2023

Identifying the optimal time series model to predict Kenyan stock prices.

Moenga, Paulyne Kerubo
Strathmore Institute of Mathematical Sciences
Strathmore University

Identifying the Optimal Time Series Model to Predict Kenyan Stock Prices

By

Paulyne Kerubo Moenga

Registration Number: 142282

Master of Science in Statistical Science

July 2023

Identifying The Optimal Time Series Model to Predict Kenyan Stock Prices

Paulyne Kerubo Moenga

Submitted in total fulfilment of the requirements for the degree of Master's of Science in Statistical Science of Strathmore University

Institute of Mathematical Sciences

Strathmore University

Nairobi, Kenya

July 2023

Declaration

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

Name: .... **Paulyne Kerubo Moenga** ....

Signature: ....

Date: .... June 9, 2023 ....

Approval

The thesis of (Paulyne Kerubo Moenga) was reviewed and approved by the following:

Dr. Thaddaeus Egondi

Supervisor,

Institute of Mathematical Sciences, Strathmore University.

Dr. Linda Chaba

Supervisor,

Institute of Mathematical Sciences, Strathmore University.

Dr. Godfrey Madigu

Dean,

Institute of Mathematical Sciences, Strathmore University.

Dr. Bernard Shibwabo

Director,

Office of Graduate Studies, Strathmore University.

Abstract

Prior research indicates that a rise in the stock market has been associated with a corresponding upsurge in economic growth. The act of investing in stock prices serves to bolster a nation's economy through the mobilization of long-term financial assets for the purpose of production, while simultaneously mitigating potential investment risks via diversification strategies. Hence, the significance of the stock market endures as government's worldwide endeavor to achieve economic advancement as a primary objective. Investing in the stock market bears inherent risk due to the heightened levels of volatility and the intricate and capricious nature of the market. In order to make informed investment decisions, investors and market analysts must diligently analyze market behavior and formulate effective purchasing or selling strategies. One of the methods for comprehending the behavior of markets is by foreseeing impending values and possessing discernment with regard to the timing of investments. Investors have endeavored to devise various models that can precisely forecast the future values of stocks. This study aims to make a noteworthy contribution to the quest of forecasting stock prices for Kenyan companies by ascertaining the most optimal time series model. It employed the ARIMA and prophet model in order to ascertain the most suitable time series model for the prediction of share prices in Kenya. It has utilized the daily data of SAFARICOM PLC, Equity Group Holdings Limited (NSE: EQTY), KCB Group Limited (NSE: KCB), East African Breweries Limited (NSE: EABL) and Co-Operative Bank of Kenya Limited (NSE: COOP) for a period of five years, starting from January 2017 and ending in December 2021. The data set consisted of 1248 trading days, which were analyzed in the current investigation. The Root Mean Square Error (RMSE) was employed for model assessment in order to determine the optimal time series model for the prediction of stock prices. It discovered that the ARIMA model exhibited superior predictive performance in comparison with the Prophet model in forecasting Kenyan stock prices. The study posits that future research endeavors may benefit from augmenting sample size and encompassing multiple industries to improve the generalizability of findings.

Table of contents

List of figures

List of tables

List of abbreviations

ARIMA	Autoregressive Integrated Moving Average	MAE	Mean Absolute Error
ADF	Augmented Dickey-Fuller test	MAPE	Mean Absolute Percentage Error
COOP	Co-operative Bank	RMSE	Root Mean Squared Error
EABL	East Africa Breweries	SAFCOM	Safaricom
EQTY	Equity Bank Holdings	KCB	Kenya Commercial Bank

Acknowledgement

Without the participation and assistance of numerous people, this research would not have been feasible. The following persons and organizations are acknowledged in our gratitude.

Sincere thanks go out to my supervisor, Dr. Thaddeus Egondi, and Dr. Linda Chaba for their advice and constructive criticism during the development of the project and the execution of this work. I am delighted to be working under their leadership and watchful eye.

I'd also like to recognize the administrative personnel at the Institute for Mathematical Sciences for giving me the opportunity to pursue a master's degree. A statistic that the institution and my parents paid for my tuition and took care of everything related to my studies and well-being. I would also like to thank the staff of the Bureau of Statistics, the library, the students, friends and colleagues of Strathmore College. I feel compelled to appreciate my wonderful family for their efforts throughout my absence during this term of study. This appreciation would be meaningless until I specifically thanked my entire household.

Dedication

This thesis is dedicated to God Almighty for giving me wisdom and good health.

Chapter 1

Introduction

1.1 Background to the study

Stock prices in Kenya are an essential indicator of the health of the nation's budget, as they indicate the overall presentation of a company's stock. The stock prices in Kenya are tracked through the Nairobi Securities Exchange (NSE), the largest securities exchange in East Africa. The NSE provides investors with a platform to buy and sell stocks of companieslisted on the exchange. Kenya's stock prices are influenced by economic factors such as inflation, GDP growth, political stability, currency exchange rates, and company performance. Global events and changes in the world financial markets also impact stock prices. By tracking the stock prices in Kenya, investors can gain insight into the performance of their investments and make informed decisions.

Investing in the stock market in Kenya allows investors to diversify their portfolio, access a wide range of stocks, and potentially generate higher returns than other asset classes. It allows investors to participate in the economy's growth while protecting their capital from inflation. Investing also provides the opportunity to take advantage of market movements and tax benefits. Investing in the Kenyan stock market can build long-term wealth, increase financial stability, and gain exposure to the global markets.

The hypothesis behind the research design of comparable accurately forecasting stock prices in Kenya is that a predictive model can accurately forecast stock prices in the Kenyan stock market. This model will accurately indicate when to buy and sell stocks to maximize profits. The research design will identify the optimal model most suitable to the Kenyan stock market and its specific characteristics. This will be achieved by fitting two time series models the autoregressive integrated moving average (ARIMA) and the Prophet model.

1.2 Statement of the Problem

Its a crucial indicator of economic growth and development for a country.A study by (Mun et al., 2008) suggested that a decrease in the stock market reflects a future recession and that an increase in the stock market is an indicator of future economic growth.The decision of investors to trade in the stock market depends on the stockbroker who executes and advises clients on which stocks to buy and when to sell.

Most stockbrokers depend on their expertise, technical analysis, and fundamental analysis to advise their clients. In the past, these methods have been applied and yielded results; however, both are criticized as being subjective and can be a hit or miss. If not well interpreted, this could cause a significant loss to the investors. Due to the lack of certainty, investors tend to be reluctant to invest in the stock market, which could, in turn, affect the economy in general.

Therefore, there is a need to have a tool that can use to guide on which stocks to buy or sell, that not only predicts the pattern in which prices will go in addition to the stock's highest realistic price worth with better accuracy as a basis for making an investment decision. Therefore, this schoolwork purposes to recognize the optimal model for predicting stock prices in Kenya.

1.3 Research Objectives

The **main objective** of this study is to formulate, evaluate and identify the optimal time series model to predict stock prices in Kenya.

specific objectives

1. To formulate a time series model that predicts stock prices.
2. To evaluate the time series models under study to find the best-performing model.
3. Identify and apply the optimal time series model to predict Kenyan stock prices.

1.4 Significance of the Study

Identifying the optimal model for predicting stock prices in Kenya can have several important implications. One of the most significant is that it can help investors make more informed and profitable trading decisions. By providing accurate and reliable predictions of future stock prices, an optimal model can help investors identify undervalued stocks that are likely to appreciate and avoid overvalued stocks at risk of losing value. In addition, an optimal model for predicting stock prices in Kenya could also be used by companies to make better-informed decisions about when to issue new shares or repurchase existing shares, as well as by banks and other financial institutions to make more accurate risk assessments when lending money to companies in the stock market.

Furthermore, it can enable the government and policymakers to evaluate the overall performance of Kenya's economy and market trend and adopt policies to regulate and promote better economic performance. Overall, an optimal model for predicting stock prices in Kenya can have a wide range of essential applications and can play a critical role in helping individuals, organizations, and governments make more informed financial decisions. Finally, it will provide a baseline of models and methods that can use to predict future stock prices, allowing investors to make more accurate predictions and better manage their portfolios.

Chapter 2

Literature review

2.1 Introduction

This section provides a literature review on the Kenyan stock market and time series predictive models and a summary conclusion of the review.

The Nairobi Security Exchange trades in both bonds and shares and are classified into treasury bonds, issued by the Kenyan government, and corporate bonds, which the trading companies issue. The shares in the NSE are classified into different sectors: Commercial, Agricultural, Financial, Allied, and Services. Nairobi Security Exchange encourages savings and investment, which is essential for the expansion of Kenya's economy.

Over the last decade, the Kenyan stock market has experienced many fluctuations. According to Sheila (2014), the market capitalization ratio for Nairobi Stock Exchange was stationaryin the late 1980s but only began to increase in 1991, to arrange 43 per cent in 1994, before deteriorating and producing a deep and comprehensive economic trough in 1995 to 2006. However, according to the (world bank, 2012), market capitalization developed considerably between 2001 and 2006, and in 2007, the stock market capitalization compared to many. Nonetheless, this was still considered comparatively low compared to many developed countries.

The period between 2014 and 2018 was characterized by enhanced trading activities in the stock market, increasing the stock market capitalization of about KES 750 billion in 2014 to realize 4.9 trillion Kenya shillings by 2018. Similarly, the number of shares traded grew from 593 million to a high of 6.33 billion in a similar period (NSE, 2019). On the other hand, the Stock Market returns have been performing poorly over the last few years, and investors have been losing money in the Kenyan stock market. For instance, according to (Global economy 2022), The average value for Kenya between 2013 and 2020 was 5.5 per cent, with a minimum of -33.18 per cent in 2009 and a maximum of 49.05 per cent in 2013. The latest value from 2020 is -4.93 per cent. For comparison, the world average in 2020 based on 88 countries is 12.02 per cent.

One of the essential models that have evolved is the random walk. It has cast severe doubts on the methods of predicting stock prices. Random walk theory states that forthcoming stock prices are unsystematic and are not subjective by previous events (Fama, 1995). Random walk theoryproponents believe that all current information is reflected in the stock price; therefore, pastmovements cannot be used to forecast future movements. The theory asserts that technical analysis and fundamental analysis, at times, can be misleading.

According to the efficient market theory, stock prices accurately represent all facts that are accessible at any given time in a liquid market. The best possible use is made of all the info presented to the market participants, and the Cost increases are as unpredictable as new knowledge is random (Lawrence, 1997). According to EMH, share prices are random, making it impossible for investors to beat the market. It suggests that stock prices reflect the knowledge and expectations of all investors, arguing that the market cannot be beaten consistently. However, this theory is highly controversial as many studies done by (Dutta et al., 2006) (and Aghababaeyan and Tamanna Siddiqui, 2011) have managed to negate the efficient market hypothesis.

In search of a better model to predict stock prices (Mondal et al., 2014) Analyzed the performance of the Autoregressive Integrated Moving Average on 56 Indian equities from seven distinct industries. They chose 23 months of data for the investigation and assessed how well the ARIMA model predicted stock prices. To choose the optimal ARIMA model, they employed AIC. The preciseness of the ARIMA model was above 85% for all the studied sectors, indicating that ARIMA provides exceptional precision in forecasting stock prices. They used AIC to select the best ARIMA model. For all the sectors under study, the accuracy of the ARIMA model was above 85 per cent, which indicated that ARIMA gives good accuracy in predicting stock prices. This study was supported by a study done by (Adebiyi et al., 2014), who developed an ARIMA model for forecasting the New York Stock Exchange prices.

Their findings exhibited that for the short-term prediction, the ARIMA model's outcome could compete relatively with new forecasting techniques (Wanjawa and Muchemi, 2014) applied Artificial Neural Network (ANN) to predict Kenyan stock prices from the Nairobi security exchange. The venture established an ANN model grounded on feedforward multi-layer perceptron (MLP) with fault back-propagation through experimentation. The outcomes displayed that ANN-based models could be secondhand to progress low RMSE systems and therefore can be employed in emerging stock market forecast software.

A study (Kihoro and Okango, 2014) used artificial neural networks (ANN) to predict stock market prices in Kenya. They fitted ARIMA models to the stock price data to identify the best input lags in the ANN model. The study concluded that artificial neural networks could effectively model local stock market prices for reliable forecasts. (Patel et al., 2015) addressed the problem of predicting the direction of movement of stock prices and stock price index for Indian stock markets. They compared four models Artificial Neural Network, Support Vector Machines Random forest and naïve-Bayes, with two approaches for input to these models.

The first approach for input data involved the computation of ten technical parameters using stock trading data (open, high, low close prices). In contrast, the second approach focused on representing these technical parameters as trend deterministic data. The accuracy for each model considering the two inputs was evaluated. The findings showed that in the first approach, which computed ten technical parameters using stock trading data, the random forest outperformed the other three models on overall performance contributed to the search for the optimal model for predicting stock prices by trying to predict the return value the of SP Bombay stock exchange (BSE) Sensex index (Latha et al., 2018). The BSE Sensex index comprises 30 listed companies, popularly known as blue-chip companies.

The study of the ARIMA model to forecast future returns values for 24 months ahead. AIC was used to evaluate and select the best ARIMA model. Root Mean Square Error and Mean Absolute Error (MAE) were used to measure the model's accuracy.(Ramteke, 2020) did a comparative study on the performance of time series models to predict stock prices of five banks in the National Stock Exchange (NSE); the time series models compare the ARIMA, Prophet and KERAS with LSTM models? The results showed that all three models had a strong potential for prediction. All three models performed better in different data samples, while LSTM was best in forecasting seasonality. (Chan, 2020) used ARIMA and the Prophet model to predict Myanmar stock prices. Both models were compared on the same dataset (daily, weekly, and monthly).

The learning establish that the Prophet model outdone the ARIMA model for three periods. However, mutually models are appropriate for short-term prediction (daily and weekly).To improveon the past studies that used ARIMA and ANN models to predict stock prices (Ayub and Jafri, 2020) compared the accuracy of ARIMA, ANN and a Hybrid model of ARIMA and ANN inpredicting the stock prices of Karachi Stock Prices that included two companies National foods (NATF) and Engro Foods (EFOOD). The study suggested that it combined the strength of ARIMA, linearity, and ANN, non-linearity, to develop a better model. Their findings showed that the Hybrid ARIMA-ANN outperformed the ANN and ARIMA, respectively.

The results also showed that ANN performed better than the ARIMA model in predicting stock prices (Saxena and Kamnge, 2020). The comparison was made on a short-term basis for ARIMA, simple Exponential smoothing, double exponential smoothing and Damped trend linear exponential smoothing (DTLES) on the daily stock price data of the NMB bank in Tanzania. The study found that the Damped trend linear, exponential smoothing outperformed the other model (Groenendijk et al., 2021). they improved on the work done by (Ayub and Jafri, 2020) by creating a hybrid of ARIMA-ANN and an adjusted hybrid of ARIMA-LSTM to predict daily and intraday stock returns. The adjusted hybrid of ARIMA-LSTM outperformed the ordinary hybrid of ARIMA-ANN.

It is in line with a study which recommends using an ARIMA-based hybrid model (a model that uses ARIMA as the base model), which dramatically minimizes total forecasting errors

(Shukor et al., 2021). It also applied the exponential smoothing models by comparing the Double exponential smoothing, Holt’s linear trend, and random walk to predict stock market prices of gold, silver, crude oil, and platinum. The results showed that Holt’s linear trend outperformed the Double exponential smoothing and Random walk (Anjikar et al., 2022) .

In the spirit of trying to find the best model to forecast stock prices (Murthi, 2021) performed an advanced forecasting method to study the effectiveness of the ARIMA model in stock price prediction and concluded that the ARIMA model could compete with other forecasting techniques. To support (Murthi, 2021) claims (Maskey, 2022) utilized ARIMA models to predict the stock price for the NEPSE index. The study used daily closing prices for eight years. The ARIMA model has a high potential for anticipating short-term market swings, which might benefit short-term investors. A comparative study (Pindiga, 2022) applied ARIMA and Prophet models on the Dow Jones industrial average price. The research found theARIMA model to have outdid the Prophet model; this is contradicted by a study done by (Chan, 2020) where the Prophet was found to be better than the ARIMA model.

