2020

A model for forex market price prediction: case of Central Bank of Kenya.

Makaya, David Nyangau
Faculty of Information Technology
Strathmore University

A Model for Forex Market Price Prediction: Case of Central Bank of Kenya

By

David Nyang'au Makiya

A Thesis Submitted to the Faculty of Information Technology in partial fulfillment of the requirements for the award of a degree in Masters of Science in Information Technology.

Masters of Science in Information Technology

Faculty of Information Technology

Strathmore University

June 2020

Declaration and Approval

I declare that this work has not been previously submitted and approved for the 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.

David Nyang'au Makiya

26th June, 2020

Approval

This thesis of David Nyang'au Makiya was reviewed and approved by:

Dr. Joseph Orero,

Dean, Faculty of Information Technology,

Strathmore University

Dr. Bernard Shibwabo,

Director of Graduate Studies,

Strathmore University

Abstract

Forex markets are full of uncertainties. The forces of demand and supply determine the price of the Kenyan shilling in the international market. The rates usually adjust depending on the prevailing status of the economy, politics and influences of the Central Bank of Kenya (CBK) policies. Forex dealers are the dominant operators within the Kenyan forex market with forex bureaus and commercial banks taking the lead amongst them. A bureau would have a different pricing of a currency against the shilling but would nonetheless be within the bid-ask (buy sell) spread of the CBK ratings. Various online forex trading platforms have been implemented to facilitate trade in the Kenyan market. Predicting forex market prices is quite complicated as a process and subjective in nature for forex dealers, economists and business persons. The potential to make loses due to poor speculative guesses is quite high for multinational organizations located in more than one economy. The aim of this study is to develop a model for forex market price prediction in the Kenyan market using the Central Bank of Kenya data. Using the Data-Driven modelling technique, a model for forex market price prediction has been developed based on historical data from the CBK. The dataset is divided into training and testing data by a splitting of 80-20 respectively. The unique behavior of each of the currency data necessitated separate implementation of the currencies on the model for increased accuracy and lower error levels hence efficiency and optimality. The prediction model is achieved by combining time series analytical techniques with resilient backpropagation neural network. Successful predictions are conducted of up-to eight months forward with accuracy levels ranging 88-98% and Sum of squared residual (SSE) of 0.496-2.667, hence showing that combining time series analytics to resilient backpropagation neural networks to create a forex market prediction model with unique implementation of each currency is optimal for forex market price prediction where more data depicts longer period predictions.

Keywords: Forex Market Price, Data-Driven Modelling, Time Series Analytics, Resilient Backpropagation Neural Network

Table of Contents

List of Tables

List of Figures

List of Equations

Abbreviations/Acronyms

ANN	–Artificial Neural Networks
BI	–Business Intelligence
CBK	–Central Bank of Kenya
CPI	–Consumer Price Index
DDM	–Data-Driven Modelling
ENXM	–Elastic Network Model for Forex Market
FAPS	–Forex Autopilot System
FOREX	–Foreign exchange
GDP	–Gross Domestic Product
GNP	–Gross National Product
kNN	–k Nearest Neighbor
LRMA	–Long-term MA
MA	–Moving Averages
MSE	–Mean Squared Error
NLP	–Natural Language processing
PPP	–Theory of Purchasing Power parity
RBA	–Robotic Automation process
RNA	–Ribonucleic Acid
rprop+	–Resilient Backpropagation Algorithm
SRMA	–Short-run MA
SSE	–Sum of Squared Error/Residual
UML	–Unified Modelling Language

USD –United States Dollar

Chapter 1 : Introduction

1.1 Background

Financial markets in Kenya have been in consistent development over the past decade. The formations of the financial markets in Kenya constitutes of capital and money markets (Amadeo, 2019). The two markets can exist as either fixed income markets, alternative investment markets, main investment segment market and the forex market (Ndeda, Financial Institutions and Markets, 2015). Forex markets are full of uncertainties and from this, Kenya has seen its currency value rapidly fluctuate over the years. The slow but steady adoption of technology has facilitated proper storage and recording of all the changes and shifts that have happened over the years.

The causes of shifts recorded in the forex rates in Kenya are attributed to a number of events, some of which include; economic variations of the country, exchange rate policies and most importantly the Balance of payments crisis from the 1990's (Ndung'u, 2000). Other than this, trade balance, Gross domestic product and prevailing interest rates can also be sited for the same (Wanjiku, 2012).

After Kenya was declared independent, the nation still used the East African shilling as the centralized national reserve board was the East African Currency Board which ensured stability of the pricing of the then currency. After 1974, Kenya adopted its own Central Bank that had the explicit oversight roles of monetary policy and forex currency implementations (Wanjiku, 2012). At that particular moment a fixed rate regime was still in use where the central bank simply set a rate and it was used across all the whole board. In 1990 however, a dual exchange rate was adopted but only affecting specific segments of the economy and the US Dollar currency (Ndung'u, 2000). After 1993 the Central bank adopted the crawl which was the floating rate regime which took considerations of the parallel markets according to Ndung'u (2000). The floating rate regime implied pegging the rates to the normal forces of demand and supply and having the Kenyan shilling compete with the other currencies in the international market for value pricing.

Forex trading systems have been adopted in common forex market place. The common characteristic of such tools is their robot like nature which makes them referred to as forex bots (Gaucan, 2011). However, the popular aspect on such trading tools is not their predictive nature rather the ability to simply facilitate trades in the forex market. Some of such programs include; FAPTurbo, Forex Megadroid, U$DBot, Forex Auto Pilot System (FAPS).

The complexity tied to predictive analytics in forex trading is based on the fact that various economic factors are different from the base currency of a given economy say Kenya for instance. It might be possible running a forex rate forecast of Kenya shillings against United states dollar using any of the common approaches which is still not automated in Kenya. However, if trying to run multiple comparisons of different currencies based on a singular base currency, the variability of the respective currencies might have different behavioral patterns that may necessitate the application of much more advanced tools that have the abilities to process big data at high speeds. This therefore demonstrates the motivation behind this study that depicts the necessity of applying machine learning techniques in forex market pricing predictions in Kenya.

1.2 Problem Statement

According to Wanjiku (2012), due to the shift from using the fixed rates to floating rates in Kenya, there is need for forecasting majorly among speculators and fiduciaries in the financial markets to hedge against foreign exchange risk. The uncertainty that comes with using the floating rates has created the foreign exchange risk whilst at the same time exposing the forex market to interest rate exposures, political risks and other additional inherent factors. It therefore makes predicting forex market prices quite complicated as a process and subjective in nature for forex dealers, economies and business persons. The potential to make loses due to poor speculative guesses is quite high for multinational organizations located in more than one economy. Financial models have created derivative products that hedge against such risks but such hedging structures like forward swaps and futures contracts are based on speculative guesses and current theories of forecasting. Such derivative products are based on the short-term money markets that usually last for within ranges of ninety (90) days to half a year (Ndeda, Financial Institutions and Markets, 2015).

The direct inter-relation of more than one currency in an economy makes it complex to predict different market exchange rates considering that one currency say for Chinese Yuan might not have the exact same variability factors as to another currency say, South African Rand. There is therefore need for a prediction model that can attempt to address the challenges posed by the insufficiencies, of current approaches, that can simulate multi-economic exchange rates over the stretch of one years' time horizon. This study addresses the need using a forex market price prediction model that utilizes CBK dataset as a solution.

1.3 Aim

The aim of this study is to develop a model for forex market price prediction in the Kenyan market using the Central Bank of Kenya data.

1.4 Specific Objectives

i. To examine the factors affecting forex market prices in Kenya

ii. To review existing forex market price prediction approaches

iii. To develop a model for forex market price prediction

iv. To validate the forex market price prediction model for efficiency and optimality on predictions in the forex markets.

1.5 Research Questions

i. What are the factors affecting forex market prices/rates in Kenya?
ii. How are predictive analytical tools/approaches used to forecast forex market rates?
iii. How can the model for forex market price prediction be developed?
iv. How can the forex market price prediction model be validated for efficiency and optimality for predictions in the forex markets?

1.6 Justification

The study is meant to inform the participants of the forex market in the economic region of Kenya. Such participants include forex traders/dealers/hedgers and most importantly facilitate a benchmark for tentative future rates for both the traders/dealers/hedgers and the Central Bank of Kenya. In the highlight of Sercu's (n.d) slide presentation on forex markets self-predictability, it is prudent that the major question of whether the forex market can predict themselves has not been fully answered. Forward rates have been used over the past to price derivatives in the secondary market, however, for an economy like Kenya's, the derivatives market is not quite well developed hence less sophisticated. With the rapid growth rate of forex trading in Kenya, the market is expected to grow in volumes of daily trades hence increased volatility and instant moments changes notice depending on the prevailing conditions of a given country/economy (Gaucan, 2011). It therefore paves way for simplicity in application of a predictive model that simply employs analytical tools of business intelligence to leverage on multiple currencies at a go.

The study ought to create an automated support tool the Central Bank of Kenya may utilize in pricing and estimating of the forex market rates. It is important to note that the forex market will always have a direct impact on prevailing inflation rates and interest rates. It therefore implies that, other than simply predicting the exchange rates, the model can also be used by other banks and financial institutions to have a tentative feel of the time value for money in the foreseeable time horizon within prediction set. According to Bauer College of Business (2019), exchange rates forecasts are necessary to also evaluate the foreign denominated cash flows from international transactions. It therefore implies that such prediction is very important in evaluation of the benefits and risks associated with the international business environment.

1.7 Scope and Limitation

The major output of the study is a predictive model that utilizes machine learning in neural networks techniques to predict forex market prices to the time horizon of an approximate one year. The basis of formulating the model is tied within the context of Kenyan study. The study is limited only to the currencies relating to the Kenya shillings directly with data available from the Central Bank of Kenya repository between the years of 2016 and 2019. The currencies are purposively sampled to three only due to time factor and data availability i.e.) United States dollar, Chinese Yuan and South African Rand. With the construction, the model is able to predict to at least 8-12 months going forward with different currencies.

Applications of predictive modelling techniques have not quite been exhaustively attempted for the forex market prices in Kenya using business intelligence techniques. The applications of the same might require the supplement of additional techniques like the stochastic models that may have been used before as in Wanjiku's (2012) attempt.

Chapter 2 : Literature Review

2.1 Overview

The purpose of this chapter is to assess the current theories, studies and approaches relating to forex markets focusing on forecasting and prediction models. The end game is to capture prediction analytical technique into the context of forex market prices.

The focus of the chapter is to investigate the various factors affecting/determining forex market pricing in Kenya and examine the current prediction models for forex market prices/exchange rates. The suitability of the prediction analytical tool for selection shall be dependent on the strengths identified in each of the examinable ones and gaps identified from the same.

The chapter builds up-to a conceptual framework that formulates the tentative guiding factor of the deliverable predictive model. In other context, the intrinsic examinations of literature facilitate the build-up into a technical analytical framework by utilizing all the identified factors from the examinations.

2.2 Forex Markets in Kenya

The term FOREX refers to the Foreign currency exchange market in which over 4,600 international banks and other small and large speculators around the world are in (Gaucan, 2011). On a daily basis this worldwide market exchanges over $1.7 trillion in dozens of different currencies. Pondicherry University (2019), defines a Forex market as one in which currencies are bought and sold. It is not considered the same as a market in which currencies are borrowed and lent.

The Foreign exchange market is categorized as an over the counter market within the secondary market trading segment (Ndeda, Economics, 2015). Usually, there is not necessarily a physical place where participants may meet and execute their trades rather it is more of an informal arrangement happening between forex trade brokers, dealers or banks operating in a financing center for purchasing and selling currencies (Pondicherry University, 2019). In Kenya, such centers have been established as Forex bureaus and commercial banks.

In this market, the value/price of one currency is normally determined by comparing it to another currency via the exchange rate (Gaucan, 2011). Some of the major currencies under trading include the United States Dollar (USD), Euro (EUR), Japanese yen (JPY) and British pound (GBP). The Central Bank of Kenya, has facilitated an environment by being the regulator of all commercial banks and setting monetary policies of the country hence influencing the performance of the Kenyan shilling on the forex market (Ndeda, Economics, 2015).

2.3 Factors Affecting Forex Market Prices in Kenya

Basically, this section hints out on the key factors that affect the foreign exchange rates comparable to the Kenya shilling. It analyses the various trading patterns and theories for better comprehension of the forex market operations. It shall also hint out as a concept on additional factors that may not directly affect the forex market through the Kenyan shilling from the case in point.

The Kenyan market has adopted the floating rate system which implies that basic forces of demand and supply would be the determinants of the prices in the forex market. Activities performed by the government in the phase of public finance and planning tend to directly influence forex exchange rates on the price of the Kenyan shilling. The Kenyan government does not have the Central Bank of Kenya (CBK) directly set/impose an exchange rate, however, through its monetary policies, the same can be realized (Central Bank of Kenya, 2019). The CBK plays a vital role in ensuring general price stability and hence having a direct impact on inflation of the country, using such techniques is when fluctuations and shifting of the forex rates of the Kenyan shilling can be realized. Other than the CBK, the national government, through the State department of treasury could also make fiscal policy moves that could directly have an impact on general pricing of the forex rates as is implied from Ndeda’s (2015) book on Economics.

2.3.1 Forex Trading Patterns and Theories

Forex trading represents a bet against or in favor of an economy. Economies performing better will always have stronger currencies while currency weakens whenever economy underperforms (Talk Business, 2019). Traders speculate based on the shape of different economies and then buy/sell a currency pair based on their analysis. If correct, currency pair would move in their direction hence profit is realized otherwise a loss is observed.

Technical trading has stipulated buy/sell interpretations based on either pattern recognition approaches or trading theories. Pattern recognition approaches include:

a. Head and Shoulders: Pattern predicts a top or bottom formed/about to form
b. Wedges: rising and falling wedges indicate bottoming or topping conditions
c. Double and triple tops and bottoms
d. Triangles: Ascending/descending triangles act as continuation patterns though triangles as reveres patterns also exist
e. Pennants: This represent continuation patterns similar to triangles. Always appear when market builds energy to break higher.

On the other side of trading theories:

a. Elliot Waves Theory: Tracks the chart of a currency pair using various patterns interpreted as waves. Traders comprehend the moves the market makes and may label them as corrective or impulsive (three wave or five-wave structures). After a sophisticated top-down analysis, such traders place a trade based on the projected forecast. This implies that the theory analyzes human behavior in the market
b. Gann Theory: The concept behind the theory by Gann is that everything moves for a reason. A technical analysis on the underlying reasons establishes a trading pattern. Tools like Gann Lines and Gann Square are focused on this theory

c. Dow Theory: The theory was developed by Charles Dow and Edward Jones. It sets the pillars of the Dow Jones Industrial Index that tracks the movements of various companies in the United States of America. The common application is on the securities prices
d. Gartley theory: It grows from Dow theory. It applies action reaction principle. Later theories saw technical traders like Pesavento apply this and Fibonacci ratios. Consequently, this resulted to the harmonic trading observed to date.

According to Talk Business (2019), despite the fact that computers changed how traders looked at the market, additional new technologies available will always affect the face of the forex industry. The behavioral patterns and trading theories have a direct impact on pricing of the forex market rates. Technical traders use indicators and oscillators that may be structured in various technological packages. They are used to track the price movement and employ various filters to plot support and resistance levels on a currency chart.

2.3.2 Monetary Policies

This are technical measures and protocols put in place/directly under oversight of the Central Bank of Kenya to have an impact on general price values hence affecting inflation in the country (Ndeda, Financial Institutions and Markets, 2015).

2.3.2.1 Trade Balance

The Central Bank of Kenya (CBK) had a deficit recording of sixty-seven (67) Billion Kenyan Shillings as of March 2015. In addition to this, records from the Kenya National Bureau of Statistics (n.d), indicates that the country has been importing more machinery as compared exporting since September, 2014. Such a high imports to exports ratio increases the demand for the foreign currency hence increased supply of the domestic currency and consequently, its price, relative to the foreign currency will go down (Wanjiku, 2012). If the CBK makes selective restrictions on imports or exports, this will definitely affect the flow of the forex.

2.3.2.2 Interest Rates

For a long time in Kenya, interest rates were hinged to the forces of demand and supply within the consumers and the commercial banks until when the parliament enacted an interest rate capping through the amended banking act of 2016. This however, did not necessarily reflect on government issued/trading bonds.

Since the forex is hinged on the forces of demand and supply, whenever there is an increasing in prevailing interest rates, relative to foreign interest rates the incentive on foreigners to purchase domestic financial assets increases considering that foreign deposits and investments attract significantly higher interest over that period (Wanjiku, 2012).

2.3.2.3 Terms of Trade

This is partially related to trade balance from section 2.3.1.1. However, in this case the terms of trade is the ration of export prices to import prices (CompareRemit, 2019). It therefore implies that, Kenya's terms of trade improve when its export prices rice at a greater rate than import price which causes higher revenue hence higher demand for the country's currency and a consequential increase in its price.

2.3.2.4 Gross Domestic Product (GDP)

In Wanjiku's (2015) analysis of the GDP factor, at ceteris paribus, an increase in GDP will lead to an increase in the demand for that local currency. In essence, it will directly reflect on the interest rates as the real money demand exceeds the real money supply.

2.3.3 Fiscal Policies

Fiscal policies are any activities that may be performed by the national government through the treasury to have an influence over either taxes levied or subsidies issued to industries in the economy (Ndeda, Economics, 2015). Such fiscal policies include:

2.3.3.1 Government Debt

This is public debt owned by the central government. Countries with government debts are less likely to acquire foreign capital, leading to inflation. CompareRemit (2019) documents that foreign investors will sell their bonds in the open market if the market predicts its government debt hence decrease in value of its exchange rate flow.

2.3.3.2 Political Climate and Stability

The nature of politics, regime shifts, power moves and legislation settings has a direct impact on forex rates. A country with less risk of political turmoil is more attractive to investors (CompareRemit, 2019). The political state is also tied to legislation process and not only cyclic electioneering periods. Having a proper financial and trade policy reduces the impact of uncertainty in Kenya’s currency.

2.3.4 Contributive Factors

Some of the important factors to be mentioned though might not necessarily always affect Kenya since among the highlighted factors the same have been considered include; Recession where country’s chances of attracting foreign capital is reduced as a result of interest rates hike (CompareRemit, 2019). Speculation is also among this factors where it is much more subjective dependent on the other prevailing currencies. If they all rise, so shall the Kenyan currency. See Figure 2.1 below that highlights in summary the factors.

Figure 2.1 : Summary of Factors Affecting Exchange Rates (CompareRemit, 2019)

2.4 Forex Price Prediction Approaches

A prediction/forecast involves an expectation about a future value or values of a particular attribute. Usually, such expectations are constructed using an information set selected by a forecaster (Bauer College of Business, 2019). The process of trying to predict a market is a complex exercise and usually requires the use of a scientific basis. Based on the splitting of this study, the subcategorization into common industry approaches (Forex related approaches) that are econometric and Machine learning based techniques were identified.

2.4.1 Approaches for Predicting Forex Market (Behavioral design)

Using this particular approach, there are two (2) methods for predicting the trends (Bauer College of Business, 2019; Gaucan, 2011) which are the fundamental and technical approaches as in the following sections utilizing econometric models.

