UNIVERSITY OF NAIROBI

A Neural Network Model for Predicting Retail Maize Prices In Kenya

By

Mayabi, Timothy Wamalwa

P52/85998/2016

A research project submitted for the partial fulfillment for the requirements for the award of Degree of

Masters in Computational Intelligence at the

School of Computing and Informatics

Of University of Nairobi

February 2018

Declaration

I hereby declare that this thesis is my own work and has, to the best of my knowledge, not been submitted to any other institution of higher learning.

Signature: ………………………………… Date: ………………………………

Name: Mayabi Timothy Wamalwa **Registration Number:** P52/85998/2016

This research project has been submitted for examination towards fulfilment for the award of degree of Masters in Computational Intelligence with my approval as the supervisor.

Name: Dr Lawrence Muchemi
School of Computing and Informatics
University Of Nairobi

Signature: ………………………………… Date: ………………………………

DEDICATION.

This research project is dedicated to my late parents Mr and MrsLuvisia Wamalwa.

ACKNOWLEDGEMENT.

ABSTRACT.

Forecasts of food prices are intended to be useful for farmers, policymakers and agribusiness industries. In the present era of globalization, management of food security in the agriculture-dominated developing countries like Kenya needs efficient and reliable food price forecasting models more than ever. Sparse and time lag in the data availability in developing economies, however, generally necessitate reliance on time series forecasting models. The recent innovation in Artificial Neural Network (ANN) modelling methodology provides a potential price forecasting technique that is feasible given the availability of data in developing economies. In this study, the superiority of ANN over linear model methodology has been demonstrated using monthly retail price (real prices) series of maize in three major counties. I.e. Kisumu, Nairobi and Eldoret. The study also portrayed the superiority of the ANN model in its univariate form over its multivariate form. The empirical analysis has indicated that ANN models are able to capture a significant number of directions of monthly price change as compared to the linear models. It has also been observed that feeding the model with lagged observation of the same variable (univariate form) leads to more accurate forecasts than its performance in its multivariate form (Feeding it with different variables). The study has aimed at developing a user-friendly ANN prototype based on the final proposed model.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS.

LIST OF ABBREVIATIONS

AI .... Artificial Intelligence, Artificial Intelligence
ANN........ Artificial Neural Network
ARIMA .... Auto-regressive Integrated Moving Average
CRISP .... Cross Industry Standard Process
DM .... Data Mining
FAO........ Food and Agriculture Organization
GB .... Gigabytes
GDP........ Gross Domestic Product
IBM .... International Business Machines
IDE........ Integrated Development Environment
JSF........ Java Server Faces
KEBS .... Kenya Bureau of Standards
MAD .... Mean Absolute Deviation
MAPE .... Mean Absolute Percentage Error
NCPB .... National Cereals and Produce Board
RAM .... Random Access Memory
ReLU........ Rectified Linear Unit
RMSE........ Root Mean Squared Error
UI .... User Interface
US$ .... United States Dollars
VCS........ Version Control System

CHAPTER 1

INTRODUCTION

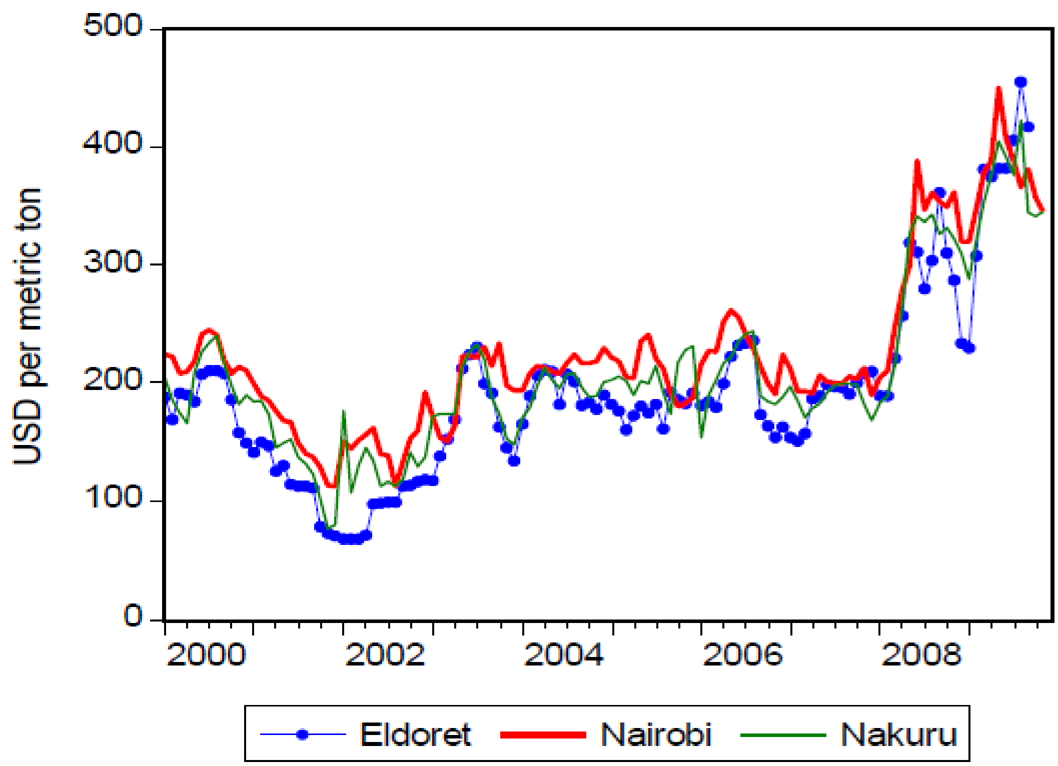
1.1. Background

Maize is the main staple food in Kenya and is an important source of calories to a large proportion of the population in both urban and rural areas. Maize consumption is estimated at 98 kilograms per person per year, which translates to roughly 30 to 34 million bags (2.7 to 3.1 million metric tons) per year and this figure has for many years served as the basis for the computation of food balance sheets and other estimates of national cereal import requirements. Maize is also important in Kenya's crop production patterns, accounting for roughly 28 percent of gross farm output from the small-scale farming sector (Jayne et al., 2001).

Kenyan policy makers have been confronted by the classic "food price dilemma." On the one hand, policy makers are under pressure to ensure that maize producers receive adequate incentives to produce and sell the crop. Rural livelihoods in many areas depend on the viability of maize production as a commercial crop. On the other hand, the food security of the growing urban population and many rural households who are buyers of maize depends on keeping maize prices at tolerable levels. For many years, policy makers have attempted to strike a balance between these two competing objectives – how to ensure adequate returns for domestic maize production while keeping costs as low as possible for consumers. Maize marketing and trade policy has been at the center of debates over this food price dilemma, including discussions over the appropriateness of trade barriers and the role of government in ensuring adequate returns to maize production. The government has pursued its maize pricing and income transfer policies through (a) the activities of the National Cereals and Produce Board (NCPB), which procures and sells at administratively determined prices, and (b) restrictions on external maize trade through a variable maize import tariff. The effects of the NCPB's activities, and government maize trade policy more generally, on maize market price levels and volatility are both controversial and poorly informed by existing analysis (Nyoro et al., 2005).

Maize grain prices in Kenya are among the highest in the Eastern and Southern Africa region. Mean wholesale market prices in the surplus zones of Nakuru and Eldoret and the capital city, Nairobi, between January 2000 and December 2009 have been US$192, US$209, and US$225 per metric ton, respectively, considerably higher than world market levels during this period (Figure 1.1). Comparing price levels in the major urban markets of Kenya, Zambia, Tanzania, Uganda, Malawi, Mozambique and South Africa, only in Malawi has mean maize prices exceeded those in Kenya (Chapoto and Jayne, 2009).

Figure 1.1 - Wholesale maize prices, nominal USD per metric ton, 2000-2009



Conventional wisdom in Kenya blames NCPB for the high maize prices, mainly for benefitting the large and politically well-connected farmers but there is little rigorous analysis to support this. Maize is therefore termed as a political crop by most Kenyans. Given the importance of maize in the Kenyan economy, empirical research on the historical effects on NCPB's activities in shaping the wholesale price movements of maize will be used in this study.

This study proposes a review of trends in price levels and forces that shape the trends as a good starting point in formulating a meaningful discussion about alternative food price policy. To further give more insight during an alternative policy formulation, the study proposes an inclusion of forecasted trends in formulating policies as opposed to the use of past and current trends only, with an aim of reducing the ramifications experienced as a result of volatile maize prices for both policy and non-policy makers. In summary price fluctuations have ramifications for every individual across the country.

Non policy makers also need to be empowered to enable them make prior adjustments involving maize and related maize products expenditure based on the predicted prices. The current popular methods of technical analysis (using trends for guessing future maize prices) that are used by non – policy makers have no pointers to the exact wholesale price for maize. A system that can guide on the most likely price for the following year and hence assist in policy formulation (for policy makers) and expenditure adjustments (for non-policy makers) is therefore missing and necessary. Such a predictive tool is therefore desirable.

It is for this reason that there is need to develop an artificial intelligence (AI) model that can be developed into a tool that can be used by both maize consumers and maize policy makers. Such a tool should not only show the price trends but also the most probable wholesale maize price for the following month. Prediction is foretelling of a future event or outcome before it occurs based on past and present data and most commonly by analysis of trends. The AI system can be able to provide long term expenditure plans on maize by providing price information for future years. Such an AI system can be based on neural networks, which are computer algorithms formulated using specific AI rules to learn from data and then be used for tasks such as prediction.

Using trending as a technical analysis, for basis of prediction is a tool that has already been employed globally through the FAO (Food and Agriculture Organization) website. The website provides information for different agricultural products for trending future prices. Research has also been done on other markets where prediction on an agricultural commodity has been attempted. (Li et al., 2010) developed a short-term price forecasting ANN model for Agro – products in China achieving an average accuracy of 95%. (Shahriary & Mir., 2015) developed an ANN model for predicting price of milk in Iran and concluded that neural networks have a higher level of prediction than ARIMA (Auto-regressive Integrated Moving Average) models.

1.2. Problem statement

In Kenya, food security has generally been taken as synonymous with maize security by policy makers and other segments of society. This is because maize is not only the main staple food but also the most common crop grown by rural poor households for food (Nyoro et al., 1999). The importance attached to maize by policy-makers in Kenya can be inferred from the emphasis laid on maize in current and past national food policies. Therefore volatility of food prices has been a problem in Kenya and such fluctuations have ramifications for every individual in Kenya. Currently non-policy makers rely on their experience, technical analysis and fundamental analysis when buying and selling maize. These methods are subjective and misleading since they are not backed by actual figures.

This study did not come across any artificially intelligent predictive tool for future maize prices in Kenya. Although FAO (Food and Agricultural Organization) provides information on maize price trends, it does not have a predictive mechanism for future maize prices in Kenya. The website provides information that points to the use of fundamental and technical analysis methods as being their basis of prediction of future wholesale maize prices. The website neither shows trends in future wholesale maize prices nor shows the actual figures of the most probable wholesale future maize prices. It is therefore desirable to have a tool that does not just point to the direction of price movement but also provides the most likely price value of whole sale maize itself. Non – policy makers that include the end consumers are the appropriate targets for such a tool since they are directly affected by volatile fluctuations in maize prices. AI methods that can actually analyze maize prices over time and gain intelligence then use this intelligence in prediction can be used to model such a tool. The predictive model shall provide information that will be a basis for consumers in making important decisions with regards to their expenditure on maize and maize related products.

1.3. Significance of Study

Currently Kenyan consumers, mostly non-policy makers, use non-AI tools which may not be effective. Additionally these current methods do not have predictive abilities. The methods mainly provide trends but not the most likely price of maize for the following month. Due to the lack of AI tools, Kenyan farmers and consumers have no option but to use their own intuition in determining the most likely prices of maize and therefore make necessary preparations in response to the anticipated outcome. By using AI models to develop tools that can also help food policy researchers' advice food policy makers in Kenya, consumers and farmers will be able to make decisions from an informed point based on actual data provided to them.

1.4. Research objective.

The research project had the following four specific objectives:

1. Investigate artificial intelligence models that are capable of food price prediction as a basis of designing a model for the Kenyan market.
2. Design an AI based model for use by both Kenyan non-policy makers and policy makers in predicting prices of maize.
3. Develop a prototype based on the design of the model.
4. Test and evaluate the performance of the model in predicting maize prices based on FAO data compiled in the five year period.

1.5. Research questions.

1. What impact does alternative staple food in the cereals category have on the models prediction capability of maize prices?
2. Does the model's prediction capability improve by supplying it with the same variable observed over a certain time period or by supplying it with different types of variables?

1.6. Assumptions.

The research made some assumptions as the basis of the study. The first assumption was that, historical monthly maize price data acquired for conducting the research has already been shaped by political activities at that time. This is because politics usually play a major role in Kenya's inflation rate. The second assumption was that maize data used in this research has already been altered by the informal, unrecorded cross-border agricultural trade activities between Kenya and her neighbors. The third assumption was that despite the maize sector's liberalization that enabled the private entities to participate in the maize market, the ministry of agriculture through NCPB is remains to be the main player in the buying and selling of maize. The fourth assumption was that maize price data has already been shaped by purchase and usage of agricultural facilities such as fertilizer, pesticides and irrigation schemes. The final assumption was that the predicted outcome is independent of possible natural calamities like floods and drought.

1.7. Limitations.

Maize prices in Kenya vary depending on the region where it is being purchased and sold. The data from various regions could be used in this study. However, to enable data consistency, the study was limited to Nairobi, Eldoret and Kisumu counties.

1.8. Structure of the project.

The first chapter introduced the topic of the Kenyan maize market. The chapter also gave the significance for the use of artificial intelligence models to build tools that both policy and non-policy makers can use as a basis of making important decisions in terms of maize production and maize trade in Kenya.