Over the years, many studies have been done, and many models have been developed in search of the optimal tool to use to predict stock prices accurately. Many studies have applied the ARIMA model, which has proved to have the ability to predict stock prices. However, this is only true if it is used in predicting short-term periods, which could be days or weeks.Many studies have also tried to combine the ARIMA model and other models to improve its efficiency in predicting stock prices, and some of these have been successful.

However, when it comes to predicting Kenyan stock prices, most studies have used machine learning models like ANN and a combination of ARIMA and ANN, which have proven to work in some cases and failed in some cases. Most recent studies have tried to compare the ARIMA and PROPHET models in search of the optimal model. However, there needs to be more clarification regarding which model performs better.

The readings that were completed concentrated on the stock price in Africa and Kenya. They concentrated on the influence of macro variables in relative to stock earnings (Saxena and Kamnge, 2020) while very few concentrated on the prediction of stock prices which has proven to be useful for investors and the economy in general.

Also, no studies have been finished on the use of ARIMA and the PROPHET models to prediction the stock price in Kenya. From the literature, the prophet is more accurate than the ARIMA in one study and vice versa in another study. In the study done in the United State by (Pindiga, 2022) the ARIMAmodel was found to have outperformed the prophet model however in the study done in Asia the Prophet model outperformed the ARIMA model. Due to this contradiction, the study will apply the Kenyan data to observe its behavior and see how the prophet model performsagainst the ARIMA model.

Therefore, this study is looking to fill these gaps and contribute to the existing literature in the search for the optimal time series model to predict Kenyan stock prices.

Chapter 3

Methodology

3.1 Introduction

The chapter describes the study approach that will be followed in the current study project. It also outlines the demographic to be investigated for the inquiry and the group of volunteers who will be used. The approaches for acquiring data will be provided. The data analysis technique to be employed and the justifications for using it are also mentioned. Data analysis software, as well as presentation tools, have been provided.

3.2 Data

The data that was used for model evaluation was the daily closing price of SAFARICOM PLC, Equity Group Holdings Limited (NSE: EQTY), KCB Group Limited (NSE: KCB), East African Breweries Limited (NSE: EABL) and Co-Operative Bank of Kenya Limited (NSE: COOP) for over five years, from January 2017 to December 2021. These five companies were selected for the study because they were ranked the largest Kenyan (Nairobi Security Exchange) stock by market cap.

This research used secondary data available in the public domain from authorized data vendors. Refinitiv is one of the known financial data providers. It is an American-British global financial market data and infrastructure provider founded in 2018. The data is publicly available; therefore, the research experiments can be conducted independently by anyone accessing this data. The data used has a period of 5 years with 1248 trading days. The Stock data included open, low, high and close prices of the above-mentioned companies. This study used the closing and opening prices to forecast stock prices. The data was downloaded from the Refinitiv website in a CSV file. R programming software was used to read and analyze the data. Missing data was eliminated by replacing them with the previous day's prices for all dates when trade did not occur during the study period. Finally, the data were transformed into time series data for analysis.

The data were analyzed in R studio to predict the stock prices one week ahead. Models that were developed for this analysis are ARIMA and Prophet Models. After that, the effectiveness of these models was appraised by means of the Root Mean Square Error (RMSE). RMSE, measurethe deviation between actual and forecast values. The smaller the values of RMSE, the closer the predicted time series values are to that of the actual value.

3.3 The Auto-Regressive Moving Average model

In order to talk about ARIMA we have to talk about the ARMA model which is a model that is stationary and consists of the moving average component and the autoregressive component. Y_t is said to follow an autoregressive moving average process of order (p,d) or ARMA (p,d) if;

$$Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_P Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}, \quad (3.1)$$

where Y_t is the time series and ε_t is an uncorrelated random error term with zero mean and constant variance.

An ARMA (p,d) can also be written using a backshift notation as shown below

$$(1 - \varphi_1 B - \dots - \varphi_P B^P) Y_t = \left(1 - \theta_1 B - \dots - \theta_q B^q\right) \varepsilon_t \varphi(B) Y_t = \theta(B) Y_t \quad (3.2)$$

3.4 Auto-Regressive Integrated Moving Average

If a series is non-stationary, then a differencing technique has to be followed to make this time series stationary. The series has to be differenced d times for it to be stationary. The number of times needed to make a series stationary is what is referred to as Integrated of order d. If no differencing is done then the series is said to be Zero integrated denoted by I (0).

The first differencing of a series Y_t is given by $W_t = Y_t - Y_{t-1}$ written as;

$$W_t = (1 - B)Y_t \tag{3.3}$$

A time series process Y_t is said to be an Auto-regressive Integrated Moving Average (ARIMA) model if a stationary ARMA is obtained after differencing the series d times.

The general form of ARIMA (p,d,q) is given by

$$\phi_p(B)(1 - B)^d Y_t = \theta_q(B)\varepsilon_t \tag{3.4}$$

where p is the number of auto-regressive terms in the model, d is the number of times the raw observations are differenced; also known as the non-seasonal differencing, and m is the number of lagged calculation errors in the approximation equation.

3.4.1 Model Assumptions

ARIMA model makes a few assumptions and one of the assumptions made by this model is that the data should be stationary meaning that the properties of the series do not depend on the time it was captured (i.e. the data must have constant variance and mean). If your data is non-stationary you have to transform it before using ARIMA.

3.4.2 Evaluation of the Model

Since the ARIMA model is a special time of regression, the typical regression metrics shall apply, including RMSE, MSE, MAE, etc Another evaluation test that can be applied is the Akaike Information Criteria (AIC) one of the advantages of the AIC test is that it considers the number of parameters used as part of the assessment because using too many parameters in a model might lead to over-fitting. The model with a lower AIC value suggests that that model would have better performance when doing forecasting.

3.5 Prophet Model

The Prophet model is a free prediction program created by (Taylor and Letham, 2018). This model was created for distinctive Facebook problems, such as foreseeing operator actions. Therefore, this model is more suitable for envisaging seasonality, singular trials, holidays,and data with outliers and varying trends (Shah and Dimitrov, 2022).

Prophet uses *y* (target) and *ds* (Date time) in the time series. Compared to outmoded exponential smoothing approaches, the Prophet can cutting tendencies and intermittent signals transversely awide assortment of time series and has no regular quantity spacing. Prophet and classic ARIMA techniques vary significantly in that they attempt to fit multiplicative regression models; in other words, they attempt to fit the shape of the curve.

The prophet algorithm effectively obtains the needed information. It is a combustible paradigm that separates a large task, including time series data forecasting, into smaller ones (Sheeba et al., 2021). The Prophet model takes on the procedure of an preservative regression which be in the right place to the Generalized Additive model (GAM) family. It considers three parameters, namely seasonality, trends and holidays, as shown below;

$$Y(t) = G(t) + S(t) + H(t) + \varepsilon_t \tag{3.5}$$

$Y(t)$ demonstrates the additive regression model $G(t)$ demonstrates non-periodic alterations in period series $S(t)$ demonstrates changes done on a regular basis, such as each week, annually, and occasionally. H(t) represents the consequence of holidays the user provides with inconsistent schedules, and t represents any exceptional adjustments.

3.5.1 Parameters Considered by the Prophet Model

The Trend Model

The overall trend of the data is modelled by the development function. The model believes that growth trends can be present at all points in the data or can be altered at the change points. Change points are defined as moments in the data where the data shift direction. The modelautomatically detects the change points or they can be tuned by the analyst. The two main trend models are the logistic growth model and the piece-wise linear model.

Piece-wise Linear Model

The model uses a set of piece-wise linear equations with differing slopes between change points. It is a simple modification of a linear model which can be written as;

$$Y = \beta_0 + \beta_1 X + \beta_2 (X - c) + \varepsilon, \tag{3.6}$$

where c is the value of the change point.

Logistic Growth

This handles the time series that has a floor in which the standards being modelled become water-logged and can't surpass a determined or minimum assessment. The progression term lookssimilar to that of the typical equation for a logistic curve except, the carrying capacity (C) varies as a function of time and growth rate (k) and (m) an offset parameter

$$G(t) = \frac{C}{1 + exp\,(-k(t - m))}. \tag{3.7}$$

Seasonality Component

The Fourier series may be used to mimic an annual seasonal component. Dummy parameters can be utilized to simulate a periodic weekly component, and the user can be provided with a list of festivals (Sah et al., 2022). The following formula gives the Fourier series;

$$S(t) = [\sum_{n=1}^{N}\left(a_n Cos\left(\frac{2\pi nt}{P}\right) + b_n Sin\left(\frac{2\pi nt}{P}\right)\right)] \text{ . (3.8)}$$

here, $S(t)$ stands for seasonality, and P refers to the period which can be taken on a monthly, weekly, daily, quarterly and even annual basis, N refers to the frequency of changes, and the parameters a_n and b_n depend upon the given N.

Holiday Component

Occasions and festivals are also addressed. Modifications can take place throughout special festivals, affecting predictions.

3.5.2 Assumptions of the Model

The prophet model consistently provides confidence intervals for the predictions. It makes an important assumption that supports these inconsistency ranges. The model indicates that the general frequency and size of trend fluctuations over time will be equivalent to past events.

3.5.3 Evaluation of the Prophet Model

The model can be evaluated by splitting the data into exercise data and challenging data. Then the model is built based on the physical activity data and used to forecast for the testing data period. Then, we can assess how large or minor the variations between the projected and real values are by comparing the expected outcomes to this real-world information.

To do this, a variety of measures may be employed to assess the discrepancy amongst the actual and predicted values. Several of the most used measures for assessing models that forecast time series. RMSE (Root Mean Square Error) MAE (Mean Absolute Error) and MAPE (Mean Absolute Percentage Error). The Smaller the value of these parameters the better the model.

3.6 Model Building

3.6.1 Parameter Estimation

ARIMA and Prophet Models need to estimate parameters to build accurate models. The ARIMA model uses information from the time series data to figure out the best numbers for the parameters (p, d, and q). To figure out these variables, we look at the patterns in how the time series data repeats itself over time. Specifically, we analyze two categories of graphs autocorrelation function and partial autocorrelation function. The AIC criterion is a way to pick the best ARIMA model by looking at the numbers estimated. The AIC looks at how well the model fits and how many parts it has. It doesn't like models that are too complicated.

The Prophet model uses a way of calculating the parameters called Bayesian. The model fitting process means finding the best numbers to make the model match the real data as meticulously as conceivable. The optimization technique tries to find the best values for three parts of the model: trend, seasonality, and holiday. A prophet can handle many different types ofpatterns in time series data, like trends, seasonal changes, and important events. The way itdoes this is very strong and reliable.

Cross-Validation

Cross-Validation means testing how well a model works by using some of the data to train it and then testing it on the remaining data. It is a way to check if a model is accurate and not just memorize the data. We often use cross-validation to check if the ARIMA and Prophet Models work well. During cross-validation, the previous data is split into training and testing sets. The machine is taught on one data set and then used to predict what will happen in another. We compared what we expected to happen with what actually happened to see if our predictions were correct.

Evaluation Metrics

Different ways to measure how well models work can be used. The most commonly used ways to measure things are:

The RMSE is a way to see how close predictions are to average values. It tells us how well a model can predict things overall.

MAE is a way to determine how different the predicted values are from the real ones. It measures the average difference between them. It's like RMSE but without squaring the differences. It is not as affected by extreme values as RMSE.

The MAPE is a way to determine how far the predicted values are from the real values by looking at the percentage difference on average. This helps compare how well models predict different things over time.

We can find out which model is better at predicting stock prices by looking at how well they do using certain measurements. They are called ARIMA and Prophet Models.

3.6.2 Building ARIMA Model

To build the ARIMA model, the following steps will be followed; Identification of the model, parameter estimations, diagnosis on the fit model and calculating consuming the fit. These stepladders are deliberated in additional aspect as follows.

Model	ACF	PACF
AR(p) MA(q) ARMA (p,q)	Dies down Cut off after lag p Dies down	Cut off after lag q Dies down Dies down

Figure 3.1: ACF and PACF table

STEP 1: Model Identification

1. **Plotting Time Series Data**

   A time series plot was plotted in order to check for any stationarity or non-stationarity, also this graph is used to observe whether there were any trends, seasonality, outliers etc so as to make any data transformation required.

2. **Tests for Stationarity of Time Series**

   This was obtained by applying the Augmented Dickey-Fuller (ADF) test. The Augmented Dickey-Fuller Unit Root Test with the null and alternate hypothesis as:

   H_0: Unit root present thus the series is non-stationary

   H_1: Unit root absent thus the series is stationary

   If the p-value is less than 0.05 we reject the null hypothesis that all the time series at the first difference is non-stationary hence confirming that our series is stationary at the first difference and vice- versa.

3. **Model Identification by ACF and PACF**

   The partial autocorrelation function (PACF) and autocorrelation function (ACF) work together to suggest a model whether it is Auto-Regressive, Moving Average or ARMA model. This is done by checking the cut-offor tail-off of the lag. This can be seen in the Figure 3.1.

STEP 2: Estimation of Parameters

Examination of the ACF, PACF and stationarity was done and the best model was suggested also the AIC was used to select the best ARIMA model.

STEP 3: Diagnostic Checking

Various combinations of Moving average and Auto-regressive were performed collectively all through the model identification stage and the best model fit for predicting time series was obtained after considering the subsequent principles or diagnostics.

1. **Akaike Information Criteria (AIC)**

   The AIC assesses a model dependent on its fitted principles to the experimental values. It suggests the best model for the available data with the minimum number of parameters. By using R software the AIC was obtained. The model with the lowest value of AIC is the most adequate and selected.

2. **Ljung-Box Q**

   In this step, a model obligation be checkered if it is adequate by considering the properties of the residuals and whether the residuals from an ARIMA model have a normal distribution and are random. The Ljung-Box Q statistic can be used for this process if the p-value associated with the Q statistic is small (p-value $< \alpha$), the model is considered inadequate, and vice versa. The Q statistic is computed as follows

$$Q = n \sum_{j=1}^{k} r_j^2 \tag{3.9}$$

where k is the maximum lag considered, n is the number of observations in the series, r_j is the estimated auto-correlation at lag j and k is any positive integer usually around 20. Following the examination and testing, the best model is employed to estimate potential outcomes and to determine whether the model is acceptable. Alternatively, we must return to the process of model development (Saxena and Kamnge, 2020).

STEP 4: Forecasting with the Model

Predictions for any number of upcoming times are selected, as are the requirements for the best-fitting model.

3.6.3 Building Prophet Model

Preparation of Data

According to the prophet model developed by Taylor and Letham, 2018 to build a Prophet model, the data has to be in the prophet format which is the date labelled as 'ds' and the output labelled as 'y'. The data sets for various stock prices were converted into the prophet format. The data was then split into a physical activity set and a testing set which is in the ratio 80:20,respectively.

Model Fitting

The prophet model is fit using the prophet package in R studio. In the previous chapter, it is observed that the Prophet model follows the following parameters; Trends, Seasonality and holidays. These parameters are tuned to fit the data and to fit the best prophet model. The best prophet model will then be used to forecast.

Prophet Model Forecast Quality Evaluation

The model precision was accomplished with cross-validation of the historic data. This was done by selecting cutoff points in this history and for each of them fitting themodel using data only up to that cutoff point. Then the forecasted values were compared tothe actual values, by default in the prophet package the initial training period is set to threetimes the horizon and cutoffs are made every half a horizon.