2.4.1.1 Fundamental Approach

This approach is based on a wide range of data considered as fundamental economic variables determining exchange rates. Such variables are derived from economic models. Some of the variables here are Gross National Product (GNP), consumption, trade balance, inflation rates interest rates, productivity indexes and unemployment which are basically macro-economic factors. The structural models are then modified to take into account statistical characteristics of the set data and the experience of the forecasters.

The fundamental approach instantiates from a model that produces a forecasting equation which can be based on a theory, PPP, combination of theories or ad-hoc experience of a practitioner. After this, the forecaster proceeds to collection of the S_t, X_t (in the case of PPP, exchange rates and Consumer Price Index - CPI data needed.)

This creates a notation for the information set available at time say t:

S_{t+T} is:

$$E_t[S_{t+T}],$$

Equation 2.1 : Notation at given time

Where $E_t[.]$ represent an expectation taken at time t.

Each forecast has an associated forecasting error, ε_{t+1}, which is defined as:

$$\varepsilon_{t+1} = S_{t+1} - E_t[S_{t+T}]$$

Equation 2.2 : Associated Forecasting Error

This forecasting error is then used to judge the quality of the forecasts with a typical metric used for this purpose being the mean squared error or MSE that is defined as:

$$MSE = [(\varepsilon_{t+1})^2 + (\varepsilon_{t+2})^2 + (\varepsilon_{t+3})^2 + ... + (\varepsilon_{t+Q})^2]/Q$$

Equation 2.3 : Mean Squared Error Dérivation

,

where Q is the number of forecasts hence say that the higher the MSE, the less accurate the forecasting model.

Having two kinds of forecasts; the in-sample and out-of-sample where the former works within the sample at hand and the latter outside of the sample. Generally, using either, it is very unlikely that the inflation rate at say a particular time, ***t*** which is one of the most fundamental variables. This therefore implies that to generate out-of sample forecasts it is necessary to make assumptions about the future behavior of fundamental variables. This highlights a key aspect of weakness of using this particular approach.

2.4.1.2 Technical Approach (TA)

The TA focuses on a smaller available subset of data. It is based on price information. The approach is technical considering that it does not depend on fundamental analysis of the underlying economic determinants of the exchange rates or asset prices rather, on extrapolations of past price trends only. The major aspect of concern that makes this approach a weakness is the fact that its dependence on science is quite minimal and it more so is seen as an artistic use of an expert's knowledge supplemented with the price information.

Common TA models are simple and rely on moving averages (MA), filters, or momentum indicators. The goal of an MA model is to smooth erratic daily swings of asset prices in order to signal major trends. A use of a simple moving average equation which is unweighted mean of previous data points:

$$SMA = (S_t + S_{t-1} + S_{t-2} + ... + S_{t-(Q-1)})/Q$$

Equation 2.4 : Simple Moving Average Expression

When the most recent past prices are included it is considered a short-run MA (SRMA) whilst of a longer series of past prices is used then it calculates a long-term MA (LRMA). In MA models, buy and/or sell signals are usually triggered when a SRMA of past rates crosses a LRMA. An instance is when a currency is moving downward, its SRMA will be below its LRMA but when it starts rising again it crosses its LRMA hence generating a buy foreign currency signal as illustrated in Figure 2.2.

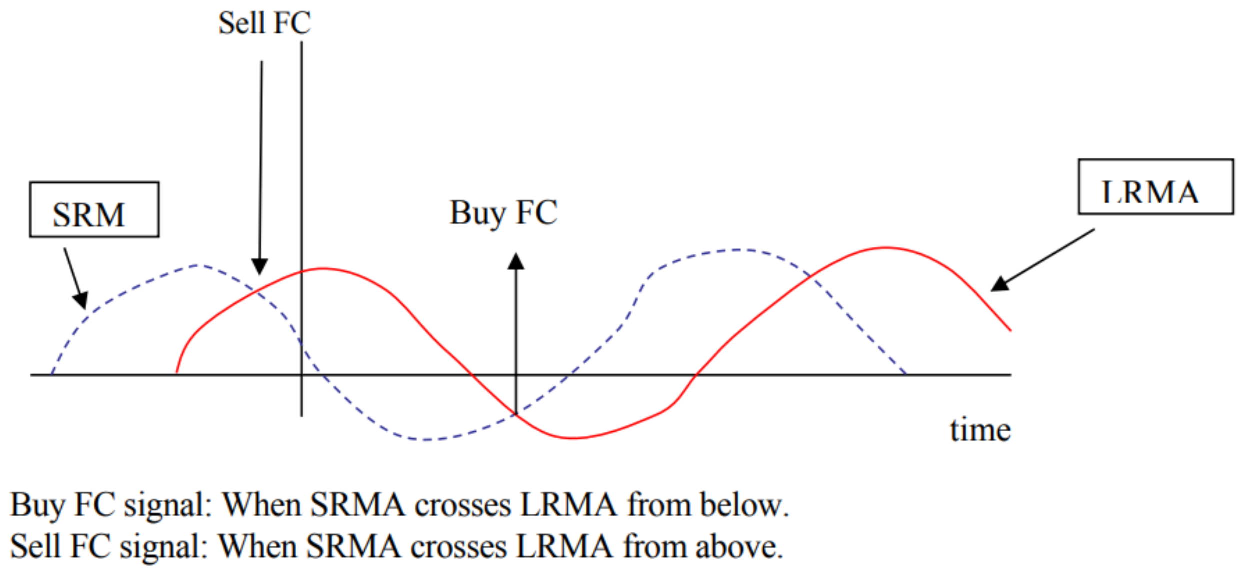


Figure 2.2 : Illustration of SRMA and LRMA Movement (Bauer College of Business, 2019)

Filter model of the TA are quite popular. It is based on the principle that asset prices show significant small autocorrelations. The system relies on determining when exchange rates start to show significant changes and not irrelevant noisy changes. Filter methods generate buy signals when an exchange rate rises by say Y percent (considered as the filter) above its most recent trough, and sell signals when it falls Y percent below the previous peak. The idea is to smooth (filter) daily fluctuations in order to detect lasting trends. The filter size, X, is typically between 0.5% and 2.0%.

Momentum models of the TA determine the strength of an asset by examining the change in velocity on movements of the asset prices. For the case of when a particular asset price goes up at a significant speed, a buy signal is issued. The challenge that may arise with this is that it is left upon the trader to decide on what constitutes significant increase in speed.

2.4.2 Prediction with Business Intelligence

Business intelligence is almost everywhere and commonly implemented as Artificial intelligence for business case. Artificial intelligence is often defined as an aspect of computer science that deals with the simulation of intelligent behavior in computers. It tends to be an overhaul that encompasses rules extraction, Robotic Process automation (RPA), Natural Language Processing (NLP) and Machine learning (Chartis Research, 2018).

Automation does not always have to be the answer, mostly if certain criteria is not met. According to Chartis Research (2018), it is important to evaluate the structuring for appropriate candidates for automation using the format as in Figure 2.3. The layout is clear that it is not just a matter of applying it, rather, where exactly is it applied at within the forex market process.

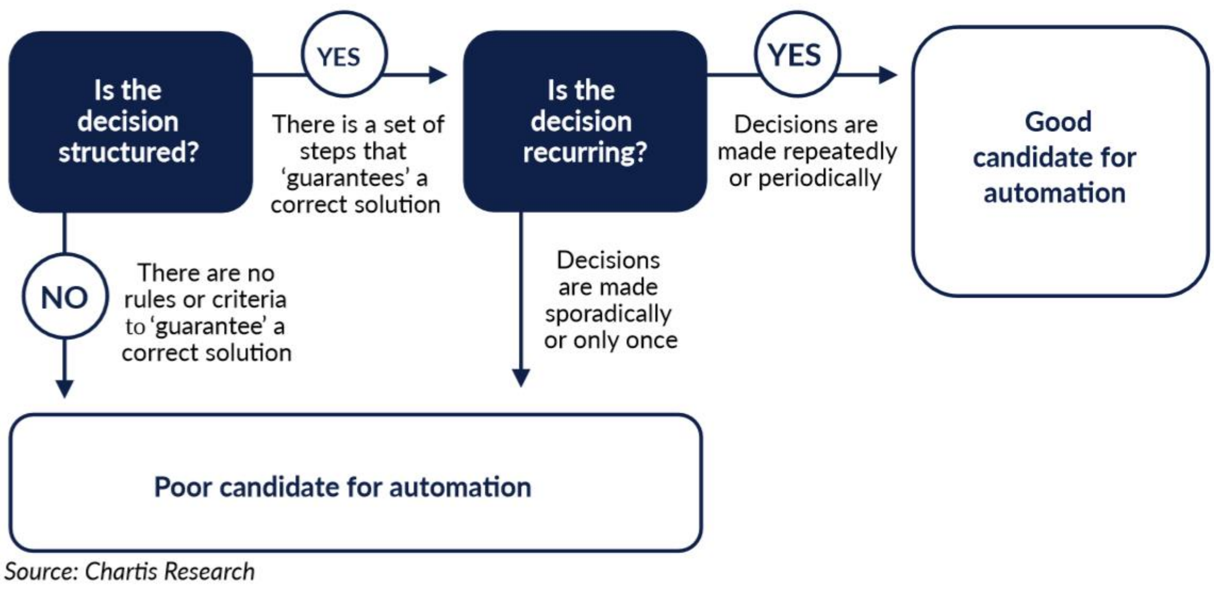


Figure 2.3 : Structure for BI Application Process (Chartis Research, 2018)

2.4.2.1 Machine learning categorizations

According to Sidana (2017), common classifications of machine learning algorithms are depicted as whether there is supervised learning or unsupervised leaning. Unsupervised learning apparently does not have definite target outputs attributed to inputs and consequent involves consequent learning from observation (Lwanga, 2018). Supervised learning on the other hand is an approach in which a software learns from input given and uses it to classify new observation. Among the common algorithms include; Linear classifiers, Nearest Neighbor, Support Vector machines, decision trees, Boosted trees, random forest and deep learning (Sidana, 2017).

One of the linear classifiers, Naïve Bayes methods, is used for clustering and classifications. It is applied based on the Thomas Bayes' theorem with an assumption of there being an independence between every pair of selected features (Lwanga, 2018). In small sample sizes, the Bayes can be very effective compared to powerful alternatives. It is employed widely in disease diagnosis, medical RNA classifications.

Support Vector machines are considered classification and regression examination techniques. It has been widely employed in text and sentiment analysis, image processing and bioinformatics. It constructs a hyperlane/set of hyperlane in a high or infinite dimensional space (Lwanga, 2018). Since the data is not linearly separable, it is possible to create a room that is transformable.

The K-nearest neighbor, commonly termed kNN takes a group of labelled points and uses the same to learn how to label other points. To create a label on a new point, all other labelled points closest to it are observed and those neighbors are weighted so the label with the most weight becomes the new point now say k, where k is the number of neighbors under observation (Sidana, 2017).

The logistic regression is a statistical method for analyzing a set of data in which there are one or more independent variables determining the outcome. The measurement of the variable is done with a dichotomous variable. The ultimate aim of a logistic regression according to Sidana (2017), is to find a best fitting model describing the relationship between the dichotomous characteristic of interest and a set of independent variables. It is better than binary classifications like kNN considering that it also explains in a quantitative manner the factors leading to classification.

In Decision tree classifiers, the predicate and conditions used imply the absence or presence of a number of variables/factors. Decision trees build classification or regression models in the form of tree structure, where it subdivides a data set into smaller and smaller subsets while the object of tree is incrementally developed. Sidana (2017) stipulates that a decision node has more than one branch and a leaf node would normally represent a decision/classification. The classifier is instrumental for both categorical and numerical data. Consider Figure 2.4 that demonstrates the inputs, outputs and algorithmic approach towards making machine learning a decision maker.

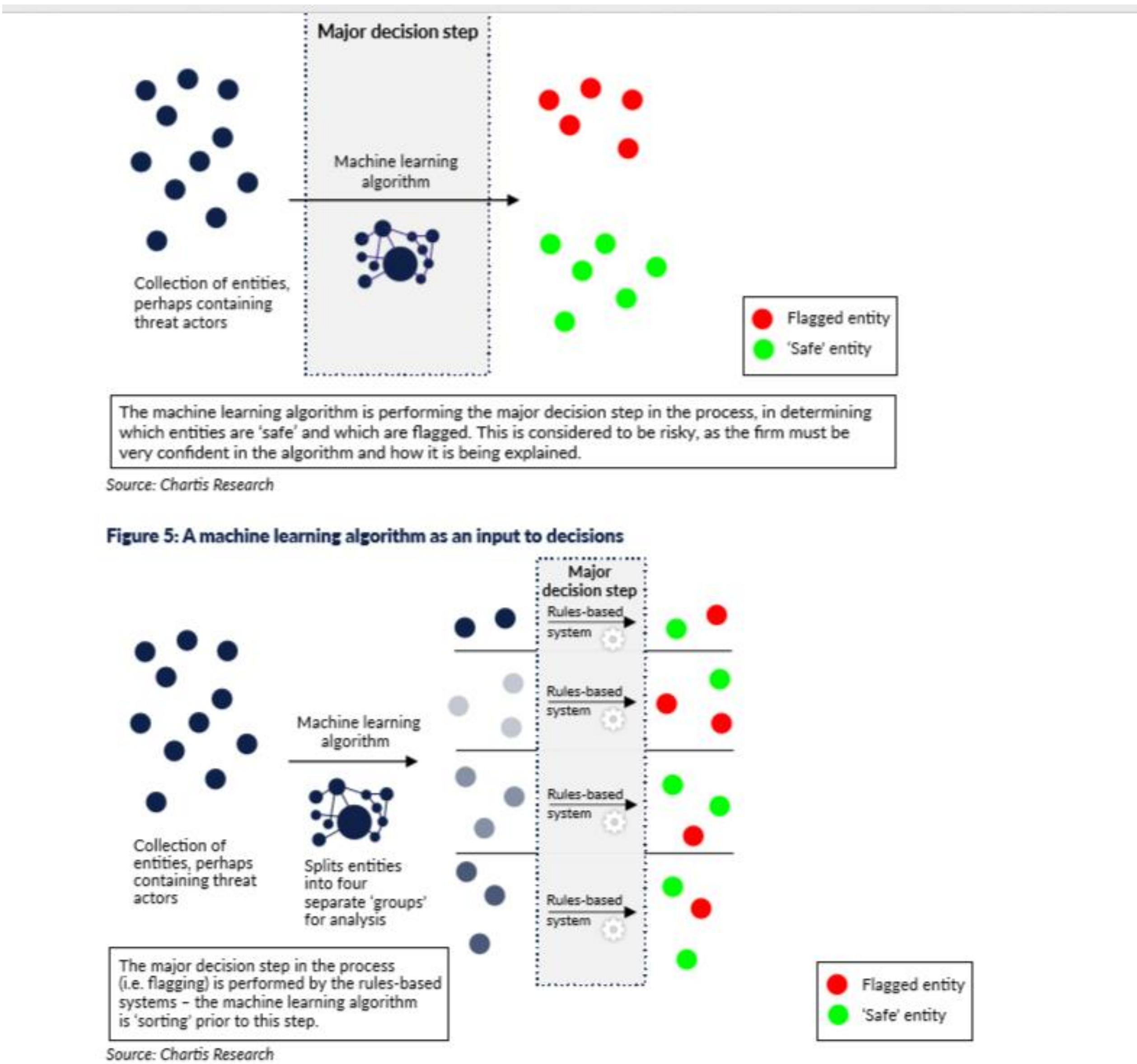


Figure 2.4 : Sample utilization of a Machine Learning Algorithm as a decision-maker (Chartis Research, 2018)

2.4.2.2 Machine Learning with Artificial Neural Networks (ANN)

ANN is one of the common ways of predicting stock market pricing that has been used in other markets like India and China (Gould, 2004). It is actually an information processing system consisting of graphs representing the processing system and various algorithms. ANN is a complex and sophisticated aspect of computing with the ability to adopt, recognize patterns, generalize and cluster data. As per Gould (2004), ANN has widely been used in the FOREX markets of sophisticated economies to increase profitability.

Taking the model of a biological neuron, an artificial neuron simply accepts signals/inputs from the other neurons and/or the surrounding environment (Lwanga, 2018). According to Uhrig (1995), when specific conditions are met, a signal is fired from the neuron hence transmitting the signal from that point to all other connected Neurons. There is an association between numerical positive and negative value which is associated with every other neuron which inhibit or excite inputs with each connection made to the artificial neuron. The neuron recollects all signals coming in by computing their net input signals as a function. The net input signals then apparently serve as input to the respective activating function that calculates the corresponding output signal of the artificial neurons. Lwanga (2018), summarizes that ANN being a layered system can contain one of many artificial neurons with the components including an input layer, hidden layer as well as an output layer. Figure 2.5 illustrates the structure of a simple neural network, back-propagation design.

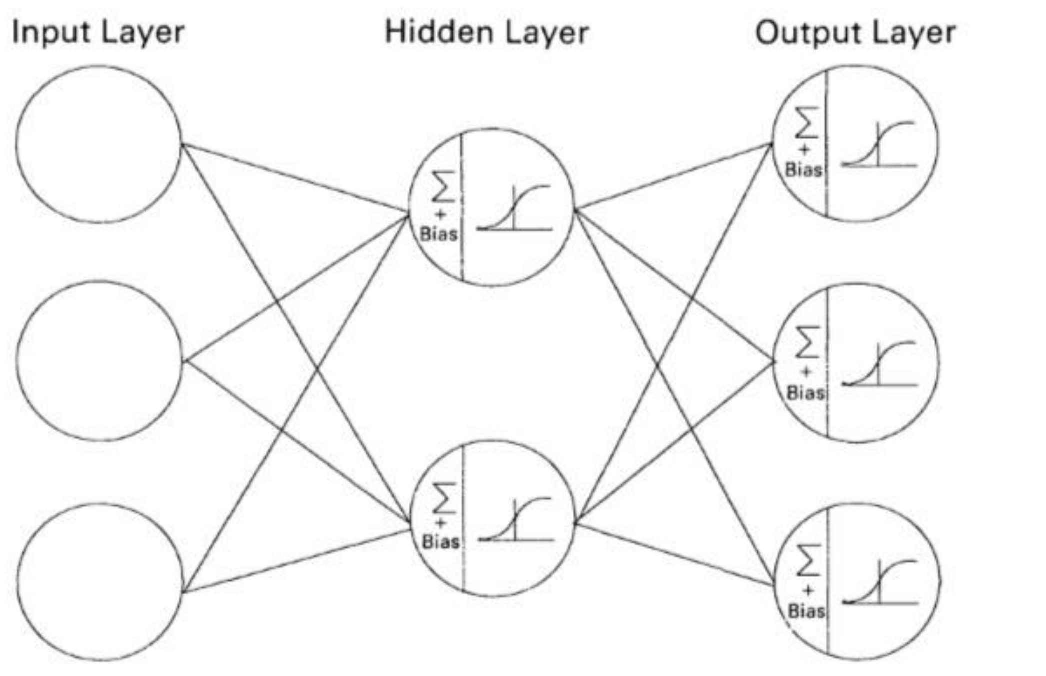


Figure 2.5 : Neural Network structure (Curram & Mingers, 1994)

Chartis Research (2018) has broken down the respective machine learning algorithms with definitions and respective weaknesses when being applied to a business case scenario of Financial Institutions as in Figure 2.6.