The second chapter provides a detailed review of what is happening in the sectors of maize production, trade, and consumption with an aim of linking the field of AI and agriculture. It also provides insights into research done in agricultural products prediction, the gaps identified and how predictive models can be constructed.

The third chapter provides details on the methodology that was used to design the model, develop the prototype and the performance metrics used to evaluate the model through experimentation. Data collection sources are also presented.

The fourth chapter provides the results and then reviews, interprets and analyzes these results.

The final chapter summarizes the research findings and suggests areas of further study.

CHAPTER 2

LITERATURE REVIEW.

2.1. Introduction.

Humans have always been and will always be curious about the future. A good example which is familiar especially in Kenya is prediction of a soccer game outcome through betting platforms like sport-pesa, m-cheza etc. The good advantage about this curiosity is the fact that there are possibilities of huge rewards in terms of money. Other examples include; predicting an upcoming election, and more complex scenarios like what the fate of our planet and the universe will be in the next couple of billion years. Therefore forecasting is the process of making predictions of the future based on past and present data and most commonly by analysis of trends.

A famous example of a successful prediction is by the German astronomer Johann Gottfried Galle, who predicted the existence of the previously unknown planet Neptune by calculations based on Sir Isaac Newton's law of gravity. Other predictions have become famous because of their extreme lack of accuracy. One example is when Thomas Watson, chairman of IBM, in 1943 made the following prediction: "I think there is a world market for maybe five computers." Another example is the reason why Decca Records Ltd. rejected the Beatles in 1962: "We don't like their sound, and guitar music is on the way out."

The last example highlights the value of predictions in economic situations. If a company always knew everything about the coming market trends and the company products, it would of course be easier to adapt to the market and optimize the company's strategy. Even if only partial information were available, it would be of high value.

Price forecasting is an integral part of commodity trading and price analysis. Quantitative accuracy with small errors, along with turning point forecasting power is important for evaluating forecasting models. Food production and prices are often random as they are largely influenced by eventualities and are highly unpredictable in case of natural calamities like droughts, floods, and attacks by pests and diseases. This leads to a considerable risk and uncertainty in the process of price modelling and forecasting. Based on a report obtained from

KEBS website, food is at the top of basic human need, Food prices play an important role in consumers' access to food as it has a direct impact to low income earners who will most likely spend a big percentage of their income on food.

Table 1.1: Commodity prices and CPI weight

Broad Commodity Group	**CPI Weight**	**Weight % Change on last month (April 2017/ March 2017)**	**% Change on same month of previous year (April 2017/ April 2016)**
Food & Non-Alcoholic Beverages	36.04	3.55	20.98
Alcoholic Beverages, Tobacco & Narcotics	2.06	0.04	3.26
Clothing & Footwear	7.43	0.04	4.01
Furnishings, Household Equipment and Routine Household Maintenance	18.3	0.63	2.94
Health	3.13	-0.19	3.05
Transport	8.66	-0.22	5.11
Communication	3.82	0.00	0.10
Recreation & Culture	2.25	0.03	1.99
Education	3.14	0.00	2.85
Restaurant & Hotels	4.48	1.10	5.65
Miscellaneous Goods & Services	4.52	0.11	3.55
Total	**100**	**1.79**	**11.48**

A good forecast model is thus defined by its reliability, ease of use, having an output that is meaningful, compatibility with other systems, timeliness of the forecast and reliable accuracy (Arienda, Asana, & Constantino, 2015).

2.2. Maize market in Kenya.

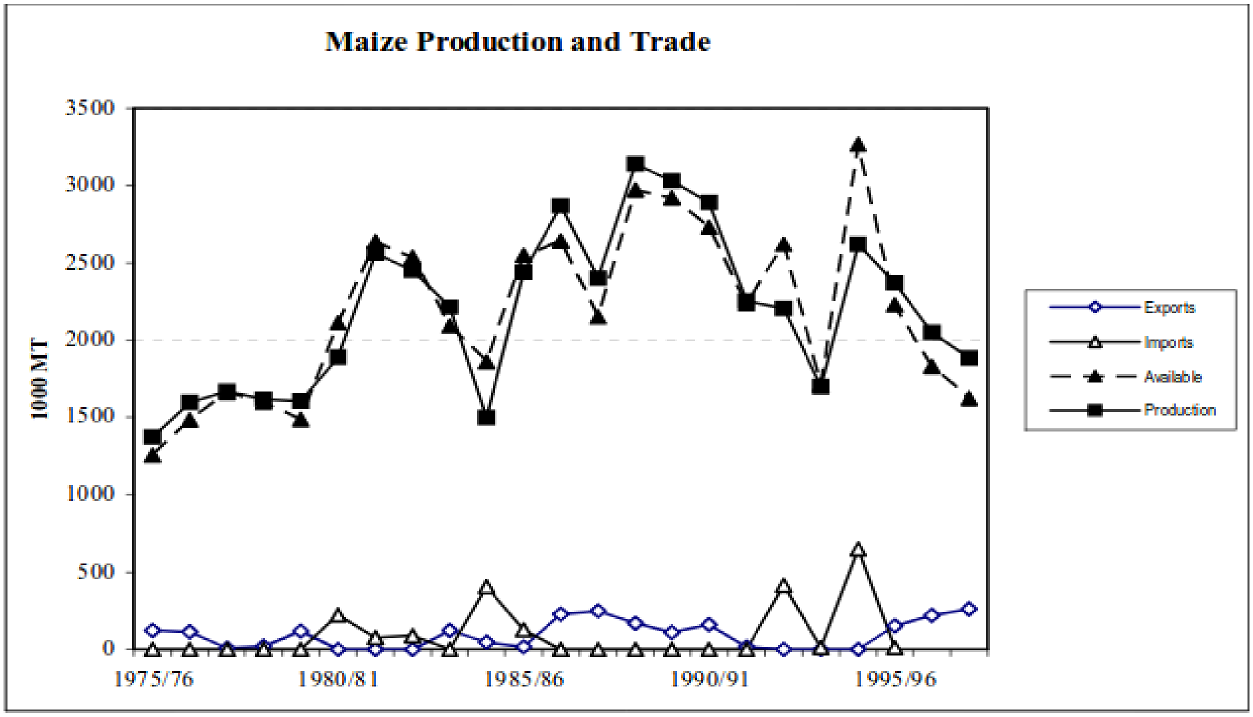
Maize is a crop that has a very healthy demand in Kenya on a daily basis. It occupies a central position in the diet of Kenyan people. In recent times, the volatility of maize prices has been a cause for concern, as it adversely affects both the producers and consumers. Producers are also affected by weather-induced volatility in production. If prices and production vary in opposite directions, then to the volatility in incomes is reduced at least to some extent.

Maize is the key food crop in Kenya, constituting 3 percent of Kenya's Gross Domestic Product (GDP), 12 percent of the agricultural GDP and 21 percent of the total value of primary agricultural commodities. (Government of Kenya, 1998).

Maize is both a subsistence and a commercial crop, grown on an estimated at 1.4 million hectares by large-scale farmers (25%) and smallholders (75%). As shown in Figure 2.1, the total average annual production of maize between 1988 and 1998 was 2.3 million metric tons fluctuating from 1.7 million metric tons in 1993/94 to 3.14 million tons in 1988/89 (Government of Kenya, 1998; Argwings-Kodhek, 1998; Nyangito, 1997). Approximately 40% of maize produced in Kenya is marketed while the balance is used for subsistence. Figure 2.1. Maize production and trade in Kenya 1976-1996. Figure 2.2 shows the main maize surplus and deficit districts of Kenya. The major maize surplus areas are in the Rift Valley Province (Nakuru, Nandi, Kericho, Uasin-Gishu and Trans Nzoia). These areas account for about 95% of the total marketed maize in Kenya. Other surplus areas include Western, Nyanza and parts of Eastern Provinces. Most arid and semi-arid lands (ASAL) of Eastern, North Eastern, Coast and Northern Rift Valley are perennial deficit areas in maize production. Figure 2.2: Maize Surplus and Deficit Districts of Kenya, 1998.

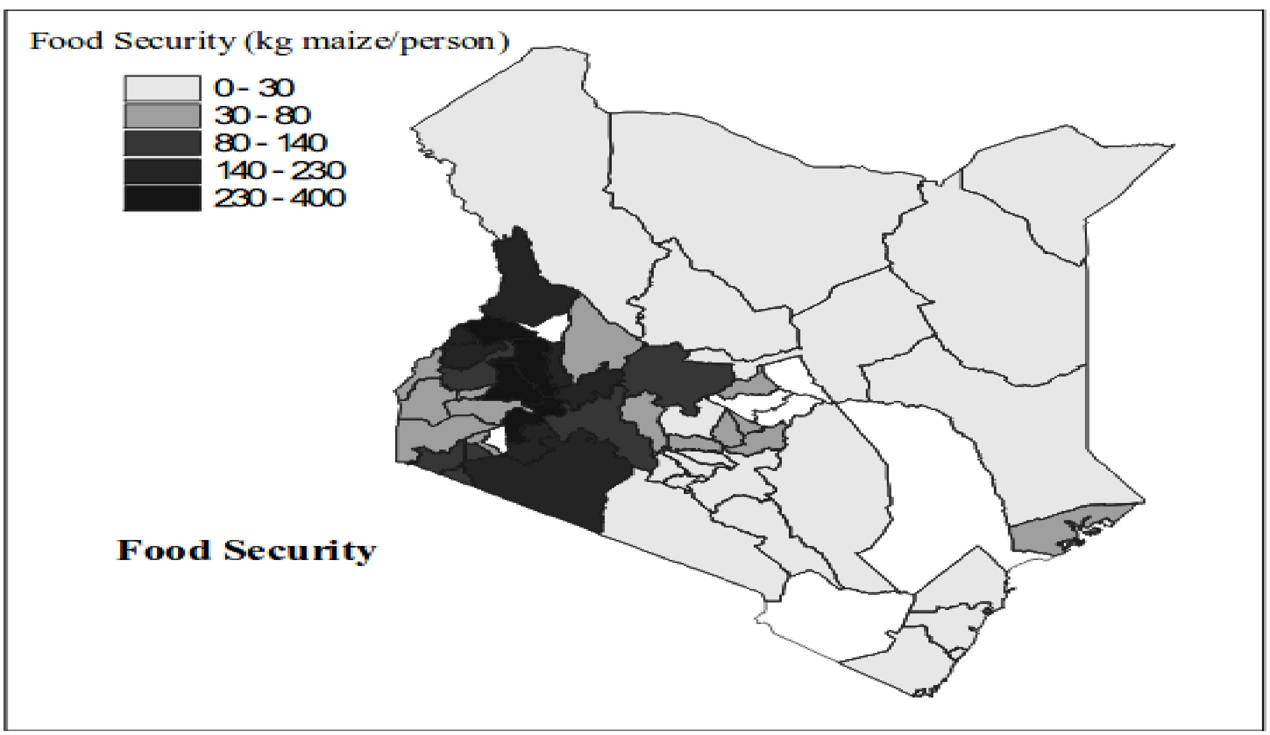
The Government strictly controlled all aspects of maize marketing until 1986 when there was a major policy shift towards liberalization that was completed in 1995. State corporations that controlled maize marketing were reduced to "buyers and sellers of last resort" and were kept for maintaining strategic reserves. The general shift in policy was a trend in Eastern and Southern Africa countries that had strictly regulated the marketing of maize.

Figure 2.1 - Maize production and trade in Kenya 1976-1996



Source: Government of Kenya, Statistical Abstracts 1976-1998.

Figure 2.2 - Maize Surplus and Deficit Districts of Kenya, 1998.



Source: Authors' design using data from Government of Kenya, Statistical Abstracts 1998.

2.2. Artificial Intelligence.

In the 21st century artificial intelligence (AI) has become an important area of research in virtually all fields: engineering, science, education, medicine, business, accounting, finance, marketing, economics, stock market and law, among others (Halal, 2003), (Masnikosa, 1998), (Metaxiotis et al, 2003), (Raynor, 2000), (Stefanuk and Zhozhikashvili, 2002), (Tay and Ho, 1992) and (Wongpinunwatana et al, 2000). The field of AI has grown enormously to the extent that tracking proliferation of studies becomes a difficult task (Ambite and Knoblock, 2001), (Balazinski et al, 2002), (Cristani, 1999) and (Goyache, 2003). Apart from the application of AI to the fields mentioned above, studies have been segregated into many areas with each of these springing up as individual fields of knowledge (Eiter et al, 2003), (Finkelstein et al, 2003), (Grunwald and Halpern, 2003), (Guestrin et al, 2003), (Lin, 2003), (Stone et al, 2003) and (Wilkins et al, 2003).

Currently there is still an influx of software that utilize elements of Artificial Intelligence (AI). Subfields of AI such as Machine Learning, Natural Language Processing, Image Processing, and Data Mining have become important for many of today's tech giants. Machine Learning is actively being used in Google's predictive search bar, in the Gmail spam filter, in Netflix's show suggestions, and in the Cleverbot chat website. Natural Language Processing exists in Apple's Siri and Google Voice. Image Processing is necessary for Facebook's facial recognition tagging software and in Google's self-driving cars. Data Mining has become a "buzz word" in the software industry due to the mass amounts of data being collected every day. Companies like Facebook and Google collect large amounts of statistics from users every second and need a way to interpret the data they receive. Artificial Intelligence has already proven to be a useful new tool in today's technology heavy culture.

Almost all of these technologies that have begun to implement facets of AI have only been around for a decade or less. Many of these aspects of AI have proven to be hugely helpful in industry, but these are merely applications of the technologies being researched. AI has greatly advanced in the last few years and there have been countless improvements within each subfield.