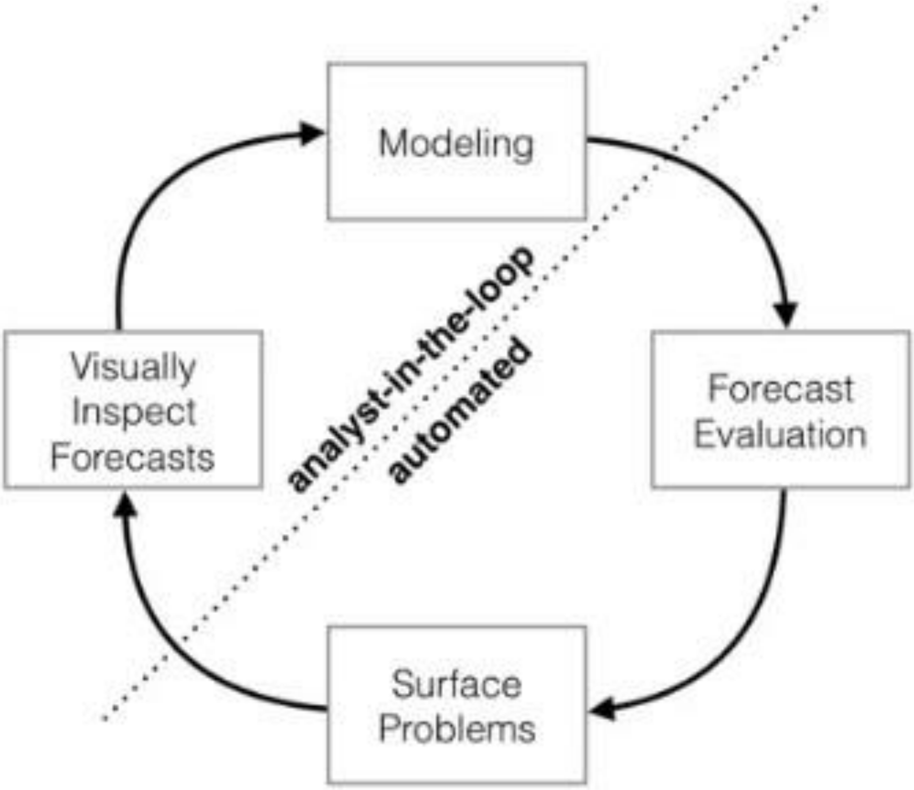

Figure 3.2: Schematic view of the analyst-in-the-loop approach to forecasting at scale, which best makes use of human and automated tasks.(Taylor and Letham, 2018)

Chapter 4

Results and Discussions

4.1 Introduction

This chapter will showcase the results as per the ARIMA model and Prophet model described in the previous chapter.

4.2 ARIMA Model Fitting On the Stock Prices for Kenyan

4.2.1 Model Identification Process

Time series plots for SAFCOM, EQTY, KCB, EABL and COOP

From Figure 4.1, we can observe different patterns e.g. trends, seasonal variations etc. For instance, it can be observed that Safaricom (SAFCOM) stocks have had an upward trend over the past few years while the East African Breweries (EABL)stocks have experienced a downward trend over the past five years. The plots also infer a non-constant variance and mean for all the series under study hence differencing must be performed to make the series stationary.

In order to make the data stationary with constant variance and zero means, data transformation was performed for all the time series and plotted as seen in Figure 4.2. By visual inspection, all the time series appear stationary in both the variance and the mean. However, to be sure and to confirm these claims the Augmented Dickey-Fuller (ADF) test was performed on the series. This was displayed in Figure 4.3.

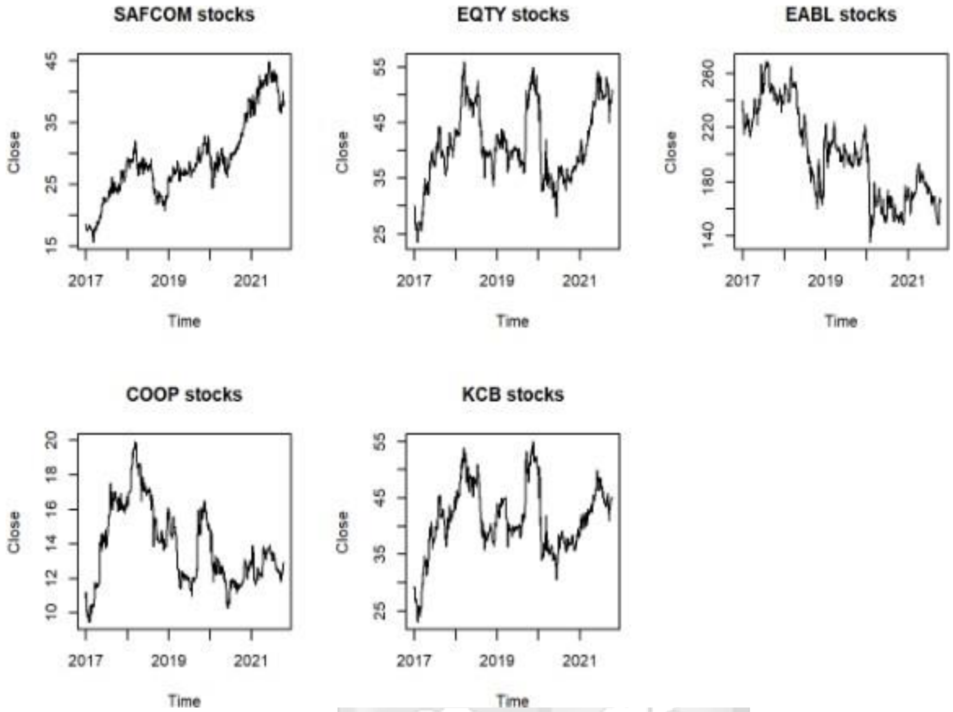

Figure 4.1: time series plots

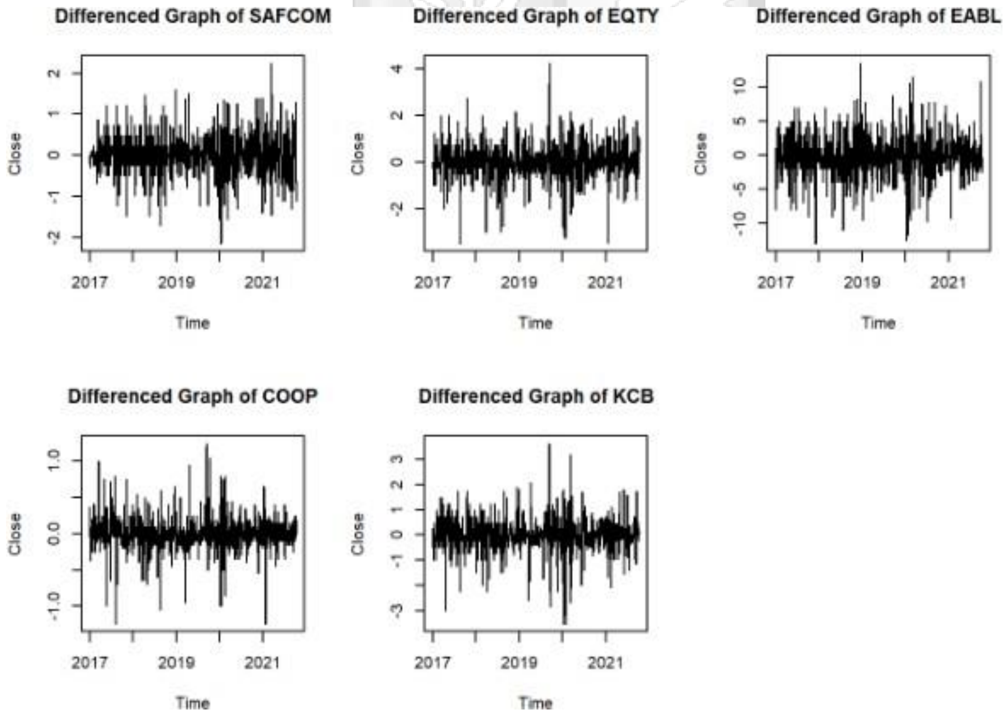

Figure 4.2: Differenced time series plot

Augmented Dickey-Fuller Test for the first order difference of the stock prices for SAFCOM, EQTY, KCB, EABL and COOP

Stocks	Dickey-Fuller	Lag order	p-value
SAFCOM	-11.638	10	0.01
EQTY	-9.9456	10	0.01
EABL	-10.87	10	0.01
COOP	-10.31	10	0.01
KCB	-9.5908	10	0.01

Figure 4.3: ADF.test

SAFCOM	EQTY	EABL	COOP	KCB
ARIMA (1,1,2)	ARIMA (0,1,1)	ARIMA (5,1,1)	ARIMA (1,1,0)	ARIMA (2,1,0)
ARIMA (2,1,2)	ARIMA (0,1,2)	ARIMA (5,1,2)	ARIMA (2,1,1)	ARIMA (1,1,1)
ARIMA (2,1,2)	ARIMA (1,1,1)	ARIMA (4,1,2)	ARIMA (2,1,0)	ARIMA (3,1,0)

Figure 4.4: ARIMA models

The Augmented Dickey-Fuller Unit Root Test with the null and alternate hypothesis as: H_0: Unit root present thus the series is non-stationary H_1: Unit root absent thus the series is stationary was performed on the differenced time series in R studio and the results were presented in Figure 4.3. Since the p-value for all the companies is less than 0.05, we reject the null hypothesis that all the time series at the first difference is non-stationary, hence confirming that our series is now stationary at the first difference.

Model Building Process

Since all the time series for the mentioned companies have been observed to be stationary ARIMA models can now be optional by watching at the ACF and PACF plots for each stock in Appendix A.1 By observing the order of the AR and the MA various models were suggested for each stock Suggested models can be seen in Figure 4.4.

In order to come up with the most parsimonious model for predicting each stock a model estimation procedure was performed for each stock based on three criteria; AIC, log-likelihood and the Ljung Box statistics.

Model Estimation for SAFCOM stocks

Based on the statistics in Figure 4.5 the ARIMA (1,1,2) model was found to be the best and most parsimonious model in the sample based on the above-mentioned criteria. The ARIMA (1,1,2) has the lowest Akaike information criterion AIC as well as the lowest RMSE suggesting that it is the best ARIMA model among the selected models. All three models meet the assumption of stationarity.

A residual diagnosis was performed using the Ljung Box test, outcomes show that the p-values are greater than 0.05 as displayed in Figure 4.5, therefore we do not reject the null hypothesis of zero autocorrelations for errors indicating that the models have captured linear autocorrelation pattern sufficiently in the training set. Considering the RMSE which approximations the standard deviation of a distinctive practical value from the fitted model's prediction ARIMA (1,1,2) is concluded as the best model for forecasting the Safaricom stocks since it has the lowest RMSE and Akaike information criterion AIC.

Statistic/ Parameter	ARIMA (1,1,2)	ARIMA (2,1,1)	ARIMA(2,1,2)
AR (1)	0.7442	0.9375	0.7812
AR (2)	-	-0.2154	-0.0423
MA (1)	-0.5624	-0.7617	-0.5982
MA (2)	-0.2208	-	-0.1802
Variance			
AIC	1429.94	1431.4	1431.86
Log-likelihood	-710.97	-711.7	-710.93
Ljung box test p-value	0.7017	0.6288	0.6537
RMSE	0.4277	0.4288	0.4278

Figure 4.5: ARIMA parameter estimate and Assessment statistics for SAFCOM stocks

Model Validation for SAFCOM

The accuracy of the fit in approximating the observed values is something we want to evaluate during model validation. Figure 4.6 shows the outcomes of this analysis, which was conducted on the stock prices of the SAFCOM series. The fitted stock prices, as illustrated in Figure 4.6, can be seen to closely match the observed data.

Model Estimation for EQTY Stocks

Based on the statistics in Figure 4.7 the ARIMA (0,1,1) model was found to be the best and most parsimonious model in the sample based on the above-mentioned criteria. The ARIMA (0,1,1) has the lowest Akaike information criterion AIC as well as the lowest RMSE suggesting that it is the best ARIMA model among the selected models. All three models meet the assumption of stationarity.

Model Validation for EQTY

In model validation, we want to calculate the accuracy of the fit in approximating the experimental principles. This was done on the stock charges of the EQTY series and the consequences are demonstrated in Figure 4.8. It can be experimental that the close-fitting stock charges track well with the experimental data as shown in Figure 4.8.

Figure 4.6: Model fit for stock Price for SAFCOM

ARIMA parameter estimate and Assessment statistics for EQTY stocks

Statistic/ Parameter	ARIMA (0,1,1)	ARIMA (0,1,2)	ARIMA(1,1,1)
AR (1)	-	-	0.0302
MA (1)	0.2606	0.2626	0.2324
MA (2)	-	0.0076	-
Variance	0.4633	0.4633	0.4632
AIC	2583.42	2585.36	2585.35
Log-likelihood	-1289.71	-1289.68	-1289.67
Ljung box test p-value	0.9452	0.9353	0.9354
RMSE	0.6803	0.6804	0.6804

Figure 4.7: ARIMA parameter estimate and Assessment statistics for EQTY stocks

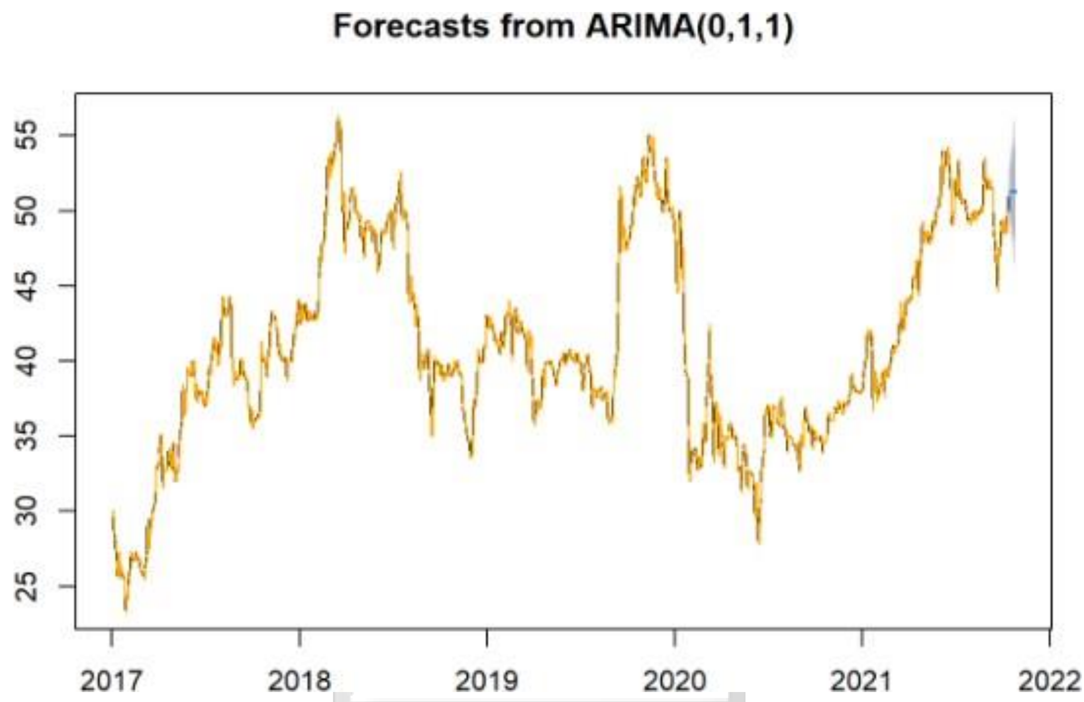

Figure 4.8: Model fit for stock Price for EQTY

Model Estimation for EABL Stocks

Based on the statistics in Figure 4.9 the ARIMA (5,1,1) model was found to be the best and most parsimonious model in the sample based on the above-mentioned criteria. The ARIMA (5,1,1) has the lowest Akaike information criterion AIC as well as the lowest RMSE suggesting that it is the best ARIMA model among the selected models. All three models meet the assumption of stationarity.

Model Validation EABL

In model validation, we want to evaluate the correctness of the fit in valuing the experiential values. This remained completed on the stock amounts of the EABL series and the fallouts are shown in Figure 4.10. It can be observed that the close-fitting stock prices pathway fine with the pragmatic data. As shown in Figure 4.10.