Naïve Bayes	A classification algorithm that assumes that the presence of a given feature in a particular class is unrelated to the presence of any other feature.	• Easy to understand. • Fast. • Performs well for categorical/ non-numeric assumptions.	• A normal distribution is assumed for numeric variables. • Results depend on an assumption of independence.
Decision trees	An algorithm that uses estimates and probabilities to calculate outcomes.	• Easy to understand/ explain. • Potential options and choices are mapped out. • Determines costs and benefits.	• Probabilities are assumptions and prone to errors. • Uses quantitative/numerical data.
Logistic regression	A probability and risk estimator used to predict a binary outcome (e.g. 1 or 0, or yes or no) given a set of independent variables.	• Relatively easy to use. • Does not assume linear relationships.	• Requires a reasonably large data set. • User needs to specify which interactions are allowed within a model.
Support vector machines	Transform linear data into a non-linear space, then map it into categories which are divided by as wide a gap as possible. Future data is mapped into the space and into one of the categories.	• Can avoid over-fitting. • Relatively easy to control.	• Requires a large amount of data. • Resource-intensive.

Figure 2.6 : Strengths and Weaknesses with Current Machine Learning Techniques (Chartis Research, 2018)

2.5 Empirical Review

The study classifies the prior studies either ones utilizing the behavioral design which entails traditional econometric models or ones utilizing techniques other than simple traditional econometric models that involves additional layers of analysis. In contextual analysis, the term forecasting forex market values has been used to imply using the traditional econometric models while prediction implies the use of additional/slightly advanced techniques. Based on this, the study shall review two works of prediction models implemented:

2.5.1 ENMX: An Elastic Network Model for Predicting Forex Market evolution

ENXM is an algorithm that was created by a research group which was inspired by behavior of macro-molecules in dissolution to model the evolution of forex market. It allows the system to escape from a potential local minimum, such that it can reproduce the unstable nature of the Forex market hence allowing the simulation get away from equilibrium (Contreras, et al., 2018).

The algorithm operates by introducing several novelties in the simulation of Forex market. It starts by enabling a user simulate the market evolution of up-to 21 currency pairs, connected to each other and thus emulating the behavior of the real world forex market. Consequently, the interaction between investors and specific quotations, may yield slight deviations from the quoted prices in the market represented by a random movement. The Monte Carlo method is then utilized to create an optimization for the model hence this ensures a good simulation of the behavior rather than valuations. The experimental results demonstrated that ENMX predicted values on the forex market better and more accurately than traditional econometric approaches. The application of the algorithm is solely focused only on returns as is common with commercial rankings for predictions. The model fails to implement risk measures to create a risk-adjusted model for the framework to account for additional deviation factors in exploration of strategies that may lead to abnormal returns in real world settings.

2.5.2 Back-propagation Neural Network for Australian dollar verses United States Dollar

Gould (2004) created an artificial neural network system that based the simulations on the Australian dollar against the USD. The work proved that back-propagation neural network can be applied for correct Forex rates prediction allowing for maximal profits. The approach utilized feed-forwarding topologies and supervised learning but limited it to a split time series forecasting structure. Gould (2004) applied different network structures and algorithms on dynamic datasets and established a network model that used LM training algorithm.

Despite the success in the forex market rates predictions, there were still limitations in relation to the results. The common problems underlying neural network model were not fully eliminated by the approach taken. The training times were high and a large number of parameters had to be selected experimentally to establish a good fit. This technically implies that it did not guarantee its longevity and long-term application considering that each time, experimental parameters always had to be redefined. The approach also hinted out that using back-propagation never always guaranteed correct weights for optimal solutions. Selection of an appropriate learning rate for the non-linear network proved to always be a challenge.

Additional dimensions and more complex data should be tested for additional data analysis using a statistical software. The overall use of the right analytical tools and methods may decrease the chance of making incorrect decisions and increase probability of profitability in foreign currency exchange rates (Gould, 2004).

2.6 Conceptual Framework

The construction of the model is hinged on feeding it with the data from the Central Bank of Kenya. The concept behind it adapts from utilizing the Data driven modelling technique as was applied in a viral load prediction system focusing on data flow and utilization (Tunduny, 2017). The data is quite explicitly defining in nature for all the variables involved. Additional features tied to predictors are defined at the feature extraction and parameter setting. The predicting accuracy of the model is refined using the output being fed back to the input point of the model in an iterative design.

The data is first taken through preprocessing for effective training of the neural network. The headers extracted from the dataset included the date variable, mean price value, sell price value, buy price value and currency type. The data types for the attributes are date-time for the date and numeric for price values which implies no additional processing is required on them. However, the currency type is then assigned numeric values within the range on 0-1 representing three currencies. The preprocessed data is then set for predictor variables and target variables based on the preceding day price values. This allows for the set to be split into training and testing data. The training set is then used to train and design the optimal neural network structure. When a future price prediction is required, a date/range of dates is fed into the trained prediction model which then returns the corresponding mean price values for the requested date/date ranges. The returned values are then fed back to the retraining process to refine the model output for increased accuracy.

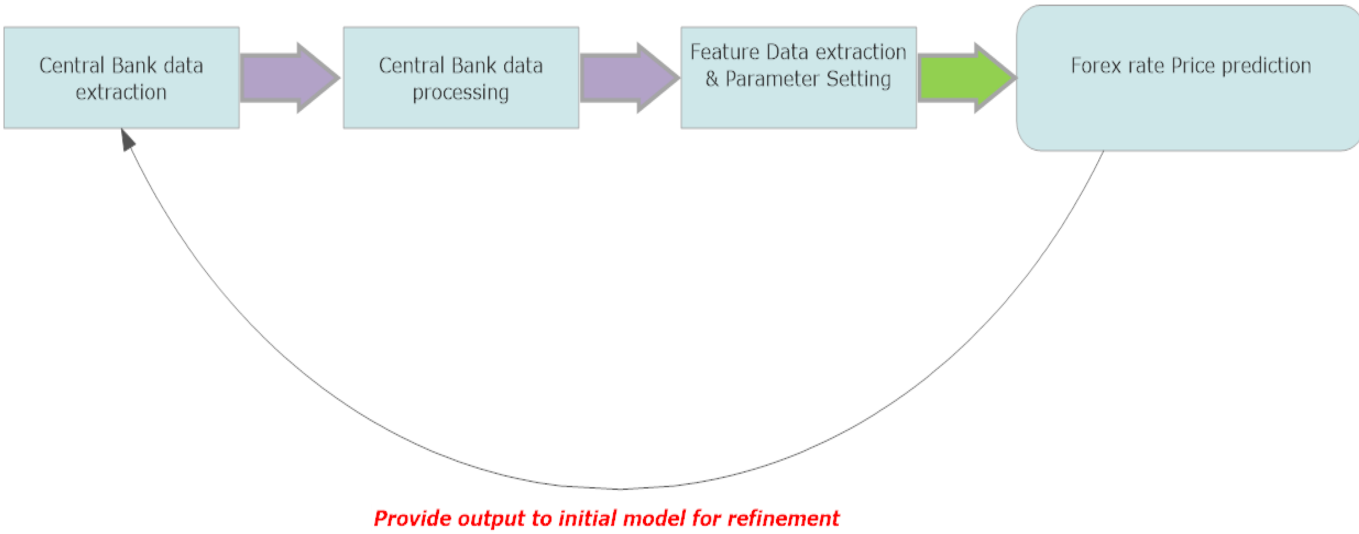


Figure 2.7 : Conceptual Framework

Chapter 3 : Research Methodology

3.1 Overview

This chapter discusses the methodology the researcher took in conducting the study. It discusses issues to do with the specification of the model, sources of data, the process, definition of variables and criteria for collection of data. In highlight, research methodology refers to the procedures by which researchers go about their work of describing, explaining and predicting phenomena (Rajasekar, Philominathan, & Chinnathambi, 2013).

3.2 Research Design

Degu and Yigzaw (2006) defined a study design as the process that guides a researcher on how to collect, analyze and interpret observations. As highlighted in the literature review, being an applied research, quantitative techniques were used to examine relationships between variables. The study adopted simulation design. This was facilitated by the fact that the researcher used quantifiable dataset from the population sample on historical foreign exchange rates. Essentially, the design entails simulations from a collection of the quantifiable data sets. The simulations were iterative in nature and focused on enabling the model set dependent and independent variables that mapped out forex market prices within their affecting factors in each simulation to create a pattern/indication for dates. The purpose for running it in this form is to ensure the most effective and optimal model is established for fitting. This was therefore a good approach as it ensured objectivity achieved and that the subjective aspect is eliminated.

3.3 Location of the Study

The research focuses on the Central Bank of Kenya (CBK) located in Nairobi, Kenya. All the current monetary policies and the forex market records are maintained by the CBK with their data centers and public web-based portal. Considering that CBK has been the regulator of the forex market, it is an ideal place to rely on for the quantitative data it has been collecting in the past. The historical records of the foreign exchange rates which are vast and numerical in nature formed the dataset in study for Kenya as a region.

3.4 Target Population and Sampling

A target population is the entire member or hypothetical sets of people, companies, objects or events that the researcher wishes to extrapolate and generalize the results of the study (Mugenda, 2003). It is essentially the specific population that the researcher is interested in, and this is the historical forex market prices. In this research, the selection entails a purposive sample population of the historical data extract from the Central Bank of Kenya on the foreign exchange rates of three (3) currencies against the Kenya shillings i.e.) United States Dollar (America), Chinese Yuan (Asia) and South African Rand (Africa). This population was purposively selected in nature to the range limit between the years 2016 and 2019. The choice of currency was informed by the behavioral factors arising from continental activity of the respective currency economies. They are the major currencies in their respective continents. The selection of the duration was based on the availability of dataset and requirements of a significant volume of data to construct a predictive model based big data.

3.5 Data Collection

This entails the entire process of gathering and measuring data, information or any variables of interest as from the conceptual framework in a standardized and established manner that will enable the researcher evaluate outcomes of the particular collection. This study employed different approaches in the collection of all research data. This shall entail:

a) Historical data. This entails the raw information/source of data for the research. Essentially, this is the historical foreign exchange rates captured from a public access portal from 2016-2019 from the Central Bank of Kenya (CBK). The dataset is of public interest hence available at no cost. The data in relation to prices is available for each day from 2016 to date except for public holidays and weekends. For purposes of this research, the last date of selection was set to 29/10/2019. The data only shows an indicative of the forex market prices which are captured at the beginning of any trading day by the CBK, however, the prices are not set by the CBK rather the forces of demand and supply. The use of the data only facilitates for an indicative forex market prices but independent dealers and brokers may opt to select any price within the buy-sell spread that will maximize their return.

The dataset is available in the format as in Table 3.1 for each day from the year 2016 to 2019.

Table 3.1 : Data for Given Range of Days

Date	Currency	Buy	Sell	Mean
11/10/2016	SA RAND	7.3196	7.3087	7.3305
11/10/2016	US DOLLAR	101.2847	101.1889	101.3806
11/10/2016	CHINESE YUAN	15.0999	15.0792	15.1206
13/10/2016	US DOLLAR	101.2953	101.1994	101.3911
13/10/2016	CHINESE YUAN	15.0693	15.0485	15.0902
13/10/2016	SA RAND	7.0759	7.0655	7.0863

Table 3.1 is speculative extract format for two (2) days data. The same was derived and utilized for the entire period under study as described in the population.

b) Document review. In-order for the research to take appropriate shape and direction of study, an in-depth understanding of operations of the forex, procedures manuals and financial systems was necessary which included the reviews of; monetary policy documents, financial institutions documentation (e.g., journals, articles, periodicals and bank's reports). The documentation in this context is from the Central Bank of Kenya which is considered an authoritative source.

3.6 Data Analysis

This is organizing of the data collected and postulating it into a form that can be manipulated or statistical inferences drawn from such. It is also done to ensure that the data is complete and consistent to ensure quality of data (Onyango, 2018). The researcher populated the data into Microsoft excel spreadsheets where all the pre-processing was done. It was then exported to csv file formats for further integral analysis. The analysis of the data was based on R statistical software package in a step by step manner. The research entailed exploratory data analysis techniques such that enabled substantive identification of preliminary selection of appropriate models prior to the formal modeling and inferences. The consequence of each analytics was always saved for graphical comparisons with proceeding consequences.

The extrapolation of the results was then fit into the model by using the outputs from the analysis. As highlighted, pricing values for weekends are not available therefore there most of the preprocessing has to clean and streamline the data-based on the required and appropriate market pricing as per the forex market. The analytical process utilizes appropriate classifiers on the application of weighting of factors and creation of additional interrelations to facilitate appropriate mapping/iterations in prioritization for the model development process.

3.7 Model Development Methodology

In order to develop an effective artificial neural network, the process entails using the dataset retrieved from the Central Bank of Kenya, have it pre-processed then model developed and validated. The preprocessing phase ensures that all blanks on trading days are appropriately accounted for. Such gaps are as a result of public holidays, weekends or special market closure days.

Tunduny (2017) applied Data-driven modeling (DDM) for a prediction system that used artificial neural networks on viral loads. DDM relies on the analysis of data available about a given case. The modeling is done not necessarily with explicit knowledge of the physical behavior of the system. This was quite significant for this research case as the behavior of the data was quite unique for each of the independent currencies.

In this context the Central Bank of Kenya (CBK) data retrieved from the collection point was split into respective variable quantities within the dataset and categorized into training, test and validation data. The data usable for the model case is a good representation of the forex market in Kenya. The training data was used for training the model to objectively minimize model related errors, while the test data was employed for performance evaluation of the model. The level of efficiency obtained from the model shall be assessed using the test and validation data and application of the sum of squared residuals/errors and R-squared as a research quality element.

The basis of all assumptions to be conducted during validation phase involved speculative data derived from any other currency or out of set selection.

3.7.1 The Modelling Approach to be applied

For the development of the model, the Data-drivel modelling approach is employed as in Figure 3.1 with the procedural layout (Solomatine, See, & Abrahart, 2008).

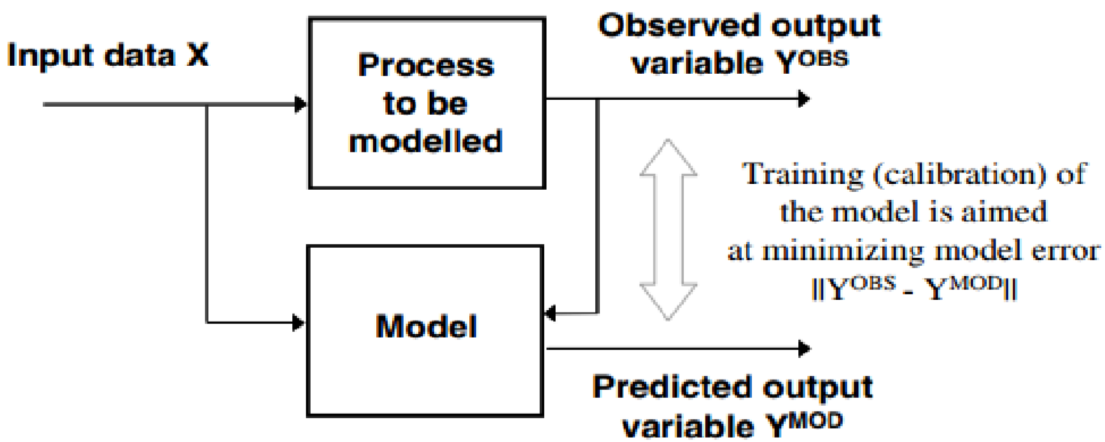


Figure 3.1 : Data-Driven Modelling Development Process (Solomatine, See, & Abrahart, 2008)

3.8 R Implementation with Tensorflow

The statistical tools that shall be used to implement and guide the research from conceptualization to modelling and having a workable model shall be R statistical software package applied with Tensorflow.

3.9 Research Quality

3.9.1 Reliability

The research was presented to experts in both the fields of Information Technology and Finance for their review. Based on this, the appropriate relevance was established of the study and findings.

The expert engagement created a general informed approximation of the extent of reliability as well as quality of research from the product point of utilization

3.9.2 Validity

The data collected for the purposes of the study were reviewed by the authorizing authority as well as other secondary sources for accuracy. The model was then validated using a prototype and the construction of a basic error-accuracy levels mappings. A confusion matrix could not be used for this research purpose considering the nature of the final model operational design. Usually such a matrix has information about actual and predicted classifications done using a classification system (Lwanga, 2018). The matrix could therefore not be used to demonstrate the rate of success in the predictions for the forex market price prediction model effectively as this was not a classification system.

3.9.3 Objectivity

The datasets from the field are based on facts on the ground. It is not subjective in nature or influenced by whatsoever opinions in compilation of the facts. The independence of the observations and collection methodology ensures its objective nature.

3.10 Ethical Considerations

In upholding the ethical expectations of a quality research, the research met the general ethical requirements as required and approved by an Independent Ethical Board as in Appendix A. A proper License certification of the research was obtained from National Commission for Science and Technology (NACOSTI) on the research as depicted in Appendix B. The researcher ensured all proper approvals from stakeholders involved in the study are obtained and authorized. The findings of the study shall be published in a reputable journal for computer science for free public access.

Chapter 4 : System Design and Architecture

4.1 Overview

The chapter elaborates the construction of the system architecture, the analytical process and design of the proposed forex market price prediction model. The system architecture is a build up from the conceptual framework stipulated in Figure 2.6. Essentially, the flow-out covers the interactions of the model from its variable inputs from users to the extrapolation of the output. This is achieved by modelling using UML diagrams: use case diagrams, dataflow diagrams and system sequence diagrams as in standard software modelling case scenarios.

4.2 Requirements Analysis

The requirements specifications for the construction of the model are based on the findings of the objectives covered from the reviews. This therefore structure out the requirements that ought to be met by the modelled system. From Bandakkanavar's (2018) requirements specifications on standard softwares are structured into three common ones that are functional requirements, non-functional requirements and usability requirements. This therefore creates a good basis for categorizing the requirements specifications as:

4.2.1 Functional Requirements

i. The system should be able to capture as an excel file input variables and archive into a cloud based database server. Additional formats for updates may be allowed
ii. The system should retrieve the variable headers as daily dates, currency, buy price, sell price and the mean
iii. The clean-up process should enable the system pre-map out the retrieved entries and enter date prices in similar format for buy, sell and mean for weekends and public holiday using the prices of the previous last working day. The same should be updated into the database
iv. The system should extrapolate in a numerical and graphical format the forex market prices for the preceding six to twelve months and categorize them on the basis of USD, RAND and YUAN.
v. The predicted values should be valid based on the inputs from the user and system validation clean

4.2.2 Usability Requirements

The system is intended for use by the Financial Institutions and markets. The main practitioners interacting with it should be economists from the Central Bank of Kenya, commercial banks and brokers in Forex bureaus. This implies that the system should streamline the objectives of the interest of the parties involved such that it ensures accuracy in the predictability levels considering that it should inform key economic and financial decisions.

