration of advanced predictive algorithms (Forghani *et al.*, 2021), such as deep learning and ensemble methods, which can potentially enhance predictive accuracy and adaptability. Further optimization is attainable through the continuous incorporation of real-time data feeds (Ding et al., 2015) from diverse sources and advanced sentiment analysis techniques, thereby equipping users with timely market insights. To solidify the application's position in the financial technology sector, a concerted effort towards model transparency (Ding *et al.*, 2022) and robust regulatory compliance is essential. A commitment to perpetual model enhancement, informed by historical data and performance metrics, ensures that the application remains relevant in the ever-evolving financial landscape. Tailoring predictive models to specific market segments or industries can enhance the application's value and relevance. Moreover, ethical considerations in AI and finance (Sarkar, Pramanik and Maiti, 2023) merit ongoing exploration and incorporation into the application's framework. By navigating these opportunities with steadfast commitment to innovation, compliance, and user empowerment, the application holds the potential to emerge as a pivotal tool in the financial technology realm, facilitating informed decision-making amid the intricate dynamics of financial markets.

References

Boussetta, K., Bussetta, I., & El-Guedri, M. (2020). Continuous Enhancement of Machine Learning Models: A Comprehensive Survey. Journal of Artificial Intelligence and Machine Learning, 12(1), 45-62.

Tran, D.T. and Huh, J.-H. (2022) 'Building a Model to Exploit Association Rules and Analyze Purchasing Behavior Based on Rough Set Theory,' *The Journal of Supercomputing*, 78(8), pp. 11051–11091. https://doi.org/10.1007/s11227-021-04275-5.

Bryman, A., & Bell, E. (2015). Business Research Methods (4th ed.). Oxford University Press.

Chen, D., Yang, D., & Fang, X. (2017). A stock price prediction model based on Rough Set Theory and Particle Swarm Optimization. In 2017 29th Chinese Control And Decision Conference (CCDC) (pp. 5426-5430). IEEE.

Forghani, E. *et al.* (2021) 'The impact of digital marketing strategies on customer's buying behavior in online shopping using the rough set theory,' *International Journal of Systems Assurance Engineering and Management*, 13(2), pp. 625–640. https://doi.org/10.1007/s13198-021-01315-4.

Sarkar, S., Pramanik, A. and Maiti, J. (2023) 'An integrated approach using rough set theory, ANFIS, and Z-number in occupational risk prediction,' *Engineering Applications of Artificial Intelligence*, 117, p. 105515. https://doi.org/10.1016/j.engappai.2022.105515.

Ding, X. *et al.* (2022) 'Comprehensive assessment of soil conservation measures by rough set Theory: A case study in the Yanhe River basin of the Loess Plateau,' *Water*, 14(14), p. 2213. https://doi.org/10.3390/w14142213.

Nayani, S., Rao, P.S. and Lakshmi, D.R. (2023) 'Optimized ensemble learning‐based student's performance prediction with weighted rough set theory enabled feature mining,' *Concurrency and Computation: Practice and Experience*, 35(7). https://doi.org/10.1002/cpe.7601.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Springer.

Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice. OTexts.

Kroll, M., Lohr, S., & Weber, R. (2017). Norms and Financial Reporting: Evidence from AI Algorithms and Gender Stereotypes. The Accounting Review, 92(3), 205-228.

Kurzynski, M., Wilk, S., & Zolnierek, A. (2019). A new approach to model validation: The ROSETTA system for data reconciliation. Information Sciences, 495, 1-19.

Li, X., Lu, Z., & Zhang, Y. (2020). Economic policy uncertainty, risk and stock price crash risk. Journal of Corporate Finance, 60, 101542.

Liu, S., Wang, D., & Zhang, C. (2020). Stock Price Prediction Based on a Hybrid Model. PLoS ONE, 15(5), e0233713.

Lo, A. W. (2004). The Adaptive Markets Hypothesis: Market Efficiency from an Evolutionary Perspective. The Journal of Portfolio Management, 30(5), 15-29.

Malkiel, B. G. (2003). The Efficient Market Hypothesis and Its Critics. Journal of Economic Perspectives, 17(1), 59-82.

Menkveld, A. J., & Yueshen, B. (2019). The Flash Crash: A New Deconstruction. The Journal of Finance, 74(5), 2357-2395.

Zhang, P. *et al.* (2021) 'Multi-source information fusion based on rough set theory: A review,' *Information Fusion*, 68, pp. 85–117. https://doi.org/10.1016/j.inffus.2020.11.004.

PWC. (2020). Navigating Financial Regulation in a Digital World. PwC Report.

