MODELLING STOCK MARKET VOLATILITY USING RANDOM FOREST

BY

TERRY NJUGUNA

A RESEARCH PROJECT SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE AWARD OF MASTER OF SCIENCE, FINANCE AT THE SCHOOL OF BUSINESS, THE UNIVERSITY OF NAIROBI

NOVEMBER 2021

DECLARATION

I, the undersigned declare that this research project is my original work and has not been presented in any other University.

Signed....TMN....................................... Date.25.11.2021.............................

Terry Njuguna

D63/36314/2020

This research project has been submitted for examination with my approval as University Supervisors.

Signed: .....DChirchir.. Date.25/11/2021.......................

Dan Chirchir

Department of Finance and Accounting

Supervisor

Signed: .... Date.November 25, 2021

Dr. Winnie Nyamute

Department of Finance and Accounting

Supervisor

DEDICATION

This project is dedicated to my family members who have always believed in my potential and encouraged me to pursue my Masters at the University of Nairobi, thanks for your prayers. To my parents who always sacrificed, supported, and inculcated relentless virtues in me. It was always your desire I be fruitful mentally and in all depths of life. To my supervisor, lecturers, and fellow students for their undying support throughout my studies.

ACKNOWLEDGEMENTS

I would like to express my sincere and utmost gratitude to my supervisors, Dan Chirchir and Dr. Winnie Nyamute for the invaluable guidance, support, patience, and tireless efforts without which I would not have completed this study. Their useful comments, remarks, assistance, and dedication throughout the process made the execution of this research possible and bearable. I thank my family who were always there for me even as I sacrificed their time and happiness for this academic work. Further, I thank the University of Nairobi for enrolling me into the program and for allowing me to utilize their facilities. Above all, I thank the Almighty God for His sustenance and grace and the wisdom He has bestowed upon me in the successful completion of this study.

ABSTRACT

In recent years, stock market volatility has had an increasing role in investment decision making on what stock to trade in, in order to reap on the returns. Owing to the high uncertainty in the stock markets; stock market volatility has become equally helpful in many micros as well as macro-economic decision-making. Since stock activity is chaotic and volatile, it is risky to invest in stock market. Traders would consider purchasing a share whose price is expected to rise in the coming days but avoid a share whose price is likely to fall in coming days. To limit the risk of losing money, one needs a lot of knowledge about how stock prices will move in the future. As a result, it is necessary to accurately estimate stock market price movements. This research project aimed at modelling stock market volatility using random forest over a period of 12 years that is from 1st Sept 2009 to 31st Aug 2021. Random Forest is an algorithm utilized for classification and regression usually derived from a set of classification and regression trees. So many models that have been developed to model stock market volatility have focused on statistical methods such as time series analysis, regression analysis, and multivariate analysis. However; most researchers have learnt that stock prediction performs exemplary when regarded as a classification problem instead of as an analytical problem. Accordingly, the study considered modelling stock market volatility using machine learning techniques and technical indicators. Random forest had the advantage of performing well in big feature sizes and efficiently handling complex data. The random forest package in R was used for analysis, and the Miss Forest package was used for imputation. On the basis of the model's classification accuracy, model diagnostics, and interpretation, conclusions were drawn.

TABLE OF CONTENTS

ABREVIATIONS AND ACRONYMS

KNN	K Nearest Neighbour
LM	Langragian Multiplier
MACD	Moving Average Convergence Divergence Oscillator
NSE	Nairobi Securities Exchange
PNN	Probabilistic Neural Networks
ROC	Price Rate of Change
RSI	Relative Strength Index
SMA	Simple Moving Averages
SVM	Support Vector Machines

CHAPTER ONE

INTRODUCTION

1.1 Background of the Study

Predicting stock market price movements is a difficult undertaking owing to the many unknowns involved and the numerous elements that impact market value on a given day, for example economic conditions, investor attitude toward a certain business, political events, and so on. As a result, stock markets are vulnerable to sudden shifts, resulting in unpredictable swings in stock prices. Based on stock market dynamic, non-parametric, chaotic, and noisy features, fluctuation is considered a random process with significant variations during short time frames. However, certain stocks tend to establish linear patterns in the long run.

The chaotic and extreme volatility of stock activity poses danger to stock markets. Thus, it is essential to offer extensive knowledge on stock price movement in the future as a way of minimizing the risks. Traders tend to purchase a stock that is projected to rise in value in the future but they are unlikely to purchase a stock whose value is projected to decline in the future. Therefore, it is essential to forecast stock market price movement to maximize profit margins and minimize loss. Among the primary techniques used to forecast stock price behaviour, the following stand out: (1) Technical Analysis and (2) Time Series Prediction (3) Data Mining & Machine Learning (Holmstromm & Hellstrom (1998)), and (4) Modelling and predicting stock volatility using differential equations (Holmstromm & Hellstrom 1998).

Modelling volatility has been a huge topic of research in observational funds, with applications ranging from resource evaluation, portfolio assignment, subsidiary estimation, all the way up to risk management executives. Volatility refers to the degree to which the arrival of a benefit or the unconditional variation of returns given by an asset change. Because variance or standard deviation is utilized as a risk metric in risk management, volatility may be defined as the degree of uncertainty or risk on the return on a certain investment. Higher volatility demonstrates that the return on an asset is stretched out over a long period of time, whereas lower volatility implies that the return moves across a narrow range or that there are no obvious exceptional variations. The cause for volatility in any financial market is simply trading. Financial time series data are based on the idea of volatility, which characterizes financial time series data.

Theoretical and empirical research on modelling and forecasting stock market volatility has been extensive. Many volatility applications need the estimate or forecasting of a volatility parameter (Brooks, 2014). Stock volatility has sparked intense arguments and discussions among economists, stock market experts, government regulators, and policymakers. This interest and discussion stems in part from the implications for market efficiency, stock market bubbles, market crashes, and recessions in the economy's financial sectors (Nyong, 2015).

The majority of conventional time series econometric techniques are concerned with modelling a random variable's conditional mean. As a result, volatility modelling, forecasting, and correlation are critical variables in risk management and financial market vulnerability evaluation. Furthermore, predicting volatility is critical to an investor's capacity to accurately anticipate the unpredictability in asset price movements and the connection between assets.

1.1.1 Stock Market Volatility

Volatility in the stock market has a risk affecting the entire jurisdiction economy. Stock market volatility can be caused by so many phenomena around the world which may include pandemic (Corona Virus 2019), political instability among nations and the change in the oil price day in day out globally.

Volatility, as a percentage fluctuation in stock returns, is a key variable in mainstream financial economics. In reality, significant empirical research has been conducted in this field. Furthermore, stock market volatility is asymmetric, concentrated, persistent, and has a long-term memory.

If the distribution or variability of returns is known, it is feasible to anticipate returns more accurately. But the question is whether stock market volatility is understood, whether it is constant over time or stochastic, and, more significantly, whether daily or monthly stock returns are normally distributed. Practitioners in the investing business are extremely smart, have a knowledge of the difficult elements of return predictability, and devote a significant amount of time and effort to projecting returns. One of the primary conclusions of this sector, as well as a consensus within academics, is that stock values are not predictable in the short to medium term, and there is a great deal of volatility and uncertainty.

Many Kenyan investors do not have that knowledge yet and hence making decisions based on mob psychology. Few people, if any are willing to share the success knowledge they have (Dunne, 2015). According to O'Neil (2009) many investors have made serious losses in the stocks market. He explains that it takes a lot of trial and error before you can nail down substantial gains in stocks. Many authors acknowledge that trading in the stocks markets is a challenging task. According to Bonde and Khaled (2011), prediction of stocks prices has always been a challenging task. They go on to say that the stock price of any specific firm does not necessarily rely on the country's economic status. It has nothing to do with the economic growth of any one country or region. Indeed, predicting stock prices has grown more challenging than in the past.

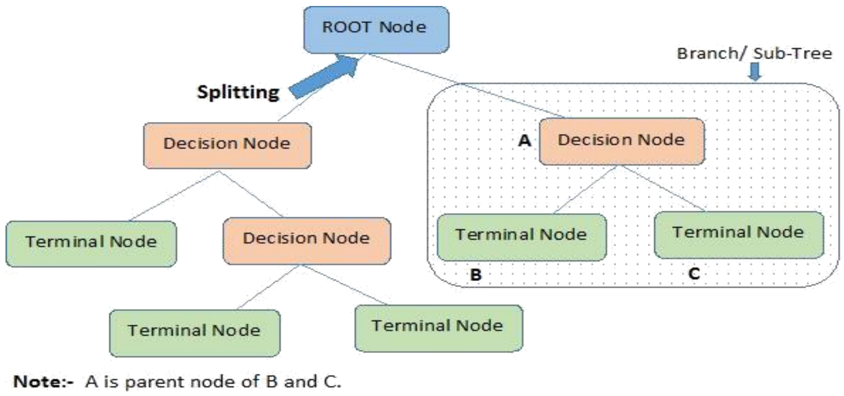
1.1.2 Machine Learning

Machine learning is a type of artificial intelligence that studies how to build computer systems that automatically learn, adapt, and improve with experience. Machine learning algorithms utilize data to find out how to solve an issue (Saahill, 2015). Machine learning techniques are classified into three namely: reinforcement, supervised and unsupervised machine learning.

Predicting outcome/dependent variables from a given collection of predictor variables is what supervised machine learning is all about. The training data in this case is a sequence of labelled instances, with each example consisting of a collection of characteristics labelled with the proper output corresponding to that feature set. This implies that the algorithm is assigned characteristics and output of one dataset (training data) and must use what it learns from that dataset to predict the outputs (labels) for another dataset (test data). Support vector machine, logistic regression, random forests methods, and KNN are some of the algorithms that are employed (k-nearest neighbour). Supervised learning is further subdivided into regression and classification issues.

Unsupervised machine learning is a form of algorithm that lacks an outcome variable, making it unsuitable for predicting. It is mostly used to divide the population into many categories. It is made up of cases in which the feature set is unlabelled. The algorithms attempt to categorize data into separate groups. Finally, reinforcement machine learning is an algorithm that exposes the machine to an environment in which it continuously trains itself from previous experience and attempts to collect the greatest available information in order to make accurate judgments.

Although machine learning is a fairly new concept in the market it is becoming go to method for researchers and analyst because it is able to build models and analyse bigger and complex data, it is able to deliver faster and more accurate results by eliminating human error and it is able to identify profitable opportunities more efficiently than other methods like traditional time series analysis. In this research project we employed the use of random forest which falls under the category of supervised learning in machine learning.



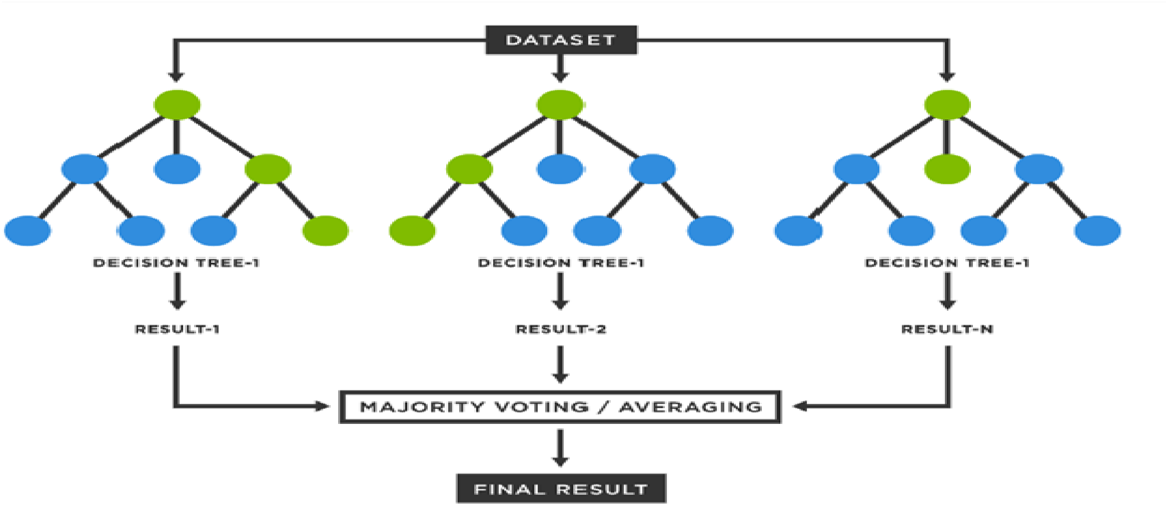
1.1.3 Random Forest

Before delving into random forest, let's first cover the fundamentals of a decision tree. Decision trees are a supervised learning approach used for categorization or determining which set of categories a given instance belongs to. It is a supervised learning approach since the classifier already knows a collection of categorized instances and learns to assign new examples based on these examples. The objective is to build a model that predicts the class of a test data set by learning basic decision rules based on data characteristics.