4.2.3 Non-functional Requirements

i. Reliability – The accuracy of the available data from the Central Bank shall greatly inform the reliability of the model. The available data shall be utilized for training and its output should facilitate the model to retrain itself
ii. Interoperability – Considering that the model shall be of interest to multiple parties in the financial institutions. The system shall facilitate multiple interactions from the kind of output that it shall be delivering
iii. Scalability – The nature of the output should not be rigid and should allow for swift future upgrade with increase in available data and its direct interaction with the real world

4.3 System Architecture

Figure 4.1 describes a build up from the conceptual framework in Figure 2.6 that formulates the architecture of the proposed model of the system. The forex market prediction system comprises of historical forex data from the Central Bank of Kenya, the analytics portal which is basically where the training and test for the neural network is performed using the input predictors that now formulate the Forex market price model. First a user would pull/load data onto the system, after which appropriate feature sets for the neural networks are set. This specifies the inputs for the dataset, training and test case. At the upload exchange some validation and clean-up shall be performed to eliminate null entries that will set public holidays and weekend criteria. Based on the neural network algorithm employed in the study, the analytics was able to create a model where a user in this case economist/forex dealer may make requests to view predicted values.

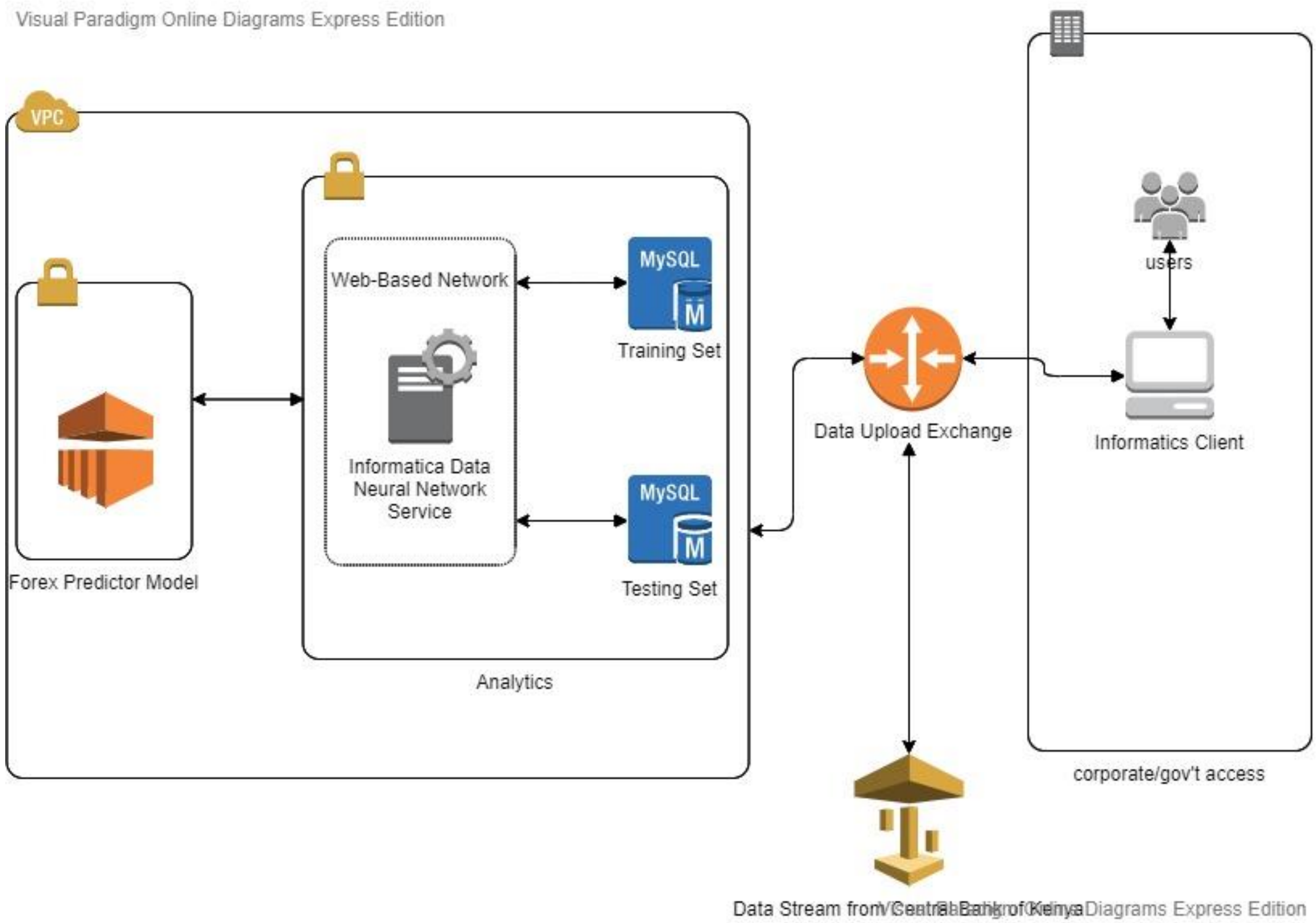


Figure 4.1 : System Architecture

4.4 Use Case diagram

This depicts the actor as something with a behavior/role in the interaction with the model. In this case, the actors are the system users and the system itself. Figure 4.2 clearly illustrates the system-actor interactions to depict prototype users as the administrator and economist. In building up the use case diagrams the following scenarios are used in Table 4.1, Table 4.2 and Table 4.3.

Table 4.1 : Use Case Scenario 1

Use Case Scenario 1	**Data Capture and standardization from file**
Primary Actor	Administrator
Precondition	Forex currency is either YUAN, USD or RAND, standardization function
Post-condition	System uploads daily forex prices and archives into MySQL database with all the required header features
Main Success Scenarios	
Actor	**System Responsibility**
1. Administrator enters dataset	
2. Administrator sets currencies	
	3. System runs data validation with standardization and corrects null entries for public holidays and weekends
	4. System saves forex prices in a cloud based MySQL database

Table 4.2 : Use Case Scenario 2

Use Case Scenario 2	**Training and Testing**	
Primary Actor	Administrator	
Precondition	Standardized Daily forex prices in Database	
Post-condition	Information data artificial neural networks with multiple layer levels	
Main Success Scenarios		
Actor		**System Responsibility**
1. Administrator creates the neural network characteristics		
2. Administrator selects iterations for epochs		
3. Administrator re-aligns the sigmoid		
4. Administrator inserts the target output for the data		
		5. System trains model based on epoch iterations and selected characters
		6. System generates neural network model
		7. System runs model testing using test data
		8. System saves tested model and generates predictive prices

Table 4.3 : Use Case Scenario 3

Use Case Scenario 3	**Accessing the Future forex market prices**
Primary Actor	Economist/Dealer
Precondition	Information data artificial neural networks with multiple layer levels, Recent forex market prices updated, Inflation index adjustment
Post-condition	Information data artificial neural networks with multiple layer levels
Main Success Scenarios	
Actor	**System Responsibility**
1. Economist/Dealer captures date range/date	
2. Economist/Dealer makes a currency selection	
	3. The system runs the model and extrapolates the prediction from available data
	4. System makes suggestive recommendations based on current economic levels adjustments, GDP and Inflation indices
5. Economist/Dealer views the predicted forex market prices	

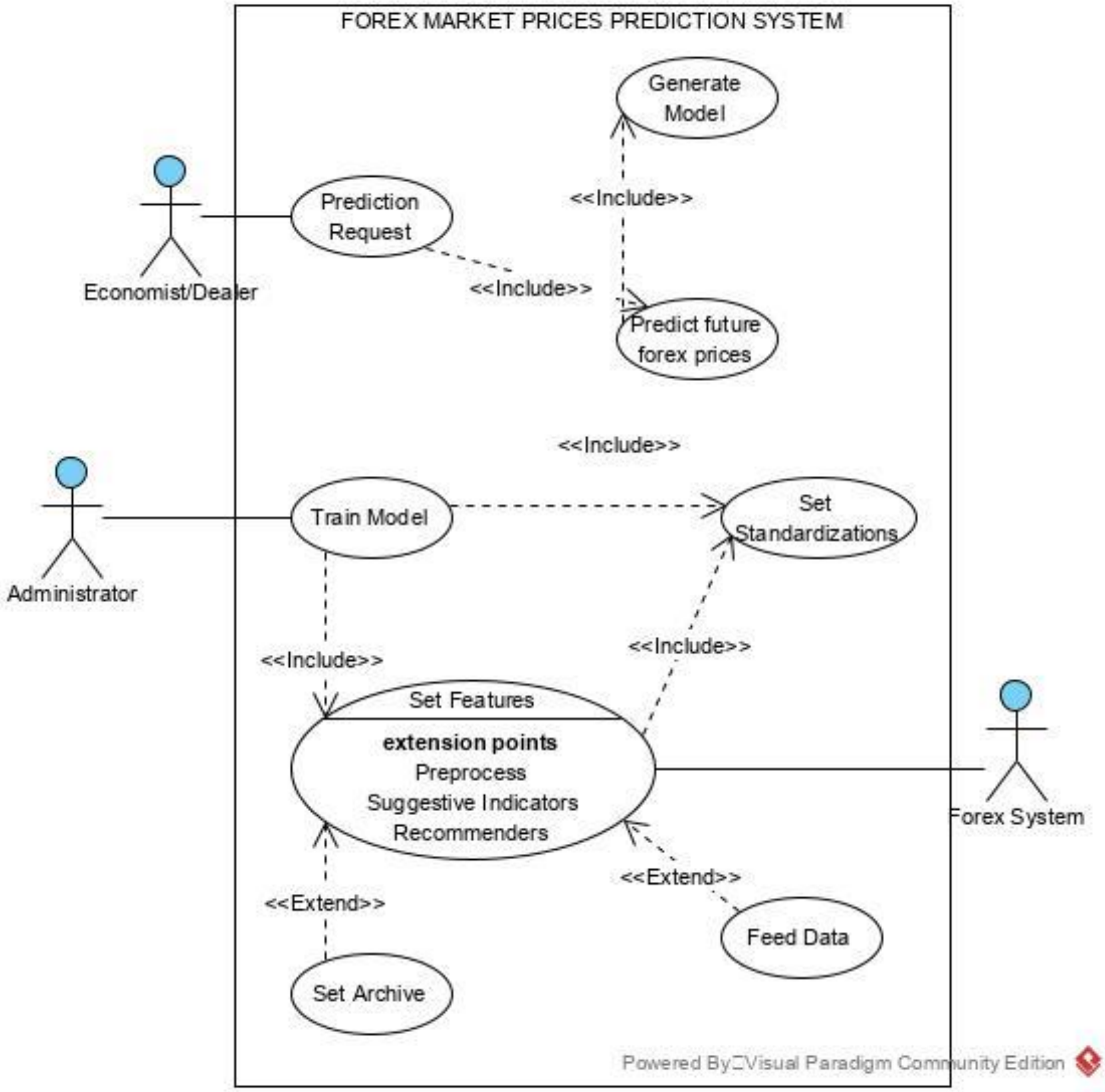


Figure 4.2 : Use Case Diagram

4.5 System Sequence Diagram

This adds to the explanation on the behavior of the system. Figure 4.3 illustrates the sequence diagram as a build-up of the use cases from Figure 4.2. The interaction should have the economist as in Figure 4.2 making a request and creating criteria for request of future prices. This involves submitting a specified future date range or a specific foreseeable date with a selection on currency as well if needed. Before the request interacts with the system, the administrator would have historical data loaded up and train the system based on current data. The system then runs analysis and creates a tentative prediction of the values within the range date request and return the same to the economist/dealer. The result level includes for recommendations and added sentiments based on the economic levels that are observed from general inflation levels as one of the inputs. Ideally, each result output should be archived by the system into the database in-order to refine the result set of the model for re-trains.

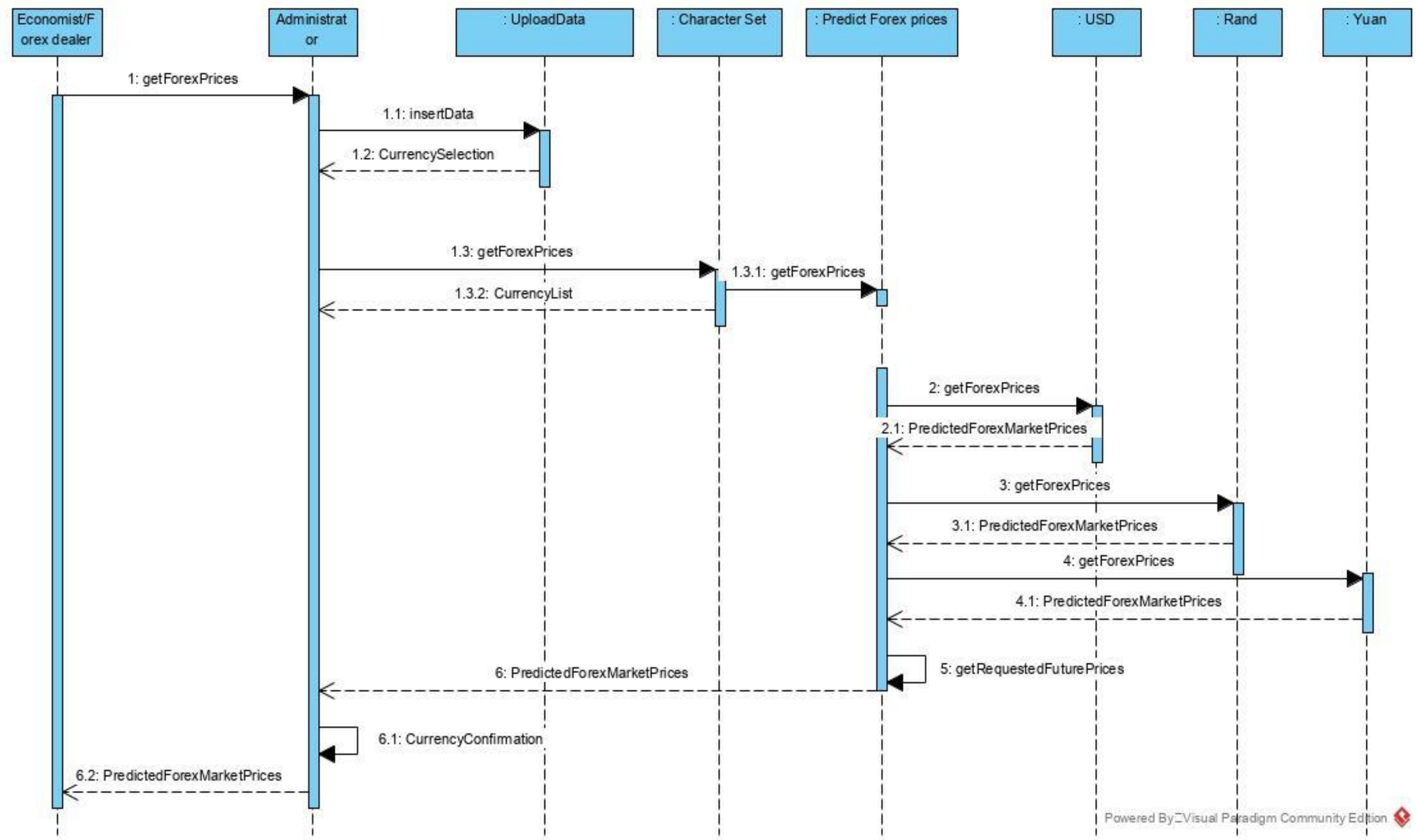


Figure 4.3 : System Sequence Diagram

Chapter 5 : Implementation and Validation

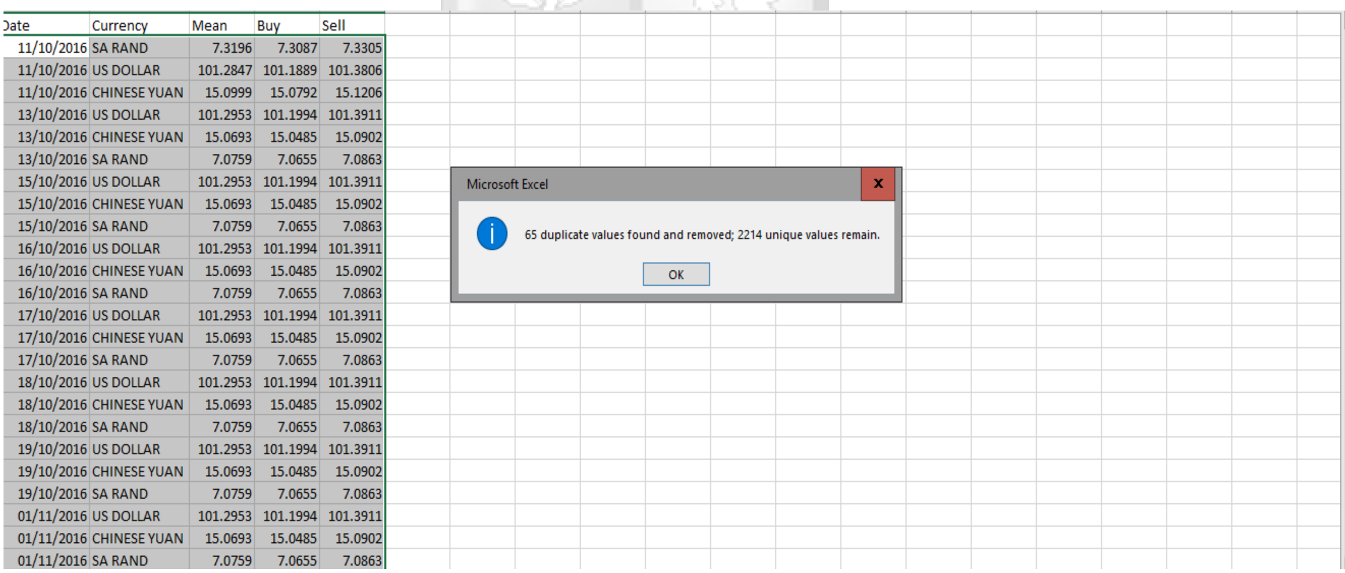
5.1 Overview

This chapter is focused on implementation and validating the model. Implementation shall structure out the whole process of development and actual functionality of the model. The various components that constitute the prototype shall be highlighted and discussed. The validation shall be performed using a prototype system of which the results shall be mapped against expectations where normal measures of Sum of Squared Residual Error (SSE), correlation R-squared shall be used to extrapolate its various levels of accuracy.