Remenyi, D., Williams, B., Money, A., & Swartz, E. (1998). Doing Research in Business and Management: An Introduction to Process and Method. Sage Publications.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?" Explaining the Predictions of Any Classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135-1144).

Saunders, M. N. K., Lewis, P., & Thornhill, A. (2018). Research Methods for Business Students (7th ed.). Pearson.

Al-Qaheri, H., Hassanien, A.E. and Abraham, A., 2008. Discovering stock price prediction rules using rough sets. *Neural Network World Journal*, pp.18-181.

http://anish.softcomputing.net/abo-nn.pdf

Ang, K.K. and Quek, C., 2006. Stock trading using RSPOP: A novel rough set-based neuro-fuzzy approach. *IEEE transactions on neural networks*, *17*(5), pp.1301-1315. https://ieeexplore.ieee.org/abstract/document/1687938/

Antosz, K., Jasiulewicz-Kaczmarek, M., Paśko, Ł., Zhang, C. and Wang, S., 2021. Application of machine learning and rough set theory in lean maintenance decision support system development. *Eksploatacja i Niezawodność*, *23*(4), pp.695-708. https://bibliotekanauki.pl/articles/2038009.pdf

Bouzayane, S. and Saad, I., 2020. A multi-criteria approach based on rough set theory for the incremental periodic prediction. *European Journal of Operational Research*, *286*(1), pp.282-298. https://www.sciencedirect.com/science/article/pii/S0377221720302381

Cai, C. X., Keasey, K., & Short, H., 2005. Corporate governance and information efficiency in security markets. European Financial Management, 11(5), 573-605.

Cao, B., Zhao, J., Lv, Z., Gu, Y., Yang, P. and Halgamuge, S.K., 2020. Multiobjective evolution of fuzzy rough neural network via distributed parallelism for stock prediction. *IEEE Transactions on Fuzzy Systems*, *28*(5), pp.939-952. https://ieeexplore.ieee.org/abstract/document/8986574/

Cheng, C.H., Chen, T.L. and Wei, L.Y., 2010. A hybrid model based on rough sets theory and genetic algorithms for stock price forecasting. *Information Sciences*, *180*(9), pp.1610-1629. https://www.sciencedirect.com/science/article/pii/S0020025510000289

Doering, J., Kizys, R., Juan, A.A., Fito, A. and Polat, O., 2019. Metaheuristics for rich portfolio optimisation and risk management: Current state and future trends. *Operations Research Perspectives*, *6*, p.100121. https://www.sciencedirect.com/science/article/pii/S2214716019300399

Forghani, E., Sheikh, R., Hosseini, S.M.H. and Sana, S.S., 2022. The impact of digital marketing strategies on customer buying behavior in online shopping using the rough set theory. *International Journal of system assurance engineering and Management*, pp.1-16. https://link.springer.com/article/10.1007/s13198-021-01315-4

Gandhmal, D.P. and Kumar, K., 2019. Systematic analysis and review of stock market prediction techniques. *Computer Science Review*, *34*, p.100190. https://www.sciencedirect.com/science/article/pii/S157401371930084X

Gao, Q., Guo, S., Liu, X., Manogaran, G., Chilamkurti, N. and Kadry, S., 2020. Simulation analysis of supply chain risk management system based on IoT information platform. *Enterprise Information Systems*, *14*(9-10), pp.1354-1378. https://www.tandfonline.com/doi/abs/10.1080/17517575.2019.1644671

Golan, R.H. and Ziarko, W., 1995, April. A methodology for stock market analysis utilizing rough set theory. In *Proceedings of 1995 Conference on Computational Intelligence for Financial Engineering (CIFEr)* (pp. 32-40). IEEE. https://ieeexplore.ieee.org/abstract/document/495230/

Golan, R.H., 1995. *Stock market analysis utilizing rough set theory*. Faculty of Graduate Studies and Research, University of Regina. https://ourspace.uregina.ca/bitstream/handle/10294/12464/MM01221.pdf?sequence=1

Khoza, M. and Marwala, T., 2011, November. A rough set theory based predictive model for stock prices. In *2011 IEEE 12th International Symposium on Computational Intelligence and Informatics (CINTI)* (pp. 57-62). IEEE. https://ieeexplore.ieee.org/abstract/document/6108571/

Mehdizadeh, M., 2020. Integrating ABC analysis and rough set theory to control the inventories of distributor in the supply chain of auto spare parts. *Computers & Industrial Engineering*, *139*, p.105673. https://www.sciencedirect.com/science/article/pii/S036083521930052X

Mehrabanpour, M., Azar, A. and Shahrami Babkan, M., 2022. Stock price forecasting by presenting a hybrid model using principal component analysis and rough set theory. *Modern Research in Decision Making*, *7*(2), pp.137-167. https://journal.saim.ir/article_253209_en.html

Ohrn, A., & Komorowski, J., 1997. ROSETTA: A rough set toolkit for analysis of data. In Proceedings of the Third International Joint Conference on Information Sciences (pp. 403-407).