Random forest falls under the umbrella of ensemble classifiers (Leo Breiman, 1996). Ensemble classifiers improve prediction performance by combining various machine learning methods. Ensemble learning is based on the concept that a single classifier is insufficient for determining the class of test data. This is due to the fact that a classifier cannot discriminate between noise and pattern based on a single sample of data. This machine learning ensemble was created to increase the accuracy and stability of machine learning models used in statistical regression and classification. It lowers variance and aids in the avoidance of overfitting in decision trees. Gerardnico (Machine Learning - Bootstrap Aggregation). Boosting and bagging are both ensemble methods in the sense that rather than learning a single classifier, many are trained, and their predictions are merged. While bagging employs an ensemble of separately trained classifiers, boosting is an iterative method that seeks to reduce earlier model prediction mistakes by predicting them with subsequent models. (H. Grunbusch) (2016, February 20).

Random forest is a supervised method that can do both regression and classification and is widely used in predictive modelling and machine learning. As the ensemble's basic learner, it employs several decision trees. Therefore, when the no. of trees in the forest is high the results are more accurate. The primary disadvantage of only utilizing decision trees is that trees must be grown quite deep to understand highly irregular patterns. This causes the training sets to be overfitted, and even little noise in the data might cause the tree to grow in an unanticipated way. The argument is based on decision trees' low bias and high variance.

Random Forest circumvents this issue by training several decision trees on various subspaces of the feature space at the expense of somewhat higher bias (khaidem, saha, Roy dey, L. (2016, April 29). Normally, decision trees are usually built using the whole data set while considering all features. In random forest, decision trees are created using a subset of the whole data while considering a particular number of features which have been selected at random and trained upon. Since the method operates by sampling input data with replacements, none of the decision trees in the forest will receive the complete training data. Ultimately a number of decision trees will be grown to the largest extent possible without pruning and each decision tree will result into a final outcome. Random forest will then compile the results of all of the decision trees to bring out the final result which is done by taking the average for regression and vote for classification.



1.1.4 Stock Market Prediction

For its importance in most international economies, securities market data is perhaps the most researched time series data (Gavrilov et al., 2000). As a result, the industry continues to be a significant focus of study. Predicting the stock market, on the other hand, is a difficult undertaking. To begin, the trend of stock prices must be modelled, which is nonlinear, and then specific aspects of the input data must be extracted in order for it to be useful (Zarandi et al., 2012). Other scholars like Adebiyi et al. (2012), believe that share price prediction is a critical issue in finance and business. They claim that many approaches are ineffective in forecasting the market. For example, fundamental, statistical and technical analyses. This conclusion was reached because these approaches are incapable of providing the "deeper analysis" required for prediction.

According to Neuro AI (2013), there exists four stock market forecasting approaches : machine learning, technical analysis, time series analysis and fundamental analysis. Technical analysis is the application of historical data to forecast future prices, such as price, volume, and price movement. Technical indicators in forecasting are meant to identify trends and patterns that will subsequently guide the direction of future pricing. Technical analysis uses previous price oscillations to forecast a trend for future trading decisions, according to Huang et al. (2011). This technique considers a company's activities and financial status when determining the company's anticipated future outlook. According to Deng et al. (2011), several prominent and effective technical indicators include Moving Average Convergence Divergence (MACD), rate of change (ROC), and bias. The difference between the current price and the price a few days ago is represented by the ROC, whereas the MACD is a signal and value indicator that predicts changes in market trend. Finally, bias examines the gap between moving averages and share prices and the moving average. All of these indications must be computed, and utilize their values in forecasting.

Fundamental analysis, a second stock market prediction approach, comprises the examination of important statistical features of individual firms, such as future projects, profitability, balance sheet and so on. Since both consider a company's intrinsic performance, there is minimal difference between Technical and Fundamental analysis. Nonetheless, fundamental analysis provides a high-level perspective of the firm and takes into account factors such as the organization's political, economic, and business environment.

Some of the fundamental indications may also appear on the list of technical factors. The overarching goal of fundamental analysis is to comprehend the firm and make decisions about its future outlook. The balance sheet, profit and loss, strategic plan, competition, return on investment, and supply and demand are all examples of financial performance that may be used to get an insight. According to Chen et al. (2014), one of the major goals of fundamental analysis is to investigate the link between financial info. together with other firm facts such as revenue growth or inventory growth. Earnings per share, price to earnings ratio, dividend yield, gross sales, and ROE are some more metrics.

Machine Learning models on the other hand, have been used in many other fields for many years. More recently, Chen et al. (2019) exploit machine learning model in the estimation of stochastic discount factor, Gu et al. (2020) showed superior performance of machine learning models for empirical asset pricing, Jiang et al. (2020) found more accurate stock return predictions based on machine learning image analysis methods, and Gu et al. (2021) introduced auto encoder model that priced assets and reported less, out-of-sample errors when compared to leading factor models.

A financial time series is a series of observations on financial data gathered over a set period of time. According to Tsay, finance time series data exploration differs from other types of time series explorations because empirical time series and financial theory have an aspect of a complex dynamic system with high volatility and a lot of noise (Tsay, 2010). According to Yonis in Cont (2000), the uncertainty and noise cause the series to display some statistical regularity, which is referred to as stylized facts.

Forecasting techniques that are not based on AI, often use short-term numbers to determine the pattern and strive to forecast future movement based on this tiny dataset. Nevertheless, it may be feasible to employ AI approaches to provide meaning to the data instead of depending on the statistics in short-term. Therefore, it should allow for some purposeful data analysis to acquire intelligence from the data, learn from the figures, and then utilize the insight to anticipate the movement of the market. Furthermore, if we can develop a system which can offer solutions in the short term, such as following day share prices, then trading and prospective rewards should be offered within a short period. As a result, the waiting period, which is usually long-term, implied by most trading methods, e.g., those proposed by Mizrahi, would be minimized (2008).

1.2 Research Problem

Share trading is major industry in many countries, including Kenya. However, it carries a significant amount of risk owing to the nature of the stock market, which is very volatile. With the amount of trade money at stake, inappropriate investing may quickly result in significant losses for investors, especially if they continue to make poor judgments. Potential investors have been hesitant to join in the market due to a lack of guaranteed profits. As a result, having a tool that can guide on future pricing (prediction) as a foundation for making any investment choice is useful.

Presently, stockbrokers who performs trades and provide advice to their clientele choose stocks based on the technical research (price patterns), or fundamental analysis or individuals' personal experience. These present techniques are very subjective and, due to their limited capabilities, are frequently short-sighted. This is where artificial intelligence (AI) comes in.

Globally, several machine learning methods, including Naive Bayesian Classifier , SVM, Linear Regression, Neural Network, KNN, and Linear Discriminant Analysis, have been employed in stock prediction. For instance, Khaidem, Sudeepa, Dey, and Saha (2016) conducted a study on predicting the direction of stock market prices using random forest. The study used ensemble learning and technical indicators as inputs to train their model. From the dataset, the study achieved accuracy rate oscillating between 85-95% for long term prediction. When they compared their findings to the models used by Dai and Zhang (2013), Random forest (96.92%) still outperformed the logistic regression (53%), gaussian discriminant analysis (56%), quadratic discriminant analysis (78%), and support vector machine (79.3%).

Nti, Opoku, Adekoya, Adebayo and Benjamin, (2019) investigated Random Forest Based Feature Selection of Macroeconomic Variables for Stock Market Prediction. Based on the findings, the study concluded that the proposed stock-market prediction with microeconomic variables provides an efficient approach to automatic identification and extraction of microeconomic variables that affect a wide range of sector stocks, as well as an accurate prediction of a stock's future price. The suggested model's low error metrics (R, MAE, RMSE, RAE, and RRSE) in table 1.1 compared to the ARIMA model revealed that the proposed model had greater prediction accuracy.

Figure 1.1: Results of LSTMRNN and ARIMA Models

	Banking		Petroleum		Pharmaceuticals		Telecommunication		Technology	
Error Metrics	A[a]	L[b]	A[a]	L[b]	A[a]	L[b]	A[a]	L[b]	A[a]	L[b]
R	0.9588	0.9803	0.7956	0.9382	0.4591	0.9992	0.8456	0.9786	0.9213	0.9968
MAE	0.2557	0.1784	0.1373	0.0559	0.001	0.0001	0.0286	0.0115	0.1802	0.0527
RMSE	0.398	0.2547	0.2126	0.1191	0.002	0.0001	0.0362	0.0157	0.2273	0.0667
RAE (%)	24.970	18.0600	63.5800	25.8900	122.400	6.4700	47.7000	19.1400	38.5100	11.2700
RRSE (%)	28.110	19.6900	61.3100	34.3300	96.610	4.6200	51.9000	22.5000	46.7000	13.7100

Elagamy, Stanier and Sharp, (2018) did a study on stock market random forest-text mining system mining critical indicators of stock market movements. The study showed how text mining paired with the Random Forest algorithm may be used to extract crucial indicators and classify connected news articles in a novel way. The results of this study demonstrated that Random Forest can outperform other classifiers and provide high accuracy, extending the present classification of vital indicators from three to eight classes. From the international studies, studies show random forest as having higher accuracy in modelling stock market volatility.

Locally, the existing evidence concerning the modelling and forecasting of the volatility of the stock market is limited. Kihoro and Okango (2014) did a study on the application of artificial neural networks in predicting stock market prices in Kenya. The researchers hypothesized that only prior prices have an impact on future prices, then fitted ARIMA models to stock price data to find the appropriate input lags for the ANN model. In terms of mean square error, the

3-3-1 network architecture performed best (.00532). Malile, Sheghu, Mwera and Kalekye, (2017) did a study on the effect of the day of the week on stock returns and volatility in Kenya. The empirical analysis was done using two different models of time varying volatility with the student t distribution. Based on our findings, the EGARCH model suggests the existence of the day of the week effect on stock returns. According to this model, the day of the week effect is present, with Thursday having the highest returns while Monday has the lowest. The results demonstrate that the day of the week effect is present on stock returns but not on stock volatility. From the global perspective, algorithms such as SVM, random forest, and ANN have been studied for their accuracy in predicting securities market and we have seen that random forest outperforms other machine learning algorithms. Locally ensemble learning methods have remained unexploited as there is very little to none research on the effectiveness of Random Forest in predicting stock market volatility. This research aimed to answer the question whether random forest could be used as an effective model of predicting stock market volatility at NSE?

1.3 Research Objective

The objective of the study was to assess the effectiveness of modelling stock market volatility using random forest.

1.4 Value of the Study

Knowledge and effects of stock market volatility impacts every individual in a given country. Decisions - informed or not, made by personnel in financial institutions as a result of volatility have always been felt by the very citizens of that country. This project provides an insight on the behaviour and nature of stock market which is helpful in understanding and making informed decision on stock market on when to purchase a stock and when to sell the stock.
The findings will be useful to investors, stockbrokers, investment banks, and bankers who wish to comprehend, correctly anticipate, and quantify stock market volatility and forecast NSE stock market movements. As a result, it will assist investors in developing methods to reduce the risks connected with financial markets.