5.2 Model Construction

5.2.1 Pre-processing and Normalization

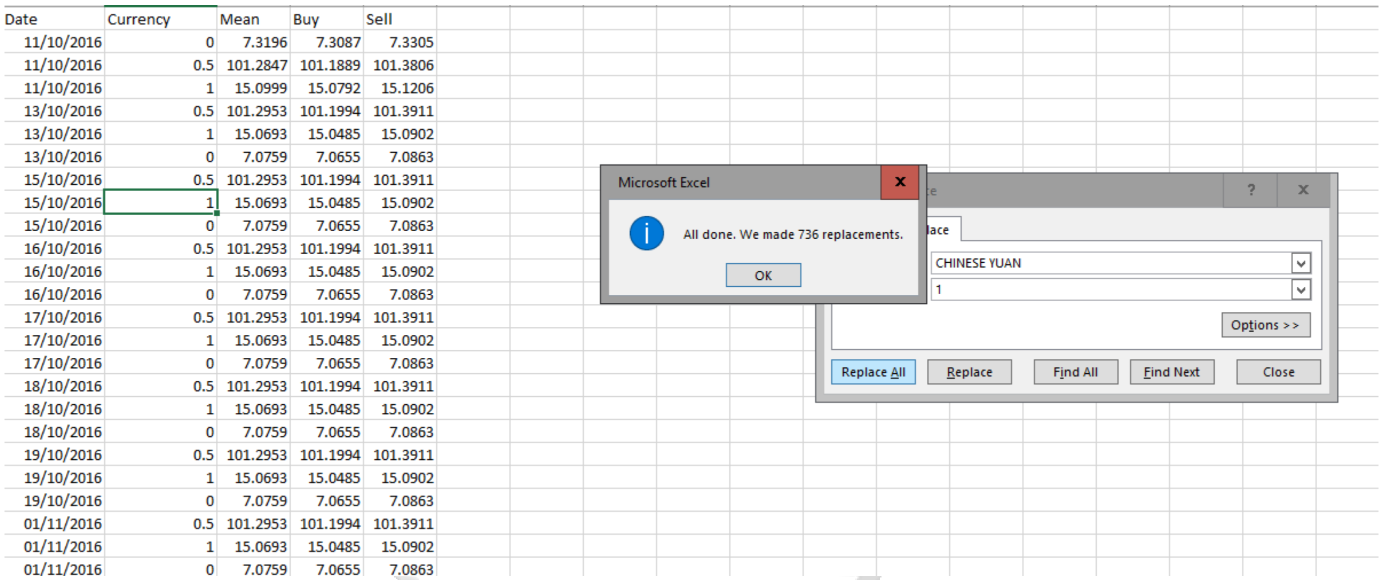
The required data in the csv file format should be cleaned up for basic normalization. Firstly, the data containing five columns: Date, Currency, Mean, Buy and Sell is checked for duplication in entries. All duplicate values respective to a day forex market price for a currency is eliminated as in Figure 5.1.



Date	Currency	Mean	Buy	Sell
11/10/2016	SA RAND	7.3196	7.3087	7.3305
11/10/2016	US DOLLAR	101.2847	101.1889	101.3806
11/10/2016	CHINESE YUAN	15.0999	15.0792	15.1206
13/10/2016	US DOLLAR	101.2953	101.1994	101.3911
13/10/2016	CHINESE YUAN	15.0693	15.0485	15.0902
13/10/2016	SA RAND	7.0759	7.0655	7.0863
15/10/2016	US DOLLAR	101.2953	101.1994	101.3911
15/10/2016	CHINESE YUAN	15.0693	15.0485	15.0902
15/10/2016	SA RAND	7.0759	7.0655	7.0863
16/10/2016	US DOLLAR	101.2953	101.1994	101.3911
16/10/2016	CHINESE YUAN	15.0693	15.0485	15.0902
16/10/2016	SA RAND	7.0759	7.0655	7.0863
17/10/2016	US DOLLAR	101.2953	101.1994	101.3911
17/10/2016	CHINESE YUAN	15.0693	15.0485	15.0902
17/10/2016	SA RAND	7.0759	7.0655	7.0863
18/10/2016	US DOLLAR	101.2953	101.1994	101.3911
18/10/2016	CHINESE YUAN	15.0693	15.0485	15.0902
18/10/2016	SA RAND	7.0759	7.0655	7.0863
19/10/2016	US DOLLAR	101.2953	101.1994	101.3911
19/10/2016	CHINESE YUAN	15.0693	15.0485	15.0902
19/10/2016	SA RAND	7.0759	7.0655	7.0863
01/11/2016	US DOLLAR	101.2953	101.1994	101.3911
01/11/2016	CHINESE YUAN	15.0693	15.0485	15.0902
01/11/2016	SA RAND	7.0759	7.0655	7.0863

Figure 5.1 : Removal of duplicate Price Entries

Secondly, assignment of text values to numeric values. Considering the currencies are only three, Table 5.1 indicates the indication of the assignment from test to numeric values. Figure 5.2 shows the translation.



Date	Currency	Mean	Buy	Sell
11/10/2016	0	7.3196	7.3087	7.3305
11/10/2016	0.5	101.2847	101.1889	101.3806
11/10/2016	1	15.0999	15.0792	15.1206
13/10/2016	0.5	101.2953	101.1994	101.3911
13/10/2016	1	15.0693	15.0485	15.0902
13/10/2016	0	7.0759	7.0655	7.0863
15/10/2016	0.5	101.2953	101.1994	101.3911
15/10/2016	1	15.0693	15.0485	15.0902
15/10/2016	0	7.0759	7.0655	7.0863
16/10/2016	0.5	101.2953	101.1994	101.3911
16/10/2016	1	15.0693	15.0485	15.0902
16/10/2016	0	7.0759	7.0655	7.0863
17/10/2016	0.5	101.2953	101.1994	101.3911
17/10/2016	1	15.0693	15.0485	15.0902
17/10/2016	0	7.0759	7.0655	7.0863
18/10/2016	0.5	101.2953	101.1994	101.3911
18/10/2016	1	15.0693	15.0485	15.0902
18/10/2016	0	7.0759	7.0655	7.0863
19/10/2016	0.5	101.2953	101.1994	101.3911
19/10/2016	1	15.0693	15.0485	15.0902
19/10/2016	0	7.0759	7.0655	7.0863
01/11/2016	0.5	101.2953	101.1994	101.3911
01/11/2016	1	15.0693	15.0485	15.0902
01/11/2016	0	7.0759	7.0655	7.0863

Figure 5.2 : Text Currency Numeric Assignment

Table 5.1 : Normalizing Text Currency Notations

Currency	Numeric Translation
SA Rand	0
United Stated Dollar	0.5
Chinese Yuan	1

Thirdly, create a basis for assumption for missing entries. Considering that the available data may not consistently have daily price data, basis assumption that pre-ceding day's price will be similar to current missing day price.

5.2.2 R Packages and Libraries

The process of the model introduces libraries of Keras, neuralnet and quantmod. Being a time-based data, this creates need for extrapolation in the form of a time series which therefore applies extensible time series (xts) library.

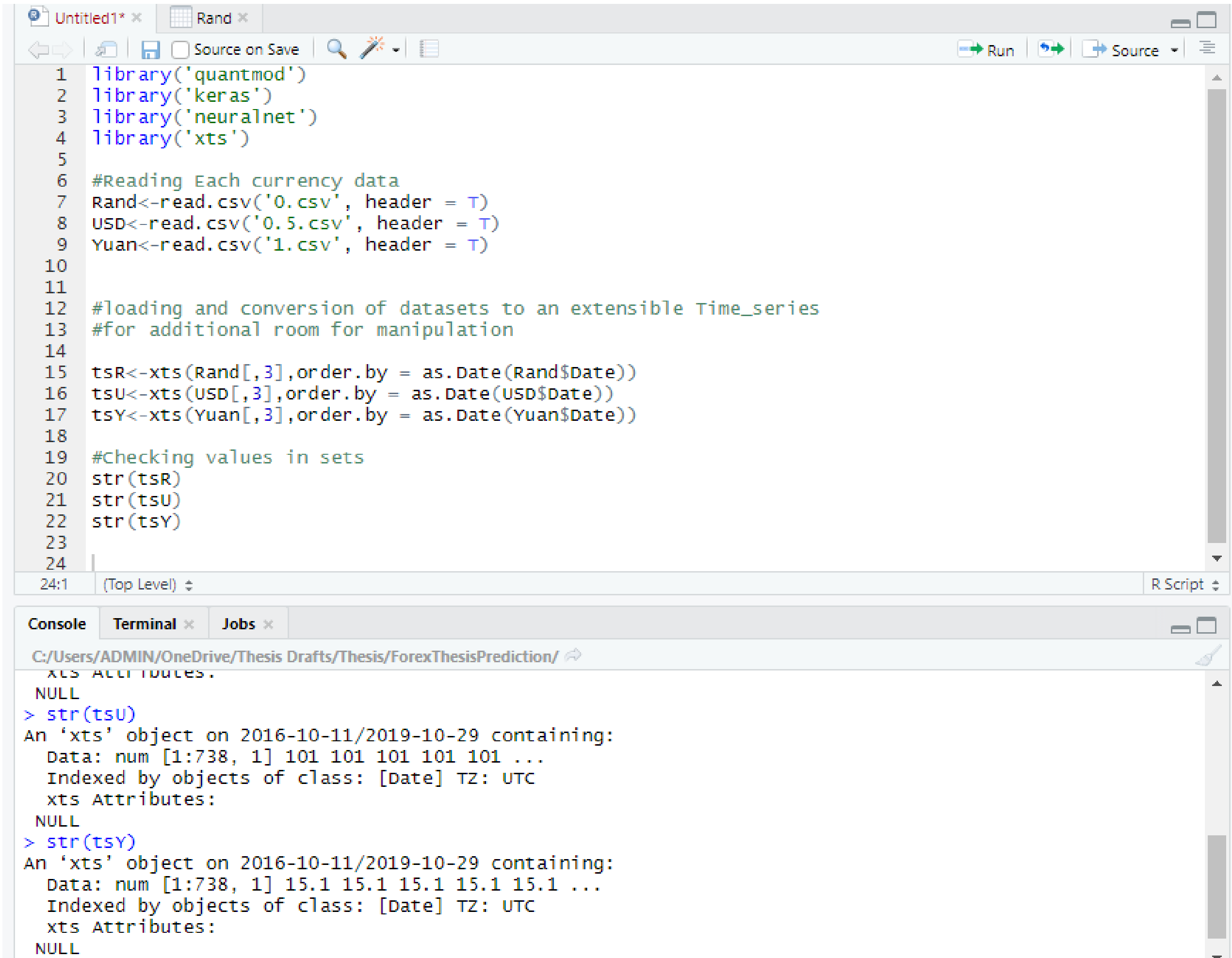


Figure 5.3 : Loading Packages and Data conversions

The data has to be indexed based on the date and ordered serially in the same manner as in Figure 5.3. Using xts, date variables are converted to a format that can be stored for manipulation as dates from string during importation.

5.2.3 Target and Predictor variable feature creation

The structure is set to allow for the mean price values be predicted to a future date. The dailyReturn() function of xts is mounted on the structure to normalize the mean values that are given from the data-set as in Equation 5.1.

$$\text{Mean} = (\text{Buy} + \text{Sell})/2$$

Equation 5.1 : The Selected Variable to be Predicted

A lag feature is introduced to define the next step in the sequence, using xts a lag of +1 day implies that a time step of a single day from the current selection considering that the target values are based on the dailyReturn() function. Setting the lag feature to -1, would imply back-tracing the timeseries, however for the current selection model construct +1 ensures a forward delay. Consider Figure 5.4. illustration. The USD dataset has a unique behavioral pattern and the data requires further standardization to reduce its noisy nature. Equation 5.2 depicts the standardization utilized to scale the USD set:

$$\text{resultant } x = (x-102.2367714)/1.246788$$

Equation 5.2 : Standardization Equation for scaling USD dataset behavior

The values in equation 5.2 from 102.2367714 is set from the average of the USD dataset random observations and 1.246 is the standard deviation of the set.

```
# Target and Predictor feature setting
#dailyReturn
rtsR<-dailyReturn(tsR)
rtsU<-dailyReturn(tsU)
rtsY<-dailyReturn(tsY)

rtsR1<-lag(rtsR,k=1)
rtsU1<-lag(rtsU,k=1)
rtsY1<-lag(rtsY,k=1)

rtsdallR<-cbind(rtsR,rtsR1)
rtsdallU<-cbind(rtsU,rtsU1)
rtsdallY<-cbind(rtsY,rtsY1)

colnames(rtsdallR)<-c('rtsR','rtsR1')
colnames(rtsdallU)<-c('rtsU','rtsU1')
colnames(rtsdallY)<-c('rtsY','rtsY1')
```

Figure 5.4 : Lag Feature setting

5.2.4 Data Partitioning

The training and test data is set based on the indexing structure of the date. The training data is captured at 80% of the total data which begins from 10/11/2016 to 12/03/2019. The remaining 20% ending on 29/10/2019 is set for testing. The xts library using window facility allows for the setting of the data partitions.

```
rtsRtrain<-window(rtsdallR, end='2019-03-12')
rtsRtest<-window(rtsdallR, start='2019-03-12')
rtsUtrain<-window(rtsdallU, end='2019-03-12')
rtsUtest<-window(rtsdallU, start='2019-03-12')
rtsYtrain<-window(rtsdallY, end='2019-03-12')
rtsYtest<-window(rtsdallY, start='2019-03-12')
```

Figure 5.5 : Data Partition Order

5.2.5 Trained Model Structure

With the captured intercepts of the data, the model neural network is trained on an experimental basis for most optimal results by trial and error. The three currencies are split and exposed to various simulations. The Rand is constructed with 2 hidden layers and 8 neurons for the hidden layers. The repetitions are set at 5 with a stepmax of 1,000,000. Figure 5.6 shows a plot of the rand neural network model utilizing the rprop+ resilient backpropagation algorithm as a first-order optimization algorithm.

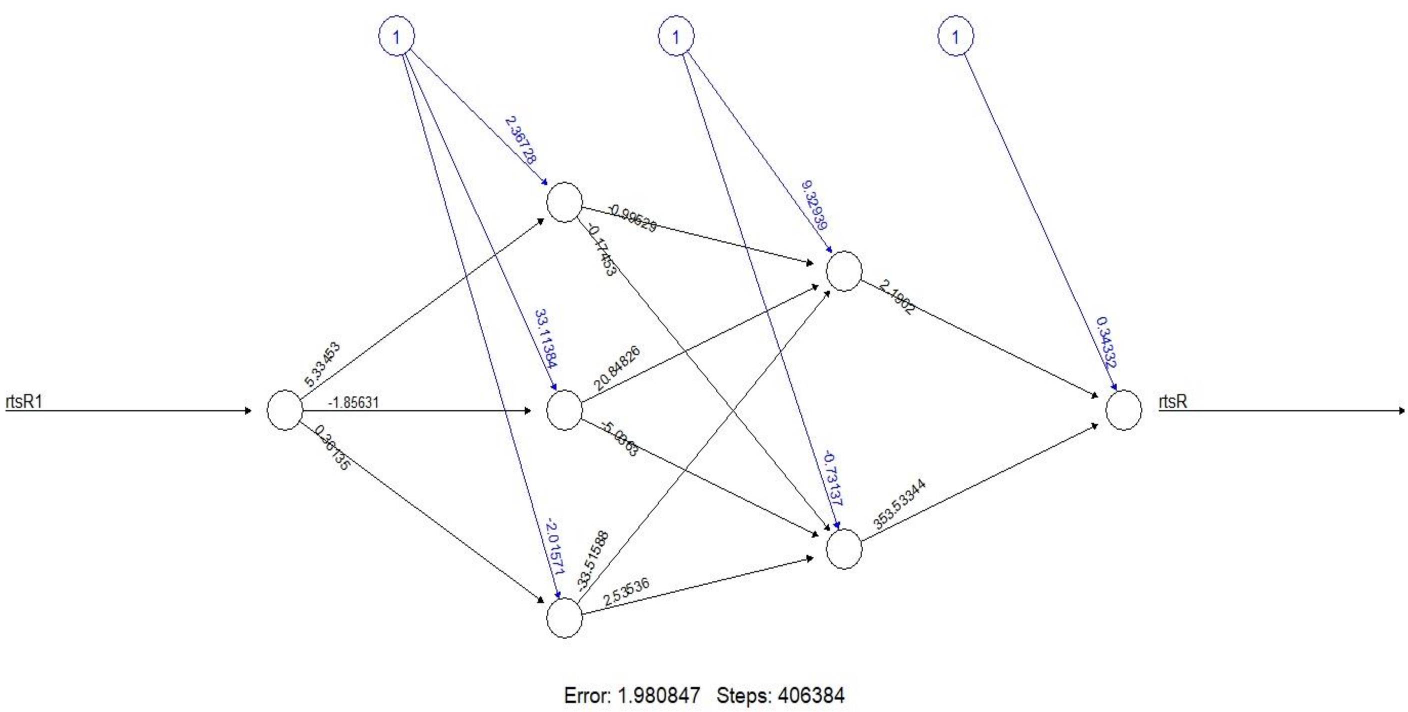


Figure 5.6 : Rand Neural Network Structure

Table 5.2 : Rand Model Construct Summary

	Length	Class	Mode
call	10	-none-	call
response	590	-none-	numeric
covariate	590	-none-	numeric
model.list	2	-none-	list
error functions	1	-none-	function
activation functions	1	-none-	function
linear.output	1	-none-	logical
data	2	data.frame	list
exclude	0	-none-	NULL
net.result	5	-none-	list
weights	5	-none-	list
generalized.weights	5	-none-	list
startweights	5	-none-	list
result.matrix	100	-none-	numeric

The USD is constructed with 2 hidden layers and 8 neurons for the hidden layers. The repetitions are set at 100 with a stepmax of 1,000,000. Figure 5.7 shows a plot of the USD neural network model utilizing the rprop+ resilient backpropagation algorithm.