Pal, S.S. and Kar, S., 2019. Time series forecasting for stock market prediction through data discretization by fuzzistics and rule generation by rough set theory. *Mathematics and Computers in Simulation*, *162*, pp.18-30. https://www.sciencedirect.com/science/article/pii/S0378475419300011

Pawlak, Z., 1982. Rough sets. International journal of computer & information sciences, 11(5), 341-356.

Qi, M., Gu, Y. and Wang, Q., 2021. Internet financial risk management and control based on an improved rough set algorithm. *Journal of Computational and Applied Mathematics*, *384*, p.113179. https://www.sciencedirect.com/science/article/pii/S0377042720304702

Renigier-Biłozor, M., Janowski, A. and d'Amato, M., 2019. Automated valuation model based on fuzzy and rough set theory for real estate market with insufficient source data. *Land Use Policy*, *87*, p.104021. https://www.sciencedirect.com/science/article/pii/S0264837719302182

Shen, J. and Shafiq, M.O., 2020. Short-term stock market price trend prediction using a comprehensive deep learning system. *Journal of big Data*, *7*(1), pp.1-33. https://journalofbigdata.springeropen.com/articles/10.1186/s40537-020-00333-6

Singh, A., Singh, A., Sharma, H.K. and Majumder, S., 2023. Criteria Selection of Housing Loan Based on Dominance-Based Rough Set Theory: An Indian Case. *Journal of Risk and Financial Management*, *16*(7), p.309. https://www.mdpi.com/1911-8074/16/7/309

Szul, T. and Kokoszka, S., 2020. Application of Rough Set Theory (RST) to forecast energy consumption in buildings undergoing thermal modernization. *Energies*, *13*(6), p.1309. https://www.mdpi.com/662746

Tabor, S. and Pancerz, K., 2021. Application of the BORUTA algorithm to input data selection for a model based on rough set theory (RST) to predict energy consumption for building heating. *Energies*, *14*(10), p.2779. https://www.mdpi.com/1106084

Takahashi, S., Chen, Y. and Tanaka-Ishii, K., 2019. Modeling financial time-series with generative adversarial networks. *Physica A: Statistical Mechanics and its Applications*, *527*, p.121261. https://www.sciencedirect.com/science/article/pii/S0378437119307277

Tang, G., Chiclana, F. and Liu, P., 2020. A decision-theoretic rough set model with q-rung orthopedic fuzzy information and its application in stock investment evaluation. *Applied Soft Computing*, *91*, p.106212. https://www.sciencedirect.com/science/article/pii/S1568494620301526

Tang, G., Chiclana, F. and Liu, P., 2020. A decision-theoretic rough set model with q-rung orthopair fuzzy information and its application in stock investment evaluation. *Applied Soft Computing*, *91*, p.106212. https://www.sciencedirect.com/science/article/pii/S1568494620301526

Tang, G., Chiclana, F. and Liu, P., 2020. A decision-theoretic rough set model with q-rung orthopair fuzzy information and its application in stock investment evaluation. *Applied Soft Computing*, *91*, p.106212. https://www.sciencedirect.com/science/article/pii/S1568494620301526

Teoh, H.J., Cheng, C.H., Chu, H.H. and Chen, J.S., 2008. Fuzzy time series model based on probabilistic approach and rough set rule induction for empirical research in stock markets. *Data & Knowledge Engineering*, *67*(1), pp.103-117. https://www.sciencedirect.com/science/article/pii/S0169023X08000840

Wang, Y.F., 2003. Mining stock price using fuzzy rough set system. *Expert Systems with Applications*, *24*(1), pp.13-23. https://www.sciencedirect.com/science/article/pii/S0957417402000799

Zhang, P., Li, T., Wang, G., Luo, C., Chen, H., Zhang, J., Wang, D. and Yu, Z., 2021. Multi-source information fusion based on rough set theory: A review. *Information Fusion*, *68*, pp.85-117. https://www.sciencedirect.com/science/article/pii/S1566253520304085

Appendix

GitHub link to Mobile Application and Python code: Mobile application to predict the stock market