The study will also be useful to policymakers, such as the Kenyan government, in understanding stock market volatility and, as a result, in developing plans for future economic development.

Finally, this study effort will provide the groundwork for future scholarly endeavors in this field. Future academics and researchers will use the study findings as a reference source, establishing the study as a foundation for future research. The study's findings will also add to our understanding of stock market volatility.

CHAPTER TWO

REVIEW OF LITERATURE

2.1 Introduction

This chapter examines the theoretical underpinnings and models that were previously utilized in modelling the stock volatility dynamics, constituting the basis of this study. Many researchers have investigated the behaviour of the stock market and volatility using financial time series models due to the nature of financial market and stock market data having a non-constant variance. Theories on stock price volatility must be thoroughly discussed from both prior researchers' and this study's perspectives.

2.2 Theoretical Review

To underpin the influence of share price volatility on stock market performance, the following finance theories were used: Efficient Market Hypothesis, Chaos theory, Dow theory, Elliot wave theory and time series technique.

2.2.1 Efficient Market Hypothesis Theory

Stock prices, according to efficient market theory, reflect all relevant information about the stock. As a result, beating the market is impossible (Fama & Malkiel, 1970). Efficient markets may be classified into three types. A powerful kind of efficient market is one in which prices completely reflect both private and public knowledge. The semi-strong variant of EMH argues that prices reflect all publicly available information as well as knowledge about predicted future events. In this arrangement, no investor can outperform the market since prices react fast to fresh public information. The last type of efficient markets is the weak form, which asserts that previous price, volume, or other statistic information cannot be utilized to forecast future stock values.

According to the Efficient Market Hypothesis, the price of Stock Market volatility represents all available information regarding its worth. If the Stock Market prices completely reflect the relevant available information, the market is considered to be efficient. Fama (1965) formulated the efficient market hypothesis argument. In essence, the EMH contends that because the market reacts so quickly to new information regarding stock market prices and the economy as a whole, no portfolio selection approach, whether fundamental or technical, can beat a strategy of just purchasing and maintaining a diversified portfolio. The data collected may be modelled using the notion of efficient market hypothesis to evaluate stock market volatility and provide forecast information. In addition to the above, efficient markets have assumptions. One of the assumptions states that the stock prices are random. Therefore, it is impossible to predict future stock prices as they do not exhibit any trend. However, many researchers have gone against this theory and have been able to predict stock prices. They actually state that there is a trend that is exhibited by the past stock prices that can be analysed to predict future stock prices. Various method has been used to predict stock prices such as regression modelling, logistic regression and machine learning.

2.2.2 Chaos Theory

Chaos is a positive Lyapunov exponent bounded deterministic system. Chaos, as defined by the Royal Society of London in 1986, is random activity that occurs in a deterministic system. A repeated experiment in a deterministic system will provide random results in a chaotic system. This may seem contradictory considering that, in the absence of random factors, knowledge the current state of the system and its development path should allow for forecasts of all future states. (Kuchta, 2012).

One distinguishing feature of chaotic systems is their sensitivity to starting circumstances. Any degree of ambiguity in the original data, as is common in measurement, will increase as the system evolves (Kuchta, 2012). Furthermore, the mistakes will spread in unforeseen ways, rendering predicting difficult. As a result, a chaotic system has both local randomness and global determinism, and these systems can be constructed or natural, and they can arise in social structures as well as in humans (Cohen, 1997).

Because of the independence of higher moments, random walk theory has the property that price fluctuations will not follow any patterns. The independence of today's knowledge and tomorrow's pricing indicates efficient markets in this formulation. If the Efficient Market Hypothesis is correct, then gains in asset markets with random characteristics may be seen, and tests for chaos and nonlinearity can be performed (Barnett and Apostolos, 1998).

If there is nonlinearity or chaos, predicting asset values becomes an intriguing prospect. However, if we can anticipate pricing for the following period, it must not be independent of the present information set, and the previous period's price was not the best estimate. Predictability will reject the Efficient Market Hypothesis, which is how the test for chaos was created in the first place (Persaran, 1992).

2.2.3 Dow Theory

The Dow Theory states that market fluctuations follow one of three patterns. It is the long-term underlying trend that indicates whether the market is going up or down that is known as the Primary Tide. Long-term trends are subject to secondary reactions, such as market corrections. Only day traders pay attention to ripples, which are the little daily changes in the value of a cryptocurrency. Three moves in the Dow Jones Industrial Average establish primary trends, each reinforcing the one before it. They build the trend. An upward trend in the Dow Jones Transportation Average and the Dow Jones Industrial Average is required to be established. Each trend's averages must close higher than the prior high to signal a bull market. For a bear market to be established, both moving averages must close lower than the preceding low.

2.2.4 Elliott Wave Theory

This hypothesis, which was first put forth by Ralph Elliott in the 1920s, states that the market moves in a sequence of waves. There are five waves to every market trend. Every correction in a bull market is followed by an impulse wave that does not return to the previous bottom. There are three of these waves in a bull market. Even though the trend reverses in a bear market, prices continue to fall. It's possible that tiny sub-waves can be found within larger waves, causing the overall trend to shift. Finding the main trend lines is difficult due to the large number of waves and sub-waves involved.

2.2.5 Time Series Method

Statistical approaches such as time series models and multivariate analysis were employed in early market forecasting models (Gencay, 1999; Timmermann & Granger, 2004; Bao & Yang, 2008). The movement of the share price was modelled as a time series function and solved as a regression problem. The time series approach forecasts future stock values by analysing previous data. Historical data is data from the past that was collected over a certain time period. The stock market is a chaotic time series event that can be predicted using chaotic time series methods. Other time series forecasting approaches include auto-regression Integrated Moving Average (ARIMA), linear regression and auto-regression. The dependency on time is a fundamental feature of time series data; hence, a present observation must rely on a previous observation in time (Neto et al., 2009).

2.3 Empirical Review

Saahil and Madge (2015) conducted a study to forecast stock price direction using Support Vector Machines, a machine learning approach (SVM). Their objective was to utilize SVM to forecast whether the price of a certain stock will be greater or lower at a given point in the future. They calculated price volatility and momentum using daily closing prices for 34 technology stocks, which they then utilized as parameters in their SVM model. According to the study's findings, their model was able to achieve a 60 percent forecast accuracy in the long run while the efficient market hypothesis held true in the short term, reducing prediction accuracy significantly.

Luckyson Khaidem, Snehanshu Saha and Sudeepa Roy Dey, (2016) utilized the random forest classier to create their predictive model in forecasting the direction of stock market prices, which showed to be quite robust in prediction. The model was tested by calculating metrics such as accuracy, precision, recall, and so on, and they obtained long-term prediction accuracy in the region of 85-95 percent. They also did a comparison study using their article on different publications such as Li, Li, and Yang (2014), Dai and Zhang (2013), Xinjie (2014), Devi, Bhaskaran, and Kumar (2015), and the random forest model beat all other models, which was ascribed to effective data processing.

In “A comparison of PNN and SVM for stock market trend prediction,” (Lahmiri, 2017) investigated the use of macroeconomic data characteristics as well as technical indicators as inputs to his models. Granger causality was utilized in the article to determine the causal link between the inputs and outcomes. According to the findings, the SVM works best when macroeconomic data is included, but the PNN performs best when just technical indicators are employed. When they attempted to incorporate both technical and macroeconomic data, the classifiers' classification rate was substantially decreased.

Gworo (2012) investigated the Nairobi Securities Exchange's market capitalization and price volatility. This was a correlational research on the firms that comprised the 20-share index as of December 31, 2011. The result was that there was a modest connection between market capitalisation and stock price volatility.

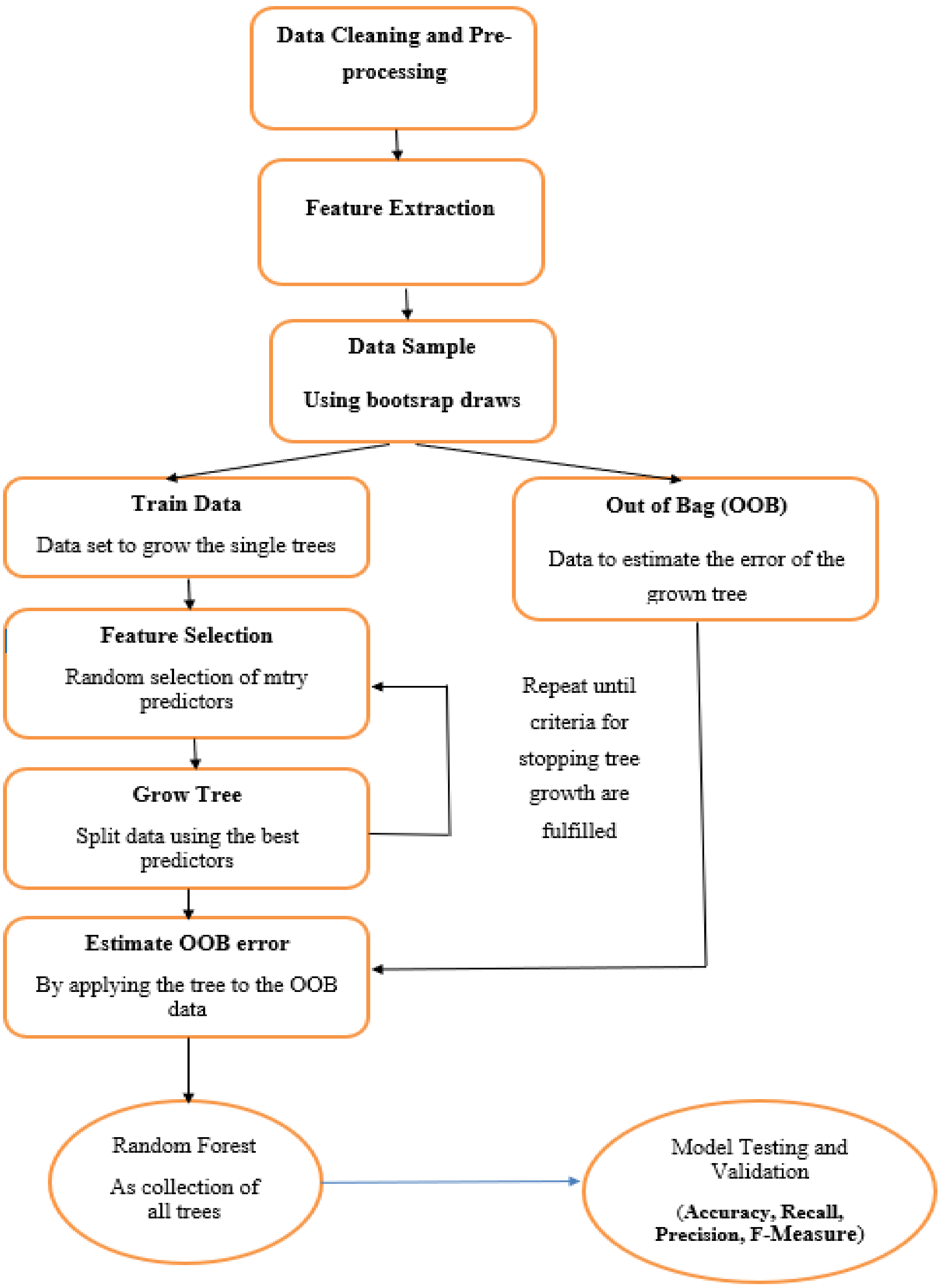
Chege et al. (2014) used monthly and weekly return series to study the changing environment of share price return and volatility in Kenya's stock market, or the NSE, between 14 January 1999 and December 2013. The GARCH-in-mean and E-GARCH models were employed in the investigation. The study's findings revealed a strong persistence of equity return shocks. The findings also indicated that fresh disturbances have no effect on present conditional variance but have an effect on historical variances. The study also discovered that big global and domestic economic events appeared to encourage market volatility.