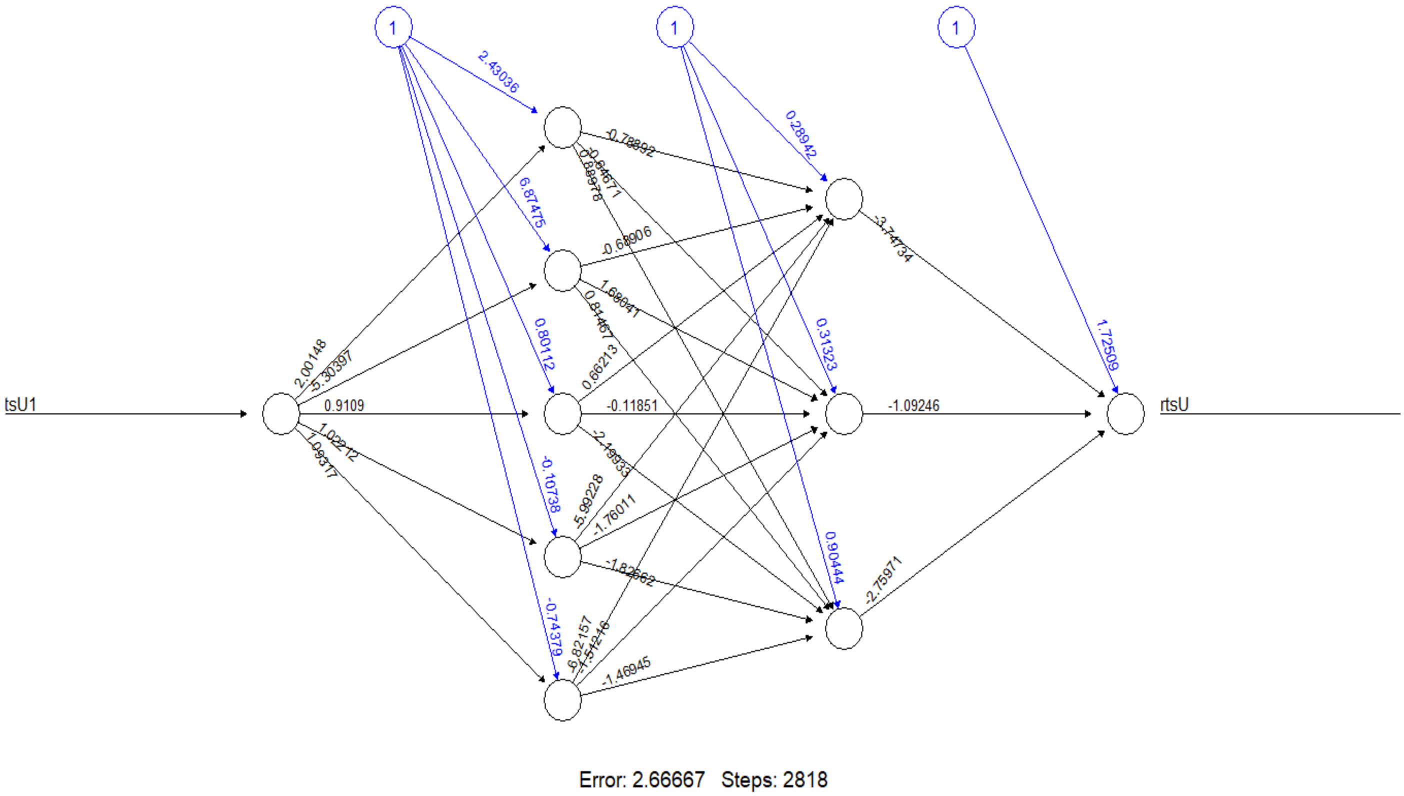


Figure 5.7 : USD Neural Network Structure

Table 5.3 : USD Model Construct Summary

	Length	Class	Mode
call	10	-none-	call
response	590	-none-	numeric
covariate	590	-none-	numeric
model.list	2	-none-	list
error functions	1	-none-	function
activation functions	1	-none-	function
linear.output	1	-none-	logical
data	2	data.frame	list
exclude	0	-none-	NULL
net.result	5	-none-	list

weights	5	-none-	list
generalized.weights	5	-none-	list
startweights	5	-none-	list
result.matrix	175	-none-	numeric

The Yuan is constructed with 2 hidden layers and 8 neurons for the hidden layers. The repetitions are set at 200 with a stepmax of 1,000,000. Figure 5.6 shows a plot of the Yuan neural network model utilizing the rprop+ resilient backpropagation algorithm as well.

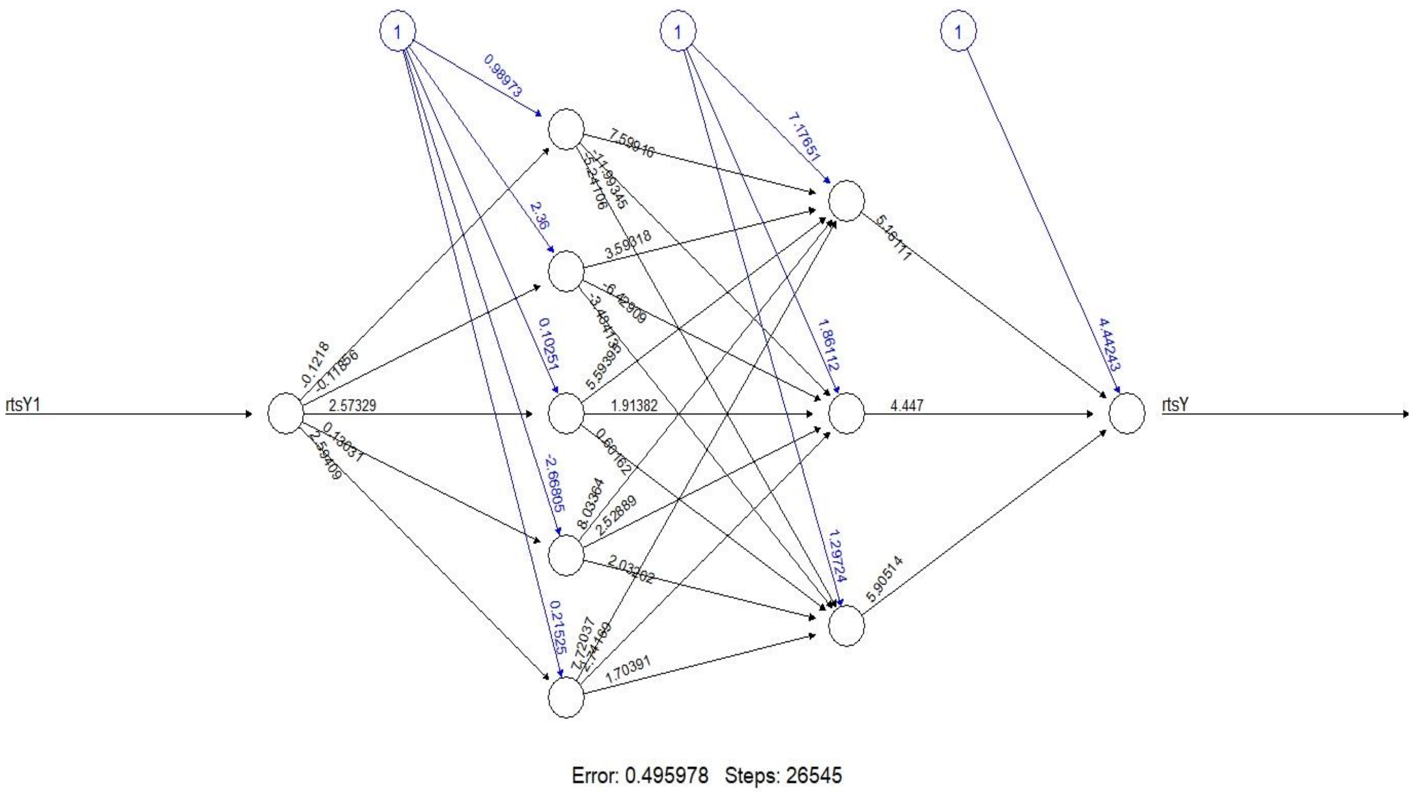


Figure 5.8 : Yuan Neural Network Structure

Table 5.4 : Yuan Model Construct Summary

	Length	Class	Mode
call	10	-none-	call
response	590	-none-	numeric
covariate	590	-none-	numeric
model.list	2	-none-	list
err.fct	1	-none-	function
act.fct	1	-none-	function
linear.output	1	-none-	logical
data	2	data.frame	list
exclude	0	-none-	NULL
net.result	15	-none-	list
weights	15	-none-	list
generalized.weights	15	-none-	list
startweights	15	-none-	list
result.matrix	525	-none-	numeric

Tables 5.2, 5.3 and 5.4 are depicting the model construct summary that shows the respective variable quantity formations from the model result to the fitted weight capacities on generalization and start weights. The weighting is significant as it creates the basis for predictions. Figures 5.6, 5.7 and 5.8 clearly depict the weighting relationships amongst the hidden layers for each of the respective currency models.

5.3 Forex Price Prediction

The existing forex market prices which include the historical and recent forex market prices are loaded into the trained model structure. The daily trained values then create a matrix of predicted forex market prices for each of the respective currencies.

5.3.1 Testing/Validating the Model

The tabulation of the tests for the model utilized 20% of the data-set, that is set from 12-03-2019 to 29-10-2019 which represents eight (8) months' worth of validation data. The Figures 5.9, 5.10 and 5.11 are consecutive samples of the predicted eight months' forex prices for the South African Rand, United States Dollar and Chinese Yuan Currencies.

```
2019-10-29
> pred.pricetestY$net.result[,1]
2019-03-12 2019-03-13 2019-03-14 2019-03-15 2019-03-18 2019-03-19 2019-03-20 2019-03-21
  14.82596   14.81359   14.93590   14.94486   14.88963   14.94073   14.99566   15.02242
2019-03-22 2019-03-25 2019-03-26 2019-03-27 2019-03-28 2019-03-29 2019-04-01 2019-04-02
  15.11624   15.06044   15.00089   15.04938   15.03409   14.99053   14.94928   15.02554
2019-04-03 2019-04-04 2019-04-05 2019-04-08 2019-04-09 2019-04-10 2019-04-11 2019-04-12
  15.02735   15.01206   15.02051   14.99526   14.97826   14.98651   15.01830   15.05219
2019-04-15 2019-04-16 2019-04-17 2019-04-24 2019-05-10 2019-05-13 2019-05-14 2019-05-15
  15.02534   15.05963   15.05440   15.05792   15.10327   14.86337   14.73760   14.69572
2019-05-16 2019-05-17 2019-05-20 2019-05-21 2019-05-22 2019-05-23 2019-05-24 2019-05-27
  14.69813   14.69903   14.63872   14.62348   14.61987   14.63120   14.65698   14.64363
2019-05-28 2019-05-29 2019-05-30 2019-05-31 2019-06-03 2019-06-04 2019-06-06 2019-06-07
  14.68689   14.67524   14.65096   14.67113   14.66842   14.65959   14.66270   14.66109
2019-06-10 2019-06-11 2019-06-12 2019-06-13 2019-06-14 2019-06-17 2019-06-18 2019-06-19
  14.65979   14.64594   14.63902   14.65166   14.65256   14.66832   14.70476   14.73007
2019-06-20 2019-06-21 2019-06-24 2019-06-25 2019-06-26 2019-06-27 2019-06-28 2019-07-01
  14.77699   14.75951   14.84557   14.82988   14.81429   14.83410   14.85080   14.87544
2019-07-02 2019-07-03 2019-07-04 2019-07-05 2019-07-08 2019-07-09 2019-07-10 2019-07-11
  14.90200   15.01518   14.98580   14.86367   14.93600   14.88942   14.92765   14.91820
2019-07-12 2019-07-15 2019-07-16 2019-07-17 2019-07-18 2019-07-19 2019-07-22 2019-07-23
  14.97101   14.97957   14.97765   14.98550   14.97544   14.98651   14.98540   15.00361
2019-07-24 2019-07-25 2019-07-26 2019-07-29 2019-07-30 2019-07-31 2019-07-31 2019-08-01
  15.06154   15.08075   15.13654   15.08638   15.06275   15.11051   15.13865   15.13865
2019-08-02 2019-08-05 2019-08-06 2019-08-07 2019-08-08 2019-08-09 2019-08-13 2019-08-14
  15.13171   14.95532   14.65126   14.63220   14.70978   14.65597   14.66792   14.58549
2019-08-15 2019-08-16 2019-08-19 2019-08-20 2019-08-21 2019-08-22 2019-08-23 2019-08-26
  14.61485   14.69211   14.67715   14.66922   14.64975   14.58709   14.55444   14.54683
2019-08-27 2019-08-28 2019-08-29 2019-08-30 2019-09-02 2019-09-03 2019-09-04 2019-09-05
  14.56095   14.45846   14.44877   14.43800   14.49562   14.49072   14.45057   14.45846
2019-09-06 2019-09-09 2019-09-10 2019-09-11 2019-09-12 2019-09-13 2019-09-16 2019-09-17
  14.51302   14.52752   14.57938   14.56275   14.58188   14.58038   14.64895   14.65778
2019-09-18 2019-09-19 2019-09-20 2019-09-23 2019-09-24 2019-09-25 2019-09-26 2019-09-27
  14.68377   14.63681   14.66330   14.63411   14.59380   14.60964   14.58889   14.57818
2019-09-30 2019-10-01 2019-10-02 2019-10-03 2019-10-04 2019-10-07 2019-10-08 2019-10-09
  14.56105   14.57677   14.53823   14.52872   14.52842   14.52182   14.52542   14.51912
2019-10-11 2019-10-14 2019-10-15 2019-10-16 2019-10-17 2019-10-18 2019-10-22 2019-10-23
  14.52372   14.54814   14.61626   14.68277   14.64343   14.61124   14.65086   14.64975
2019-10-24 2019-10-25 2019-10-28 2019-10-29
  14.63220   14.64524   14.60623   14.66109
```

Figure 5.9 : Sample Predicted Rand Currency Prices

```
> pred.pricetestU$net.result[,1]*1.246+102.24
2019-03-12 2019-03-13 2019-03-14 2019-03-15 2019-03-18 2019-03-19 2019-03-20 2019-03-21
  99.61618   99.58528  100.19294  100.24767  100.17423  100.30377  100.66628  100.85442
2019-03-22 2019-03-25 2019-03-26 2019-03-27 2019-03-28 2019-03-29 2019-04-01 2019-04-02
 101.00728  100.78386  100.71019  100.96524  100.95718  100.88151  100.75660  100.78660
2019-04-03 2019-04-04 2019-04-05 2019-04-08 2019-04-09 2019-04-10 2019-04-11 2019-04-12
 100.84896  100.90633  100.80617  100.72387  100.64154  100.64986  100.84681  101.09198
2019-04-15 2019-04-16 2019-04-17 2019-04-24 2019-05-10 2019-05-13 2019-05-14 2019-05-15
 100.95446  100.99117  100.99758  101.05428  101.52813  101.13089  101.07177  101.06749
2019-05-16 2019-05-17 2019-05-20 2019-05-21 2019-05-22 2019-05-23 2019-05-24 2019-05-27
 101.09198  101.10414  101.11629  101.10822  101.10880  101.16500  101.21119  101.22039
2019-05-28 2019-05-29 2019-05-30 2019-05-31 2019-06-03 2019-06-04 2019-06-06 2019-06-07
 101.23400  101.23184  101.27109  101.39294  101.34781  101.28033  101.27384  101.34227
2019-06-10 2019-06-11 2019-06-12 2019-06-13 2019-06-14 2019-06-17 2019-06-18 2019-06-19
 101.29292  101.19975  101.27001  101.37526  101.38022  101.53994  101.81481  102.00790
2019-06-20 2019-06-21 2019-06-24 2019-06-25 2019-06-26 2019-06-27 2019-06-28 2019-07-01
 102.02747  101.90382  101.87215  101.87795  101.91583  102.20238  102.18007  102.31141
2019-07-02 2019-07-03 2019-07-04 2019-07-05 2019-07-08 2019-07-09 2019-07-10 2019-07-11
 102.37464  102.73103  103.00443  102.30744  102.64399  102.62203  102.70394  102.76030
2019-07-12 2019-07-15 2019-07-16 2019-07-17 2019-07-18 2019-07-19 2019-07-22 2019-07-23
 102.92803  102.92755  102.95901  103.03030  103.05396  103.05910  103.10309  103.21420
2019-07-24 2019-07-25 2019-07-26 2019-07-29 2019-07-30 2019-07-31 2019-07-31 2019-08-01
 103.60917  103.75821  104.01154  103.79512  103.81428  104.05575  104.17458  104.17466
2019-08-02 2019-08-05 2019-08-06 2019-08-07 2019-08-08 2019-08-09 2019-08-13 2019-08-14
 104.14288  103.22939  103.00879  103.19923  103.26868  103.43243  103.36857  103.06803
2019-08-15 2019-08-16 2019-08-19 2019-08-20 2019-08-21 2019-08-22 2019-08-23 2019-08-26
 103.25010  103.25992  103.32785  103.37690  103.28807  102.97373  102.96569  103.05988
2019-08-27 2019-08-28 2019-08-29 2019-08-30 2019-09-02 2019-09-03 2019-09-04 2019-09-05
 103.17723  103.31856  103.44321  103.47878  103.54501  103.58236  103.63286  103.81980
2019-09-06 2019-09-09 2019-09-10 2019-09-11 2019-09-12 2019-09-13 2019-09-16 2019-09-17
 103.81596  103.87987  103.81980  103.72382  103.66046  103.77312  103.78413  103.76925
2019-09-18 2019-09-19 2019-09-20 2019-09-23 2019-09-24 2019-09-25 2019-09-26 2019-09-27
 103.78958  103.85414  103.90590  103.82659  103.82659  103.85795  103.82305  103.85581
2019-09-30 2019-10-01 2019-10-02 2019-10-03 2019-10-04 2019-10-07 2019-10-08 2019-10-09
 103.81162  103.87909  103.85141  103.85795  103.85581  103.80935  103.83514  103.79077
2019-10-11 2019-10-14 2019-10-15 2019-10-16 2019-10-17 2019-10-18 2019-10-22 2019-10-23
 103.77590  103.75214  103.76318  103.78522  103.69553  103.70944  103.69553  103.72333
```

Figure 5.10 : Sample Predicted USD Currency Prices

```
 pred.pricetest$net.result[,1]
019-03-12 2019-03-13 2019-03-14 2019-03-15 2019-03-18 2019-03-19 2019-03-20 2019-03-21
 6.867112   6.956357   7.027280   7.004378   6.925842   6.971510   6.975135   6.965503
019-03-22 2019-03-25 2019-03-26 2019-03-27 2019-03-28 2019-03-29 2019-04-01 2019-04-02
 7.017787   7.103835   7.013571   7.054948   7.017020   6.916926   6.881638   6.883621
019-04-03 2019-04-04 2019-04-05 2019-04-08 2019-04-09 2019-04-10 2019-04-11 2019-04-12
 7.111472   7.143431   7.111085   7.123664   7.153618   7.148863   7.167021   7.256639
019-04-15 2019-04-16 2019-04-17 2019-04-24 2019-05-10 2019-05-13 2019-05-14 2019-05-15
 7.209949   7.234837   7.222436   7.169353   7.120663   7.062839   7.113697   7.078640
019-05-16 2019-05-17 2019-05-20 2019-05-21 2019-05-22 2019-05-23 2019-05-24 2019-05-27
 7.114761   7.093788   7.069773   7.069195   7.040528   7.027280   7.054756   7.003325
019-05-28 2019-05-29 2019-05-30 2019-05-31 2019-06-03 2019-06-04 2019-06-06 2019-06-07
 7.041489   7.009644   6.929829   6.937902   6.853740   6.914367   7.025074   6.828939
019-06-10 2019-06-11 2019-06-12 2019-06-13 2019-06-14 2019-06-17 2019-06-18 2019-06-19
 6.754637   6.748788   6.865981   6.918633   6.818441   6.839076   6.891746   6.908209
019-06-20 2019-06-21 2019-06-24 2019-06-25 2019-06-26 2019-06-27 2019-06-28 2019-07-01
 7.019225   7.024307   7.116599   7.112246   7.104898   7.130346   7.170714   7.225755
019-07-02 2019-07-03 2019-07-04 2019-07-05 2019-07-08 2019-07-09 2019-07-10 2019-07-11
 7.255172   7.308787   7.294458   7.247640   7.321165   7.231418   7.274067   7.244512
019-07-12 2019-07-15 2019-07-16 2019-07-17 2019-07-18 2019-07-19 2019-07-22 2019-07-23
 7.318413   7.389407   7.381910   7.422292   7.371754   7.358945   7.393847   7.422983
019-07-24 2019-07-25 2019-07-26 2019-07-29 2019-07-30 2019-07-31 2019-07-31 2019-08-01
 7.485346   7.454142   7.489807   7.361309   7.255954   7.347424   7.351165   7.351165
019-08-02 2019-08-05 2019-08-06 2019-08-07 2019-08-08 2019-08-09 2019-08-13 2019-08-14
 7.328438   7.124826   6.937997   6.940373   6.980193   6.860988   6.867677   6.815444
019-08-15 2019-08-16 2019-08-19 2019-08-20 2019-08-21 2019-08-22 2019-08-23 2019-08-26
 6.759934   6.749995   6.781529   6.803187   6.752037   6.711022   6.777522   6.815350
019-08-27 2019-08-28 2019-08-29 2019-08-30 2019-09-02 2019-09-03 2019-09-04 2019-09-05
 6.813571   6.804029   6.775286   6.741089   6.801037   6.810389   6.844525   6.889100
019-09-06 2019-09-09 2019-09-10 2019-09-11 2019-09-12 2019-09-13 2019-09-16 2019-09-17
 6.998637   7.014433   7.043891   7.057161   7.062936   7.047351   7.103931   7.147408
019-09-18 2019-09-19 2019-09-20 2019-09-23 2019-09-24 2019-09-25 2019-09-26 2019-09-27
 7.069292   7.046774   7.103931   7.106928   6.974467   6.976566   6.984108   6.947315
019-09-30 2019-10-01 2019-10-02 2019-10-03 2019-10-04 2019-10-07 2019-10-08 2019-10-09
 6.945508   6.876164   6.863343   6.801224   6.831754   6.856187   6.902812   6.870034
019-10-11 2019-10-14 2019-10-15 2019-10-16 2019-10-17 2019-10-18 2019-10-22 2019-10-23
 6.821814   6.853834   7.005144   7.041393   6.977712   6.944652   7.002177   7.009835
019-10-24 2019-10-25 2019-10-28 2019-10-29
```

Figure 5.11 : Sample Predicted Yuan Currency Prices

Table 5.5 shows the performance of the respective sum of square error performance and co-efficient of correlation that estimates the accuracy of the prediction results on the currencies. By definition, the sum of squared errors is a measure of the difference between the predicted and actual result (Günther & Fritsch, 2010; Kaya & Hajimirza, 2018). The correlation values are usually on a range of 0 to 1, the closer the correlation value to 1, the more accurate the result of the predictions (Brownlee, MachineLearningMastery, 2019; Mindrila & Balentyne, 2013).