Alan Turing, the "father" of AI, wrote Computing Machinery and Intelligence (Turing, 1950). He attempted to answer the question "Can machines think?" by developing a type of "Imitation Game" between two subjects. This game is called the Turing Test and it involves written communication between two subjects without being able to see, hear, or otherwise sense the other subject. The first subject, a human, will attempt to figure out if the second subject is a machine or another human simply from written communique. If the first subject cannot tell, or chooses incorrectly, then Turing declared that his Turing Test proved that machines can think. However, there has been some doubt that just because a computer can respond coherently to a user questions or statement doesn't mean the computer can actually think. Does the computer really understand the meaning behind the words, or is it simply regurgitating symbols? This was somewhat addressed in Turing's paper, but more formally covered in a paper 30 years later by John Searle. In 1980, Searle published a thought experiment called The Chinese Room (Searle, 1980) that addressed the idea that the machine in the Turing Test is simply throwing symbols together without actually understanding the concepts. The Chinese Room uses the analogy of a native English speaker with no knowledge of how to speak, write, or read Chinese who is given a couple of sets of rules in English. These rules correlate input in Chinese to coherent output also in Chinese, even though the "translator" only speaks English. This question has been the topic of various research topics in Machine Learning and Natural Language Processing.

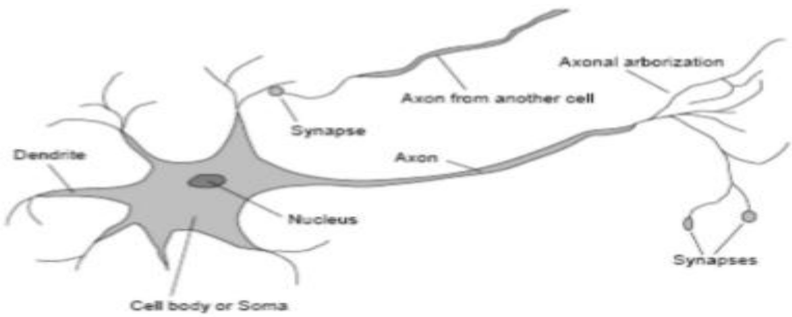
There are abundant complications when trying to create an intelligent system. Much of the old or simple AI is a list of conditions for what reaction to have based on expected stimuli. But this is arguably not intelligence, and imitating true intelligence requires an understanding of how the input relates to the output, as well as a large interdisciplinary effort among most AI subfields along with psychology and linguistics (Zeng, et al, 2009). Many complications involve Human-Machine Interaction because of the complexity of human interaction. A lot of the communication that happens between humans cannot be coded facts a machine could simply recite. There are hundreds of subtle ways that humans interact with each other that affect communication. Intonations in voices, body language, responses to various stimuli, emotions, popular culture facts, and slang all affect how two people might communicate. This is hard to model in a machine that does not have a basic common sense model already in place that can learn or make inferences.

Fuzzy Logic, which is modeled after humans' excellent ability of making approximations without any real values, poses many complications (Zadeh, 1989). Computation, by definition, require numbers and not words or concepts. Complications arise when trying to imitate human intuition or common sense. The amount of background information that is taken for granted by humans is immense and hard to replicate in machines (Havasi et al, 2009). There is difficulty in trying to imitate human emotion because of how complex and subjective they can be, especially when multiple emotions are expressed (Devillers et al, 2005), (Zeng, et al, 2009). When using a Machine Learning approach, the system will process conversations that have been labeled by humans, but these labels are not always consistent (Devillers et al, 2005)]. Image Processing also has complications with recognizing different locations from photos on the Internet because of the variability in images. Modeling the world from Internet photos is difficult because of how much the average Internet photo varies. Generally, image processing requires data to be somewhat consistent, but that obstacle will have to be overcome to render 3D models of popularly photographs locations on Earth (Snavely et al, 2008). Simply detecting what an image contains is tricky process. Handling large amounts of inconsistent data is another complication, because inconsistent data is inevitable but difficult to process. But being able to take in a large amount of data and analyze the underlying concepts would be necessary to do something like summarize a novel, which is something that is currently not possible (Zadeh, 1989). Ethical concerns arise when building a machine that can be sent into the military that could use lethal force (Sharkey, 2008). Although this is a scary concept, it has a high priority for research by the United States government force (Sharkey, 2008). Finally, using all of the subfields of AI to develop Strong AI (Artificial Intelligence equal or better than human intelligence) is incredibly complicated. Developing a system that has sentient thought would require us to fully understand how the brain and consciousness work, which we do not. There are a multitude of difficult complications within AI research. AI is a complex field, but much progress has been made in the last few years.

2.3. Artificial neural network.

Artificial neural networks (ANN), usually referred to as Neural Network (NN), is an algorithm that was originally motivated by the goal of having machines that can mimic the brain. A neural network consists of an interconnected group of artificial neurons. They are physical cellular systems capable of obtaining, storing information and using experiential knowledge. Like human brain, the ANN's knowledge comes from examples that they encounter. In human neural system, learning process includes modifications to the synaptic connections between the neurons. In a similar way, ANNs adjust their structure based on output and input information that flows through the network during the learning phase.

Figure 2.3 - Representation of human neuron.

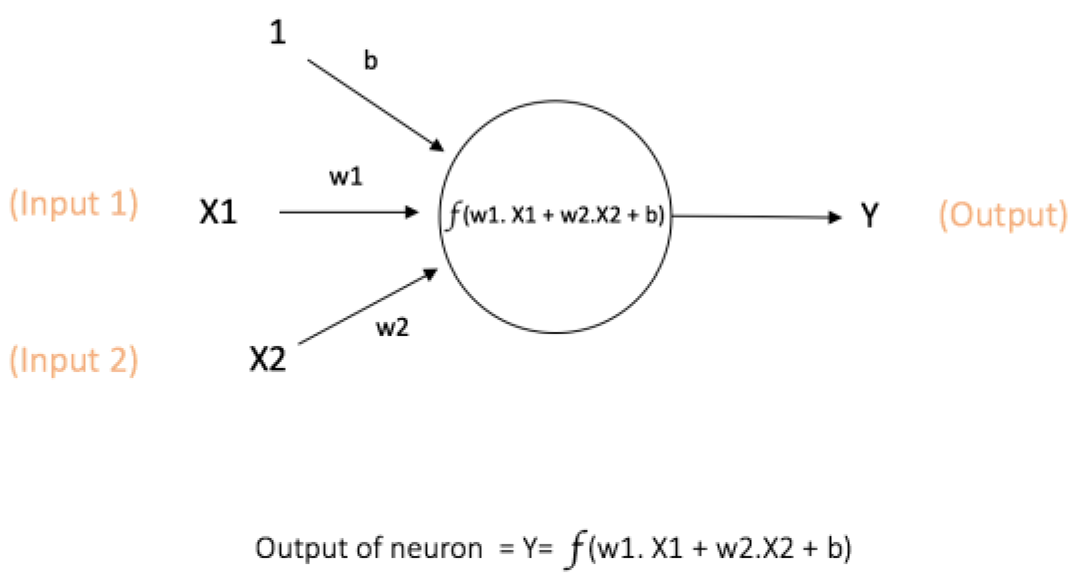


Source: (Bethard. 2008)

2.3.1. A single neuron.

The basic unit of computation in a neural network is the neuron, often called a node or unit. It receives input from some other nodes, or from an external source and computes an output. Each input has an associated weight (w), which is assigned on the basis of its relative importance to other inputs. The node applies a function *f* (defined below) to the weighted sum of its inputs as shown in Figure 2.4

Figure 2.4 - Representation of a single neuron.



Source: Author.

The above network takes numerical inputs X1 and X2 and has weights w1 and w2 associated with those inputs. Additionally, there is another input 1 with weight b (called the **Bias**) associated with it. We will learn more details about role of the bias later.

The output Y from the neuron is computed as shown in the Figure 1. The function f is non-linear and is called the Activation Function. The purpose of the activation function is to introduce non-linearity into the output of a neuron. This is important because most real world data is non-linear and it is important that neurons to learn these non-linear representations.

Every activation function (or non-linearity) takes a single number and performs a certain fixed mathematical operation on it as shown on figure 2.5. There are several activation functions that can be used. They include:

a) **Sigmoid:** takes a real-valued input and squashes it to range between 0 and 1

$$\alpha(x) = 1/(1 + \exp(\text{-x}))$$

Equation 1.1: Sigmoid function

b) **tanh:** takes a real-valued input and squashes it to the range [-1, 1]

$$tanh(x) = 2\sigma(2x) - 1$$

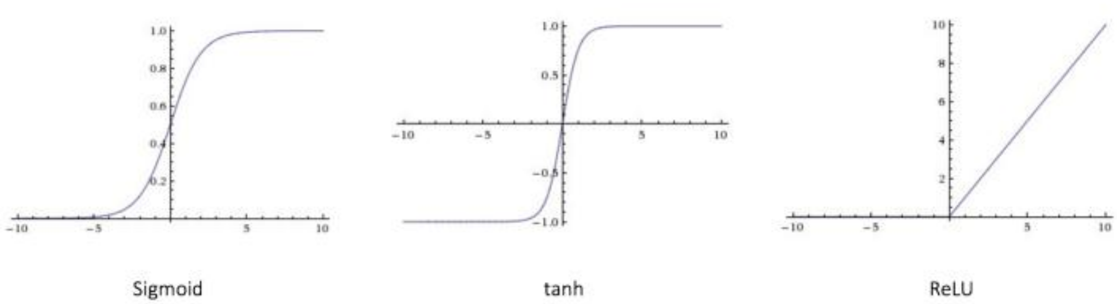
Equation 2.2: Tanh function

c) **ReLU**: ReLU stands for Rectified Linear Unit. It takes a real-valued input and thresholds it at zero (replaces negative values with zero)

$$f(x) = \max(0, x)$$

Equation 3.3: ReLU function

Figure 2.5- Activation functions



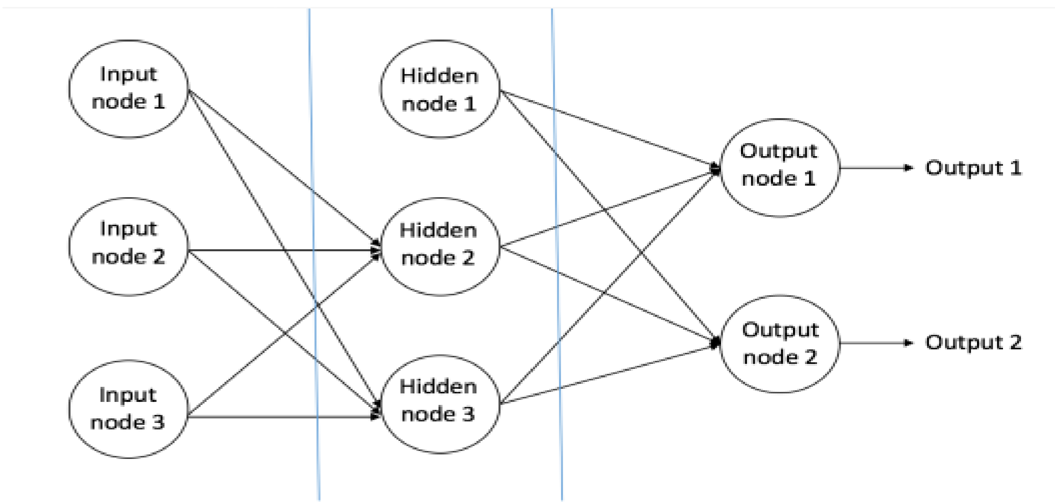
Source: Author.

The main function of Bias is to provide every node with a trainable constant value (in addition to the normal inputs that the node receives).

2.3.2. Feed Forward Neural Network

The feed forward neural network was the first and simplest type of artificial neural network devised. Figure 2.6. It contains multiple neurons (nodes) arranged in layers. Nodes from adjacent layers have connections or edges between them. All these connections have weights associated with them.

Figure 2.6 – Feed forward neural network.



Source: Author.

A feed forward neural network can consist of three types of nodes:

1. **Input Nodes –** The Input nodes provide information from the outside world to the network and are together referred to as the "Input Layer". No computation is performed in any of the Input nodes – they just pass on the information to the hidden nodes.
2. **Hidden Nodes –** The Hidden nodes have no direct connection with the outside world (hence the name "hidden"). They perform computations and transfer information from the input nodes to the output nodes. A collection of hidden nodes forms a "Hidden Layer". While a feed forward network will only have a single input layer and a single output layer, it can have zero or multiple Hidden Layers.

3. **Output Nodes –** The Output nodes are collectively referred to as the "Output Layer" and are responsible for computations and transferring information from the network to the outside world.

In a feed forward network, the information moves in only one direction – forward – from the input nodes, through the hidden nodes (if any) and to the output nodes. There are no cycles or loops in the network.

Two examples of feed forward networks are given include:

1. **Single Layer Perceptron** – This is the simplest feed forward neural network and does not contain any hidden layer.
2. **Multi-Layer Perceptron** – A Multi-Layer Perceptron has one or more hidden layers. We will only discuss Multi-Layer Perceptron below since they are more useful than Single Layer Perceptons for practical applications today.

2.3.3. The multilayer perceptron.

The neuron (or node) is the basic unit of a neural network. In the case of the MLP, it includes an input layer (that does not do any processing), one output layer and at least one hidden layer. The layers consist of a set of nodes; in the case of the hidden layer its inputs come from units in the previous layer and send its outputs to the next layer. The input and output layers indicate the flow of information during the training phase where the learning algorithm is implemented. The MLP generally learns by means of a back propagation algorithm, which is basically a gradient technique. It has also been implemented variants of the algorithm to work on the problem of slow convergence (Haykin, 1994). Once the trained process is carried out, the network weights are frozen and can be used to compute output values for new input samples. The network learning is a process in which the weights, w, are adapted by a continuous interaction (k) with the environment, in such a way that

$$w_{nj}(k+1) = w_{nj}(k) + \Delta w_{nj}(k)$$

Equation 4.4: Network learning process with weight adjustments

Where $w(k)$the previous value of the weight is a vector and $\omega(k+1)$is the updated value. The learning algorithm is a set of rules to solve the learning problem and determine the values $w_{nj}(k)$.