Statistic/ Parameter	ARIMA (5,1,1)	ARIMA (5,1,2)	ARIMA(4,1,2)
AR (1)	1.0139	0.9422	0.5205
AR (2)	0.0623	0.1390	-0.4449
AR (3)	-0.0867	-0.0821	0.0372
AR (4)	0.0077	0.0005	0.1004
AR (5)	-0.0717	-0.0759	-
MA (1)	-0.9312	-0.8591	-0.4331
MA (2)	-	-0.0707	0.5519
Variance	7.002	7.002	7.044
AIC	5979.86	5981.85	5987.27
Log-likelihood	-2982.93	-2982.92	-2986.63
Ljung box test p-value	0.8665	0.8577	0.8331
RMSE	2.645	2.648	2.653

Figure 4.9: ARIMA parameter estimate and Assessment statistics for EABL stocks

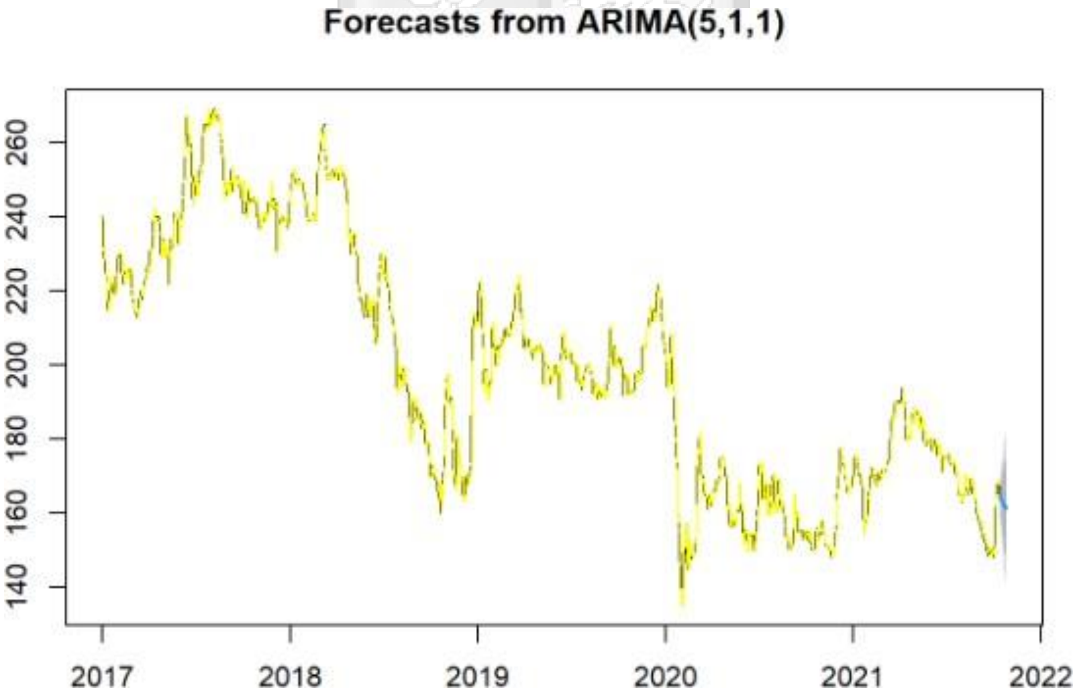

Figure 4.10: Model fit for stock Price for EABL

Statistic/ Parameter	ARIMA (1,1,0)	ARIMA (2,1,1)	ARIMA(2,1,0)
AR (1)	0.1609	-0.6710	0.1592
AR (2)	-	0.1651	0.0106
MA (1)	-	0.8295	-
Variance	0.0426	0.0425	0.0426
AIC	-392.25	-392.33	-390.39
Log-likelihood	198.13	200.17	198.19
Ljung box test p-value	0.9601	0.9872	0.9563
RMSE	0.2063	0.2060	0.2063

Figure 4.11: ARIMA parameter estimate and Assessment statistics for COOP stocks

Model Estimation for COOP Stocks

Based on the statistics in Figure 4.11 the ARIMA (2,1,1) model was found to be the best and most parsimonious model in the sample based on the above-mentioned criteria. The ARIMA (2,1,1) has the lowest Akaike information criterion AIC as well as the lowest RMSE suggesting that it is the best ARIMA model among the selected models. All three models meet the assumption of stationarity.

Model Validation for COOP

The accuracy of the fit in approximating the observed values is something we want to evaluate during model validation. Figure 4.12 illustrates the outcomes from this analysis of the stock prices of the COOP series. According to Figure 4.12, it can be seen that the stock prices that were fitted closely match the observed data.

Model Estimation for KCB Stocks

Based on the statistics in Figure 4.13 the ARIMA (2,1,0) model was found to be the best and most parsimonious model in the sample based on the above-mentioned criteria. The ARIMA (2,1,0) has the lowest Akaike information criterion AIC as well as the lowest RMSE

Statistic/ Parameter	ARIMA (1,1,0)	ARIMA (2,1,1)	ARIMA(2,1,0)
AR (1)	0.1609	-0.6710	0.1592
AR (2)	-	0.1651	0.0106
MA (1)	-	0.8295	-
Variance	0.0426	0.0425	0.0426
AIC	-392.25	-392.33	-390.39
Log-likelihood	198.13	200.17	198.19
Ljung box test p-value	0.9601	0.9872	0.9563
RMSE	0.2063	0.2060	0.2063

Figure 4.12: Model fit for stock Price for COOP

Statistic/ Parameter	ARIMA (3,1,0)	ARIMA (1,1,1)	ARIMA(2,1,0)
AR (1)	0.2670	0.0291	0.2675
AR (2)	-0.0761	-	-0.0781
AR (3)	-0.0075	-	-
MA (1)	-	0.2372	-
Variance	0.3768	0.3772	0.37678
AIC	2329.66	2329.11	2327.73
Log-likelihood	-1160.83	-1161.56	-1160.86
Ljung box test p-value	0.7059	0.6343	0.698
RMSE	0.6137	0.6139	0.6136

Figure 4.13: ARIMA parameter estimate and Assessment statistics for KCB stocks

suggesting that it is the best ARIMA model among the selected models. All three models meet the assumption of stationarity.

Model Validation for KCB

In model validation, n estimating the observed values, we wish to evaluate how accurately the fit is. The outcomes are shown in Figure 4.14 and were applied to the stock prices of the KCB series. As can be seen in Figure 4.14, the fitted stock prices closely match the observed data.

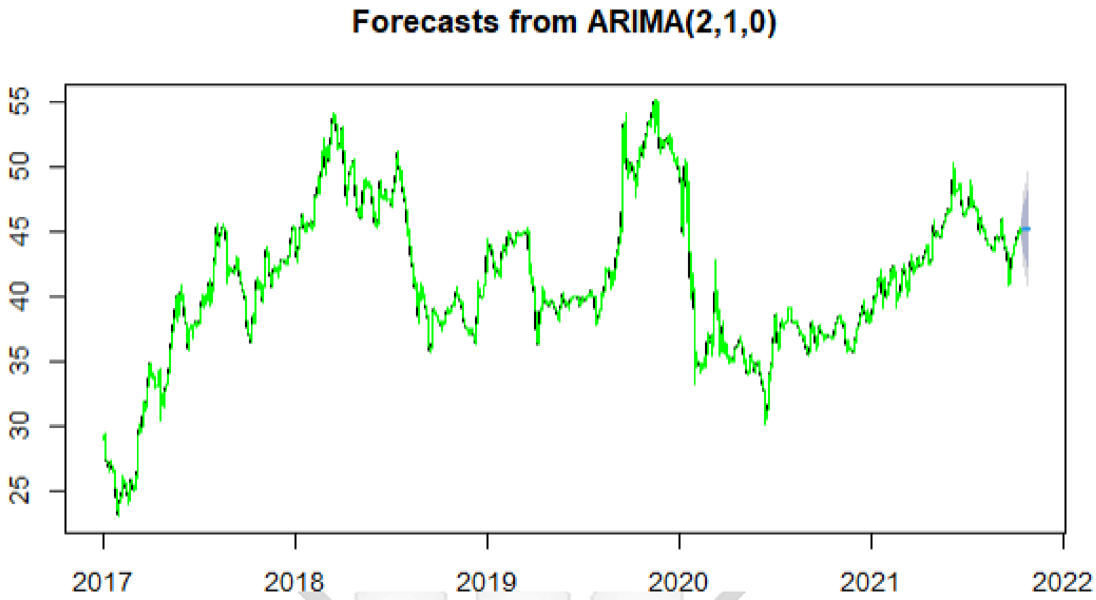

Figure 4.14: Model fit for stock Price for KCB

4.3 Prophet Model Fitting on the Stock Prices for Kenyan Companies

4.3.1 Preparation of Data

According to the prophet model developed by Taylor and Letham (2018) to build a prophet model, the data has to be in the prophet format which is the date labelled as 'ds' and the output labelled as 'y'. For this study, the stock price data for SAFCOM, EQTY, EABL, COOP and KCB was converted into the prophet format in r studio. The data was then split into a keeping fit set and a testing set which is in the ratio 80:20, respectively.

4.3.2 Model Fitting for SAFCOM

The prophet model is fit using the prophet package in r studio. In the previous chapter, it is observed that the prophet model considers the following parameters; Trends, Seasonality and holidays this can be observed from the following Figure 4.15. One particularly strong feature

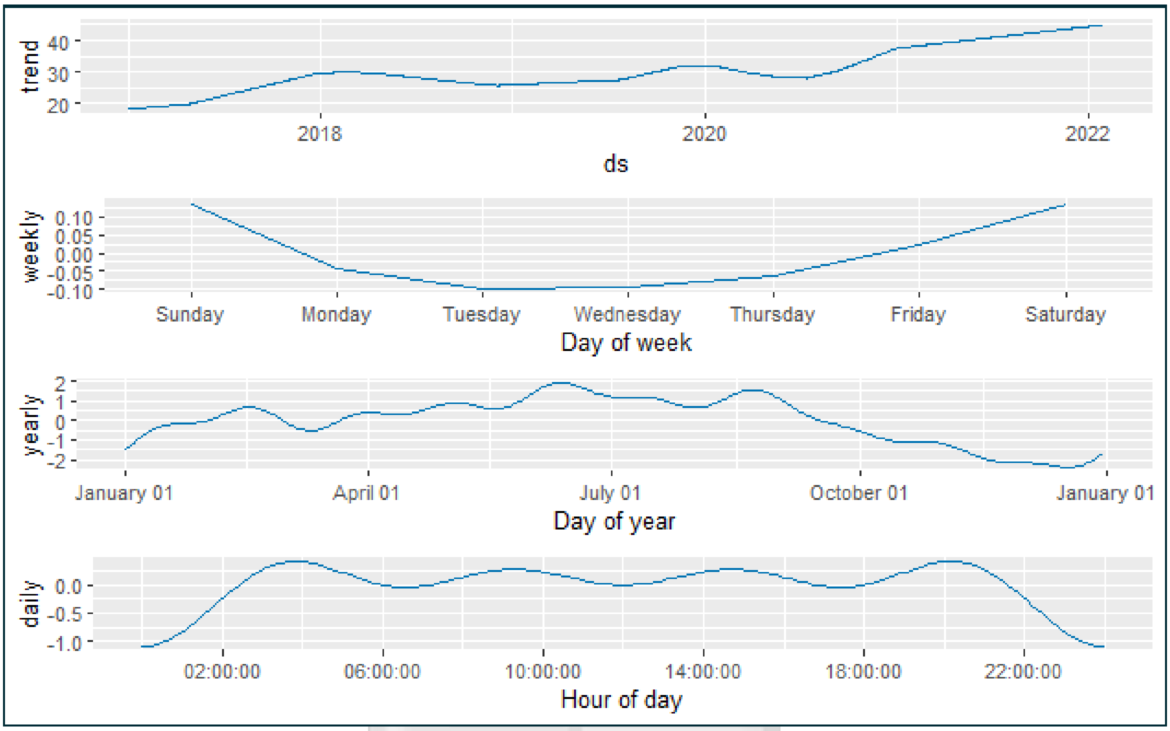

Figure 4.15: Components for the SAFCOM prophet model

of the Prophet is its capability to return the mechanisms of our predictions. Which can help disclosehow daily, weekly and yearly patterns of the time series underwrite to the complete estimatedstandards. The component plot provides interesting insights. The first plot shows that the dailyprices of SAFCOM stocks have been linearly cumulative over time. The second graph showsthat the Safaricom closing stock prices are lowest on Tuesdays and Wednesdays. The yearlygraph shows that Safaricom stocks are highest between June and September and are lowestat the end of the year that is in December.

In our prophet model, we use a linear trend since we can observe an increasing trend which is linear and does not follow a logistic growth. The change points for this series remained at the default 25. The seasonality S(t) is used as an additive for the yearly seasonality with the default Fourier order. Once the model has been created we forecasted the next 10 days and the fitted values can be seen in Figure 4.16.

The addition of the above three components along with the residuals would sum up the fit created by the prophet model. Assessment is done on the validation set of the train data in

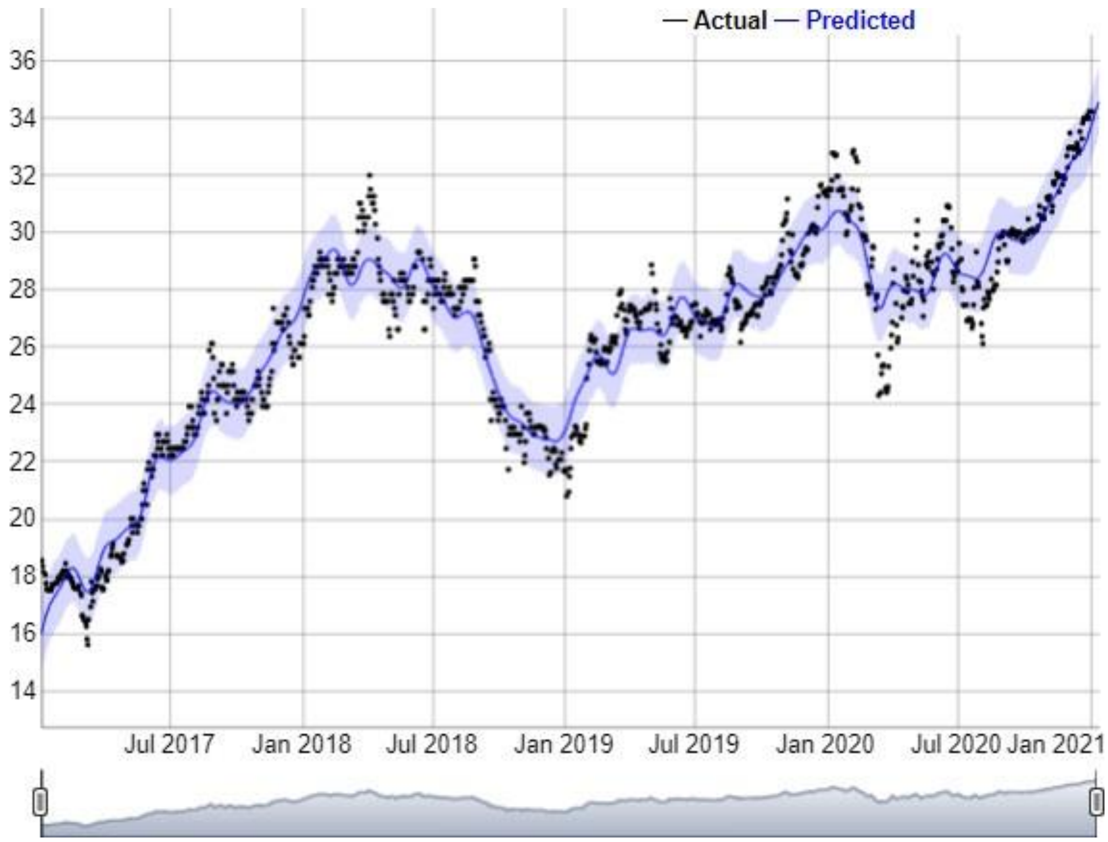

Figure 4.16: Components for the SAFCOM prophet model

order to determine the performance of the model. RMSE, MAE, and MAPE are the metrics used to assess the Prophet model (see Table 4.1).