At the NSE, Kirui et al. (2014) examined three main macroeconomic factors. Economic growth, exchange rates, and inflation rates were investigated to determine how they connect to market returns. The relationship between returns and the mentioned macroeconomic variables was determined using the Engle-Granger two-step method, and the leverage effects and volatility persistence at the Nairobi Securities Exchange were determined using the Threshold Generalized Autoregressive Conditional Heteroscedasticity (TGARCH) model.

To summarize, the financial markets of developing nations in general, and African countries in particular, have received less attention in terms of volatility modelling. According to the existing literature, the NSE, like other African stock markets, had been understudied, which left a gap that needs to be addressed.

2.4: Conceptual Framework

A conceptual framework is an analytical tool that may be applied to various situations and can be used in multiple fields where a comprehensive view is required. It's utilized to arrange thoughts and make conceptual distinctions. Strong conceptual frameworks convey something actual in an easy-to-remember and apply format. This conceptual framework was developed to demonstrate how random forest could be applied in stock market volatility modelling.



2.5 Summary and Conclusion

From the above discussion, it is clear that researchers, academicians and financial analyst are moving away from traditional methods of stock prediction to new and more effective methods by combining a number of theories found in the theoretical section of this chapter. We see that machine learning is the recent phenomena and it contains various algorithms which tackle the problem of prediction in their own way.

We may infer from this empirical and academic review that ensemble learning techniques have remained underutilized in stock market prediction challenge. To create the prediction model, we employed the Random Forest ensemble learning approach. Random forest is a collection of decision trees whose output is the mode of the individual trees' outputs. Furthermore, we have shown that the final prediction outcome is impacted not just by the prediction algorithm utilized, but also by the representation of the input variables. In order to enhance the forecast accuracy of our prediction model, we aimed to select just the relevant characteristics using technical indicators.

CHAPTER THREE
RESEARCH METHODOLOGY

3.1 Introduction

This section outlines the research techniques that were used in this study project. It included the research design, the study's population, the data collecting process, and the data analysis procedure, which includes the analytical model and the significance test. This project employed applied research to develop a specific random forest model and put it to the test on a real-world situation. Chapter 2 emphasized on the necessity for such a model that stockbrokers may use to advise on stock investing and the existing unavailability of such a tool. As a result, this study applied a specific underlying Random Forest Model to a real subject, in this case, the Nairobi Securities Exchange. The direct beneficiary of this application is the Kenyan stockbrokers. The goal of applied research is to solve a "societal or commercial" problem (Kothari, 2004).

3.2 Research Design

The descriptive research design was used in this study. The decision was required by the use of quantitative data, which was analysed, and conclusions drawn based on the results. The descriptive study methodology also provided a thorough picture of the stock market's volatility in relation to share price behaviour.

3.3 Population of the Study

The term population refers to the complete group of people who are the focus of the study within the scope of the investigation (Mugenda and Mugenda, 1999). In statistics, the target population is also the population about whom information is sought. The research population consisted of an index and a single stock to allow for comparison analysis. Based on the stock performance ratings of the listed firms the research population consisted of the top twenty trading businesses on the NSE, known as the NSE 20 index, and the Safaricom stock, which offered a good representation of the whole stock market. For the objectives of establishing and assessing the model, the research required data from a typical stock market, in this instance the Nairobi Stock Exchange, from September 1st, 2009 to August 31st, 2021. (www.nse.co.ke, Aug, 2021). As a result, the research gathered data on the performance of the NSE 20 index and Safaricom stock traded on the NSE. A 12-year record of returns by day of the week using the Nairobi Securities Exchange (NSE) 20 Price Index and Safaricom was used for this study.

3.4 Data Collection

Secondary data was used in the study. This information came from a variety of sources, including the Nairobi Securities Exchange, NCBA Investment Bank and the Kenya Central Bank. The NSE provided data on daily share price and stock market indexes. Between 2009 and 2021, data was collected.

3.5 Data Analysis and Preparation

Data obtained from the secondary data was analysed by using a statistical software package R; which is a language for statistical computing and graphics; following the steps which are shown below.

3.5.1 Data Pre-processing

The stock market data that we got was raw data which could not be used directly to train our random forest model. Stock price is an example of non-stationary data which is quite random in nature. This is because it can show trends, cycles, random walks or a combination of both at any particular time. As a result, using the stock price as an input variable is very undesirable and makes forecasting very hard. To overcome this problem some type of data pre-process (cleaning, integration, transformation, and reduction) was required to make the raw data become stationary and have statistical characteristics.

3.5.2 Feature Extraction Using Technical Indicators

Technical indicators are numerical characteristics derived from price and market data in the past. Feature extraction is the most important section in Stock Market Volatility. Technical indicators are the most commonly used tool by investors for feature extraction as they can be easily used to predict the future movement of stocks. These technical indicators were used in the random forest algorithm. The following 5 technical indicators were used in our study.

Price Rate of Change (ROC)

This technical indicator is a momentum indicator that calculates the percentage difference in price between the present price and the price n periods ago. It is classified as a momentum/ velocity indicator since the rate of change reflects the intensity of price momentum. The ROC is displayed against zero to distinguish between positive and negative values. Positive numbers represent purchasing pressure/upward momentum, whereas negative values represent selling pressure/downward momentum. Rising numbers on each side indicate rising momentum, whereas falling values on either side indicate falling momentum.

Furthermore, the ROC is utilized as a trend indicator (commonly as a divergence indicator). Divergence happens when a stock's price moves in the opposite direction of the ROC. Finally, the ROC reveals a stock's overbought and oversold situations. Overbought circumstances are indicated by values larger than positive 30, while oversold conditions are indicated by values less than negative 30.

The Moving Average Convergence Divergence Oscillator (MACD)

The convergence and divergence of two moving averages are its defining characteristics. Convergence happens when two MAs move in the same direction, whereas divergence occurs when they move in opposite directions. MACD transforms two trend-following indicators into a momentum oscillator by subtracting the longer MA from the shorter MA. MACD oscillates above and below the zero line as moving averages converge, intersect, and diverge.

MACD = 12day EMA – 26 day EMA

Signal line = 9 day EMA of MACD line.

MACD histogram = MACD line – signal line.

The MACD line's 9-day EMA is displayed with the indicator to function as a signal line and detect turns. The shorter MA is quicker and accounts for the majority of MACD changes. The lengthier MA is slower and less responsive to changes in the underlying security's price. As the shorter EMA diverges more from the longer EMA, positive values grow. This indicates that the upward momentum is rising. As the shorter EMA diverges further below the longer EMA, negative values grow, indicating that downward momentum is increasing.

Relative Strength Index (RSI)

It is calculated using the speed and direction of a stock's price change. The RSI is commonly used to identify overbuying and overselling circumstances. Values more than positive 70 indicate that a stock is overbought, implying that it is likely to be sold, whilst values less than positive 30 suggest that a stock is oversold, implying that it is likely to be purchased. It is computed in two stages. The first step is to determine the average losses and gains for the time period in question. The second step is to calculate the relative strength, which is the ratio of the values.

Williams %R

This is a momentum indicator that determines when a market is overbought or oversold. It's also utilized to determine market exit (selling) and entrance (buying) points.
It compares a stock's closing price to the high-low range over a set of days, usually 14 days.

$$W\ \%R = \frac{Highest\ high - closing\ price}{Highest\ high - lowest\ price} * \ -100$$

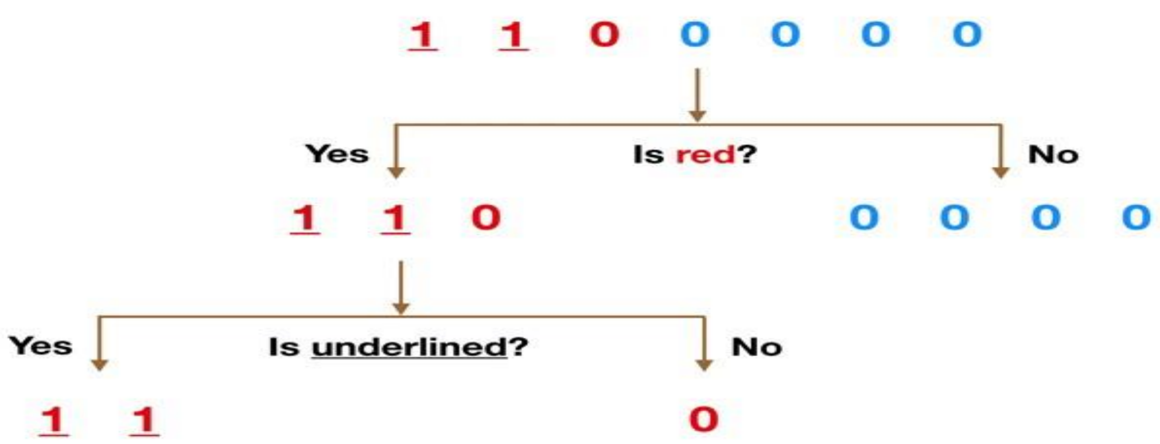
Williams %R oscillates from 0 to -100. Values between 0 and -20 indicated overbought conditions which will lead to selling of the stock while values between -80 to -100 indicated an oversold condition that led to buying of the stocks.

Stochastic Oscillator

A stochastic oscillator is a momentum indicator that provides a comparison of security's closing price to a range of prices over time. The oscillator's susceptibility to market changes can be reduced by changing the time period or taking a moving average of the outcome. The charted stochastic oscillator really consists of two lines: percent K, which represents the indicator, and percent D, which represents the three-day simple moving average (SMA) of percent K. When these two lines cross, it indicates that a trend shift is on the way.

3.5.3 The Random Forest Algorithm

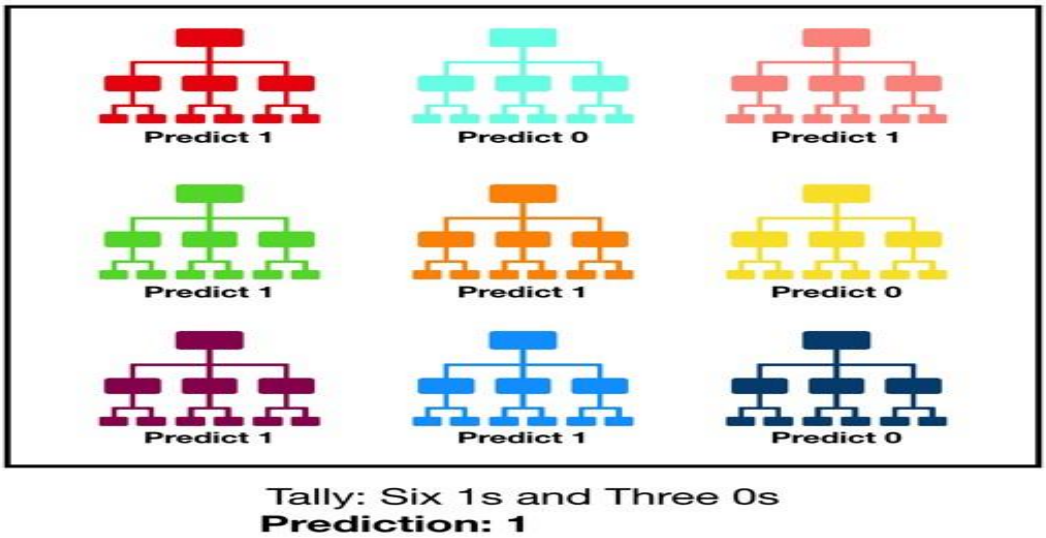
The random forest model is easily described by looking at decision trees, which are the random forest model's building parts. Fortunately, they are rather simple. Most individuals, whether intentionally or unknowingly, have utilized a decision tree at some point in their life.