One of the most important algorithm is that of the error correction. Consider the n-th neuron in the iteration. Let y_n be the response of this neuron; $x(k)$ is the vector of environment stimuli and $\{x(k), k_n(k)\}$ is the pair of training. Define the following error signal equation:

$$e_n(k) = d_n(k) - y_n(k)$$

Equation 5.5: Signal Error

The objective is to minimize the cost function (criterion) which takes into account this error. After selecting the criteria, the problem of error correction learning becomes one of optimization. Consider a function $\epsilon(\omega)$, which is a continuously differentiable function of a weight vector. The function $\epsilon(\omega)$ transforms the elements from w to real numbers. We need to find an optimal solution ω^* that satisfies the condition:

$$\epsilon(\omega *) < \epsilon(\omega)$$

Then it is necessary to solve an optimization problem without constraints posed as: the cost function minimization e(ω) with respect to the weight vector. The necessary condition for optimality is given by:

$$\Delta\epsilon(\omega *) = 0$$

Equation 6.6: Cost function minimization with respect to weight

where ∇ is the gradient operator. An important class of optimization algorithms without constraints is based on the idea of iterative descent (gradient descent method and Newton's method). Starting with an initial condition $\omega(0)$, it generates a sequence $\omega(1)$, $\omega(2)$, . . ., such that the cost function $\epsilon(\omega)$decreases in every algorithm iteration. It is desirable that the algorithm converge in to the optimal solution in such away that

$$\epsilon(\omega(k+1)) < \epsilon(\omega(k))$$

Equation 7.7: Cost function optimal solution

In the descent gradient method, the successive adjustments are applied to the weight vector, in the direction of the gradient descent. For convenience, we will use the following notation:

$$\mathrm{g} = \Delta\epsilon(\omega)$$

The gradient descent algorithm can be written formally as:

$$w(k+1) = w(k) - \eta g(k)$$

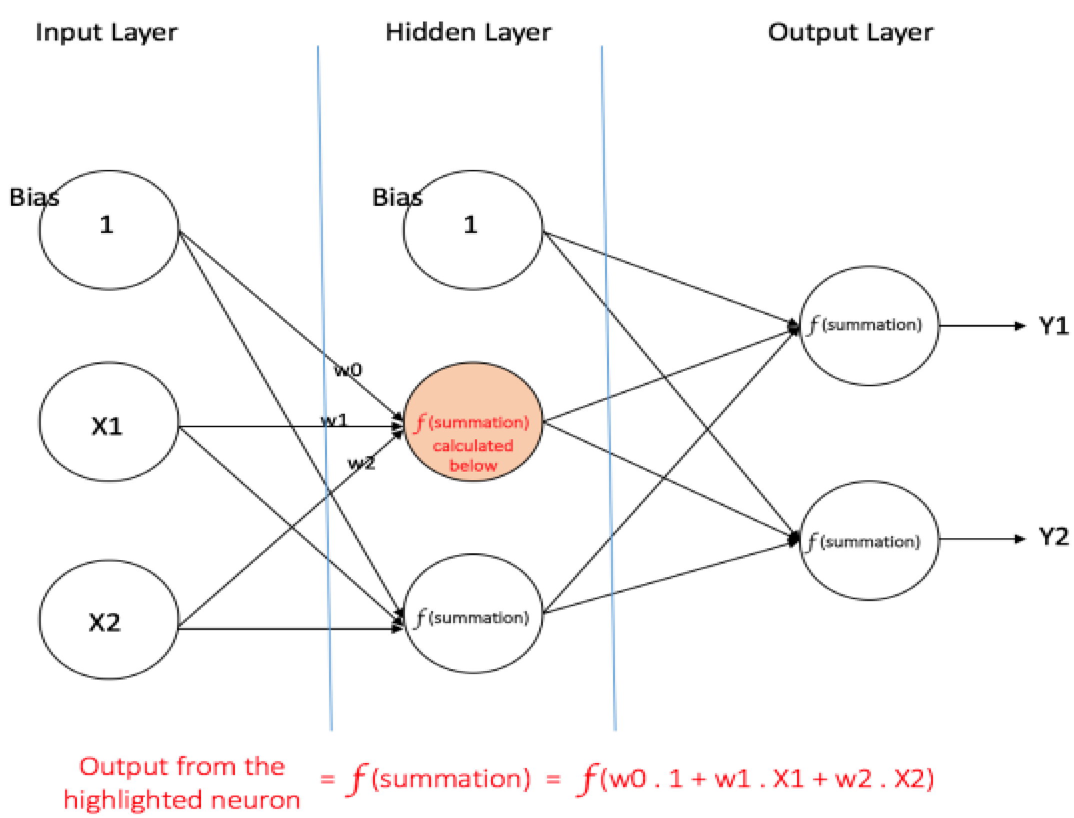
Equation 8.8: Gradient descent algorithm where η is a positive constant called the learning rate, and $g(k)$ is the gradient vector evaluated at $\omega(k)$. Therefore, the correction applied to the weight vector can be written as:

$$\Delta w(k) = w(k+1) - w(k) = -\eta g(k)$$

Equation 9.9: Weight correction on gradient vector

This method converges slowly to an optimal solution w ∗. However, the learning rate has a larger impact on this convergent behavior. When η is small, the path of$\omega(k)$, over the plane W is smooth. When η is large, the path of $\omega(k)$over the plane w is oscillatory, and when η exceeds a certain critical value, the path $\omega(k)$ over the plane W becomes unstable. Thus back propagation algorithm is a technique to implement the method of gradient descent in a weight space for a multilayer network. The basic idea is to efficiently calculate the partial derivatives of an approximate function of the behavior by the neural network with respect to all elements of the adjustable vector of parameters ω for a given value of the input vector x.

Figure 2.7 – Multilayer perceptron.



Source: Author.

2.4. Agricultural Price Prediction and Time series modelling.

Studies of different literatures reveal that grains markets are often analyzed by researchers. Based on the efficient market hypothesis, the futures price is an impartial predictor of the cash price, for a given time period (Fama 1970, 1991). Expert forecast should contain no prognostic information other than that contained in the futuristic market forecast as explained by the efficient market hypothesis. Gardner (1976) proposed to reflect the market's estimate of next period's cash price using a future market price. (Rauser and Just, 1981) described that forecasts were generally incongruent to the related future market prices.

(Garcia, Hudson, and Waller, 1988) suggested mixed evidence regarding pricing efficiency of agricultural forecasting models and if they can improve the forecast performance of future. (According to Brandt, 1985) forecasting by models or individuals can prove to be more accurate as compared to future market, hence packers and producers can benefit from this information. Using vector auto-regression (Bessler and Brandt, 1992) showed that future prices of cattle are inefficient forecasts of actual cash prices. On the other hand, expert forecast and hog futures are almost equal. (Irwin, Gerlaw and Liu 1992) when washing over a period of the first quarter of 1980 to the fourth quarter of 1991, agreed that they did not find many significant variations between the US department of agriculture and the forecast accuracy of live cattle futures prices and live hog. (Kasten, Jones and Schread, 1998) forecasted the five competing commodities i.e. feeder cattle, slaughter steers, cull cows, slaughter hogs, and sows over the period of 1987-96. (Zulauf and Irwin, 1997) discussed that available data on efficient market is mainly consistent with individual-generated forecasts.

Agricultural price modelling in general is different from modelling of non-farm goods and services due to certain special features of agricultural product markets. The characteristic features of agricultural crops include seasonality of production, derived nature of their demand, and price-inelastic demand and supply functions. The biological nature of crop production plays an important role in agricultural product price behavior. (Girish and Kanchan, 2013).

There are two basic approaches of forecasting, namely structural and time series models. The structural models proceed from the first principles of consumer and producer theory to identify the demand and supply schedules and the equilibrium prices resulting from their intersection. The structural modelling techniques provide valuable insights into the determinants of commodity price movements. The computational and data demands of structural price forecasting generally far exceed than what are routinely available in the developing countries. Consequently, researchers often rely on parsimonious representations of price processes for their forecasting needs. Contemporary parsimonious form of price forecasting relies heavily on time series modelling. The time series modelling requires less onerous data input for regular and up-to- date price forecasting.

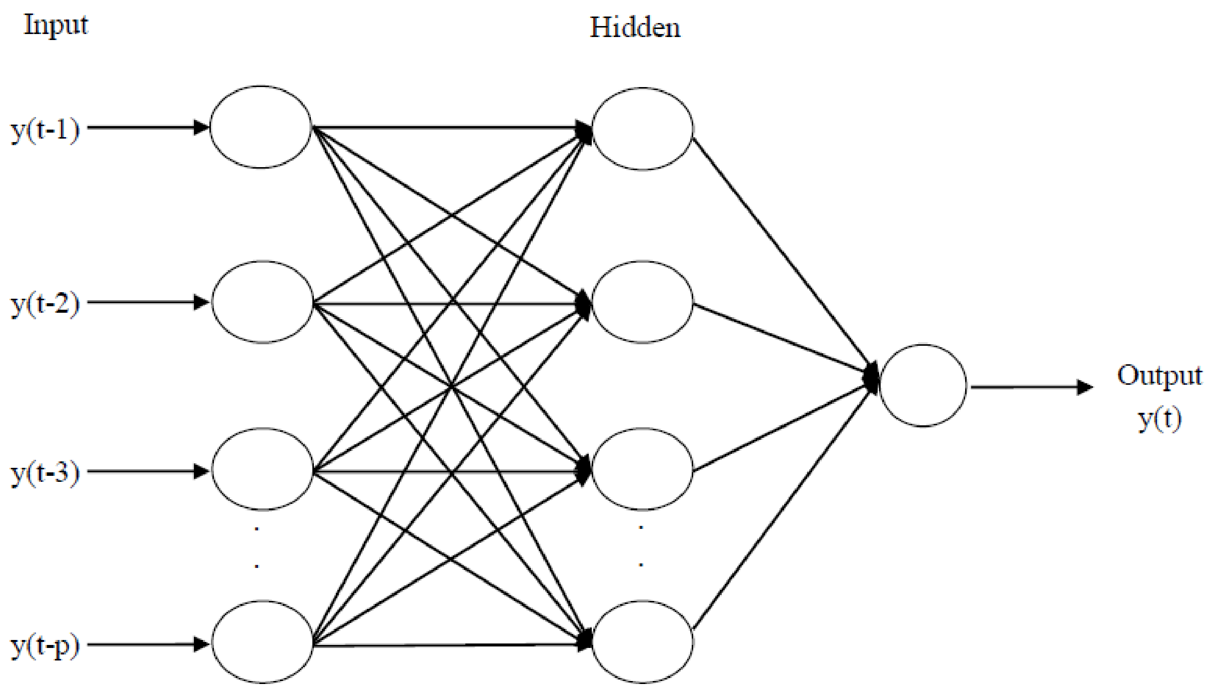
In time series modelling, past observations of the same variable are collected and analyzed to develop a model describing the underlying relationship. During the past few decades, much effort has been devoted to the development and improvement of time series forecasting models. One of the most important and widely used time series models is the Auto-Regressive Integrated Moving Average (ARIMA) model. The popularity of ARIMA model is due to its statistical properties as well as use of well-known Box-Jenkins methodology in the model building process. Recently, Artificial Neural Network (ANN) modelling has attracted much attention as an alternative technique for estimation and forecasting in economics and finance (Zhang et al., 1998; Jha et al., 2009). ANN is a multivariate non-linear non-parametric data driven self-adaptive statistical method. The main advantage of neural network is its flexible functional form and universal function approximator. With ANN, there is no need to specify a particular model form for a given data set. ANN has found applications in fields like biology, engineering, economics, etc and its use in economics has been surveyed by (Kuan and White 1994).

2.5. Conceptual model.

Review of the literature enabled the research to develop the conceptual model shown in figure 2.5. This research is optic to predict the prices of retail maize prices using ANN. The aim of this paper is not to prove that ANN methods of prediction are best, but it is to show that they are applicable to daily life problems. This research also seeks to provide answers to the research questions raised in chapter one.

The time series data can be modelled using ANN by providing the implicit functional representation of time, where-by a static neural network like multilayer perceptron is bestowed with dynamic properties (Haykin, 1999). Further literature review suggested that a neural

network can be made dynamic by embedding either long-term or short-term memory, depending on the retention time, into the structure of a static network. One simple way of building short–term memory into the structure of a neural network is through the use of time delay, which can be implemented at the input layer of the neural network. An example of such an architecture is a Time-Delay Neural Network (TDNN) which has also been used by (Girish and Kanchan, 2013).
Figure 2.8 – Conceptual model.



Source: Author

2.6. Chapter summary.

Chapter provides a detailed review of how forecasting has been previously used in other fields and what is happening in the sectors of maize production, trade, and consumption with an aim of linking the field of AI and agriculture. It also provides insights into research done in agricultural products prediction, the gaps identified and how predictive models can be constructed.

CHAPTER 3

RESEARCH METHODOLOGY.

3.1. Introduction.

Research is defined as a fact finding activity that involves a scientific investigation or thorough study of a given subject matter of particular interest. A research can be exploratory, descriptive or diagnostic in nature and thus qualitative or quantitative approaches are applied as per the research design. Research has been proved to be a vital tool that provides the basis for economic decision making by government institutions and policy makers (Mackey and Gass, 2013).