RSME	MAE	MAPE
0.390	0.361	0.0107

Table 4.1: Statistics metrics for SAFCOM.

4.3.3 Model Fitting for EQTY

From the EQTY component graph, the trend graph which is the first graph shows an increasing trend from 2017 to mid-2018 and then a decreasing trend from mid-2018 to 2020 then a decreasing trend from 2020. From the second graph, it can be observed that the EQTY stocks are lowest in October are highest in January and February.

In our prophet model for the EQTY stocks, we use a linear trend since we can observe an increasing and decreasing trend which is linear and does not follow a logistic growth. The change points for this series remained at the default 25 which yielded the lowest RMSE for this stock. The seasonality S(t) is used as an additive for the yearly seasonality with the default Fourier order. Once the model has been created we forecasted the next 10 days and the fitted values can be seen in Figure 4.17.

Assessment is done on the validation set of the train data in order to determine the performance of the model. RMSE, MAE, and MAPE are the metrics used to assess the Prophet model (see Table 4.2).

RSME	MAE	MAPE
1.43	1.25	0.035

Table 4.2: Statistics metrics for EQTY.

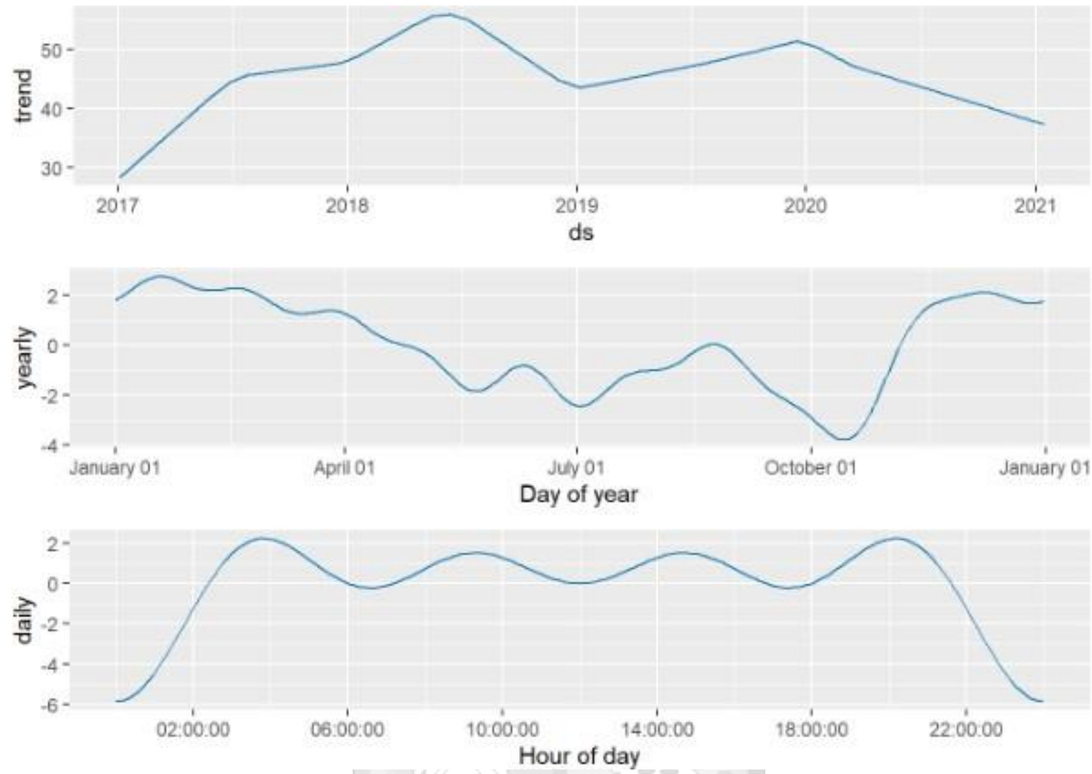

Figure 4.17: Components for the EQTY prophet model

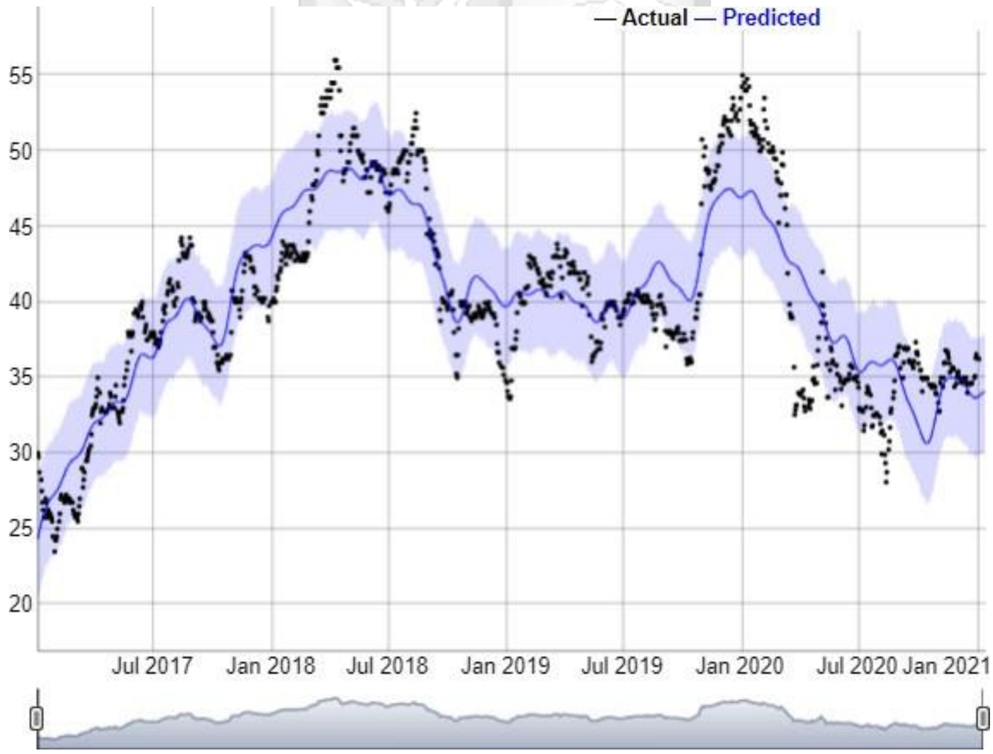

Figure 4.18: EQTY Actual values vs Predicted values using prophet model

Model Fitting for EABL

The East African Breweries stocks have registered a downward trend in the past five years as it ac be observed in figure 4.19 in the second graph we can observe a significant rise in the price of the stocks between January and March. The stocks are lowest between November and December.

In our prophet model for the EABL stocks, we use a linear trend since we can observe a decreasing trend which is linear and does not follow a logistic growth. The change points for this series were set at 40 after try and error which yielded the lowest RMSE for these stocks. The seasonality S(t) is used as an additive for the yearly seasonality with the default Fourier order. Once the model was created we forecasted the next 10 days and the fitted values can be seen in Figure 4.20. Assessment is done on the validation set of the train data in order to determine the performance of the model. RMSE, MAE, and MAPE are the metrics used to assess the model (see Table 4.3).

RSME	MAE	MAPE
3.21	1.95	0.0126

Table 4.3: Statistics metrics for EABL.

Model Fitting for COOP

The COOP stocks registered a decreasing trend between 2018 and 2021 as displayed in Figure 4.21 The yearly seasonality graph shows that the COOP stock prices are highest between January and April and they are the lowest in July.

In our prophet model for the COOP stocks, we use a linear trend since we can observe a decreasing trend which is linear and does not follow a logistic growth. The change points for this series were set at 40 after try and error which yielded the lowest RMSE for these stocks. The seasonality S(t) is used as an additive for the yearly seasonality with the default Fourier order. Once the model was created we forecasted the next 10 days and the fitted values can be seen in Figure 4.22.

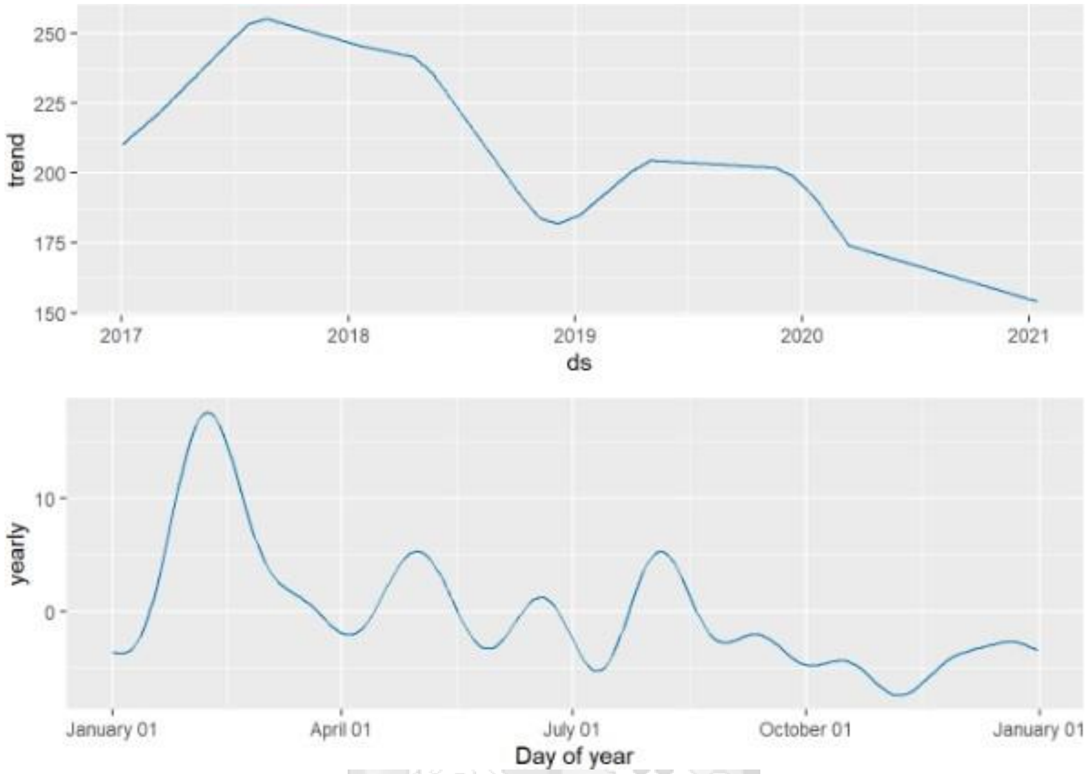

Figure 4.19: Components for the EABL prophet model

Assessment is done on the validation set of the train data in order to determine the performance of the model. RMSE, MAE, and MAPE are the metrics used to assess the Prophet model (see Table 4.4).

RSME	MAE	MAPE
0.262	0.209	0.0172

Table 4.4: Statistics metrics for COOP.

Model Fitting for COOP

As displayed in the forecast component graph for KCB the trend chart shows an upward trend for the KCB stocks between 2017 and 2021. The second graph in the component chart in Figure 4.23 shows that the KCB stock prices are highest between January and March and are lowest in July and October.

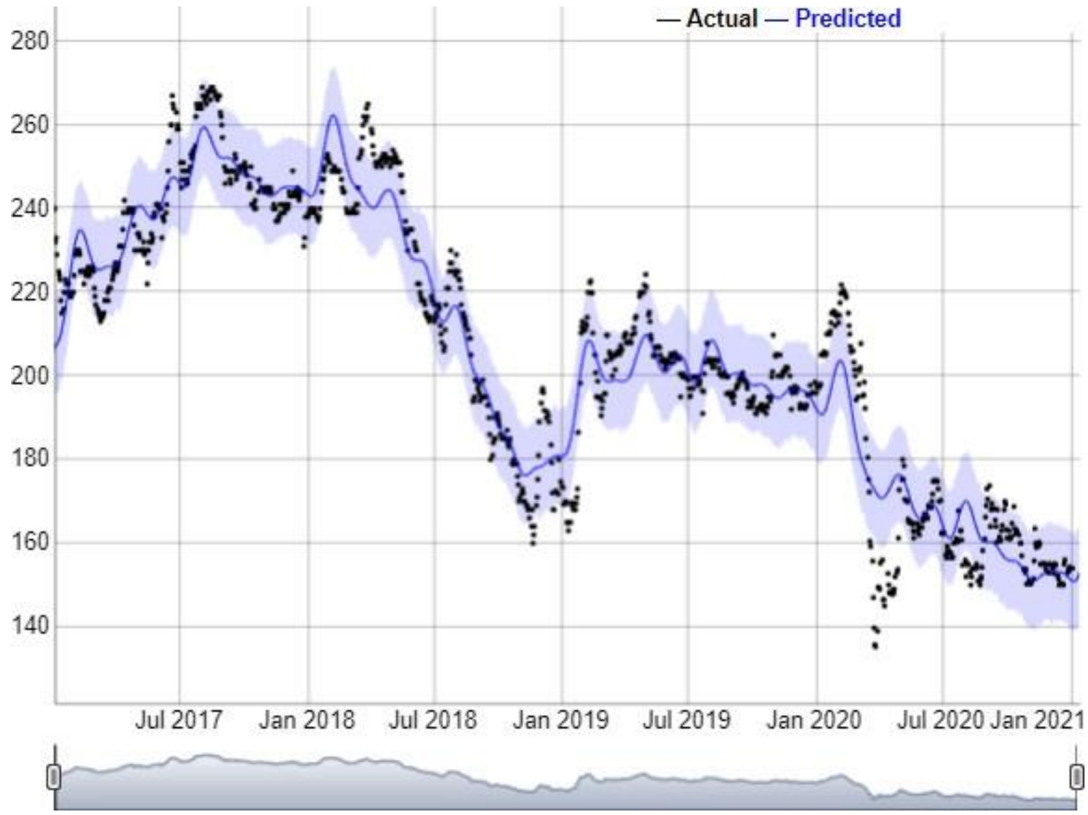

Figure 4.20: EABL Actual values vs Predicted values using prophet model

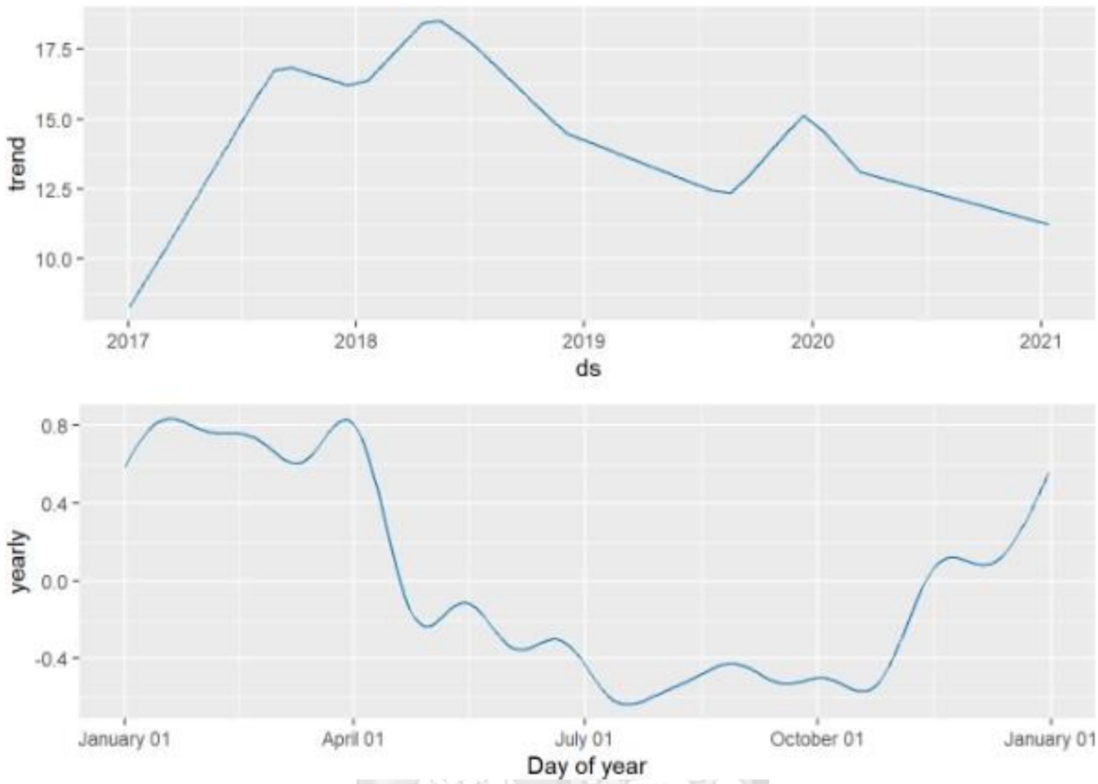

Figure 4.21: Components for the COOP prophet model

In our prophet model for the KCB stocks, we use a linear trend since we can observe a decreasing trend which is linear and does not follow a logistic growth. The change points for this series were set at 40 after trial and error which yielded the lowest RMSE for these stocks. The seasonality S(t) is used as an additive for the yearly seasonality with the default Fourier order. Once the model was created we forecasted the next 10 days and the fitted values can be seen in Figure 4.24.