3.5.3.1 The Random Forest Classifier

A random forest is composed of a significant number of individual decision trees that coordinate as an ensemble. Every tree in the random forest provides a class prediction, and the class with the most votes becomes the prediction of our model.

Visualization of a Random Forest Model Making a Prediction



The wisdom of crowds is a simple yet powerful notion at the heart of random forest. A committee made up of a large number of highly uncorrelated models (trees) will outperform any of the constituent models separately.

The crucial point is the uncorrelation amongst models. Poorly correlated models can produce ensemble predictions that are more accurate than any of the individual projections, much as how low-correlation assets (such as stocks and bonds) combine to form a portfolio that is larger than the sum of its parts. The reason for this amazing effect is because the trees shield each other from their individual faults (as long as they don't all err in the same direction all of the time). While some trees may be incorrect, many others will be accurate, allowing the trees to travel in the proper direction as a group. As a result, the following conditions must be met for random forest to function well:

1. The features must have some significant signal in order for models developed with them to outperform random guessing.

2. The predictions (and thus the errors) made by the individual trees must have minimal correlations with each other.

3.5.4 Model Testing and Validation

3.5.4.1 Out-of-sample validation

Out of sample validation refers to the use of unseen data that was not included in the model's training dataset. This is frequently seen to be the greatest approach for determining how successful a model is in predicting results on new data that hasn't been seen before.

Cross-validation is usually referred to as "out of sample testing." The model is constructed on a subset of the data (the "training" set) and then tested on data that was not used to build it (the hold-out set). Because we can apply the model to what is practically "unseen" data, we can examine how good the model is at predicting results for new data. We have the advantage of knowing what the true, "real-life" outcome is for each data point because we are using the hold-out set from the original data, which means we can assess the model's accuracy by comparing anticipated and actual outcomes.

To do so, the study will divide the research population into 2 groups, the training dataset which will be the data for the first 9 years and the testing dataset which will be the data for the last 3 years. The training dataset will be used to train our model and the testing dataset will be used to validate the outcome of the model.

3.5.4.2 Error Metrics

To measure the predictability of our random forest model based on the out of sample test above, we will employ the use of the following standard error metrics for a classification problem:

1. **Accuracy**

Accuracy measures the portion of all testing samples classified correctly. It is calculated by the formula below:

$$\textit{Accuracy} = \frac{tp+tn}{tp+tn+fn+fn}$$

2. **Recall**

Recall (also known as sensitivity) measures the ability of a classifier to correctly identify positive labels. It is calculated by the formula below:

$$\textit{Recall} = \frac{tp}{tp+fn}$$

3. **Specificity**

Specificity measures the classifier's ability to correctly identify negative labels. It is calculated by the formula below:

$$\textbf{\textit{Specificity}} = \frac{tn}{tn+fp}$$

4. **Precision**

Precision measures the proportion of all correctly identified samples in a population of samples which are categorized as positive labels. It is calculated by the formula below:

$$\textbf{\textit{Precision}} = \frac{tp}{tp+fp}$$

5. **F Measure**

F Score (also known) as F-Measure is the mean of precision and recall. It is calculated by the formula below:

$$\textbf{\textit{F measure}} = \frac{2*(precision*recall)}{(precision + recall)}$$

Where:

tp = Number of true positive values

tn = Number of true negative values

fp = Number of false positive values

fn = Number of false negative values

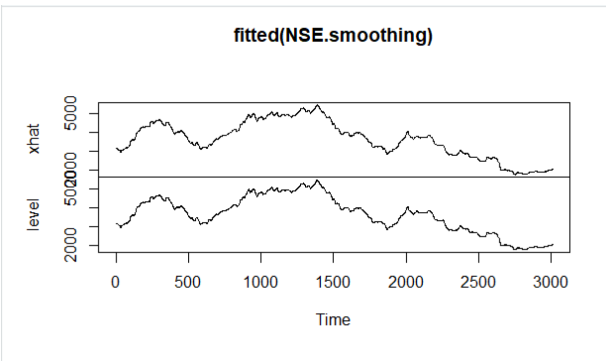
CHAPTER FOUR:
DATA ANALYSIS, RESULTS AND DISCUSSION

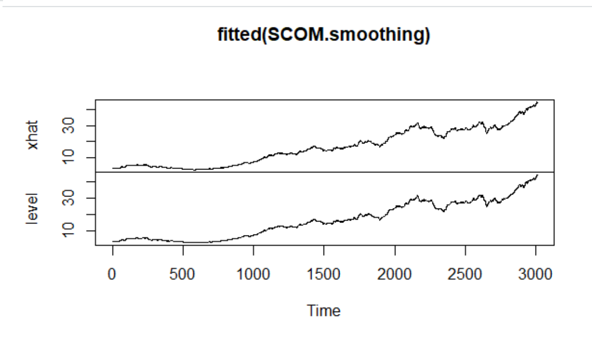
4.1 Introduction

This chapter outlines the results of the data analysis and the findings. This research aimed at finding out the effectiveness of using random forest model in modelling stock market volatility in NSE. Technical analysis which is the use of charts and graphs to study stock prices and volume data for the purpose of forecasting future trends, was used as the main technique of analysis. This paper used R software for the data analysis which is a language for statistical computing and graphics.

4.2 Data Cleaning & Pre-Processing

Secondary Data was obtained from Nairobi Securities Exchange comprising of 3200 daily closing prices of NSE-20 Index and Safaricom stock. The number of years under consideration 2009-2021(by date) were entered in the columns and the share prices were entered into the rows. Data cleaning was done to remove all the missing values for the days that trading did not occur. The data was then converted into an extensible time series (xts) which allowed for smoothing using exponential moving average. An alpha of 0.2 was used as the smoothing factor which assigned more weight to the recent share prices and exponentially less weights to the past share prices.





4.3 Feature Extraction

Technical indicators are significant variables determined from stock time - series data which provide insights to the expected stock price behaviour in future. In this study five technical indicators namely, RSI, ROC, MACD, William % R and Stochastic Oscillator were used to train the random forest. These technical indicators were used to extract the trend in the form of buy and sell signals from the NSE-20 and Safaricom time series. To calculate the features, we looked at every trading date from 2009 through 2021.

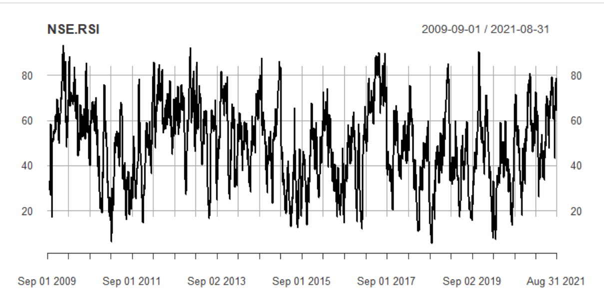
The table below shows a sample of the values of the technical indicators that we obtained from NSE-20 index and Safaricom:

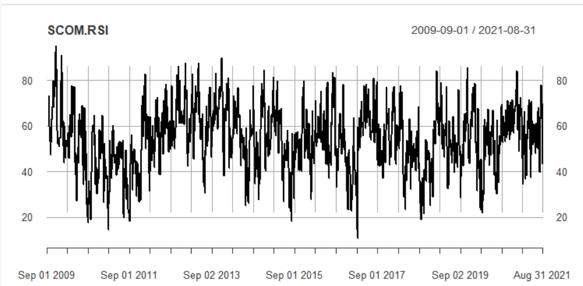
```
> tail(get_indicators(NSE,20)$first.df,20)
                RSI  StoFASTK  StoFASTD  StoSLOWD  WilliamPR      MACD MACDSignal PriceRateOfChange
2021-07-06 65.65018 100.00000  93.85275  95.54370   0.000000 0.5551918  0.4929937          1.277560
2021-07-07 68.73891 100.00000  93.85275  93.67368   0.000000 0.6036607  0.5151271          1.365863
2021-07-08 70.77766 100.00000 100.00000  95.90183   0.000000 0.6580787  0.5437174          1.508929
2021-07-09 72.89301 100.00000 100.00000  97.95092   0.000000 0.7190551  0.5787850          2.044422
2021-07-12 66.68755  87.69432  95.89811  98.63270 -12.305678 0.7272838  0.6084847          1.259513
2021-07-13 70.13876 100.00000  95.89811  97.26540   0.000000 0.7644455  0.6396769          2.288226
2021-07-14 70.85898 100.00000  95.89811  95.89811   0.000000 0.7932809  0.6703977          2.276846
2021-07-15 75.24378 100.00000 100.00000  97.26540   0.000000 0.8654944  0.7094170          3.157267
2021-07-16 75.93545 100.00000 100.00000  98.63270   0.000000 0.9222167  0.7519769          3.672506
2021-07-19 79.08019 100.00000 100.00000 100.00000   0.000000 1.0076685  0.8031153          4.232878
```

```
> tail(SCOM.first, 20)
                RSI  StoFASTK  StoFASTD  StoSLOWD  WilliamPR      MACD MACDSignal PriceRateOfChange
2021-07-06 52.85967  46.15385  51.72414  64.98674  -53.84615 0.4966127  0.5828359         0.8510690
2021-07-07 53.95305  53.84615  51.28205  57.55968  -46.15385 0.4822669  0.5627221         0.9708814
2021-07-08 53.95305  53.84615  51.28205  51.42941  -46.15385 0.4655071  0.5432791         0.3629768
2021-07-09 51.86050  42.30769  50.00000  50.85470  -57.69231 0.4180830  0.5182399         0.0000000
2021-07-12 50.45555  34.61538  43.58974  48.29060  -65.38462 0.3568417  0.4859602        -0.7263954
2021-07-13 51.16788  38.46154  38.46154  44.01709  -61.53846 0.3144630  0.4516608        -0.9661911
2021-07-14 58.27459  80.76923  51.28205  44.44444  -19.23077 0.3839348  0.4381156        -0.4778982
2021-07-15 66.41051 100.00000  73.07692  54.27350    0.00000 0.5973764  0.4699678         2.3754086
2021-07-16 66.41051 100.00000  93.58974  72.64957    0.00000 0.7567801  0.5273302         3.2203140
2021-07-19 66.41051 100.00000 100.00000  88.88889    0.00000 0.8722485  0.5963139         3.0991753
```

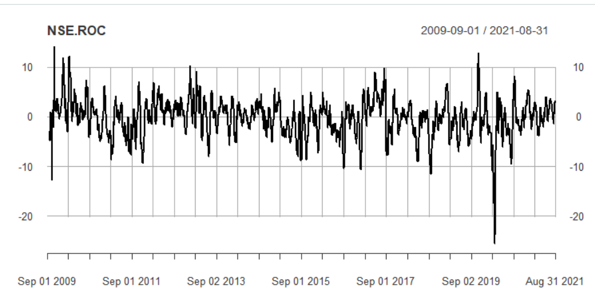
Interpretation and Visualization of the Technical Indicators

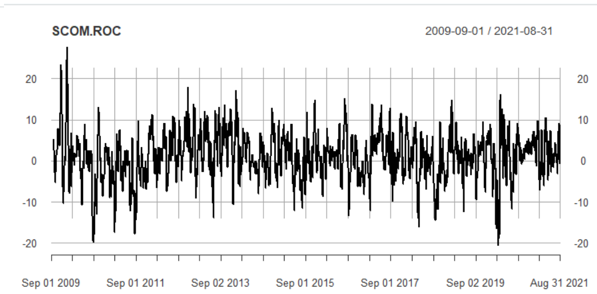
RSI – It oscillates from 0 to 100. An oversold stock is shown by a value of 30 and below which gives the Buy signal while an overbought stock is shown by a value of 70 and above which gives the Sell Signal otherwise Hold.