This project used both exploratory and applied research to create a specific artificial intelligence tool based on a model and tested its performance on a practical problem. The exploratory part of the research was attempting to identify new insights into the possibility of incorporating more diverse predictors for maize price forecasting. The predictors used are variables which were identified to be in the staple food category from the literature review. Chapter 2 indicated the need for such a tool that can be used by both policy and non-policy makers in the field of agriculture in Kenya. Therefore this research applied a particular underlying AI method to a practical field, in this case, the maize market prices in Nairobi, Eldoret and Kisumu counties. Applied research aims at solving a 'societal or business' problem (Kothari, 2004).

3.2. Research Design.

Research design is a work plan detailing how the research was be undertaken, type of data to be collected, tools and techniques that was be employed to obtain data and the method of data analysis to be used (Wyk, 2012). The research design adopted for the study was an exploratory-quantitative research approach (Mackey and Gass, 2013).

Exploratory research seeks to help identify new hidden factors that might be found in data and their significance to the study under research. This type of research approach creates a foundation for further future in-depth research from a scientific point of view with the intent of finding new insights and ideas about the topic under study. The exploratory design was adopted to investigate the performance of the neural network model when supplied with a lagged input of the same variable (historical monthly maize prices) over time compared to supplying it with different variables (prices of other staple foods, precipitation, inflation rate and maize production data) when predicting maize price of the next month.

An AI model was designed based on the artificial neural network algorithm which was then developed into a working prototype for the purposes of testing. The model employed the use of a feed forward ANN with multilayer perceptron using back propagation and trained using supervised learning. The programming language environment was Java 8. Other java components included the java server faces (JSF) framework, neuroph library and primefaces library with ajax components to provide some user interface (UI) capabilities.

3.2.1. Research Data.

For purposes of designing and evaluating the model, the research needed historical monthly maize price data (data from Nairobi, Kisumu and Eldoret counties), country wide yearly price data for agricultural products identified as staple foods in this study and yearly maize production data.

3.2.2. Data Sources.

Data collection method was not experimental and included collection of secondary data from relevant publications and the internet. The Food and Agriculture Organization (FAO), a specialized agency of the United Nations that leads international efforts to defeat hunger provided monthly maize price data from Nairobi, Kisumu and Eldoret counties through their website. Country wide yearly price and production was obtained from the knoema website, a free to use public and open data platform for users with interests in statistics and data analysis. Data was later imported into MySQL database for ease of retrieval during model development and testing.

3.2.3. Sample Split.

This involved identification of the data to be used in training the model, test data to test model training and the validation data to measure the output error.

3.2.4. Model Training

This process involved giving inputs to the model for processing in order to train the model on the type of input data and the expected output of the training session. The training data was fed into the ANN model via the identified model neurons. The training process was done over a number of iterations between 5000-50000 with each iteration targeted at reducing the error rate and adjusting the input weights.

3.2.5. Model Testing and validation.

This process involved the use of a test data set to check whether the system was properly trained by observing the actual model output versus the expected output. By using the validation data set, any disparities in the output captured by error performance measures were used to adjust the weights of the neurons for purposes of fine tuning the model.

3.2.6. Volume of data.

The volume of data implied a period of twelve years for the monthly maize price data and twenty four years for the yearly price data of other staple foods. In other studies of agricultural products prediction, other researchers considered periods that varied from 10 years to 30 years. Girish and (Kanchan, 2013) did a test over a thirty year period while (Shahriary and Mir, 2015 did a test over 14 year period. Due to these considerations, the project decided on a twelve years for the monthly maize price data and twenty four years for the yearly price data of other staple foods.

3.2.7. Performance measure.

The performance of the ANN tool was analyzed based on its accuracy in maize prices for a continuous range of dates beyond the last date of its training. Training effectiveness was determined using the root mean square error (RMSE), over the range of training cycles. RMSE was also used to compare the performance of model, where a lower comparative RMSE value would imply a better prediction. However, the testing phase was measured on the basis of mean absolute deviation error (MAD), to determine exactly how far the actual and predicted values were. Results with lower MAD were more close to them actual values. In other research, (Girish and Kanchan, 2013) also based the measurement of the performance of their model on the selected crops (soybean and rapeseed-mustard) on the basis of MAD and RMSE while (Shahriary and Mir, 2015) based the performance measurement of their model on livestock milk using MAD and RMSE The formula for RMS error (RMSE) is given by Equation 3.1 while that for MAD is given by Equation 3.2 below:

$$RMSE = \sqrt{\frac{\sum_{i=0}^{i}(y'_i - y_i)^2}{n}}$$

Equation 3.1: Root mean squared error

Where:

n = number of observations

y'i= predicted value

yi= actual value

$$MAD = \frac{\sum_{i=0}^{i}(y'_i - y_i)}{n}$$

Equation 3.2: Mean absolute deviation

Where:

n = number of observations

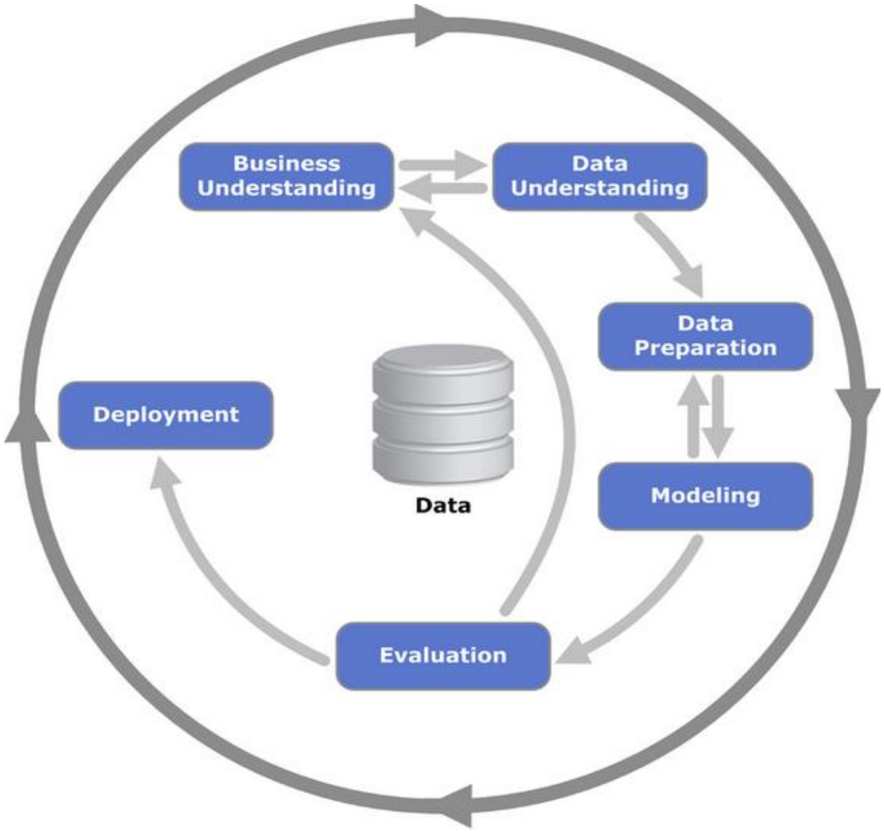
y'i= predicted value

yi= actual value

3.3. Designing the proposed model.

The system design approach for developing the model was the CRISP-DM methodology.
It incorporates six design phases that comprehensively cover the model development process. As used by (Cortez, 2010), the methodology works well with Artificial Neural Network for predictive purposes. The Figure 3.1 illustrates the development cycle of the methodology.

Figure3.1 - CRISP-DM Methodology



Source (Wirth & Hipp, 2000).

3.4. Chapter summary.

Chapter 3 provides details on the methodology that was used to design the model, develop the prototype and the performance metrics used to evaluate the model through experimentation. Data collection sources are also presented.

CHAPTER 4

MODEL ANALYSIS AND DESIGN.

4.1. Introduction.

The process of model analysis and design encompasses defining the model parameters, data understanding and pre-processing for use by the model algorithm. This is captured by defining the model requirements analysis. It is a process that describes the process of studying and developing the business and the user needs to arrive at a definition of the problem domain and model requirements. It is the most critical aspect of the study and determines the goals and functions of the developed model (Dennis, Wixom, & Roth, 2012).

4.2. Data understanding.

The model involved feeding it with the same input variable and multiple variables in order to evaluate its results based on two scenarios that included: model performance when supplied with the same variable with a particular time lag over a period of time, model performance when supplied with multiple variables over a period of time. Therefore the model performance was evaluated both in its univariate and multivariate state. The univariate data involved retail maize price (real) data while the multivariate data involved staple food prices which are an alternative to maize, Kenya's precipitation levels and inflation rate.

4.3. Data preparation.

This phase covers all activities that will be involved in constructing the final dataset (data that will be fed into the model) from the initial raw data. The step also involved deciding on the best method of data retrieval for the model and data partitioning for both univariate and multivariate scenarios of the model. Data partitioning involved segmenting the sample data into two sets for purposes of training and testing the model. The training set constituted the largest portion of the sample data and was used by the ANN model to learn the patterns present in data. 70% of the data was used as the training set while the remaining 20% was used as the testing set. Thus the sample data was partitioned as follows:

1. Training set = 70% of total sample size.
2. Testing set = 30% of total sample size.

4.3.1. Data preprocessing.

The data preprocessing refers to analyzing and transforming the input and output variables to minimize noise, highlight important relationships, detect trends, and flatten the distribution of variables to assist both traditional and neural network models in the relevant pattern. The first step in time series analysis is to plot the data. Data transformation is crucial in ANNs for achieving a good prediction performance by removing the bias and correlations between the inputs and making them statistically independent (Oancea & Ciucu, 2014). Previous studies have concluded that data transformation through normalization speed up training time by starting the training process for each feature within the same scale. It is especially useful for modelling application where the inputs are generally on widely different scales (Jayalakshmi & Santhakumaran, 2011).

The normalization approach was by decimal scaling that involved moving the decimal point of values of attribute A. The number of decimal points moved depends on the maximum absolute value of A. A value v of A is normalized to v’ by computing: v’ = (v / 10 power j) where j is the smallest integer such that Max(|v’|)<1. (Saranya & Mandikandan, 2013).

4.4. Designing the proposed model.

The model design describes the organization of the ANN and defined the number of input neurons. The input neurons were used to capture each independent variable. Model design also involved determining the number of ANN hidden layers, the number of neurons in the hidden layer, number of output neurons and the ANN activation function. The first step was to be the formulation of a baseline model as per the proposals by other researchers. The third step would then be the development of the new proposed model based on the baseline model after change of parameters as determined by experimentation.

4.4.1. Formulating the baseline model.

The model needed to be dynamic in a way that it can be trained in both univariate and multivariate modes. Therefore the model input varied depending on the training mode.

The two modes (univariate and multivariate) of the baseline model were developed for experimental purposes so as to evaluate the performance of the model and help in determining the best mode so as to come up with a new model. Based on other research, the hidden neurons should be 1 layers. (Girish and Kanchan, 2013) . The first experiment was done on the model in its univariate mode using real prices of monthly maize data from 2006 to 2018. The baseline model, with a configuration of 4:8:1 was subjected to training using 70% data (Jan-2006 to Apr-2018) and the balance for testing. The number of training repetitions was set at 50,000 maximum, after which the training was stopped. The results obtained for each of the monthly prices is shown in Appendix 7. The figure 4.2 below shows the graph of MAPE for the monthly price data. The second experiment was done on the model in its multivariate mode. The data used included yearly wheat price, rice price, rainfall, inflation rate and maize production data from 1992 to 2016. The baseline model, with a configuration of 5:8:1 was subjected to training using 70% data and the balance for testing. The number of training repetitions was set at 5,000 maximum, after which the training was stopped. The results obtained for each of the monthly prices is shown in Appendix 8. The figure 4.3 below shows the graph of MAPE for the monthly price data.

4.4.1.1. Univariate baseline model.

In its univariate mode the model comprised of input of 4 monthly maize prices is used with the aim of predicting the 5th price in the series i.e. 4 inputs and 1 output.

The baseline model was therefore formulated with the following configuration:

a) Number of inputs = 4

b) Number of hidden layers = 1

c) Number of neurons per hidden layer shall = 8

d) Number of outputs = 1

e) Bias per layer = 1

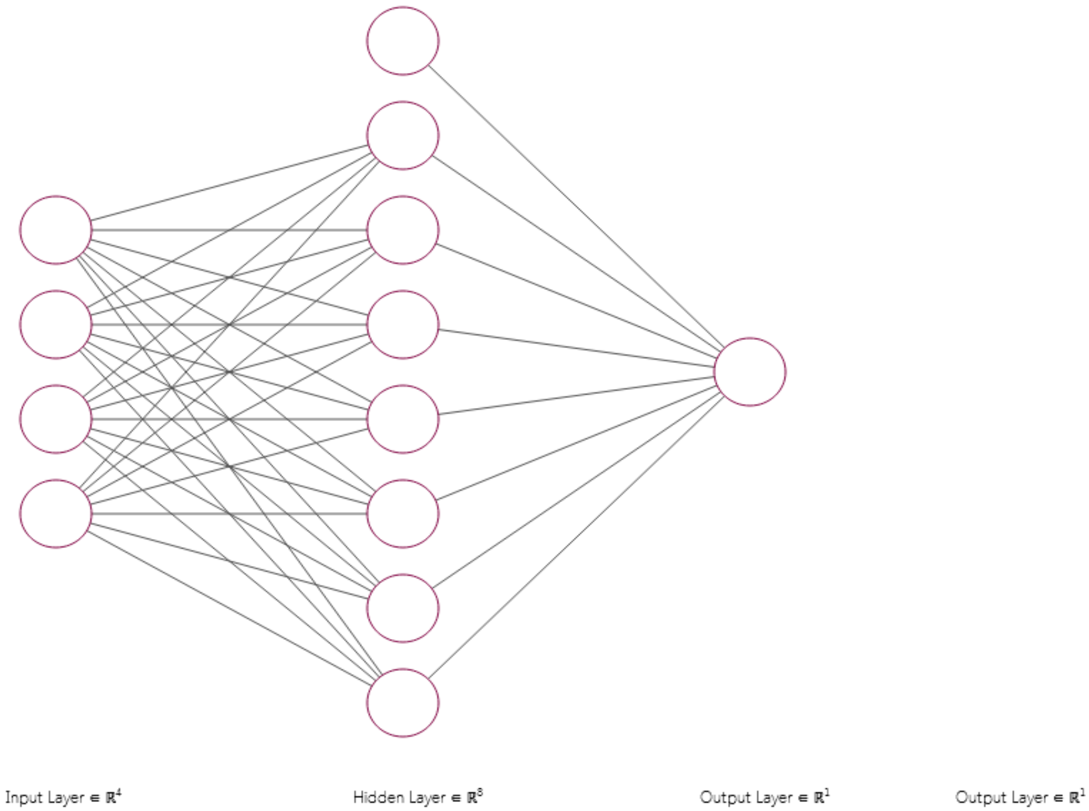
f) Transfer function = sigmoid.

g) Maximum error = 0.01

h) Learning rate = 0.5

g) Maximum iterations = 50000

Figure 4.1 - ANN univariate baseline model



Source: Author.