Assessment is done on the validation set of the train data in order to determine the performance of the model. RMSE, MAE, and MAPE are the metrics used to assess the Prophet model (see Table 4.5).

RSME	MAE	MAPE
0.729	0.62	0.0162

Table 4.5: Statistics metrics for KCB.

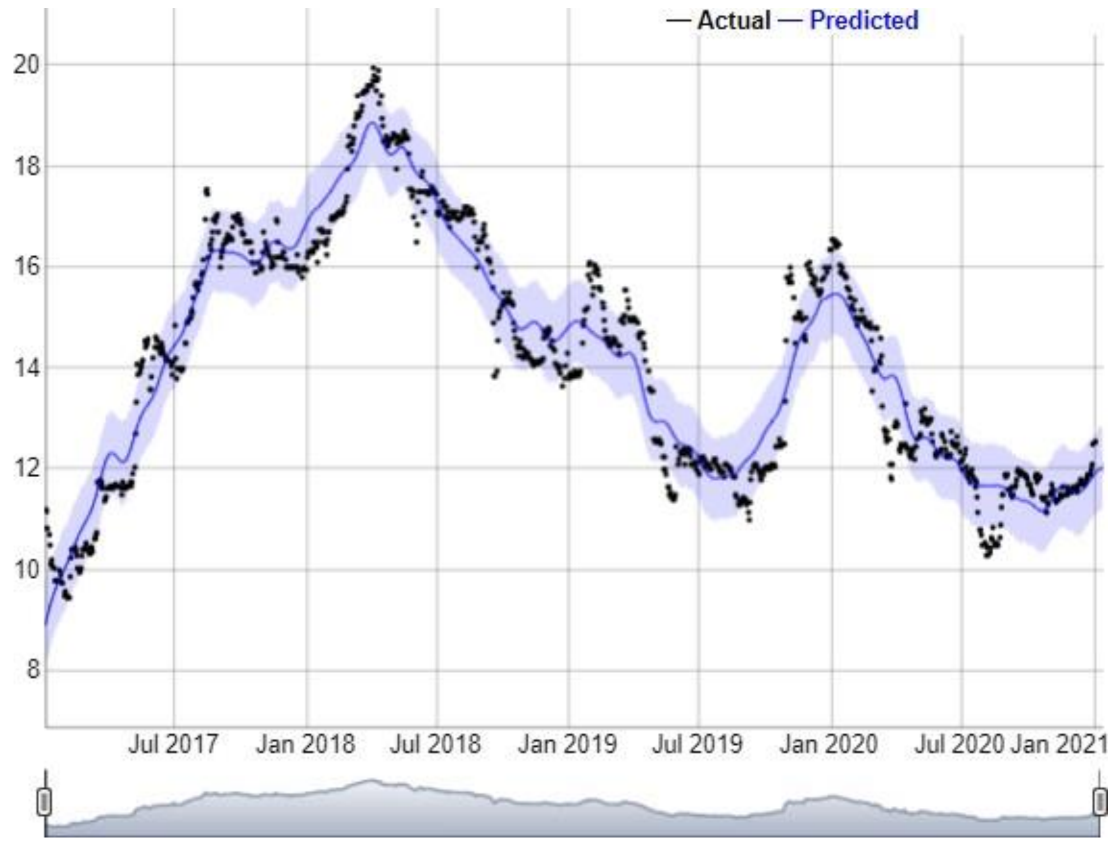

Figure 4.22: COOP Actual values vs Predicted values using prophet model

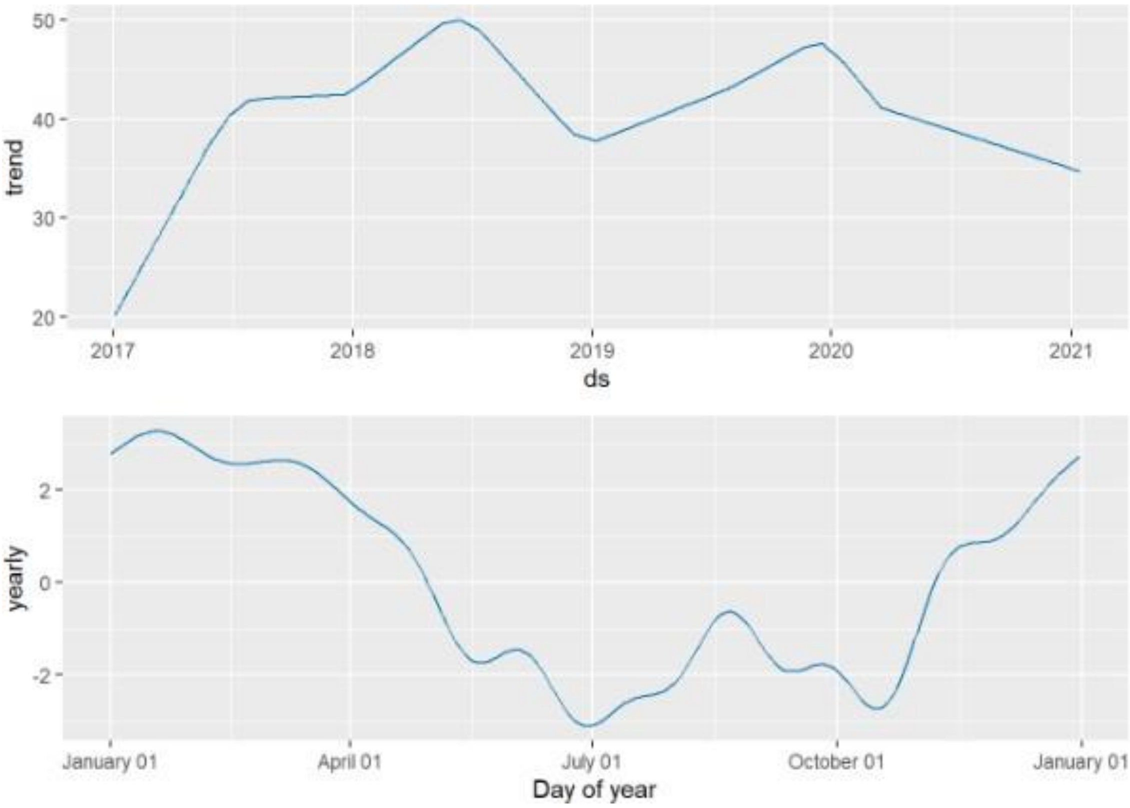

Figure 4.23: Components for the KCB prophet model

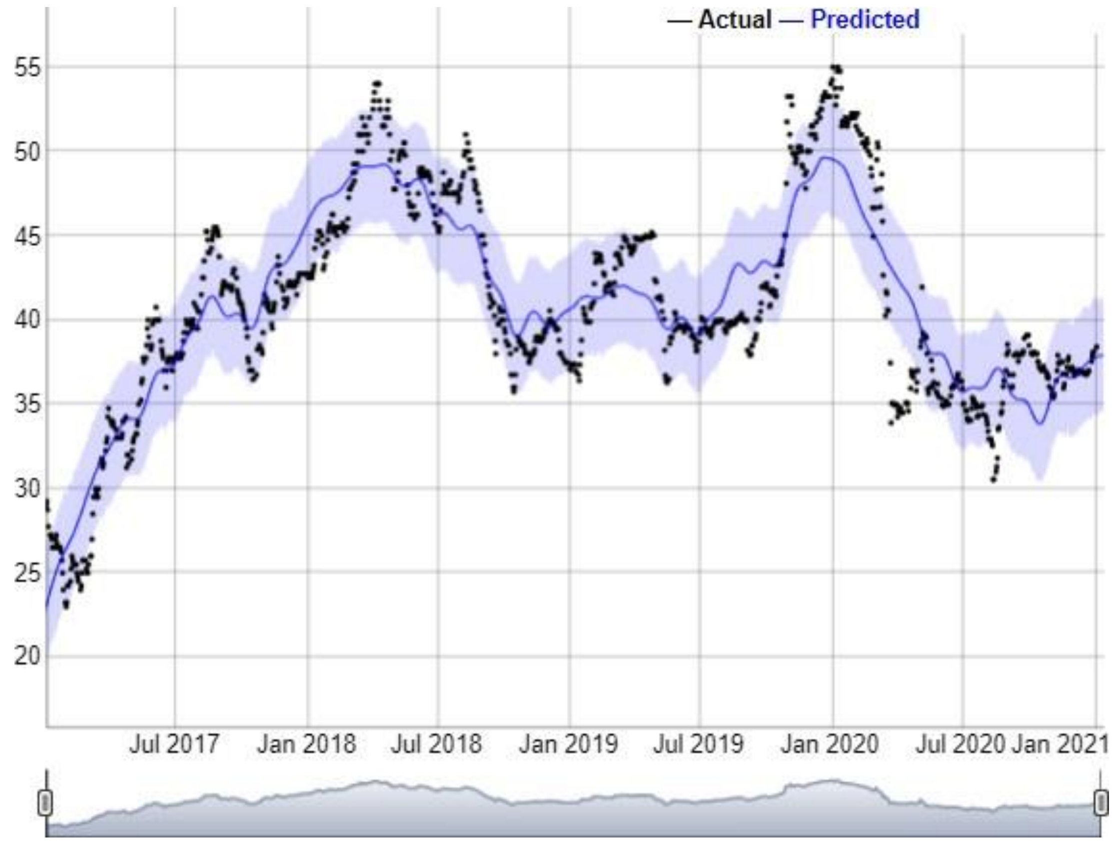

Figure 4.24: KCB Actual values vs Predicted values using prophet model

4.4 ARIMA and Prophet Model Comparison

To compare the two models and how they performed with the different stock predictions we will compare their RMSE and the actual vs predicted values for both models.

Table 4.6 Evaluation of the statistical metrics for SAFCOM, EQTY, EABL, COOP KCB stocks

Stocks	Prophet model (RMSE)	ARIMA model (RMSE)
SAFCOM	0.390	0.427
EQTY	1.43	0.680
EABL	3.21	2.645
COOP	0.262	0.206
KCB	0.729	0.614

Table 4.6: RMSE comparison

Comparing the RMSE of the forecasting model of the SAFCOM stocks which are the Prophet and the ARIMA (1,1,2) model. It is evident that the prophet model outperformed the ARIMA (1,1,2) model since it has a lower RMSE. However, for the other predicting models of the EQTY, EABL, COOP and KCB the ARIMA model outperformed the prophet model since it produced lower RMSE compared to those produced by the prophet model.

Both models were used to generate out-of-sample forecasts of daily stock prices of SAFCOM, EQTY, EABL, COOP, and KCB for the period starting from 31- Dec - 2021 up to 13 - Jan - 2022 This was further used for comparison; we will look at the actual values versus the predicted values for each model. As displayed in the actual vs predicted values for SAFCOM the ARIMA model predictions appear to be closer to the actual value compared to the prophet model predictions in Figure 4.26 This is also the case for the EQTY stock the ARIMA model appears to have predicted values that are closer to the actual values as compared to the prophet model see table. In the case of the EABL again the ARIMA model appears to have better predictions for the future values of the East African Breweries as compared to the Prophet model which seems to have predicted too high see Figure 4.27.

Date		ARIMA (1,1,2)	Prophet
	Actual values	Predicted	Predicted
2021-12-31	37.95	37.48	43.27
2022-01-03	38.15	37.56	43.29
2022-01-04	39.05	37.63	43.31
2022-01-05	39.05	37.68	43.33
2022-01-06	40.00	37.72	43.35
2022-01-07	39.90	37.74	43.37
2022-01-10	38.50	37.76	43.39
2022-01-11	38.45	37.78	43.41
2022-01-12	38.00	37.79	43.43
2022-01-13	37.95	37.80	43.44

Figure 4.25: Actual vs predicted values for SAFCOM stocks with the selected models

Moving forward to the COOP stocks both models appear to have performed well in the prediction of the future values of Cooperative bank however the values predicted by the ARIMA model appear to be closer to the actual values compared to those of the Prophet model see Figure 4.28 This is the same case for the KCB stocks where both models appear to have performed well in predicting future values however the values predicted by the ARIMA model appear to be closer to the actual values than the prophet model predictions which are seen to be a little higher than the actual values see Figure 4.29

4.5 Discussions

Different procedures were performed in this study in order to develop the best model to accurately forecast the stock prices for Kenyan Companies using ARIMA models and the Prophet model.

In order to develop ARIMA models for different stocks the data was visualized using the time series plot which showed that the time series for all the stocks was unstable in variance and mean. To accomplish stationarity in the variance and mean differencing was pragmatic to the data and the ADF test was used to confirm the stationarity of the data. For the prophet model, the data was transformed into that prophet format mentioned in 4.

Date		ARIMA (0,1,1)	Prophet
	Actual values	Predicted	Predicted
2021-12-31	52.75	52.59	56.57
2022-01-03	52.75	53.40	56.66
2022-01-04	53.25	54.00	56.71
2022-01-05	53.00	54.45	56.75
2022-01-06	53.00	54.87	56.80
2022-01-07	53.00	55.24	56.85
2022-01-10	52.75	55.58	56.89
2022-01-11	52.00	55.90	56.89
2022-01-12	50.25	56.19	56.94
2022-01-13	49.55	56.47	56.98

Figure 4.26: Actual vs predicted values for EQTY stocks with the selected models

Date		ARIMA (5,1,1)	Prophet
	Actual values	Predicted	Predicted
2021-12-31	165.00	164.89	176.20
2022-01-03	165.50	164.17	176.23
2022-01-04	165.00	163.64	176.27
2022-01-05	163.75	163.08	176.29
2022-01-06	160.75	162.72	176.33
2022-01-07	164.75	162.39	176.37
2022-01-10	163.25	162.13	176.40
2022-01-11	161.00	161.91	176.43
2022-01-12	160.00	161.73	176.46
2022-01-13	151.50	161.58	176.50

Figure 4.27: Actual vs predicted values for EABL stocks with the selected models

Date		ARIMA (2,1,1)	Prophet
	Actual values	Predicted	Predicted
2021-12-31	12.95	12.89	13.46
2022-01-03	13.00	12.90	13.47
2022-01-04	12.95	12.89	13.47
2022-01-05	13.00	12.90	13.48
2022-01-06	13.00	12.89	13.48
2022-01-07	12.95	12.90	13.48
2022-01-10	12.80	12.89	13.49
2022-01-11	12.80	12.90	13.49
2022-01-12	12.75	12.90	13.49
2022-01-13	12.55	12.90	13.50

Figure 4.28: Actual vs predicted values for COOP stocks with the selected models

Date		ARIMA (2,1,0)	Prophet
	Actual values	Predicted	Predicted
2021-12-31	45.45	45.19	47.45
2022-01-03	45.55	45.18	47.48
2022-01-04	45.05	45.18	47.50
2022-01-05	45.50	45.18	47.52
2022-01-06	45.90	45.18	47.55
2022-01-07	46.00	45.18	47.57
2022-01-10	46.40	45.18	47.60
2022-01-11	45.85	45.18	47.62
2022-01-12	45.45	45.18	47.65
2022-01-13	45.25	45.18	47.67

Figure 4.29: Actual vs predicted values for KCB stocks with the selected models

To formulate a time series model that predicts stock prices the traditional Box-Jenkins approach was used to fit ARIMA models. The results of this analysis showed that the best fit for daily stocks prices of SAFCOM, EQTY, EABL, COOP and KCB from Jan 2017 to Dec 2021 was ARIMA (1,1,2), ARIMA (0,1,1), ARIMA (5,1,1), ARIMA (2,1,1) and ARIMA (2,1,0) respectively this was achieved by analyzing the AIC, log-likelihood, Ljung box statistics and RMSE for all models.