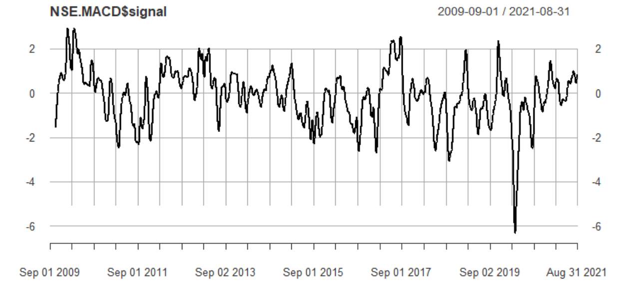


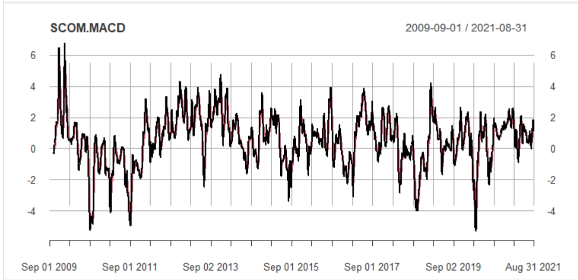
ROC – Positive values indicates an oversold stock hence gives the Buy signal and negative values indicates an overbought stock hence gives the Sell Signal.



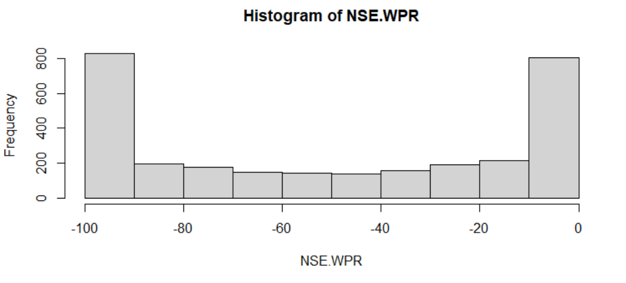


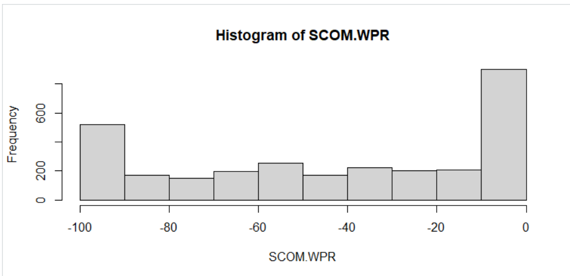
MACD - A probable sell signal is generated when the MACD crosses below the zero line, and a probable purchase signal is formed when it crosses above the zero line.



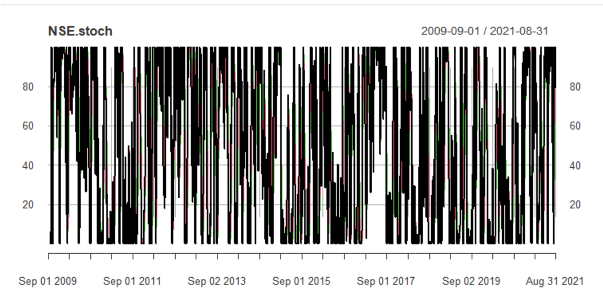


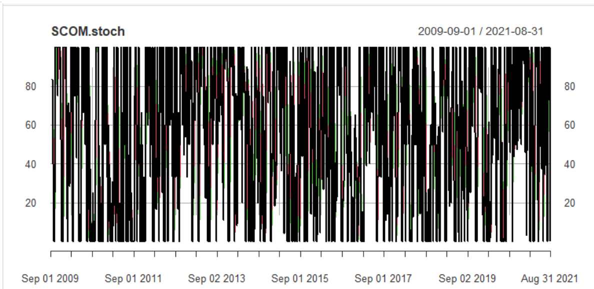
William % R – It oscillates from 0 to -100. A value below -80 indicates an oversold stock hence gives the Buy signal and a value above -20 indicates an overbought stock hence gives the Sell Signal otherwise Hold.





Stochastic Oscillator - It oscillates from 0 to 100. A value below 20 indicates an oversold stock hence gives the Buy signal and a value of above 80 indicates an overbought stock hence gives the Sell Signal otherwise Hold.





The table below shows a sample of the trend (Buy, Hold, Sell) signal that was extracted from the above technical indicators.

```
> tail(get_indicators(NSE,20)$second.df,20)
           RSIBuyorSell StoBuyorSell WilliamPRBuyorSell MACDBuyorSell PROCBuyorSell Response
2021-07-06         Hold         Sell               Sell           Buy           Buy      Buy
2021-07-07         Hold         Sell               Sell           Buy           Buy      Buy
2021-07-08         Sell         Sell               Sell           Buy           Buy     Sell
2021-07-09         Sell         Sell               Sell           Buy           Buy     Sell
2021-07-12         Hold         Sell               Sell           Buy           Buy      Buy
2021-07-13         Sell         Sell               Sell           Buy           Buy     Sell
2021-07-14         Sell         Sell               Sell           Buy           Buy     Sell
2021-07-15         Sell         Sell               Sell           Buy           Buy     Sell
2021-07-16         Sell         Sell               Sell           Buy           Buy     Sell
2021-07-19         Sell         Sell               Sell           Buy           Buy     Sell
```

```
> tail(SCOM.second, 20)
           RSIBuyorSell StoBuyorSell WilliamPRBuyorSell MACDBuyorSell PROCBuyorSell Response
2021-07-06         Hold         Hold               Hold           Buy           Buy     Hold
2021-07-07         Hold         Hold               Hold           Buy           Buy     Hold
2021-07-08         Hold         Hold               Hold           Buy           Buy     Hold
2021-07-09         Hold         Hold               Hold           Buy          Hold     Hold
2021-07-12         Hold         Hold               Hold           Buy          Sell     Hold
2021-07-13         Hold         Hold               Hold           Buy          Sell     Hold
2021-07-14         Hold         Sell               Sell           Buy          Sell     Sell
2021-07-15         Hold         Sell               Sell           Buy           Buy      Buy
2021-07-16         Hold         Sell               Sell           Buy           Buy      Buy
2021-07-19         Hold         Sell               Sell           Buy           Buy      Buy
```

4.4 Data Analysis & Findings

Implementation of the Random Forest Algorithm

Evaluation of the random forest model was done by the use of the trend extracted above. These five technical indicators were used as the model's features while the response variable was either Buy, Sell or Hold. The data was then partitioned into two, first 9 years and the last 3 years: that is from (1st Sept 2009 – 31st Aug 2018) and (1st Sept 2018 – 31st Aug 2021). The first 9 years of the data was used to train the random forest model while the last 3 years of the data was used to test the trained model.

Output of the trained model was as follows:

```
 randomForest(formula = Response ~ ., data = NSE.df.train, importance = TRUE,      proximity = TRUE, mtry = 2)
               Type of random forest: classification
                     Number of trees: 500
No. of variables tried at each split: 2

        OOB estimate of  error rate: 1.15%
Confusion matrix:
     Buy Hold Sell class.error
Buy  968    0    0  0.00000000
Hold  26  780    0  0.03225806
Sell   0    0  485  0.00000000
```

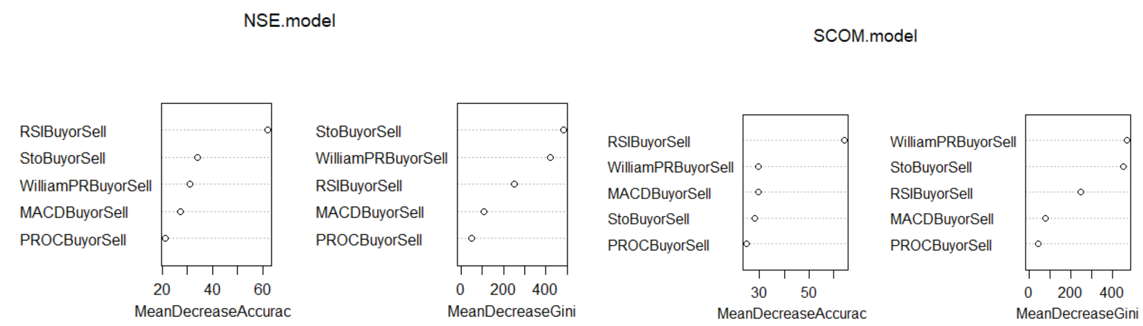
```
 randomForest(formula = Response ~ ., data = SCOM.df.train, importance = TRUE,      proximity = TRUE, mtry = 2)

               Type of random forest: classification
                     Number of trees: 500
No. of variables tried at each split: 2

        OOB estimate of  error rate: 1.2%
Confusion matrix:
     Buy Hold Sell class.error
Buy  751    0    0  0.00000000
Hold  27 1024    0  0.02568982
Sell   0    0  457  0.00000000
```

NSE-20 model had an out of bag error of 1.15% which meant the model was trained at 98.85% accuracy rate while Safaricom model had an OOB error of 1.2% which meant that the model was trained at 98.8% accuracy. From the confusion matrix above, we can see that the “Hold” signal was the only response variable with a class error for both NSE-20 and Safaricom model.

For both models, random forest created 500 classification trees using sample with replacement and predicted the response variable based on majority vote. The number of trees used was inversely proportionate to the OOB error. It was determined that the optimal number of variables to be used to split each tree with the least error was 2 features. Each tree was split using Gini impurity based on the importance and weight of each feature. For both NSE-20 and Safaricom model, RSI was the most significant feature and was used as the root node while price rate of change was the least significant feature and could be done away with. The table below shows the importance of each feature for the 2 models.



Validation of the Random Forest Algorithm

To validate the above NSE and Safaricom model, the last 3 years of the data (1^{st} Sept 2018 – 31^{st} Aug 2021) was used to test the trained model. The below output was extracted from the predicted model:

```
> NSE.pred <- predict(NSE.model, NSE.df.test[,c("RSIBuyorSell", "StoBuyorSell",
+                                                "WilliamPRBuyorSell", "MACDBuyorSell", "PROCBuyorSell",
+                                                "Response")])
> confusionMatrix(NSE.pred, NSE.df.test$Response)
Confusion Matrix and Statistics

          Reference
Prediction Buy Hold Sell
      Buy  314    0    0
      Hold   4  191    0
      Sell  38    0  197

Overall Statistics

               Accuracy : 0.9435
                 95% CI : (0.9245, 0.959)
    No Information Rate : 0.4785
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9128

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: Buy Class: Hold Class: Sell
Sensitivity              0.8820      1.0000      1.0000
Specificity              1.0000      0.9928      0.9305
Pos Pred Value           1.0000      0.9795      0.8383
Neg Pred Value           0.9023      1.0000      1.0000
Prevalence               0.4785      0.2567      0.2648
Detection Rate           0.4220      0.2567      0.2648
Detection Prevalence     0.4220      0.2621      0.3159
Balanced Accuracy        0.9410      0.9964      0.9653
```

```
> SCOM.pred <- predict(SCOM.model, SCOM.df.test[,c("RSIBuyorSell", "StoBuyorSell",
+                                                  "WilliamPRBuyorSell", "MACDBuyorSell", "PROCBuyorSell",
+                                                  "Response")])
> confusionMatrix(SCOM.pred, SCOM.df.test$Response)
Confusion Matrix and Statistics

          Reference
Prediction Buy Hold Sell
      Buy  155    0    0
      Hold   0  322    0
      Sell 147    0  113

Overall Statistics

               Accuracy : 0.8005
                 95% CI : (0.7698, 0.8288)
    No Information Rate : 0.4369
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.7018

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: Buy Class: Hold Class: Sell
Sensitivity              0.5132      1.0000      1.0000
Specificity              1.0000      1.0000      0.7644
Pos Pred Value           1.0000      1.0000      0.4346
Neg Pred Value           0.7474      1.0000      1.0000
Prevalence               0.4098      0.4369      0.1533
Detection Rate           0.2103      0.4369      0.1533
Detection Prevalence     0.2103      0.4369      0.3528
Balanced Accuracy        0.7566      1.0000      0.8822
```

An N x N matrix that is used to assess the effectiveness of a classification model, with N denoting the number of target variables is known as the confusion matrix. The matrix compares the actual goal values to the predictions of the machine learning model, giving us a complete view of how well our classification model is performing and the types of errors it makes.