4.4.1.2. Multivariate baseline model.

In its multivariate mode the model comprised of input of 5 variables that included yearly rainfall amount (mm), wheat price, rice price, inflation rate and maize production (tones) with the aim of predicting maize price per tonne for a particular year i.e. 5 inputs and 1 output.

The baseline model was therefore formulated with the following configuration:

a) Number of inputs = 5

b) Number of hidden layers = 1

c) Number of neurons per hidden layer shall = 8

d) Number of outputs = 1

e) Bias per layer = 1

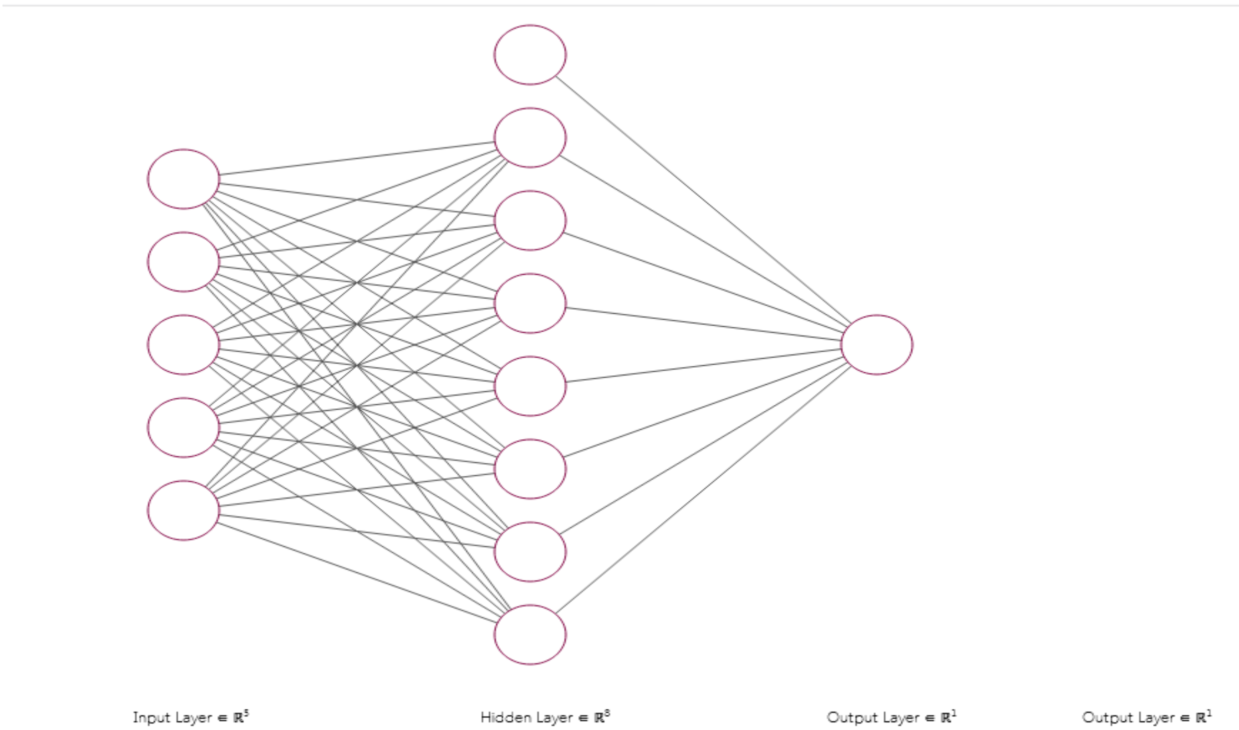
f) Transfer function = sigmoid.

g) Maximum error = 0.01

h) Learning rate = 0.5

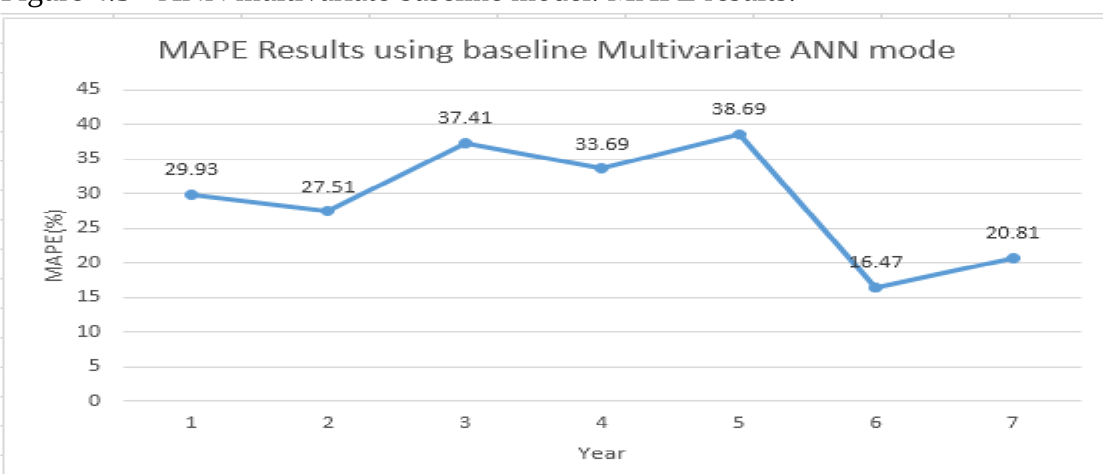
g) Maximum iterations = 50000

Figure 4.2 - ANN multivariate baseline model



Source: Author.

Figure 4.3 - ANN multivariate baseline model. MAPE results.



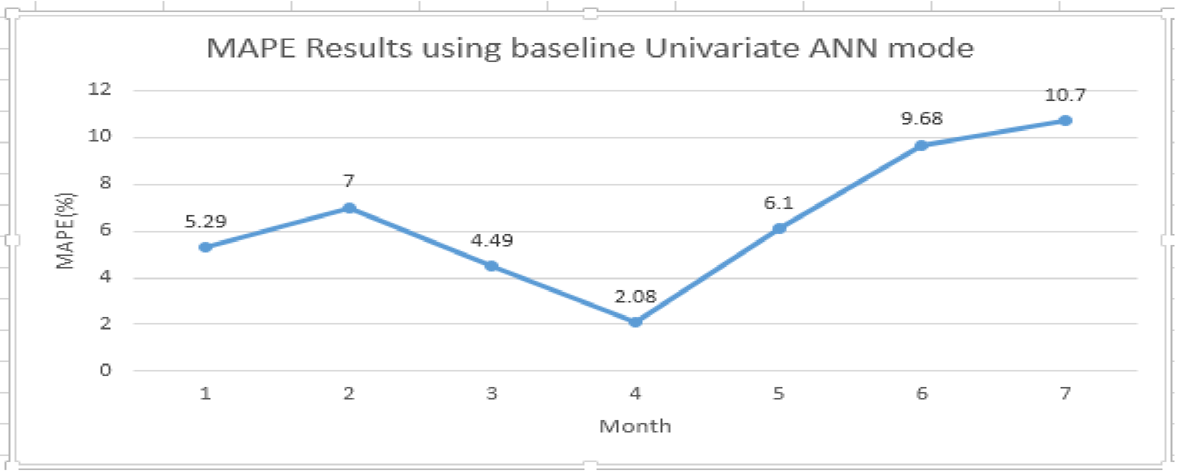
Source: Author.

Total RMSE = 3234.629086

Network MAPE = 9.928903%

Network MAD = 2610.453578

Figure 4.4 - ANN univariate baseline model. MAPE results



Source: Author.

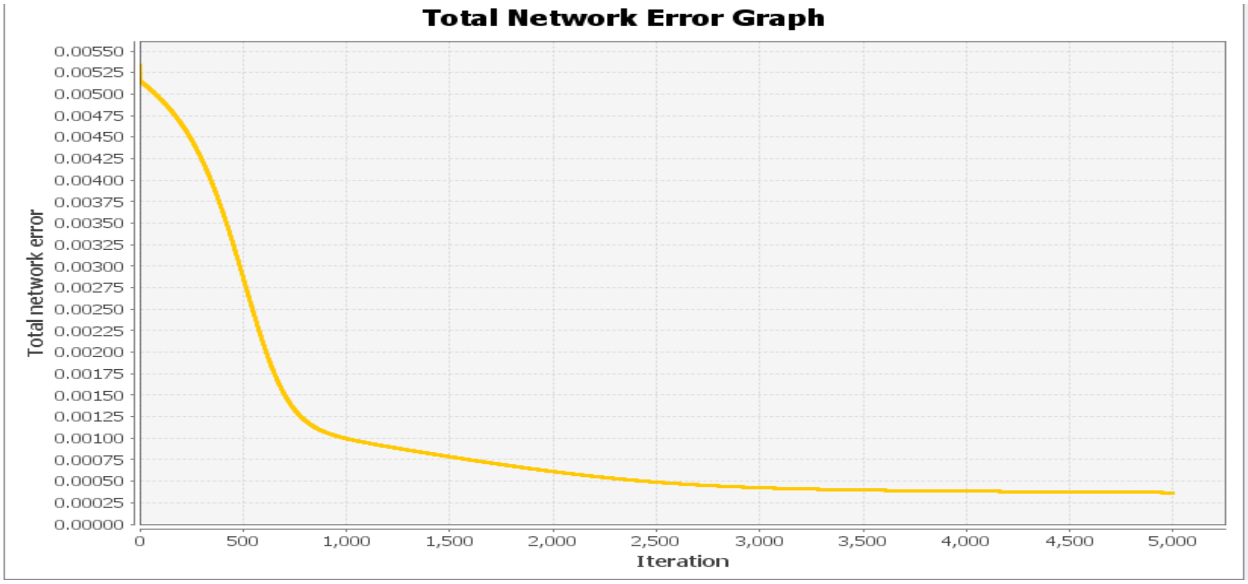
Total RMSE = 2.993883

Network MAPE = 7.342249%

Network MAD = 0.64954

The study preferred to use the ANN model in its univariate mode to develop the actual model based on the RMSE, MAD and MAPE results obtained during the formulation of the baseline model.

Figure 4.5 - ANN network error for different iterations



Source: Author

4.5. Model Parameters.

4.5.1. Number of Hidden Layers and Neurons.

The ANN structure for a particular problem in time series prediction includes determination of number of layers and total number of nodes in each layer. It is usually determined through experimentation as there is no theoretical basis for determining these parameters. It has been proved that neural networks with one hidden layer can approximate any non-linear function given a sufficient number of nodes at the hidden layer and adequate data points for training. In this study, we have used neural network with one hidden layer. In time series analysis, the determination of number of input nodes which are lagged observations of the same variable plays a crucial role as it helps in modelling the autocorrelation structure of the data. In this study, one output node has been used. The univariate model was initiated with a baseline of settings 4:8:1, using 70% data for testing at 50,000 training cycles. This baseline model was then tuned, through experiment, to determine an optimum number of neurons on the hidden layer. The number of neurons on the hidden layer was progressively adjusted, followed by a series of training and testing phases. This test kept the number of hidden layers fixed as 1. The inputs and output were also fixed at 4 and 1 respectively. The data volume was 70% for training (2006 to 2018) and

30% test data i.e. The results obtained for each setting are shown on Table 4.1 below, based on raw data shown on Appendix 8.