Also, a prophet model was fit for all the daily stock all models prices of SAFCOM, EQTY, EABL, COOP and KCB for the same period to achieve the best prophet fit adjustments were made on the change points of the models to improve their accuracy. To select optimalmodels to predict the stock prices for the different stocks which are SAFCOM, EQTY, EABL, COOP and KCB their RMSE was compared and the results showed that the ARIMA model outperformed the prophet model in the prediction of the EABL, EQTY, COOP and KCB stocks prices however the prophet model outperformed the ARIMA in the prediction of the SAFCOM stock prices.

Further assessment of the model was done by comparing the actual values and the predicted values of the selected ARIMA and prophet models and again the predictions made by the ARIMA model appeared to be closer to all the actual stock prices compared to those predicted by the prophet model.

Chapter 5

Conclusions and Recommendations

5.1 Introduction

This chapter summarizes the study, in the following order: conclusion and recommendation

5.2 Conclusions

This study analyzes two time series models ARIMA and prophet model in search of the optimal time series model to predict stock prices for Kenyan companies. The Safaricom, Equity bank holding, East African breweries cooperative bank, and KCB stock prices were used to formulate a time series model that predicts stock prices and their accuracy was studied using the RMSE of each model and the model with the lowest RMSE was selected as the best model. The study found the ARIMA model to be better at predicting Kenyan stock prices.

5.3 Recommendations

The contemporary study has induced restrictions that can be enhanced. The Box-Jenkins approach used to forecast stock prices assumed that the stock prices only depend on time. However, stock prices can be affected by other factors like political events, natural disasters, speculations about a certain stock etc. Therefore, future studies could improve the accuracy of the model by incorporating these covariates into the model. By considering the fact that these components influence stock prices other than the time the model can be modified and the effects of these events can be anticipated. In addition, this study only focused on the top five stocks in Kenya currently therefore future studies can be expanded by adding more stocks within each industry to observe if the results obtained are a norm or an exception.

References

Aghababaeyan, R. and 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*, pages 13–17.

Adebiyi, A. A., Adewumi, A. O., & Ayo, C. K. (2014). Stock price prediction using the ARIMA model. *Proceedings - UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, UKSim 2014*, 106–112. https://doi.org/10.1109/UKSim.2014.67

Anjikar, A., Ramteke, K., Chavhan, S., and Dhabarde, S. (2022). Implementation of the arima model for stock value prediction in india. In *AIP Conference Proceedings*, volume 2424, page 060009. AIP Publishing LLC.

Ayub, S. and Jafri, Y. Z. (2020). Comparative study of an ann-arima hybrid model for predicting karachi stock price. *American Journal of Mathematics and Statistics*, 10(1):1–9.

Chan, W. N. (2020). Time series data mining: comparative study of arima and prophet methods for forecasting closing prices of myanmar stock exchange. *Journal of Computer Applications and Research*, 1(1):75–80.

Dutta et al., . (2006). Artificial neural network models for forecasting stock price index in the bombay stock exchange. *Journal of Emerging Market Finance*, 5(3):283–295.

Fama, E. F. (1995). Random walks in stock market prices. *Financial analysts journal*, 51(1):75–80.

Groenendijk, J., Wel, M. van der, & Vermeulen, S. (2021). *Predicting intraday stock returns using a hybrid ARIMA and long short-term memory neural network model*. https://thesis.eur.nl/pub/60908/Master-Thesis-Groenendijk.pdf

Kihoro, J. and Okango, E. (2014). Stock market price prediction using artificial neural network: an application to the kenyan equity bank share prices. *Journal of Agriculture, Science and Technology*, 16(1):160–171.

Latha, C. M., Nageswararao, K. S., and Venkataramanaiah, M. (2018). Forecasting time series stock returns using arima: Evidence from s & p bse sensex. *International Journal of Pure and Applied Mathematics*, 118(24):1–21.

Lawrence, R. (1997). Using neural networks to forecast stock market prices. *University of Manitoba*, 333:2006–2013.

Maskey, A. (2022). Predicting nepse index using arima model.

Mondal, P., Shit, L., & Goswami, S. (2014). Study of Effectiveness of Time Series Modeling (Arima) in Forecasting Stock Prices. *International Journal of Computer Science, Engineering and Applications*, *4*(2), 13–29. https://doi.org/10.5121/ijcsea.2014.4202

Murthi, M. V. N. (2021). Comparing Two ARIMA Models for Daily Stock Price Data. *International Journal for Research in Applied Science and Engineering Technology*, *9*(11), 1606–1610. https://doi.org/10.22214/ijraset.2021.38943

Patel, J., Shah, S., Thakkar, P., and Kotecha, K. (2015). Predicting stock market index using fusion of machine learning techniques. *Expert Systems with Applications*, 42(4):2162–2172.

Pindiga, S. N. (2022). Time-Series Forecasting: Predicting Stock Index Using Arima and Facebooks Prophet Model. *International Journal for Research in Applied Science and Engineering Technology*, *10*(6), 4832–4839. https://doi.org/10.22214/ijraset.2022.45073

Ramteke, M. (2020). *Stock market price prediction using time series models*. PhD thesis, Dublin, National College of Ireland.

Sah, S., Surendiran, B., Dhanalakshmi, R., Mohanty, S. N., Alenezi, F., and Polat, K. (2022). Forecasting covid-19 pandemic using prophet, arima, and hybrid stacked lstm-gru models in india. *Computational and Mathematical Methods in Medicine*, 2022.

Saxena, K. and Kamnge, J. S. (2020). Modelling and forecasting daily stock price of nmb bank in tanzania.

Shah, V. and Dimitrov, S. (2022). A comparative study of univariate time-series methods for sales forecasting. *International Journal of Business and Data Analytics*, 2(2):187–216.

Sheeba, S. L., Gupta, N., Ragavender, R. A., and Divya, D. (2021). Time series model for stock market prediction utilising prophet. *Turkish Journal of Computer and Mathematics Education*, 12(6):4529–4534.

Shukor, S. A., Sufahani, S. F., Khalid, K., Abd Wahab, M. H., Idrus, S. Z. S., Ahmad, A., and Subramaniam, T. S. (2021). Forecasting stock market price of gold, silver, crude oil and platinum by using double exponential smoothing, holt's linear trend and random walk. In *Journal of Physics: Conference Series*, volume 1874, page 012087. IOP Publishing.

Taylor, S. J. and Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72(1):37–45.

Wanjawa, B. W. and Muchemi, L. (2014). Ann model to predict stock prices at stock exchange markets. *arXiv preprint arXiv:1502.06434*.

Appendices

Appendix A: ACF and PACF graphs

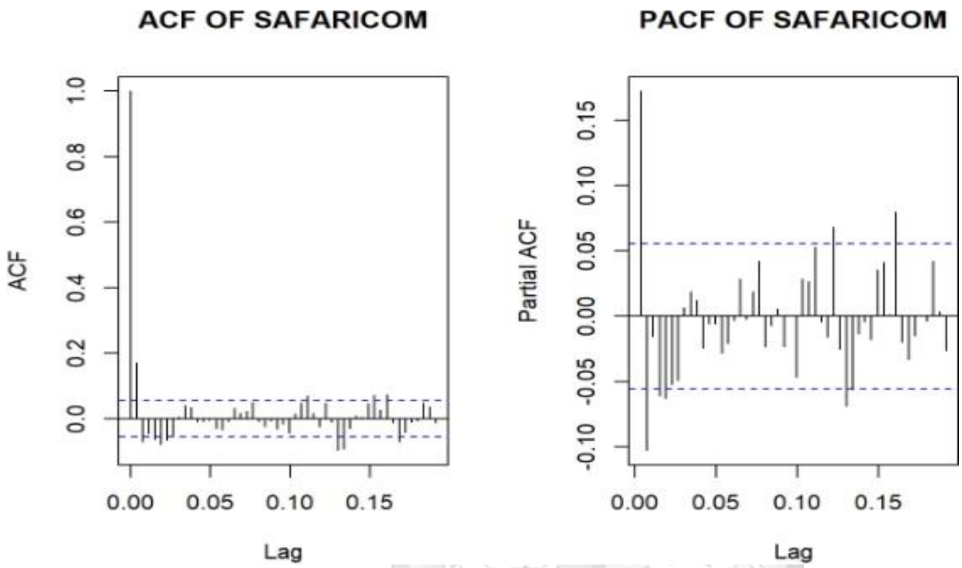

Figure A.1: Sample Autocorrelation Function for First Difference Stock Price for Safaricom

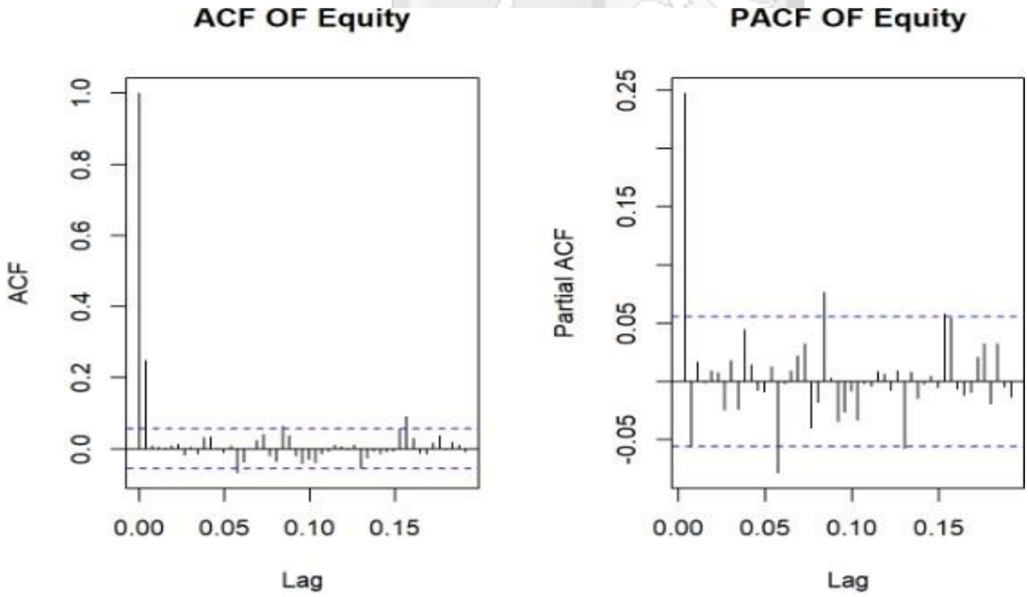

Figure A.2: Sample Autocorrelation Function for First Difference Stock Price for Equity

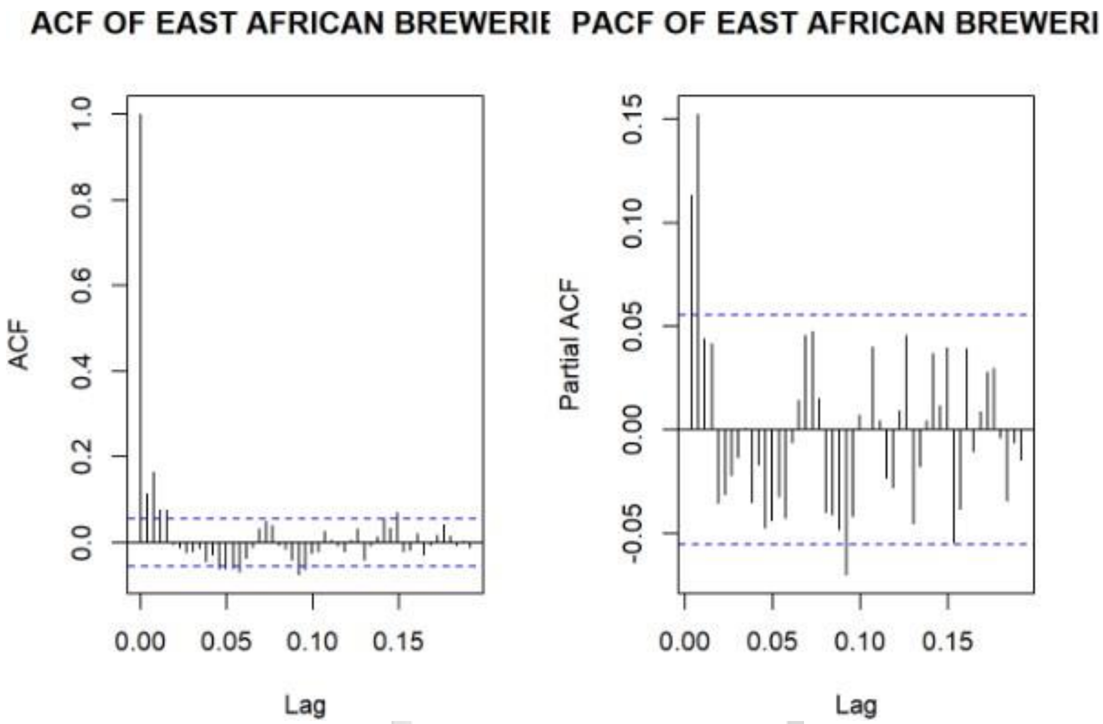

Figure A.3: Sample Autocorrelation Function for First Difference Stock Price for EABL

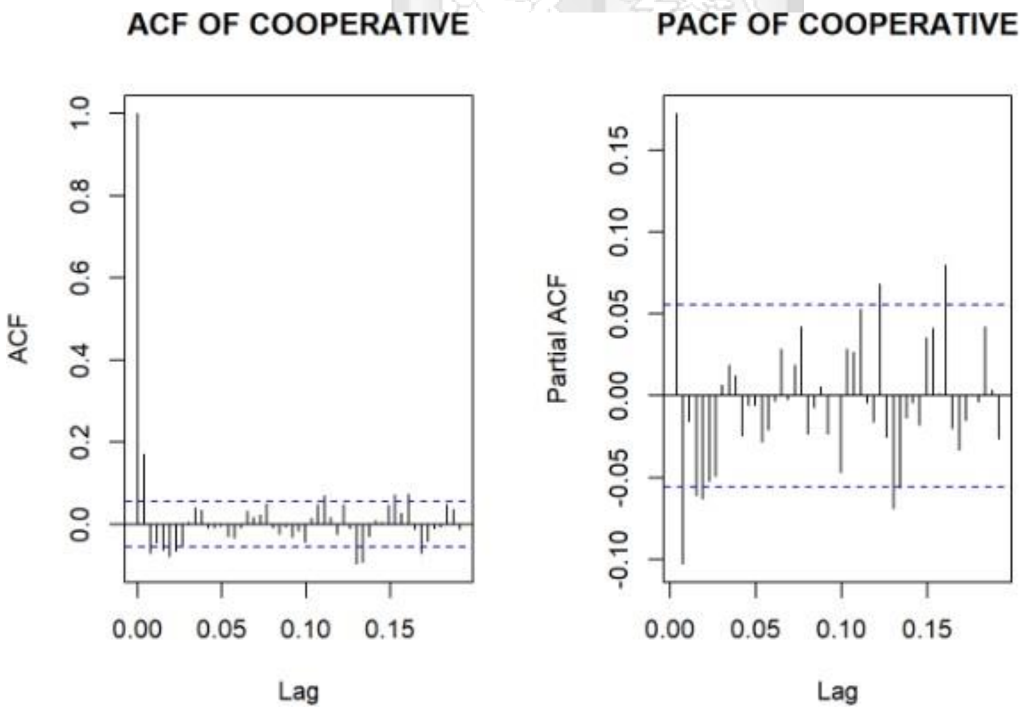

Figure A.4: Sample Autocorrelation Function for First Difference Stock Price for COOP