Table 5.5 : Sum of Squared Errors (SSE) and Correlations for Model Accuracy

Currency	SSE/R	Correlation
Rand	1.980847	0.9415805
USD	2.666667	0.9909911
Yuan	0.495978	0.9642068

Based on the model plots of the predictions, Figures 5.12, 5.13 and 5.14 show the predicted values verses the actual for each of the three currency values in flow over the 8-month time horizon.

Table 5.6 : Chart Labels Predicted vs Actual Forex prices

Red	**Actual**
Black	**Predicted**

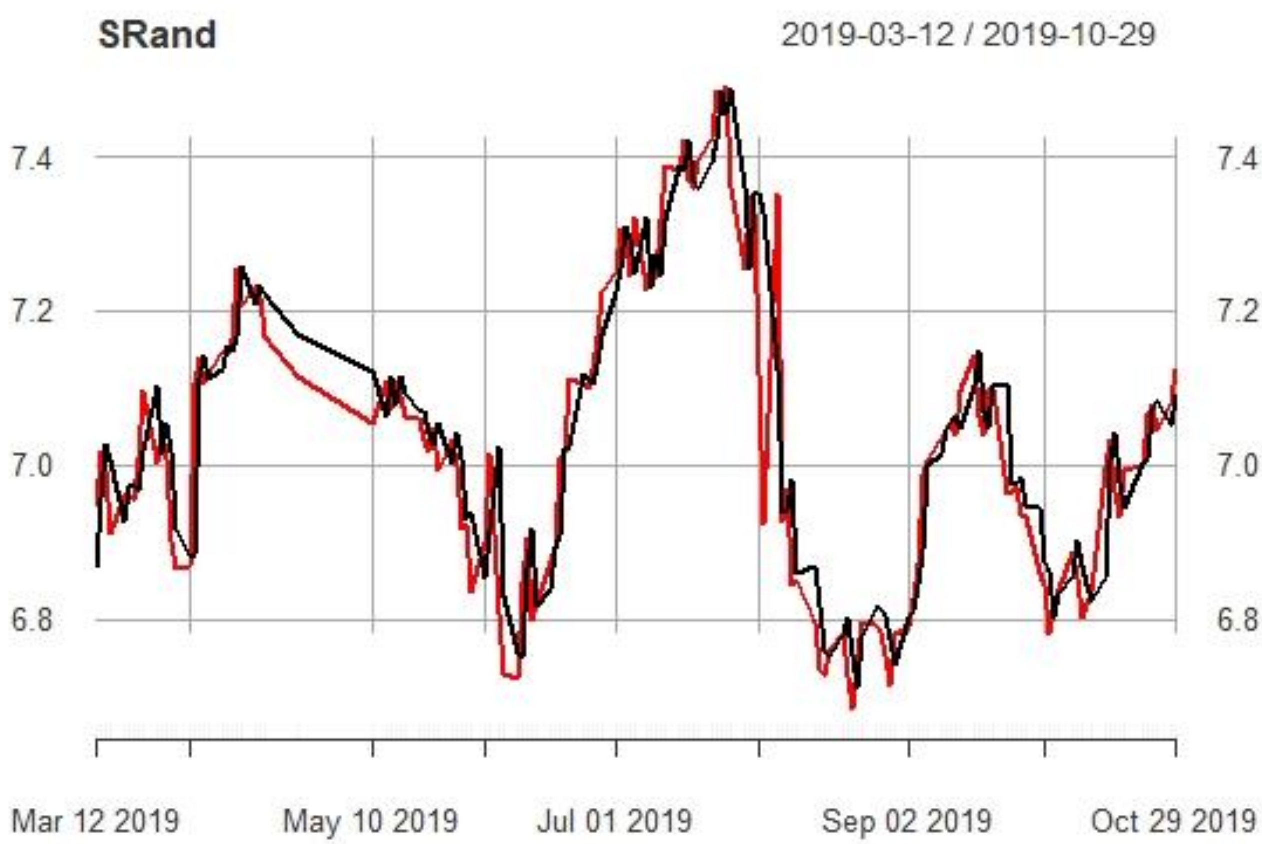


Figure 5.12 : South African Rand Actual Vs Predicted Prices

Figure 5.12 depicts the performance of the Rand where the red line are the actual values while the black shows how the prices that the neural network model has predicted. The black line is moving almost in the same flow-out of the red line.

Figure 5.13 : United States Dollar Actual Vs Predicted Prices

Figure 5.13 depicts the performance of the USD where the red line are the actual values while the black shows how the prices that the neural network model has predicted. The time horizon in this test scenario is set from March, 2019 to Oct, 2019. The black line is moving almost in the same flow-out of the red line.

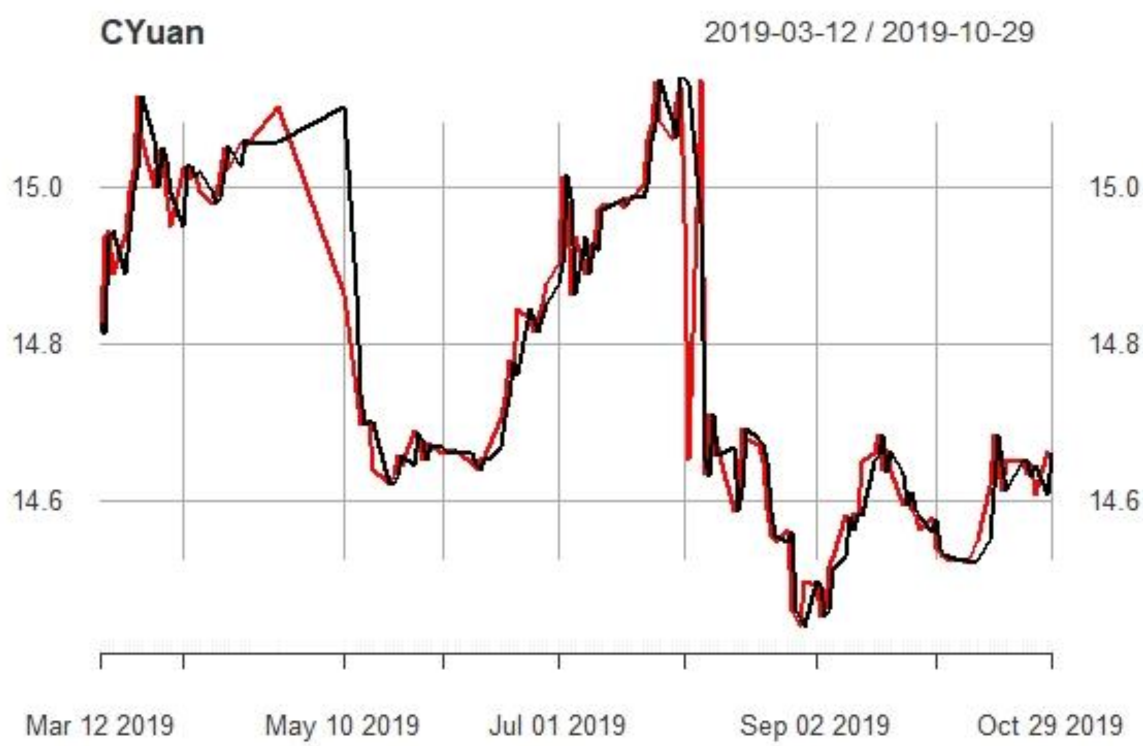


Figure 5.14 : Chinese Yuan Actual Vs Predicted Prices

Figure 5.14 depicts the performance of the Yuan where the red line are the actual values while the black shows how the prices that the neural network model has predicted. The time horizon in this test scenario is set from March, 2019 to Oct, 2019. The black line is moving almost in the same flow-out of the red line.

Chapter 6 : Discussions

6.1 Overview

This chapter reviews the optimal model outcomes based on the highlighted specific objectives of the research. The main aim was to develop a model for forex market price prediction in the Kenyan market using the Central Bank of Kenya data. The data had headers of the dates, currency name, mean, buy and sell price of each currency on a daily basis. The CBK data provided an indicative rate of the Forex market currency values historically. The indication of such data based on buy-sell spread did not imply the CBK sets the market prices, rather forces of demand and supply in the market. The model was supposed to predict each of the respective currency values accurately. In order to achieve this, it was clear that independent implementation of each respective currency to the model design is required for higher efficiency and optimality. This was one of the key findings of the study.

The construction therefore resulted to three separate implementations of a neural network, each unique to a currency. The currency prediction model is constructed utilizing the resilient backpropagation (rprop+) multi-layer neural network algorithm all of which are combined with the extensible time series analysis. Each of the simulations for the currencies had its own neural network size designed for optimality. The performance of the model was evaluated on the SSE/R and R squared for accuracy. Being a numeric prediction problem and not a classification one, the use of a confusion matrix to assess the accuracy would not be sufficient and efficient (Lantz, 2019). Therefore, measuring the strength of correlation of the predicted forward price and true values is likely to be a useful measure of accuracy. Applying the rprop+ neural network algorithm on extensible time series not only improves the accuracy of the forex market price predictions significantly but also makes the process faster and allows for additional data plug-in.

6.2 Model Optimality

The simulation approach in establishing the most optimal design to implement on the time series neural network involved making the simulations on the three currencies separately while varying the algorithm and number of iterations. Ultimately, the USD performed at 98.21% accuracy based on the R^2 followed by the Yuan at 92.27% and Rand at 88.67%. From the output format as in the

Table 6.1, with increased number of layers and neurons for the resilient backpropagation algorithm minimized its error continuously. Each of the currency predictions had a unique error level of which the USD had the highest SSE of 2.667 then the Rand at 1.981 and finally the Yuan at 0.496.

Table 6.1 : rprop+ algorithm error and maximum steps to completion

(layers, neurons)	**(1, 2)**	**(2, 2)**	**(3, 2)**	**>(3, 3)**
error	63.146000056	5.002325e-01	4.976547e-01	4.942660e-01
reached.threshold	0.007342536	9.375301e-03	9.991103e-03	9.126804e-03
Steps	55.000000000	1.142500e+04	3.034240e+05	2.403000e+04

With the sag algorithm, there was increased training time required of up-to 2.84 minutes and more iterations to the end step were needed for better accuracy to be reached. Even so, the accuracy only measured at 88.66%.

Table 6.2 : Sag algorithm errors and maximum steps to completion

(layers, neurons)	**(1, 2)**	**(2, 2)**	**(3, 2)**	**>(3, 3)**
error	1.981802e+00	1.981127e+00	1.981581e+00	1.981787e+00
reached.threshold	9.758662e-03	723058e-03	6.886672e-03	9.928044e-03
Steps	5.675300e+04	2.091400e+04	1.064300e+04	1.648870e+05

Figure 6.1 illustrates a scenario where all combined currencies are loaded onto the single model design which is significantly less optimal for implementation as a model. There are lesser steps (112) to achieve the model but a huge error factor of 203.370043 with an accuracy level of

0.3364%. Even with an increased size of the neural net, the error factor reduces slightly however the accuracy for prediction remains low on each currency. This therefore implies that for a significant times series prediction structure, it is reasonable to split the respective currencies and have each of them implemented independently using the data-driven modelling process.

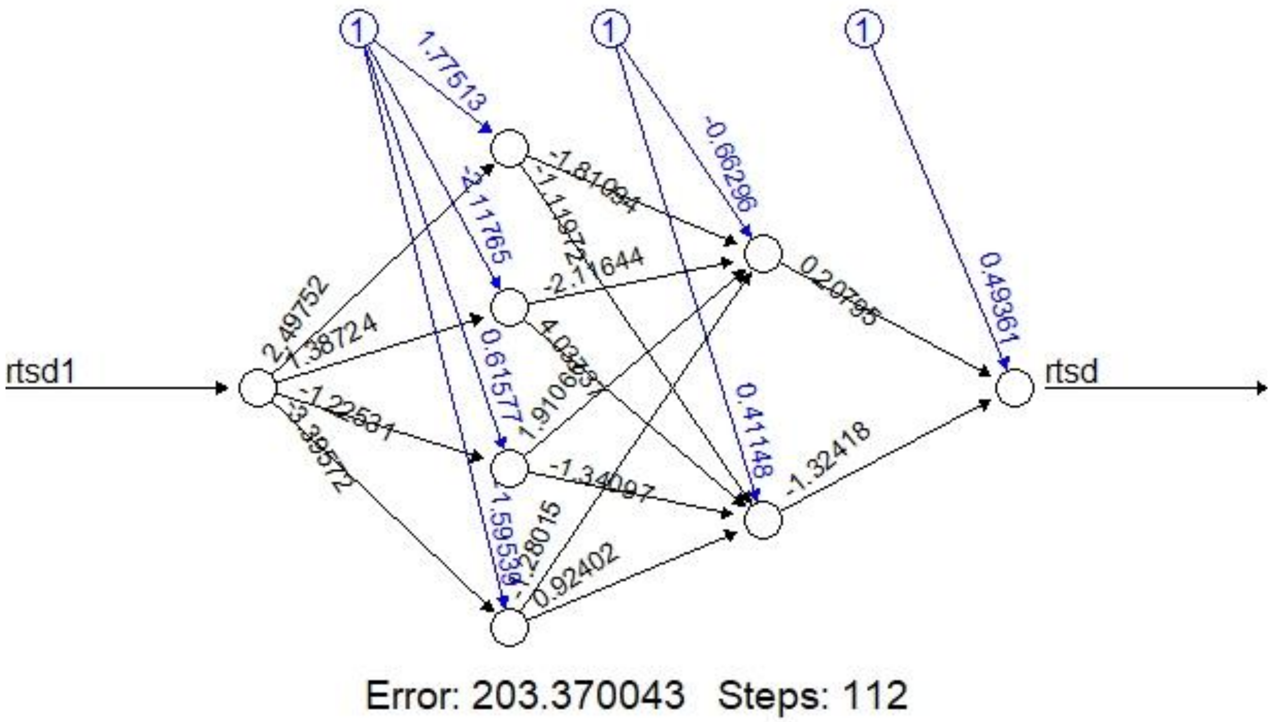


Figure 6.1 : Combined Currencies Model Structure

Common approaches employ recurrent neural networks and split the data format elements of the date into normalized variables of day, month and year into separate components. With this implementation, the future learning capacity and trend analysis of the model is significantly lost. For consistency and higher accuracy, utilizing an extensible time series proved to be effective, as it creates increased room for additional manipulations of the time series variables based on the index of date. It also makes the model easy to export and apply as a function into other applications or models.

6.2.1 Minimizing Model Error and Improving Accuracy

From the study, trying to use a single implementation neural net model for general predictions on a time based data set realized high error levels and lowest accuracies. This component is based on the different behavioral patterns of each of the respective currencies on a singular day while at the same time the existence of multiple currency values on a single day made it more difficult to train a structured time-based model. In consequence, separating each of the currencies models and understanding the data behavioral needs separately facilitated for much a more accurate model design with lowest error levels and highest accuracies as depicted in Table 5.5.

With normalization of the price value columns, the model error levels are significantly lower as opposed to not normalizing the price values. dailyReturn () was the function this study used to normalize the price values. The function basically converts the price value data-set to the range of 0-1 by averaging each of the random observations as a calculation of the daily returns. Despite a reduction in error to 0.035, the accuracy level is still significantly low due to a low correlation coefficient of 0.0631. This can be explained by the need to revert the predicted data to the standard format and compare it with the standard version. By doing this, the accuracy level goes to 93% by measure of R^2 as observed with the Yuan currency.

The nature of the USD data-set called for an additional form of regression standardization. The distribution of this particular set proved to be too noisy in the initial stages. Increasing the size of the network structure had no impact towards actually training the model. The standardization ensures that the neural net can actually learn and assign optimal weights. Such standardization is a technique of data scaling (Brownlee, MachineLearningMastery, 2019). The standardized data has to be reverted by a reverse function for actual understandable results similar to the de-normalization process. From the observations in section 6.2, the error for the USD is registered as the highest at 2.667. Despite the highest accuracy it registered, the process of scaling and unscaling causes quite an increase in the error levels from the behavior of this dataset. Comparing to the rest of the currencies, additional scaling only increased the error level but did not have any impact on the training process or the accuracy of the predictions.

6.3 Comparative Review

The model design makes a combined application of rprop+ with an extensible time series that is based on bivariate data for training and validation. Unlike the study by Gould (2004) which focused on a single currency, the final design creates a basis for Forex market price prediction on multiple currencies that can be utilized by CBK economists and Forex dealers. Wanjiku's (2012) output being stochastic in nature creates an econometric support for analyzing the USD against the Kenya shilling only. This research however, demonstrates how a neural network model can be designed for multiple currencies. It takes the view beyond simple econometrics and the focus of the predictor no longer becomes the factors of influencing forecasts or what happens underneath rather, the predicted values themselves.

Chapter 7 : Conclusions and Recommendations

7.1 Research Conclusions

The main output of the research is a model for Forex market price prediction that is based on CBK data. The research has achieved it by use of artificial neural network time series resilient backpropagation algorithm. To achieve this, four major deliverables were exploited based on the research questions. In review, we check whether they were conclusively addressed and answered.

What are the factors affecting forex market prices/rates in Kenya? Section 2.2 and 2.3 of Chapter 2 reviews the market factors conclusively and addresses the forex market operations in Kenya and various factors that may influence the forex prices. The study termed forex market prices in Kenya to leverage on the price of the Kenyan shilling against foreign exchange currencies. This helped the researcher understand the behavioral performance of the forex market price predictions.