The NSE-20 model and Safaricom model both generated its confusion matrix which was used in the discussion of the findings below.

4.5 Summary and Interpretation of Findings

The use of a confusion matrix was employed to analyse the robustness of the model.

NSE-20 model was able to predict the 'Buy', 'Sell' and 'Hold' signal from the test data with an accuracy of 94.35%. Analysing NSE-20 confusion matrix, it was able to predict accurately a buy signal for 314 days, a sell signal for 197 days and a hold option for 191 days. This model, however, was not 100% accurate and it resulted into a few misclassifications i.e., NSE 20 model predicted a buy signal for 4 days that an investor was supposed to hold and 38 days when an investor was supposed to sell.

The table below show some extracted predicted vs actual shares from R output.

```
> NSE.mets <- data.frame(date=as.Date(names(NSE.pred)),actual=NSE.df.test$Response,predicted=NSE.pred)
> tail(NSE.mets, 10)
                 date actual predicted
2021-08-09 2021-08-09   Sell      Sell
2021-08-10 2021-08-10   Hold      Hold
2021-08-11 2021-08-11    Buy       Buy
2021-08-12 2021-08-12    Buy       Buy
2021-08-13 2021-08-13   Hold      Hold
2021-08-16 2021-08-16   Hold      Hold
2021-08-17 2021-08-17    Buy      Sell
2021-08-18 2021-08-18    Buy      Sell
2021-08-19 2021-08-19    Buy      Sell
2021-08-20 2021-08-20   Sell      Sell
```

NSE-20 Actual Vs Predicted Excerpt

The following error metrics were used to evaluate the NSE-20 model as extracted from above output:

Accuracy

$$=0.9435 = 94.35\%$$

Average Sensitivity for the model:

$$= \frac{0.8820+1.0000+1.0000}{3} = 0.9607 = 96.07\%$$

Average Specificity for the model:

$$= \frac{1.0000+0.9928+0.9305}{3} = 0.9744 = 97.44\%$$

Average Precision of the model:

$$= \frac{1.0000+0.9795+0.8383}{3} = 0.9393 = 93.93\%$$

F-Measure of the model:

$$\boldsymbol{F\ measure} = \frac{2*(\boldsymbol{precision}*\boldsymbol{recall})}{(\boldsymbol{precision}+\boldsymbol{recall})}$$

$$= \frac{2*(0.9393*0.9607)}{(0.9393+0.9607)} = 0.9499$$

F measure oscillates between 0 and 1 and it gives us the information on the classifier's precision (the number of instances it properly classifies) and robustness (it does not miss a significant number of instances) hence a model with an F-Measure that is 1 or close to 1 is preferred to the latter. For the case of the NSE-20 model, it had an F-Measure of 0.9499 which meant that the NSE-20 model was able to accurate predict the buy, hold, and sell instances and at the same time it did not a significant number of instances in this case the buy, hold and sell signal.

SCOM model was able to predict the 'Buy', 'Sell' and 'Hold' trend from the test data with an accuracy of 80.05%. Analysing its confusion matrix, it was able to predict accurately a buy signal for 155 days, a sell signal for 113 days and a hold option for 322 days. The 20% inaccuracy rate of the model was a result of the misclassifications i.e., SCOM model predicted a buy signal for147 days that an investor was supposed to sell the stock.

The table below show some extracted predicted vs actual shares from R output.

```
> SCOM.mets <- data.frame(date=as.Date(names(SCOM.pred)),actual=SCOM.df.test$Response,predicted=SCOM.pred)
> tail(SCOM.mets, 10)
                 date actual predicted
2021-07-29 2021-07-29   Hold      Hold
2021-07-30 2021-07-30   Hold      Hold
2021-08-02 2021-08-02    Buy       Buy
2021-08-03 2021-08-03    Buy       Buy
2021-08-04 2021-08-04    Buy       Buy
2021-08-05 2021-08-05    Buy       Buy
2021-08-06 2021-08-06   Hold      Hold
2021-08-09 2021-08-09    Buy      Sell
2021-08-10 2021-08-10    Buy      Sell
2021-08-11 2021-08-11    Buy      Sell
```

Safaricom Actual Vs Predicted Excerpt

The following error metrics were used to evaluate the SCOM model as extracted from above output:

Accuracy

=0.8005 = 80.05%

Average Sensitivity for the model:

$$= \frac{0.5132+1.0000+1.0000}{3} = 0.8377 = 83.77\%$$

Average Specificity for the model:

$$= \frac{1.0000+1.0000+0.7644}{3} = 0.9215 = 92.15\%$$

Average Precision of the model:

$$= \frac{1.0000+1.0000+0.4346}{3} = 0.8115 = 81.15\%$$

F-Measure of the model:

$$\boldsymbol{F\ measure} = \frac{\mathbf{2}*(\boldsymbol{precision}*\boldsymbol{recall})}{(\boldsymbol{precision}+\boldsymbol{recall})}$$

$$= \frac{2*(0.8115*0.8377)}{(0.8115+0.8377)} = 0.8244$$

For the case of SCOM model, it had an F-Measure of 0.8244 which meant that the SCOM model was able to accurately predict the buy, hold, and sell instances and at the same time it did not miss a notable number of instances in this case the buy, hold and sell signal.

One of the objectives of this paper was to conduct a comparison analysis on how machine learning of an index would compare to that of a single stock. An index fund is a type of investment that tracks the performance of a component of the financial market index. This said, NSE-20 model had an accuracy rate of 94.35% with an F-Measure of 0.9499 while SCOM model had an accuracy rate of 80.05% with an F-Measure of 0.8244. We saw that NSE-20 model was a more robust model when compared to SCOM. This was as result of the diversification aspect offered by an index which greatly reduces the volatility as compared to SCOM which was highly susceptible to the market changes and movements. Historically, we know that most passive traders invest in indexes while the day traders invest in single stocks. Although NSE 20 model outperforms the SCOM model, I believe that day traders would be the ideal target users of these predicting models and can still go ahead to utilize this model.

This research aimed to answer the question whether random forest could be used as an effective model of predicting stock market volatility at NSE? From the above findings, we have been able to answer this research question and yes, random forest can be used as an effective model of predicting stock market volatility at NSE.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATIONS

5.1 Introduction

This chapter is a summary of the complete study, and it includes a summary of the research findings, an explanation of the findings that corresponds to the main objective, conclusions, and recommendations based on the findings.

5.2 Summary & Conclusions

The main objective of this study was to assess the effectiveness of modelling stock market volatility using random forest. We saw that ensemble learning models remained unexploited in the field of predicting stock market using machine learning. The prediction performance of the Random Forest was analysed based on 12 years (2009-2021) of historical data of NSE-20 index and Safaricom share price obtained from NSE.

5 technical indicators were used to extract trend from the past historical prices and this trend was used to train the random forest model. It was discovered that using technical indicators that were highly correlated led to data leakage and to curb this Stochastic Oscillator and William's % R were added, and simple moving average removed to the list of technical indicators. Choosing the right technical indicators to be used as the input parameter was the most crucial part of this study. The concept of GIGO, Garbage In, Garbage Out came into play here, where the quality of the inputs used determined the quality of output that was obtained from both the NSE-20 model and SCOM model. The models' output was discrete in nature that is it worked by predicting when to 'BUY, 'HOLD' or 'SELL' NSE-20 and the Safaricom stocks.

NSE-20 Model Results:

Accuracy	94.35%
F-Measure	0.9499

SCOM Model Results:

Accuracy	80.05%
F-Measure	0.8244

The results from our model were quite impressive as it was able to achieve 94.35% & 80.05% prediction accuracy rate for the NSE-20 and Safaricom daily share prices movement. NSE-20 model was found to be more robust than SCOM. This was attributed to the volume of stocks traded behind NSE 20 which allowed for the detection of possible trends. Also, NSE 20 was less volatile because it is a weighted average of the 20 single stocks. With the high accuracy rate, this model can actually be deployed in real time for stock's trend prediction and can be utilised by traders and investors to devise new trading strategies which will make investments more profitable and secure.

In chapter 2, we extensively discussed the empirical reviews of research papers on random forest modelling that had been done by other scholars. We saw that Luckyson Khaidem, Snehanshu Saha and Sudeepa Roy Dey, (2016) utilized the random forest classier to create their predictive model in forecasting the direction of stock market prices, where they obtained long-term prediction accuracy in the region of 85-95 percent. Also, Seyed Alavi, Hasanali Sinaei, Elham Afsharirad (2015), conducted a study on predicting the trend of stock prices using machine learning techniques any they were able to get the accuracy rate of 91.48% for random forest, 84.27% for SVM and 81.38% for KNN. In comparison, to this study's outcome, we can see that the accuracy rate of the random forest from the 3 papers ranges from 80% to 95% ascertaining the efficiency of a random forest model.

However, a word of caution should be sounded off in that although investors can use random forest to predict whether to buy, hold or sell shares on a specific day, they have a 5.65% & 19.95% (NSE-20 & Safaricom model) chance of inaccurate predictions, but the trade-off between the money one could make from using the random forest model outweighs the risk of the money one would lose due to incorrect predictions.

5.3 Recommendations

The study established that the NSE-20 and SCOM model had a high accuracy and next day prediction capabilities. It, therefore, recommends traders and investors to utilize these tools as they can be certain of reaping high returns by making well informed investments decisions on when to buy, hold or sell a stock which in turn it will boost the growth of the economy

The study also found out that predicting an index gave more impressive results when compared to the single stocks as the NSE 20 model outperformed Safaricom model with a significant margin. The study, therefore, recommends that since indices cannot be traded, brokerage companies in Kenya should create funds that closely track the NSE-20 index that will give traders an opportunity to utilize the high accuracy of the predictive tool.

5.4 Limitations of the study

One of the challenges that faced the study was data leakage which occurred when information from the testing dataset was made available to the model in the training dataset. This leakage although tiny and undetectable had a significant impact on performance of the models. To overcome this, new technical indicators were used which greatly reduced the correlation between the training dataset and testing dataset.

The main limitation of the study is that we cannot tell how long a trained random forest model will remain valid and effective in prediction before it requires to be retrained. More research is needed to be done to ascertain this.

5.5 Areas of Further Research

The study found out that the model inputs was the most critical aspect in model building which in this study was the technical indicators. The study, therefore, recommends future researchers to use multiple technical indicators as predictor variables to see if the number of technical variables affects the accuracy of the model. Also, they should incorporate other indicators such as the macroeconomic indicators and sentiment data.

The study, also, recommends to future researchers to use multiple machine learning models when modelling same dataset so that they can be able to have a comparison basis. It would be prudent to compare the accuracy and output between two models such as comparing the performance of random forest with the performance of KNN or SVM to allow cross validation.

REFERENCES

Aamir, M., & Shah, S. Z. A. (2011). Dividend announcements and the abnormal stock returns for the event firm and its rivals. *Australian Journal of Business and Management Research, 8*, 72-76.

ABC Capital. (2013). *Research*. Retrieved September 8, 2013, from http://www.abccapital.co.ke/research.php

Adebiyi, A. A., Ayo, C. K., Adebiyi, M. O., & Otokiti, S. O. (2012). Stock Price Prediction Using Neural Network with Hybridized Market Indicators. *Journal of Emerging Trends in Computing and Information Sciences, 3* (1), 1-9.

Adhikari, R., & Agrawal, R. K. (2013). A Combination of Artificial Neural Network and Random Walk Models for Financial Time Series Forecasting. *Neural Computing and Applications*.