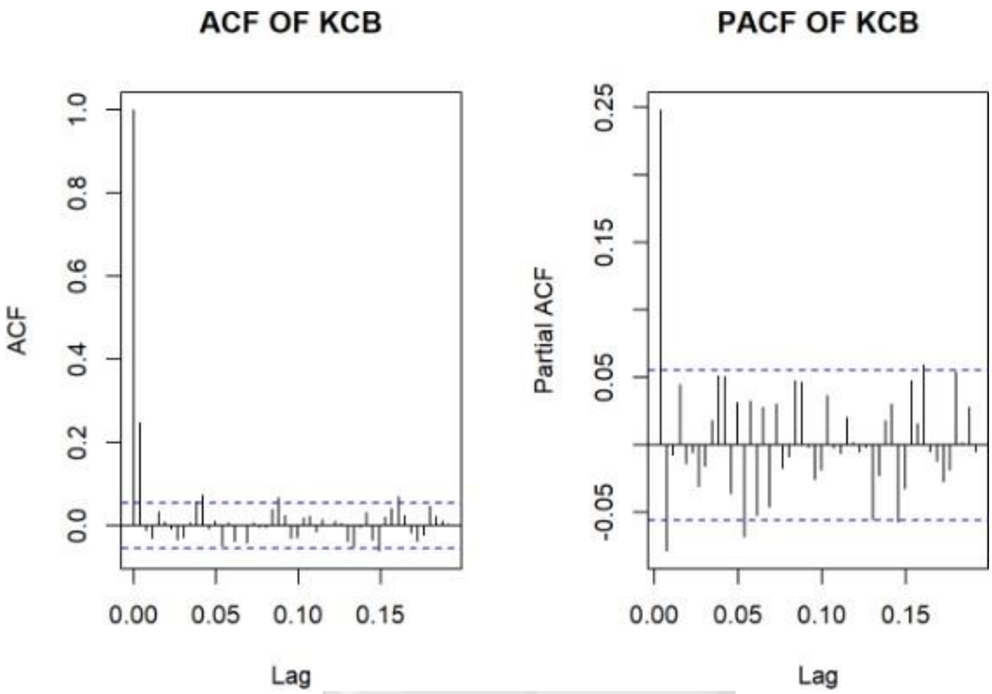

Figure A.5: Sample Autocorrelation Function for First Difference Stock Price for kcb

Appendix B: Similarity report

PAULYNE KERUBO

ORIGINALITY REPORT

24%	19%	8%	14%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	repository.udom.ac.tz Internet Source	3%
2	su-plus.strathmore.edu Internet Source	3%
3	www.researchgate.net Internet Source	2%
4	Submitted to RMIT University Student Paper	2%
5	Submitted to The University of Dodoma Student Paper	1%
6	Submitted to Liverpool John Moores University Student Paper	1%
7	Submitted to University College London Student Paper	1%
8	Submitted to Queen's University of Belfast Student Paper	1%
9	towardsdatascience.com Internet Source	1%

Figure A.6: Originality test

Appendix C: Ethical Clearance Confirmation

Strathmore UNIVERSITY

27th March 2023

Ms Moenga Paulyne Kerubo,
paulyne.moenga@strathmore.edu

Dear Ms Moenga,

RE: Identifying the Optimal Time Series Model to Predict Kenyan Stock Prices

This is to inform you that SU-ISERC has reviewed and **approved** your above **SU- master's** research proposal. Your application reference number is **SU-ISERC1621/23.** The approval period is from **27th March 2023 to 26th March 2024.**

This approval is subject to compliance with the following requirements:

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

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

Yours sincerely,

for: **Dr Ben Ngoye,**
Secretary; SU-ISERC

Cc: Mr Ambrose Rachier,
Chairperson; SU-ISERC

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

Ole Sangale Rd, Madaraka Estate. PO Box 59857-00200, Nairobi, Kenya. Tel +254 (0)703 034000
Email admissions@strathmore.edu www.strathmore.edu

Figure A.7: Ethics approval document

Appendix D: R code

```
***********LOADING DATA *************
Safaricom<-read_excel("D:/MASTERS/SEMESTER 5/THESIS/data/Safaricom 2.xlsx")
View(Safaricom)
safaricom2<-Safaricom[1:1248,2]

safaricom3<-ts(safaricom2,start = c(2017,1),frequency = 261)

plot.ts(safaricom3,main="SAFCOM  stocks")

adf.test(safaricom3)

#splitting the data saf_trainig<-
safaricom3[1:999,]
saf_test<-safaricom3[1000:1248,]
fit_saf11<-arima(saf_trainig,order  =  c(1,1,2))
fit_saf22<-arima(saf_trainig,order  =  c(2,1,1))
fit_saf33<-arima(saf_trainig,order = c(2,1,2))

###### make data stationary
safaricom4<-diff(safaricom3)
plot(safaricom4,main="Differenced Graph of Safaricom")
par(mfrow=c(1,2))
acf(safaricom3,lag.max = 50,main="ACF OF SAFARICOM")
```

```
pacf(safaricom3,lag.max = 50, main= "PACF OF SAFARICOM")
par(mfrow=c(1,2))
acf(safaricom4,lag.max = 50,main="ACF OF SAFARICOM")
pacf(safaricom4,lag.max = 50, main= "PACF OF SAFARICOM")
adf.test(safaricom4)

###### Identifying ARIMA model
fit_saf1<-arima(safaricom3,order = c(1,1,2))
fit_saf2<-arima(safaricom3,order = c(2,1,1))
fit_saf3<-arima(safaricom3,order = c(2,1,2))

summary(fit_saf1)
summary(fit_saf2)
summary(fit_saf3)

######### Forecasting using the selected model
saf_fit1<-forecast(    fit_saf1,h=10)
summary(saf_fit1)
plot(saf_fit1)
lines(fitted(fit_saf1),col="green")

saf_fit2<-forecast(fit_saf2,h=10)
summary(saf_fit2)

saf_fit3<-forecast(fit_saf3,h=10)
summary(saf_fit3)

######### residuals
checkresiduals(fit_saf1,lag = 20)
checkresiduals(fit_saf2,lag = 20)
```

```
checkresiduals(fit_saf3,lag = 20)

*****************LOADING DATA*************
Equity<-read_excel("D:/MASTERS/SEMESTER 5/THESIS/data/Equity 1.xlsx")

Equity2<-Equity[1:1248,2]
View(Equity)
Equity3<-ts(Equity2,start = c(2017,1),frequency = 261)

plot.ts(Equity3,main="EQTY stocks")

adf.test(Equity3)

########### make data stationary
Equity4<-diff(Equity3)
plot(Equity4,main="Differenced Graph of Equity")
par(mfrow=c(1,2))
acf(Equity3,lag.max = 50,main="ACF OF Equity")
pacf(Equity3,lag.max = 50, main= "PACF OF Equity")
par(mfrow=c(1,2))
acf(Equity4,lag.max = 50,main="ACF OF Equity")
pacf(Equity4,lag.max = 50, main= "PACF OF Equity")
adf.test(Equity4)

######identifying ARIMA model
auto.arima(Equity2,trace = T)
fit_Equity1<-arima(Equity3,order = c(0,1,1))
fit_Equity2<-arima(Equity3,order = c(0,1,2))
fit_Equity3<-arima(Equity3,order = c(1,1,1))
summary(fit_Equity1)
```

```
summary(fit_Equity2)
summary(fit_Equity3)

########### Forecasting using selected model
Equity_fit1<-forecast(fit_Equity1,h=10)
summary(Equity_fit1)
plot(Equity_fit1)
lines(fitted(fit_Equity1),col="orange")
Equity_fit2<-forecast(fit_Equity1,h=10)
summary(Equity_fit2)
Equity_fit3<-forecast(fit_Equity1,h=10)
summary(Equity_fit3)

######residuals
checkresiduals(fit_Equity1)
checkresiduals(fit_Equity2)
checkresiduals(fit_Equity3)

EAB<-read_excel("D:/MASTERS/SEMESTER 5/THESIS/data/EAB.xlsx")
View(EAB)
EAB2<-EAB[1:1248,2]
EAB3<-ts(EAB2,start = c(2017,1),frequency = 261)
plot.ts(EAB3,main="EABL  stocks")
adf.test(EAB3)

#make data stationary
EAB4<-diff(EAB3)

plot(EAB4,main="Differenced Graph of EAST AFRICAN BREWERIES")
par(mfrow=c(1,2))
```

```
acf(EAB3,lag.max = 50,main="ACF OF EAST AFRICAN BREWERIES")
pacf(EAB3,lag.max = 50, main= "PACF OF EAST AFRICAN BREWERIES")
par(mfrow=c(1,2))
acf(EAB4,lag.max = 50,main="ACF OF EAST AFRICAN BREWERIES")
pacf(EAB4,lag.max = 50, main= "PACF OF EAST AFRICAN BREWERIES")

adf.test(EAB4)

######### Identifying ARIMA model
auto.arima(EAB2,trace = T)
fit_EAB1<-arima(EAB3,order = c(5,1,1))
fit_EAB2<-arima(EAB3,order = c(5,1,2))
fit_EAB3<-arima(EAB3,order = c(4,1,2))
summary(fit_EAB1)
summary(fit_EAB2)
summary(fit_EAB3)

###### Forecasting using selected model
EAB_fit1<-forecast(fit_EAB1,h=10)
summary(EAB_fit1)
plot(EAB_fit1)
lines(fitted(fit_EAB1),col="yellow")

EAB_fit2<-forecast(fit_EAB1,h=10)
summary(EAB_fit2)

EAB_fit3<-forecast(fit_EAB1,h=10)
summary(EAB_fit3)

######## Residuals
```

```
checkresiduals(fit_EAB1,lag = 250)

checkresiduals(fit_EAB2,lag = 250)
checkresiduals(fit_EAB3,lag = 250)
***************LOADING DATA********

Cooperative<-read_excel("D:/MASTERS/SEMESTER 5/THESIS/data/Cooperative.xlsx")
View(Cooperative)
Cooperative2<-Cooperative[1:1248,2]

Cooperative3<-ts(Cooperative2,start = c(2017,1),frequency = 261)

plot.ts(Cooperative3,main="COOP  stocks")

adf.test(Cooperative3)

#make data stationary
Cooperative4<-diff(Cooperative3)

plot(Cooperative4,main="Differenced Graph of Cooperative")

par(mfrow=c(1,2))
acf(Cooperative3,lag.max = 50,main="ACF OF Cooperative")
pacf(Cooperative3,lag.max = 50, main= "PACF OF Cooperative")

par(mfrow=c(1,2))
acf(safaricom4,lag.max = 50,main="ACF OF COOPERATIVE")
pacf(safaricom4,lag.max = 50, main= "PACF OF COOPERATIVE")

adf.test(Cooperative4)
```

```
######### identifying ARIMA model
fit_coop1<-arima(Cooperative3,order  =  c(1,1,0))
fit_coop2<-arima(Cooperative3,order  =  c(2,1,1))
fit_coop3<-arima(Cooperative3,order = c(2,1,0))

summary(fit_coop1)
summary(fit_coop2)
summary(fit_coop3)


#######forecasting using selected model
coop_fit2<-forecast(fit_coop2,h=10)
summary(coop_fit2)
plot(coop_fit2)
lines(fitted(fit_coop2),col="green")

coop_fit2<-forecast(fit_coop1,h=10)
summary(coop_fit2)

coop_fit3<-forecast(fit_coop1,h=10)
summary(coop_fit3)

#residuals
checkresiduals(fit_coop1)
checkresiduals(fit_coop2)
checkresiduals(fit_coop3)

**************LOADING DATA***********
KCB<-read_excel("D:/MASTERS/SEMESTER  5/THESIS/data/KCB.xlsx")
```

```
View(KCB)
KCB2<-KCB[1:1248,2]

KCB3<-ts(KCB2,start = c(2017,1),frequency = 261)

plot.ts(KCB3,main="KCB stocks")

adf.test(KCB3)

########make data stationary
KCB4<-diff(KCB3)

plot(KCB4,main="Differenced Graph of KCB")

par(mfrow=c(1,2))
acf(KCB3,lag.max = 50,main="ACF OF KCB")
pacf(KCB3,lag.max = 50, main= "PACF OF KCB")

par(mfrow=c(1,2))
acf(KCB4,lag.max = 50,main="ACF OF KCB")
pacf(KCB4,lag.max = 50, main= "PACF OF KCB")

adf.test(KCB4)
#####identifying ARIMA model

fit_KCB1<-arima(KCB3,order = c(2,1,0))
fit_KCB2<-arima(KCB3,order = c(1,1,1))
fit_KCB3<-arima(KCB3,order = c(3,1,0))

summary(fit_KCB1)
```

```
summary(fit_KCB2)
summary(fit_KCB3)

#forecasting using selected model
KCB_fit1<-forecast(fit_KCB1,h=10)
summary(KCB_fit1)
plot(KCB_fit1)
lines(fitted(fit_KCB1),col="green")

KCB_fit2<-forecast(fit_KCB1,h=10)
summary(KCB_fit2)

KCB_fit3<-forecast(fit_KCB1,h=10)
summary(KCB_fit3)

accuracy(KCB_fit1)

######residuals
checkresiduals(fit_KCB1)
checkresiduals(fit_KCB2)
checkresiduals(fit_KCB3)

***************Prophet model *******

safaricom22<-Safaricom[1:1248,1:2]

names(safaricom22) <- c('ds', 'y')

train_day<-safaricom22[1:999,]
```

```
test_day<-safaricom22[1000:1248,]

im <- (prophet(
df =safaricom22 ,    # Dataframe containing the history
growth = "linear",     # trend change/growth can't be logistic or flat for this TS
n.changepoints = 25, #more than default, might overfit
changepoint.range = 0.80, # Proportion of history in which trend changepoints will
yearly.seasonality = TRUE, # Default Fourier Order
weekly.seasonality = FALSE, # no evidence for temp change for days of a week
daily.seasonality = FALSE, # Daily seasonality locked
as off as data is daily
leveled
holidays = NULL, # no evidence that holidays affect temp
seasonality.mode = "additive", # by observation
seasonality.prior.scale = 10, # default
holidays.prior.scale = 10, # default for regressors too
changepoint.prior.scale = 0.05, # default
mcmc.samples = 0, # default
interval.width = 0.80, #default
uncertainty.samples = 1000, #default
fit = TRUE
))

future <- make_future_dataframe(im, periods=10)
# prediction
fcst_im <- predict(im,future) # creating forecast for 10 days
#tail(fcst_im[c('ds', 'yhat', 'yhat_lower', 'yhat_upper')]) #observing tail
observations tail(fcst_im,10)
dyplot.prophet(im,fcst_im , uncertainty = TRUE) # creating interactive plots for
the forecast
```

```
# plotting components of the forecast
prophet_plot_components(
im,
fcst_im,
uncertainty = TRUE,
plot_cap = TRUE,
yearly_start = 0,
render_plot = TRUE
)

# out of sample (validation set) assessment
RMSE_im = sqrt(mean((tail(train_day$y,10) - tail(fcst_im$yhat,10))^2))
MAE_im = mean(abs( tail(train_day$y,10) - tail(fcst_im$yhat,10) ))
MAPE_im = mean(abs( (tail(train_day$y,10) - tail(fcst_im$yhat,10)) /
tail(train_day$y,10) ))

tibble("RMSE"= c(round(RMSE_im,4)), "MAE" = round(MAE_im,4),
"MAPE" = round(MAPE_im,4))
```

VT OMNES VNVM SINT