How can predictive analytical tools/approaches be used to forecast forex market rates? Section 2.4 and 2.5 of Chapter 2 descriptively address the previous research works, techniques and findings from the technical and theoretical perspective of traditional stochastic designs and business intelligence based techniques. It ensured the researcher understood the study context fully and could cover gaps in the pre-existing approaches.

How can the model for forex market price prediction be developed? Based on the conceptual framework build-up of section 2.6 of Chapter 2, the basis for developing the model was formed. Chapter 4 and 5 show how combining the time series analysis technique to the resilient backpropagation algorithm. The systematic build-up of the development process was highlighted in Chapter 3 from the DDM and sequentially implemented in Chapter 5.

How will the forex market price prediction model be tested for efficiency and optimality for predictions in the forex markets? In evaluation of the model, Chapter 5 demonstrates the accuracy of the predictions by stipulating in a graphical manner the comparisons to the actual values. Other than this, the optimal nature of the currency model established by setting off the most accurate of the simulations against the least error scenario.

7.2 Research Contributions

CBK economists may use the model to study and determine inflation levels in the country at that particular moment and in the foreseeable short-term time horizon. In addition, dealers being business oriented market participants whose focus shall be on the hedging characteristics of a particular currency, can use the estimations of the predicted indicative prices to price secondary market derivatives such as forwards and futures contracts. Such prices can only be made possible over three to twelve months' time horizons since such market commodities are considered money markets which are short-term in nature. The model also technically improves the accuracy on traditional economic models by ensuring constant adaptation with continuous learning, that is faster and effective by applying it in context of the market.

7.3 Research Challenges

The major challenge with conducting the study was pertaining to the available data. The data obtained was historical in nature, contained too many duplications and missing days' prices that were not necessarily weekends/public holidays. Furthermore, the available data was not sufficient to create training and testing sets that could accurately predict long term horizon values (i.e. more than 12-24 months).

7.4 Recommendations

i. Price related adjustments. It is important that after each and every prediction, attention is paid to the indicative inflation to the time horizon on prediction and appropriate functional adjustment is made.

7.5 Further Study

The study can be advanced in a number of ways:

i. Creating a prediction model using multi-variate time series data, which could capture additional variables, not only for improved accuracy but for reliable decision support in the financial markets
ii. Creation of a unique R library that facilitates for multivariate time series regression based on date variable on neural networks

iii. Other than simply predicting forex market prices, further research can be done on currency classification before/after predication of the price. It is important that intelligent systems be able to understand the currency that has been predicted after being given say twenty currencies in one data-set.

iv. Further research on predicting additional forex currencies is important, considering the unique behavioral patterns attributed to each.

v. Mounting the model on a live data-stream which can be used to train and retrain the model in real time.

References

Amadeo, K. (2019, June 25). *An Introduction to the Financial Markets*. Retrieved from The balance: https://www.thebalance.com/an-introduction-to-the-financial-markets-3306233

Bauer College of Business. (2019). *Chapter V: Forecasting Exchange Rates.* Retrieved from Bauer College of Business, University of Houston: https://www.bauer.uh.edu/

BharatiDWConsultancy. (2017, October 22). Data Science Machine Learning Do it Yourself. Retrieved from https://drive.google.com/drive/folders/0ByQlW_DfZdxHeVBtTXllR0ZNcEU

Brownlee, J. (2016, September 21). *MachinelearningMastery*. Retrieved from How To Improve Deep Learning Performance: https://machinelearningmastery.com/improve-deep-learning-performance/

Brownlee, J. (2019, February 04). *MachineLearningMastery*. Retrieved from How to use Data Scaling Improve Deep Learning Model Stability and Performance?: https://machinelearningmastery.com/how-to-improve-neural-network-stability-and-modeling-performance-with-data-scaling/

Central Bank of Kenya. (2019, October). *Forex*. Retrieved from Central Bank of Kenya: https://www.centralbank.go.ke/forex/

Central bank of Kenya. (2019). *The Exchange Rate and the Performance of the Kenya Shilling.* Nairobi: Central Bank of Kenya.

Chartis Research. (2018). *Demystifying Artificial Intelligence in Risk and Compliance.* Chartis Research Ltd 2018.

CompareRemit. (2019, February 08). *8 Key Factors that Affect Foreign Exchange Rates*. Retrieved from CompareRemit: https://www.compareremit.com/money-transfer-guide/key-factors-affecting-currency-exchange-rates/

Contreras, A. V., Llanes-Castro, A., Pérez-Bernabeu, A., Navarro, S., Pérez-Sánchez, H., Lopez-Espin, J. J., & Cecilia, J. M. (2018). ENMX: An elastic network model to predict the FOREX market evolution. *Simulation Modelling Practice and Theory 86*.

Curram, S. P., & Mingers, J. (1994, April). Neural Networks, Decision Tree Induction and Discriminant Analysis: An Empirical Comparison. *The Journal of the Operational Research Society, 45*(4), 440-450. Retrieved November 18, 2019, from https://www.jstor.org/stable/2584215

Gaucan, V. (2011). Introduction to the Foreign Exchange Market. *Journal of Knowledge Management, Economics and Information Technology*.

Gould, J. H. (2004, December). Forex Prediction using an Artificial Intelligence System. Beijing, Beijing, China.

Günther, F., & Fritsch, S. (2010). neuralnet: Training of Neural Networks. *The R Journal, 2*(1). Retrieved from https://journal.r-project.org/archive/2010/RJ-2010-006/RJ-2010-006.pdf

Hansen, P. R., & Timmermann, A. (2012). *Choice of sample split in out-of-sample forecast* .

Kaya, M., & Hajimirza, S. (2018). *Rapid Optimization of External Quantum Efficiency of Thin Film Solar Cells Using Surrogate Modeling of Absorptivity*. Texas: Texas A&M University. doi:10.1038/s41598-018-26469-3

Lantz, B. (2019). Expert Techniques for predictive modelling. In B. Lantz, *Machine Learning with R* (Third ed., p. 235). Mumbai, India: Packt Publishing. Retrieved from https://books.google.co.ke/books?id=iNuSDwAAQBAJ&pg=PA235&lpg=PA235&dq=use+a+R+neural+model+function+to+actually+compute+unseen+future+values&source=bl&ots=O78Yxc8yR-&sig=ACfU3U0JExrPRE9-sHhXc4bLRzSOHYuZIA&hl=en&sa=X&ved=2ahUKEwiF4vHJ85rnAhVMPcAKHQUTDfIQ6

Lwanga, V. K. (2018, April). Stock market price prediction using sentiment analysis: a case study of Nairobi stock exchange market . Nairobi, Nairobi, Kenya.

Mindrila, D., & Balentyne, P. (2013). *Scatterplots and Correlation.* New York: Freeman and Company. Retrieved from https://www.westga.edu/academics/research/vrc/assets/docs/scatterplots_and_correlation_notes.pdf

Mugenda, O. M. (2003). Research Methods: Quantitative and qualitative Approaches. Nairobi, Kenya: African Centre for Technology Studies.

Ndeda, S. P. (2015). *Economics.* Nairobi: Manifested Publishers.

Ndeda, S. P. (2015). *Financial Institutions and Markets.* Nairobi: Manifested Publishers.

Ndung'u, N. (2000). *The Exchange Rate and Monetary Policy in Kenya* . Nairobi: African Development Bank.

Onyango, M. M. (2018, April). A Prototype to forecast trends in rainfall amounts in Kenya. Nairobi, Nairobi, Kenya: Strathmore University. Retrieved from https://su-plus.strathmore.edu/handle/11071/6000

Pondicherry University. (2019, October 31). *Pondicherry University: Foreign Exchange Markets.* Retrieved from Pondicherry University: http://www.pondiuni.edu.in/storage/dde/downloads/ibiv_forex.pdf

Rajasekar, S., Philominathan, P., & Chinnathambi, V. (2013). *Research Methodology.* Tamilnadu. Retrieved from https://arxiv.org/pdf/physics/0601009.pdf

Sercu, P. (n.d.). Do FX Markets Themselves? *Internationl Finance: Putting Theory Into Practive*.

Shigemizu, D., Abe, T., Morizono, T., A. Johnson, T., A. Boroevich, K., & Hirakawa, Y. (2014, 03 20). *Outline of the risk prediction model construction and validation*. Retrieved from PLOS One: https://doi.org/10.1371/journal.pone.0092549.g001

Sidana, M. (2017, February 28). *Types of classification algorithms in Machine Learning*. Retrieved from Medium: https://medium.com/@Mandysidana/machine-learning-types-of-classification-9497bd4f2e14

Sima, L. (2017, December 4). *3 Ways Predictive Analytics and Big Data Can Help Forex Brokers*. Retrieved from SmartDataCollective: https://www.smartdatacollective.com/3-ways-predictive-analytics-big-data-help-forex-brokers/

Solomatine, D., See, L. M., & Abrahart, R. J. (2008). Computational Intelligence and Technological. *Practical Hydroinformatics*, 17-29.

Talk Business. (2019, December 03). *Patterns and trading theories in the Forex market*. Retrieved from Talk Business For the Entrepreneur By the Entrepreneur: https://www.talk-business.co.uk/2017/11/30/patterns-trading-theories-forex-market/

Tunduny, T. K. (2017). A HIV/AIDS viral load prediction system using artificial neural networks. Nairobi, Nairobi, Kenya.

Uhrig, R. (1995). Introduction to artificial neural networks. *Proceedings of the 1995 IEEE IECON 21st*, (pp. 33-37).

Wanjiku, W. L. (2012, November). Forecasting foreign exchange rates in Kenya using stochastic models. Nairobi, Nairobi, Kenya.

Appendix

Appendix A: Strathmore University Ethical Approval Certificate

RHInnO Ethics - - 1 of 1

Final Decision Certificate

This document certifies that the study:

"A Model for Forex Market Price Prediction: Case of Central Bank of Kenya"

Principal Investigator: Mr. Makiya, David Nyang'au
Reference number: SU-IERC0596/19

Was reviewed and received the following status:
"approved"

Appendix B: Research License

REPUBLIC OF KENYA

NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Ref No: **783098**

Date of Issue: **16/January/2020**

RESEARCH LICENSE

This is to Certify that Mr.. David Makiya of Strathmore University, has been licensed to conduct research in Nairobi on the topic: A Model for Forex Market Price Prediction: Case of Central Bank of Kenya for the period ending : 16/January/2021.

License No: **NACOSTI/P/20/3361**

783098

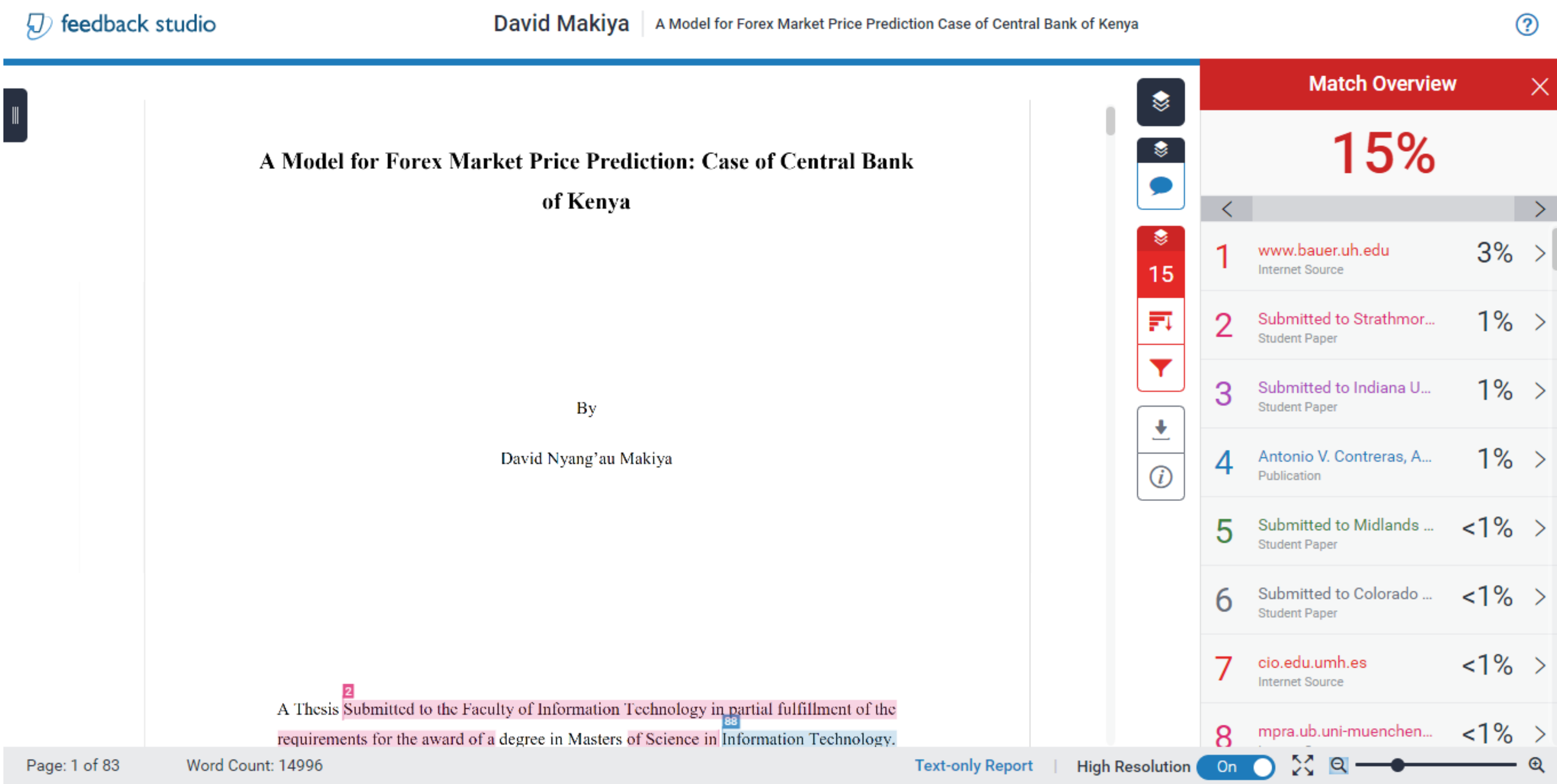
Applicant Identification Number

Director General
NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Verification QR Code

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

Appendix C: Originality Summary report



Appendix D: R Implementation Code

```
library('quantmod')

library('keras')

library('neuralnet')

library('xts')


#Reading Each currency data

Rand<-read.csv('0.csv', header = T)

USD<-read.csv('0.5.csv', header = T)

Yuan<-read.csv('1.csv', header = T)


#loading and conversion of datasets to an extensible Time_series

#for additional room for manipulation


tsR<-xts(Rand[,3],order.by = as.Date(Rand$Date))

tsU<-xts(USD[,8],order.by = as.Date(USD$Date))

tsY<-xts(Yuan[,3],order.by = as.Date(Yuan$Date))


#Checking values in sets
```

```
str(tsR)

str(tsU)

str(tsY)



# Target and Predictor feature setting

#dailyReturn

rtsR<- dailyReturn (tsR)

rtsU<- dailyReturn (tsU)

rtsY<- dailyReturn (tsY)


rtsR1<-lag(rtsR,k=1)

rtsU1<-lag(rtsU,k=1)

rtsY1<-lag(rtsY,k=1)


rtsdallR<-cbind(rtsR,rtsR1)

rtsdallU<-cbind(rtsU,rtsU1)

rtsdallY<-cbind(rtsY,rtsY1)


colnames(rtsdallR)<-c('rtsR','rtsR1')
```

```
colnames(rtsdallU)<-c('rtsU','rtsU1')

colnames(rtsdallY)<-c('rtsY','rtsY1')


rtsdallR<-na.exclude(rtsdallR)

rtsdallU<-na.exclude(rtsdallU)

rtsdallY<-na.exclude(rtsdallY)


rtsRtrain<-window(rtsdallR, end='2019-03-12')

rtsRtest<-window(rtsdallR, start='2019-03-12')

rtsUtrain<-window(rtsdallU, end='2019-03-12')

rtsUtest<-window(rtsdallU, start='2019-03-12')

rtsYtrain<-window(rtsdallY, end='2019-03-12')

rtsYtest<-window(rtsdallY, start='2019-03-12')


#Optimal Model Selection Creation test sets

gnR<-neuralnet(rtsR~rtsR1, data=rtsRtrain, hidden=c(3,2), err.fct = "sse", lifesign = 'full', rep = 5, threshold=0.01,linear.output = TRUE, stepmax = 1000000)

gnU<-neuralnet(rtsU~rtsU1, data=rtsUtrain, hidden=c(5,3), err.fct = "sse", lifesign = 'full', rep = 100, threshold=0.01,linear.output = TRUE, stepmax = 1000000)

gnY<-neuralnet(rtsY~rtsY1, data=rtsYtrain, hidden=c(5,3), err.fct = "sse", lifesign = 'full', rep = 200, threshold=0.01,linear.output = TRUE, stepmax = 1000000)
```

```
?neuralnet


plot(gnR)

plot(gnU, rep=3) #Successful on rep 3/5

plot(gnY) #Successful on rep 11/15



gnR$net.result

gnY$net.result

gnU$net.result


gnR$result.matrix

gnU$result.matrix

gnY$result.matrix


pred.pricetest<-neuralnet::compute(gnR, rtsRtest)

pred.pricetest$net.result


pred.pricetestU<-neuralnet::compute(gnU, rtsUtest[,0])
```

```
pred.pricetestU$net.result<-pred.pricetestU$net.result[,1]*1.246+102.24 #Returing from the standadized Values

pred.pricetestY<-neuralnet::compute(gnY, rtsYtest)

pred.pricetestY$net.result

write.csv(pred.pricetestU$net.result, "testU1.csv")

write.csv(pred.pricetestY$net.result, "testY.csv")

write.csv2(rtsYtest$net.result, "testact.csv")

write.csv(pred.pricetest$net.result, "testR.csv")

write.csv2(rtsRtest$net.result, "testact.csv")

TestvsActualR<-read.csv('testR.csv', header = T)

SRand<-xts(TestvsActualR[2:3],order.by = as.Date(TestvsActualR$Date))

TestvsActualU<-read.csv('testU1.csv', header = T)

USD<-xts(TestvsActualU[2:3],order.by = as.Date(TestvsActualU$Date))
```

```
TestvsActualY<-read.csv('testY.csv', header = T)

CYuan<-xts(TestvsActualY[2:3],order.by = as.Date(TestvsActualY$Date))



library(ggplot2)

library(dplyr)


str(TvA)

str(CYuan)

str(USD)


plot(TvA)

plot(CYuan)

plot(SRand)

plot(USD)

#Saving each of the trained and evaluated Neural nets


saveRDS(gnR,file="Rand.RDS")

saveRDS(gnU,file="USD.RDS")
```

```
saveRDS(gnY,file="Yuan.RDS")

actualpreditR<-data.frame(predR<-pred.pricetest$net.result, actualR<-rtsRtest[,1])

cor(actualpreditR)

actualpreditY<-data.frame(predY<-pred.pricetestY$net.result, actualY<-rtsYtest[,1])

cor(actualpreditY)

actualpreditU<-data.frame(predU<-pred.pricetestU$net.result, actualU<-rtsUtest[,1])

cor(actualpreditU)
```