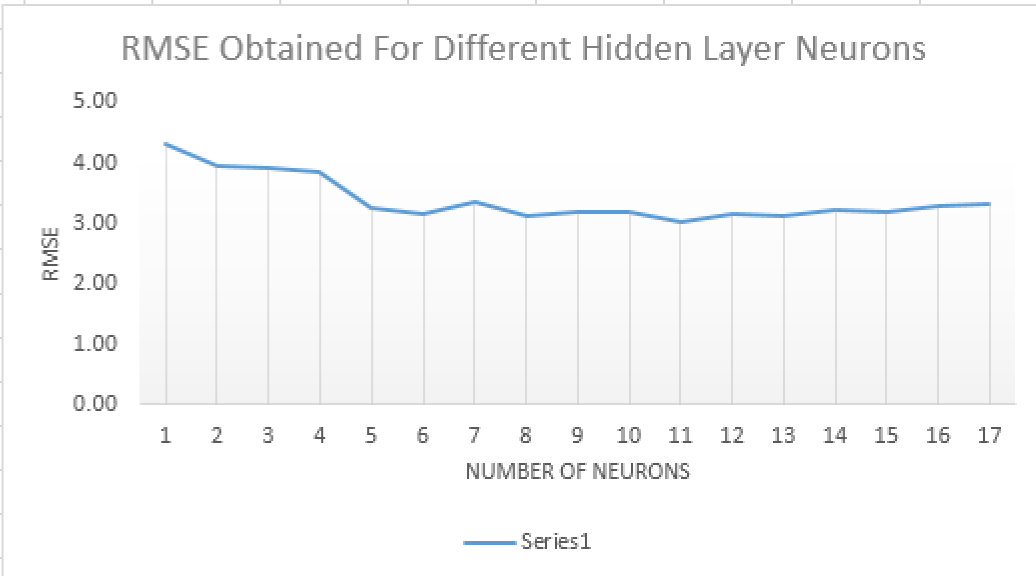
Table 4.1: Results for varying number of hidden neurons

Input Neurons	Hidden Layer (No of neurons)	Output Layer(No of neurons)	MAD	MAPE(%)	RMSE
4	1	1	0.968536	10.907103	4.303911
4	2	1	0.868631	9.676098	3.962422
4	3	1	0.84744	9.39759	3.908902
4	4	1	0.82501	9.110555	3.851209
4	5	1	0.71868	8.150653	3.266547
4	6	1	0.689011	7.805539	3.143046
4	7	1	0.737323	8.372456	3.348758
4	8	1	0.68232	7.725737	3.118642
4	9	1	0.695783	7.883295	3.175013
4	10	1	0.699363	7.926704	3.188698
4	11	1	0.658314*	7.445803	3.026113
4	12	1	0.690763	7.829555	3.153427
4	13	1	0.680143	7.698813	3.111811
4	14	1	0.707742	8.026973	3.225744
4	15	1	0.699848	7.93225	3.190799
4	16	1	0.721472	8.187287	3.284805
4	17	1	0.730329	8.293067	3.323696

*Lowest value

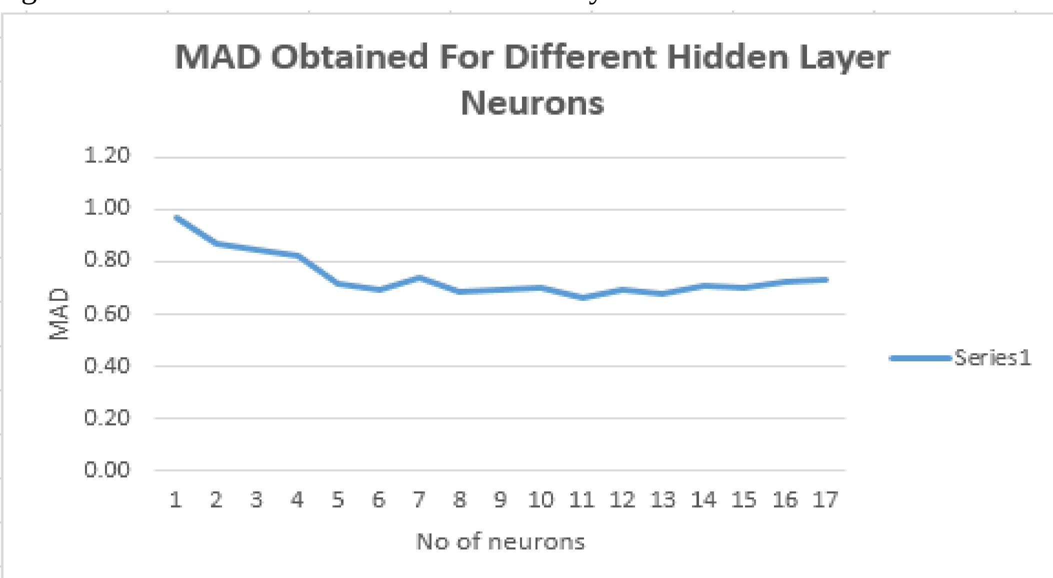
These results are also shown graphically on Figure 4.6 and Figure 4.7 below.

Figure 4.6 - RMSE for neurons in the hidden layer.



Source: Author

Figure 4.7 - MAD for neurons in the hidden layer.

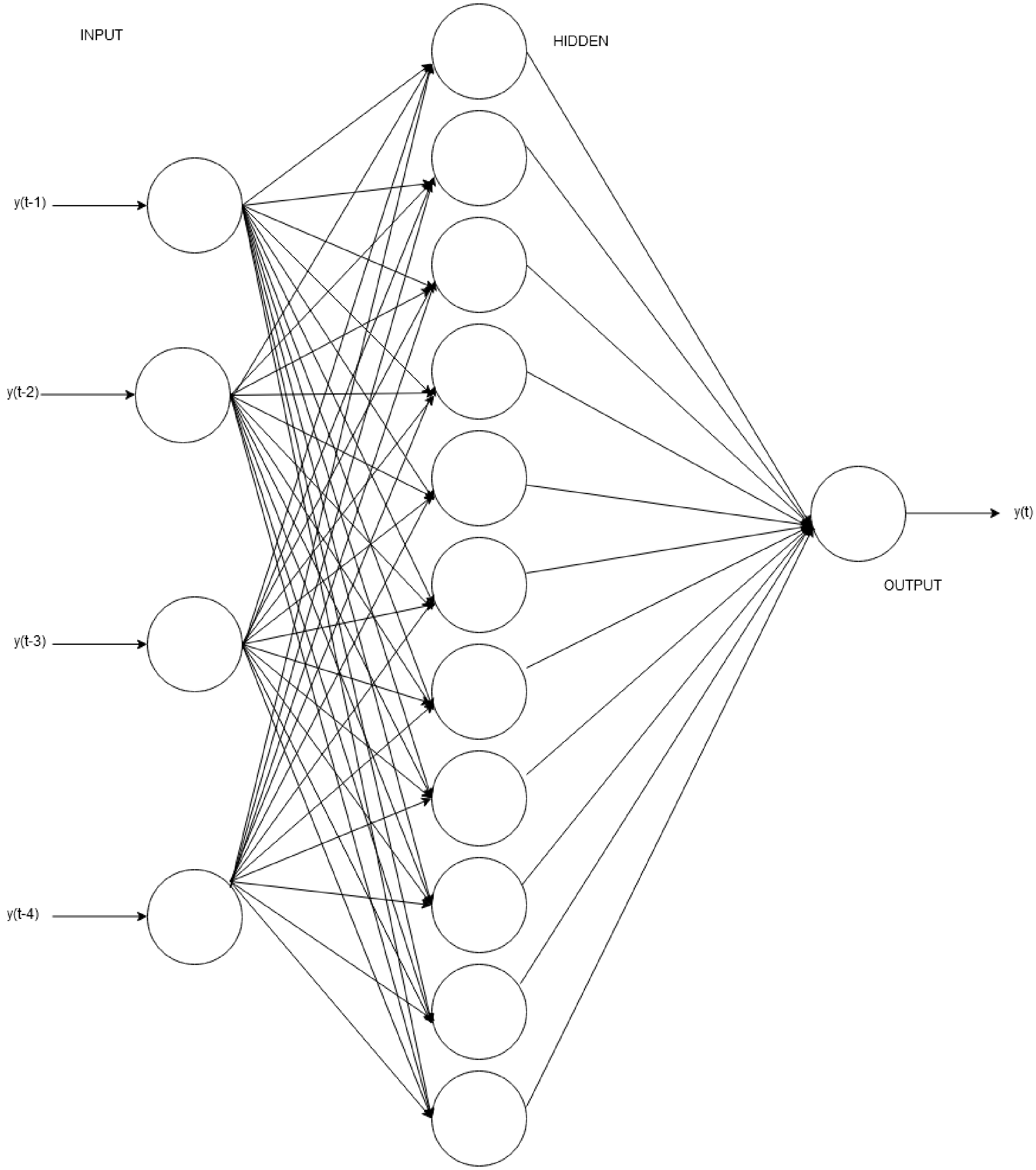


Source: Author.

The following two observations were made – firstly, the low rates of MAD was obtained with neurons per hidden layer of 8, 13 and 11. Secondly, based on MAD, the best configuration was obtained for 11 neurons per hidden layer i.e. 4:11:1. Based on this determination, the configuration of 4:11:1 was therefore the optimal ANN configuration.

The new developed model and its flow chart diagram are shown in Figure 4.8 and 4.9.

Figure 4.8 - Final proposed model.



Source: Author.

The final model implements a time delay approach whereby the output is determined by lagged observations of the last four months in the series. The general expression for the final output value y_{t+1} is given by equation 4.1

$$y_{t+1} = g[\sum_{j=0}^{q} \alpha_j \ f(\sum_{i=0}^{p} \beta_{ij} y_{t-i})]$$

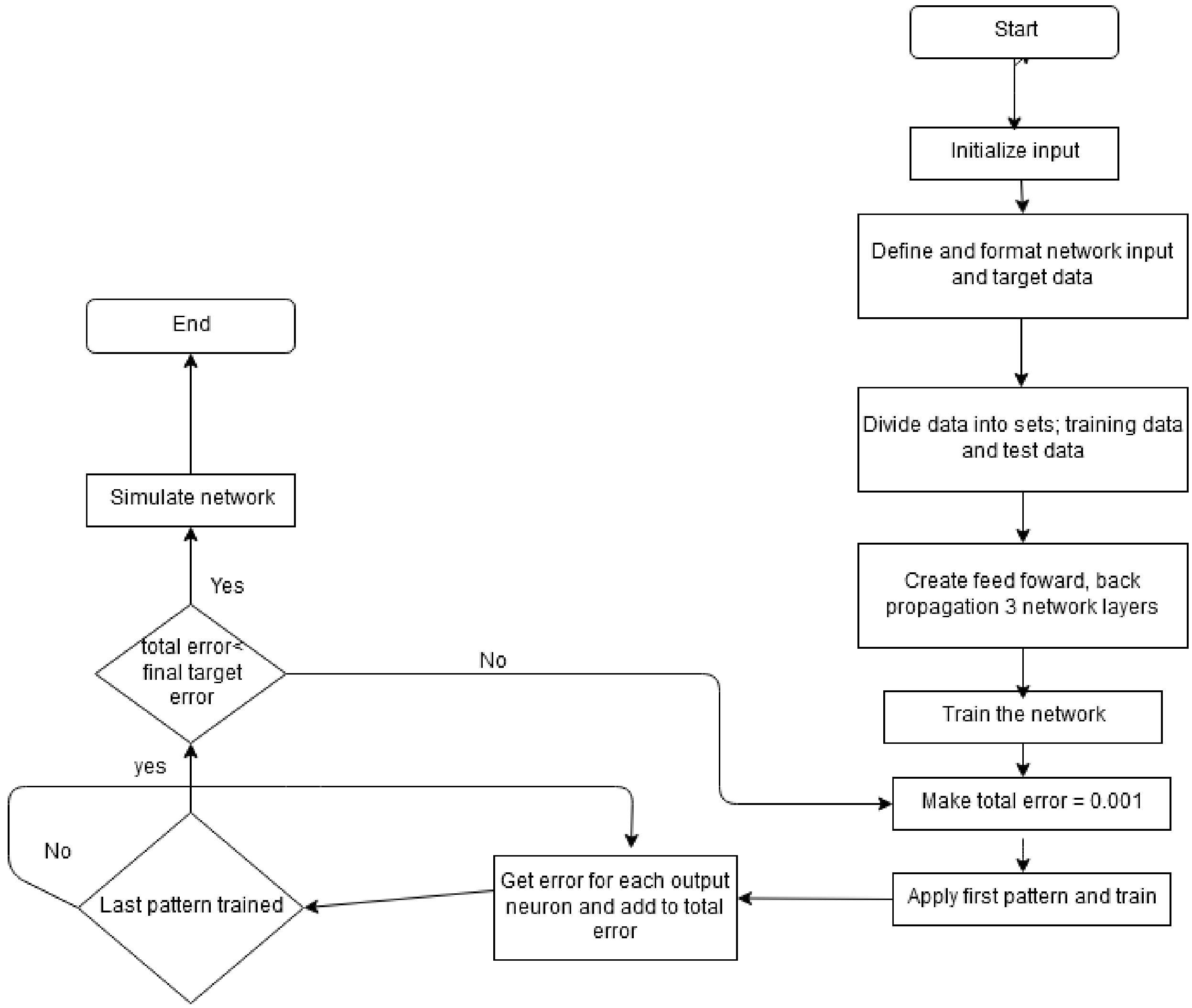
Equation 4.1: Final model equation

where, f and g denote the activation function at the hidden and output layers, respectively; p is the number of input nodes (tapped delay); q is the number of hidden nodes; $\beta_{ij}$is the weight attached to the connection between ith input node to the jth node of hidden layer; $\alpha_j$is the weight attached to the connection from the jth hidden node to the output node; and $y_{t-i}$is the j^{th}th input (lag) of the model. Each node of the hidden layer receives the weighted sum of all the inputs including a bias term for which the value of input variable will always take a value one. This weighted sum of input variables is then transformed by each hidden node using the activation function f which is usually a non-linear sigmoid function. In a similar manner, the output node also receives the weighted sum of the output of all the hidden nodes and produces an output by transforming the weighted sum using its activation function g. In time series analysis, f is often chosen as the Logistic Sigmoid function and g, as an Identity function. For a univariate time series forecasting problem, the past observations of a given variable serves as input variables. The final model with therefore attempt to map the following function.

$$y_{t-1}= f(y_t, y_{t-1}, \dots \dots . y_{t-1+p}, w) + \varepsilon_{t+1}$$

Where, y_{t+1}pertains to the observation at time $t + 1$, p is the number of lagged observation, w is the vector of network weights, and ε_{t+1} is the error-term at time $t + 1$.