Africa Alliance. (2013). *Research.* Retrieved April 3, 2013, from Africa Alliance: http://www.africanalliance.com/securities/research

African Fine Coffees Association (AFCA). (2013). *Kenya's Sasini in KShs. 349 million*

Aghababaeyan, R., & TamannaSiddiqui, N. (2011). Forecasting the Tehran Stock Market by Artificial Neural Network. *International Journal of Advanced Computer Science and Applications, Special Issue on Artificial Intelligence*.

Aharony, J., &Swary, I. (1980). Quarterly dividend and earnings announcements and stockholders' returns: An empirical analysis. *The Journal of Finance, 35(1),* 1 -12.

AIB Capital Ltd. (2013). *Research*. Retrieved September 8, 2013, from AIB Capital Ltd: http://www.afrikainvestmentbank.com/index.php?option=com_content&view=article &id=54 &Itemid=55

Akram K & Abrar M (2015). Stock Price Volatility and Dividend Policy in Pakistan. *International Journal of Scientific and Research Publications,* Volume 5, Issue 2 pp 1-6.

Allen, Dave E and Rachim, Veronica S.*" Dividend policy and stock price volatility: Australian evidence*. “Applied Financial Economics, 1996, 6, 175-188

Apex Africa Capital Ltd. (2013). *Apex Africa Capital Ltd.* Retrieved September 8, 2013, from http://apexafrica.com/?cat=5

Asquith, P., & Mullins, D. W. (1983). The impact of initiating dividend payments on shareholders wealth. *Journal of Business*, 56(1), 77-96.

Bahrammirzaee, A. (2010). A Comparative Survey of Artificial Intelligence Applications in Finance: Artificial Neural Networks, Expert System and Hybrid Intelligent Systems. *Neural Computing and Applications, 19* (8), 1165-1195.

Baskin, J. (1989). Dividend Policy and the Volatility of Common Stock. *Journal of portfolio Management*, 15 (3), 19-25.

Battacharya, S. (1979), Imperfect information & dividend policy and the 'bird in hand' fallacy, *The Bell Journal of Economics*, Vol. 10, pp. 259-70.

Ben-David, S., Blitzer, J., Crammer, K., Kulesza, A., Pereira, F., & Vaughan, J. W. (2010). A theory of learning from different domains. *Machine Learning, 79* (1-2), 151-175.

Bethard, S. (2008). *Statistical Learning Methods – Introduction to Artificial Intelligence.* Retrieved April 12, 2013, from Lecture notes online: http://clear.colorado.edu/~bethard/teaching/csci3202_2008/slides/learning-methods.pdf

Black, F., & Sholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy,* 81(4), 637-654.

Borodin, A., El-Yaniv, R., & Gogan, V. (2004). Can We Learn to Beat the Best Stock? *Journal of Artificial Intelligence Research, 21*, 579-594.

Brealey, R. A., Myers, S. C., & Allen, F. (2011). Principles of corporate finance (10th Ed.). New York: The McGraw Hill/Irwin.

Butler, M., & Daniyal, A. (2009). Multi-objective Optimization with an Evolutionary Artificial Neural Network for Financial Forecasting. *Proceedings of the 11th Annual conference on Genetic and evolutionary computation*, (pp. 1451-1457).

Capital Markets Authority. (2013). *Licensed Stockbrokers*. Retrieved April 12, 2013, from http://www.cma.or.ke/index.php?option=com_content&view=article&id=208&Itemid=283

Central Bank of Kenya. (2011). *Financial Sector Stability Report 2011.* Retrieved April 12, 2013, from http://www.centralbank.go.ke/index.php/kenya-financial-sector-stability

Central Bank of Kenya. (2012). *Annual Report 2012.* Retrieved April 10, 2013, from http://www.centralbank.go.ke/index.php/cbk-annual-reports

Central Bank of Kenya. (2012, December). *Monthly Economic Review, December 2012.* Retrieved April 3, 2013, from http://www.centralbank.go.ke/index.php/monthly-economic-review

Central Bank of Kenya. (2012, June). *Statistical bulletin.* Retrieved April 3, 2013, from http://www.centralbank.go.ke/index.php/statistical-bulletin

Central Bank of Kenya. (2013, March 28). *Weekly bulletin.* Retrieved April 12, 2013, from http://www.centralbank.go.ke/index.php/weekly-bulletin

Central Depository and Settlement Corporation Limited (CDSC). (2013). *About Us*. Retrieved April 3, 2013, from http://www.cdsckenya.com/about-us/company-profile/

Cerna, L., & Chytry, M. (2005). Supervised Classification of Plant Communities with Artificial Neural Networks. *Journal of Vegetation Science, 16*, 407-414.

Chang, P., & Fan, C. (2008). A Hybrid System Integrating a Wavelet and TSK Fuzzy Rules for Stock Price Forecasting. *IEEE Transactions on Systems, Man and Cybernetics – Part C: Applications and Reviews, 38* (6), 802-815.

Chase Bank. (2013). *Investing in Kenya*. Retrieved April 3, 2013, from http://www.chasebank.co.ke/product/investing-kenya

Chen, M., Lin, C., & Chen, A. (2007). Constructing a Dynamic Stock Portfolio Decision-Making Assistance Model: Using the Taiwan 50 Index Constituents as an Example. *Soft Computing, 11* (12), 1149-1156.

Chen, Y., & Cheng, C. (2007). Forecasting Revenue Growth Rate Using Fundamental Analysis: A Feature Selection Based Rough Sets Approach. *Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007), 3*, pp. 151-155.

Chen, Y., Wu, Q., & Chen, F. (2007). An IP and GEP Based Dynamic Decision Model for Stock Market Forecasting. *ISNN '07 Proceedings of the 4th international symposium on Neural Networks: Advances in Neural Networks*, (pp. 473-479).

Code Project. (2011). *Financial Predictor via Neural Network*. Retrieved August 6, 2013, from http://www.codeproject.com/Articles/175777/Financial-Predictor-via-Neural-Network

Dechow, P. M., & Skinner, D. J. (2000). Earnings management: Reconciling the views of accounting academics, practitioners, and regulators. *Accounting Horizons*, (June), 235

Gordon, M. (1963). Optimal Investment and Financing Policy. *The Journal of Finance*, 264-272.

Gordon, M. J. (1963). Optimal investment and financing policy. *Journal of Finance*, 18, 264-272.

Groenewold, N., O'Rourke, G., & Thomas, S., (1997). Stock returns and inflation: A macro Analysis. *Applied Financial Economics,* 7(2), 127-136.

Hussainey, K., Mgambe, C. O., &Chijoke-Mgbame, A. M. (2010). Dividend policy and share price volatility: UK Evidence. *Journal of Risk Finance*, 12 (1), 57-68.

ImadZeyad Ramadan Dividend Policy and Price Volatility. Empirical Evidence from Jordan International *Journal of Academic Research in Accounting, Finance and Management Sciences* Vol. 3, No.2, April 2013, pp. 15–22

Jensen, M. C. (1986). Agency costs of free cash flow, corporate finance, and takeovers. *American Economic Review,* 76(2), 323-329.

Khaled, H., Chijoke, O. M., & Aruoriwo, M. C. (2011). Dividend policy and share price volatility: UK evidence. *Journal of risk finance*, 12 (1), 57 – 68.

Kibet T (2016) Effects of Dividend Policy on Share Price of Firms Listed at the Nairobi Securities Exchange, Kenya. *Research Journal of Finance and Accounting,* Vol.7, No.8, 2016 pp 220-230.

Kibet, B., Joel, T., Cheruiyot, T. K., Maru, L., &Kipsat, M. (2010). The level of corporate dividend pay-out to stockholders: Does optimal dividend policy exist for firms quoted at the Nairobi Stock Exchange? *International Business & Economics Research Journal*, *3*, 71- 84.

Kinder, C. (2002). Estimating Stock Volatility: mimeo.

Laabs, D. S., & Bacon, F. W. (2013). The impact of increased dividend announcements on stock price: A test of market efficiency. *ASBBS Annual Conference: Las Vegas*, *20(1)*, 664-670.

Laabs, D. S., & Bacon, F. W. (2013). The impact of increased dividend announcements on stock price: A test of market efficiency. *ASBBS Annual Conference: Las Vegas*, *20(1)*, 664-670.

Litner, J. (1962). Dividends, Earnings, Leverage, Stock Prices, and the Supply of Capital to Corporations. *Review of Economics and Statistics,* 243-269.

Litzenberger, R. & Ramaswamy, K. (1979). The Effects of Personal Taxes and Dividends on Capital Asset Prices: Theory and Empirical Evidence, *Journal of Financial Economics,* 163- 195.

Mahmood S, Sheikh M &Ghaffari A(2011) Dividend Announcements and Stock Returns: An event study on Karachi stock exchange, Interdisciplinary Journal Of Contemporary Research In Business Volume 3, No 8 pp;972-981.

Marekia A (2015). The effect of dividend announcement on share prices of firms listed at the NSE. Unpublished MBA project, University of Nairobi.

Miller, M.H. and Modigliani, F. (1961), Dividend policy, growth and the valuation of shares, *The Journal of Business,* Vol. 34, pp. 411-33.

Mohammad Hashemijoo, ArefMahdaviArdekanihe Impact of Dividend Policy on Share Price Volatility in the Malaysian Stock Market *Journal of Business Studies* Quarterly 2012, Vol. 4, No. 1, pp. 111-129

Morck, R., Yeung, B., & Yu, W. (1997). The information content of stock markets: Why do emerging markets have so little firm-specific risk? *Research Seminar in International Economics: The University of Michigan*.

Muhammad Mahbubur Rahman (2011). Impact of Dividends on Stock Price Volatility – Evidence from Some Selected listed Companies in Bangladesh, Abasyn Journal of Social Sciences. Vol: 7 Issue: 1 pp;47-56

Mukora, M. Y. W. (2014). The effect of dividend announcement on stock returns of firms listed at the NSE, Unpublished MBA project, University of Nairobi.

Muriuki, J. (2010). Beta firm size, book-to-market equity and stock returns: Evidence from the Nairobi Stock Exchange, Unpublished MBA project, University of Nairobi.

Nishat, M., &Irfan, C. M. (n.d.). Dividend policy and stock price volatility in Pakistan. University of Karachi.

Onsunu Z & Onchiri M (2014). The Effect of Dividend Policy on Stock Price Volatility: A Kenyan Perspective*, the International Journal of Business & Management* Vol 2 Issue 7 pp 115-120.

Oyuga C (2013). The effects of earnings announcement on the share price for firms listed at the NSE. Unpublished MBA project, University of Nairobi, Kenya.

Pandey, I. M. (2011). Financial Management (10th ed.): Issues in divided policy. New Delhi: Vikas Publishing House PVT Ltd.

Poornima S & Chitra V (2015) Stock Market Volatility during Dividend Announcement a Case Of Selected Banks In India, Asia Pacific Journal of Research, Volume 1, Issue XXXIV pp 133-138

Rashid, Afzalur; Rahman, A Z M Anisur (2008) Dividend Policy and Stock Price Volatility: Evidence from Bangladesh, *The Journal of Applied Business and Economics,* pp: 71-81.

Reilly F. & Brown C. (1997). Investment Analysis and Portfolio Management 5th Edition. New Jersey: Prentice Hall Inc.

Ross, S. (1977). The Determination of Financial structure: The Incentive- Signalling Approach. *Bell Journal of Economics*, 23-40

Sitthipongpanich, T. (2011). Understanding the Event Study. *Journal of Business Administration*, *34(130),* 59-68.

Vazakidis, A., &Athianos, S. (2010). Do dividend announcements affect the stock prices in the Greek Stock Market? *International Journal of Economic Sciences and Applied Research*, *3(2)*, 57-77.

Walter, J. E. (1956). Dividend policies and common stock prices. *The Journal of Finance,* 11(1), 29-41.