Figure 4.9 - Final proposed model flow chart.



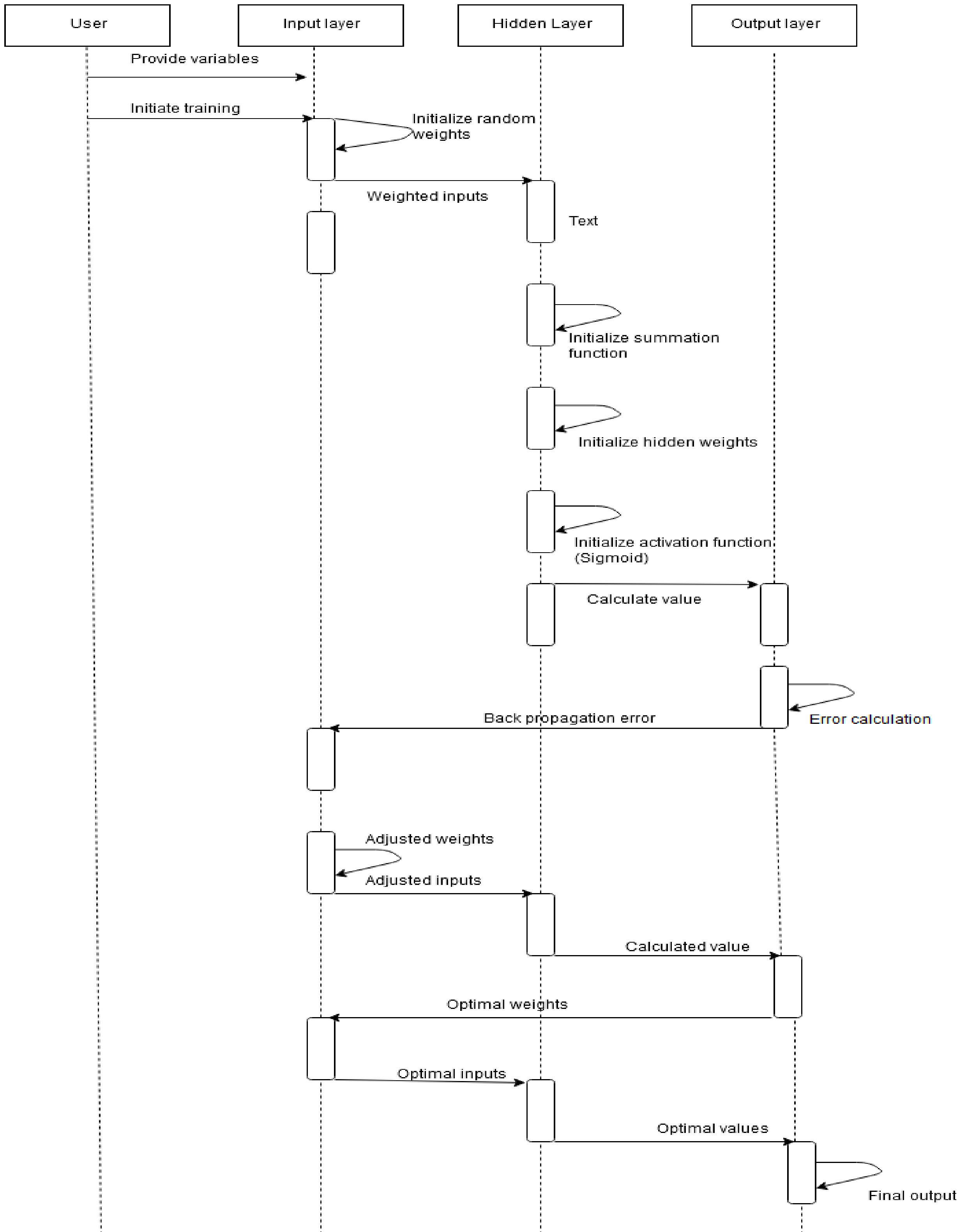
Source: Author.

4.6. Designing and Developing the Prototype.

4.6.1. Prototype Design.

The prototype was based on the model of configuration 4:11:1, using 70% of the available data (Jan 2006 to Apr, 2018) for training, with at least 50,000 repetitions. The ANN layout is multilayer perceptron (input layer, hidden layers, output layer, each with a number of neurons), the network connectivity is feed forward network (connectivity is between input neurons towards output neurons), weight adjustment is by error back propagation (resilient) and training is by supervised learning. The design tool that was used was a sequence diagram.

Figure 4.9.1 - Prototype sequence diagram



Source: Author.

4.6.2. Prototype development.

Requirements for prototype development included:

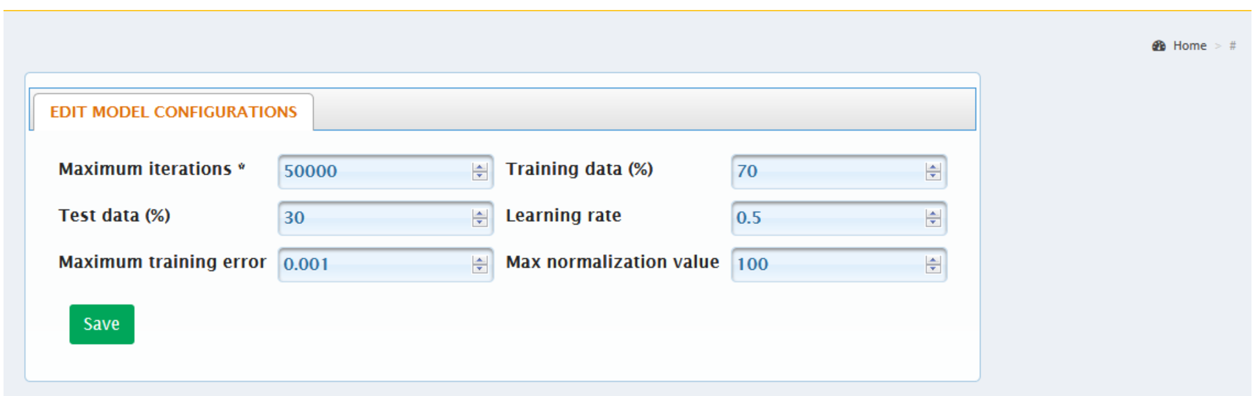
1. Laptop or desktop with at least 4GB RAM.
2. Linux or windows operating system.
3. Java 8.
4. Mysql server 5.3
5. Glassfish 4 application server.
6. Netbeans IDE.
7. Primefaces library (Has ajax components for user interface experience).
8. Neuroph library.
9. Github version control. (VCS).
10. Python numpy and scikit learn libraries.

Development of the prototype was done in java 8 programming environment. The web framework used was JSF (Java Server Faces). The user interface made use of bootstrap and primefaces library with ajax components that enabled ease of navigation for a non-technical user. Data was obtained from FAO and knoema websites in csv format and uploaded to a database created in mysql server. The source code link is indicated in **Appendix 4**.

Configurations page.

This page contains the network configurations as shown in figure 4.9.2. The configurations include; maximum iterations, training and test data percentage, learning rate, allowable network training error and maximum normalization value.

Figure 4.9.2 - Model configurations.

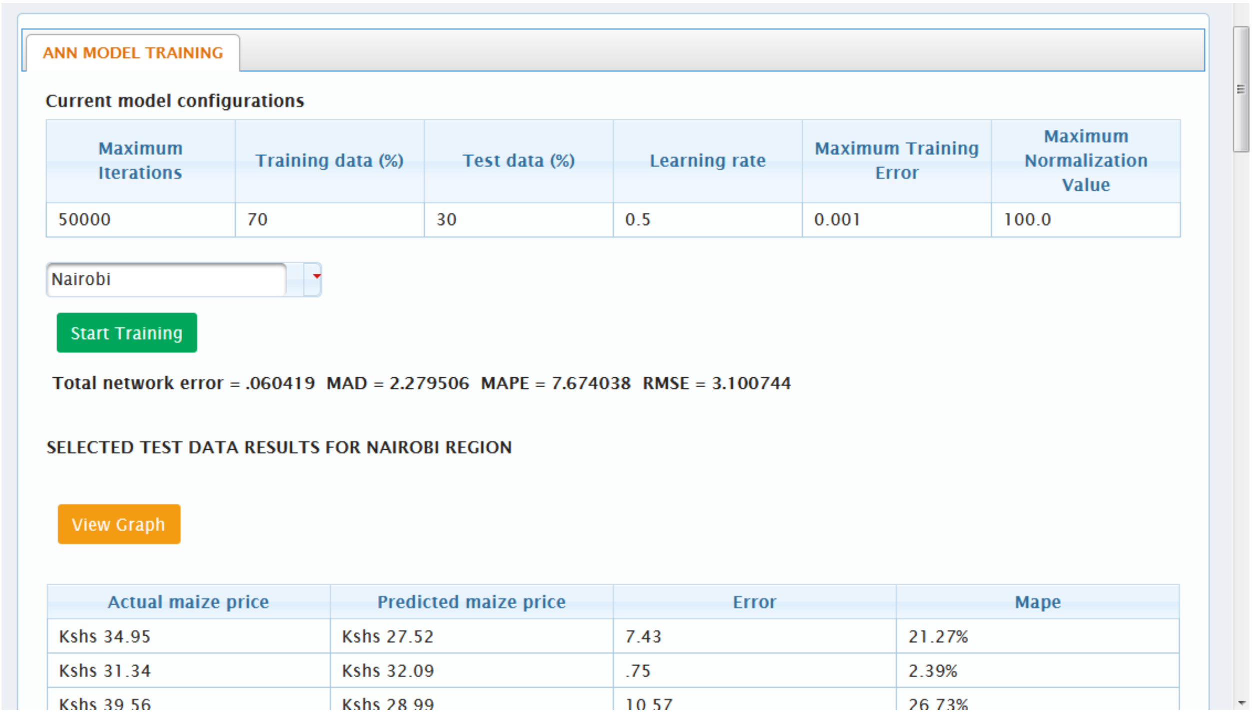


Source: Author.

Training page.

Data that was used in this study covered Eldoret, Nairobi and Kisumu counties, therefore the model training page provides a drop down option menu for the user to select and train the model using data from a particular county.

Figure 4.9.3 - Model training.

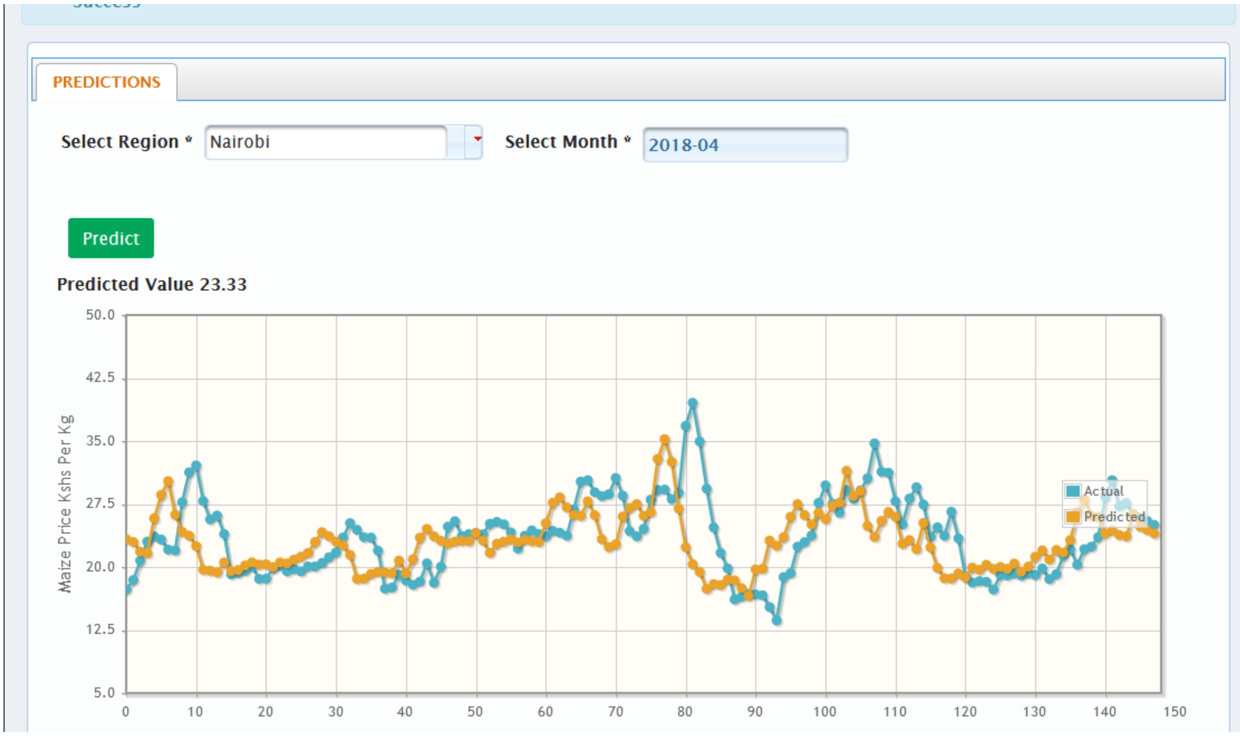


Source: Author.

Dashboard.

This is the landing page once a user logs in to the prototype. It gives the user the ability to predict maize price in a particular county based on the county selection and date.

Figure 4.9.4 - Model dashboard.



Source: Author.

4.7. Model Evaluation.

Evaluation of the ANN model was done using experimental methods. The tests were done using the developed prototype to evaluate its performance along various horizons that included; 1, 2, 3 and 4 month predictions. The model was also evaluated against an ARIMA model along the same periods. The results are shown in table 4.2. Table 4.3 shows test data results of the ANN's four month prediction while figure 4.9.5 is a graphical representation of the ARIMA model's four month prediction results.

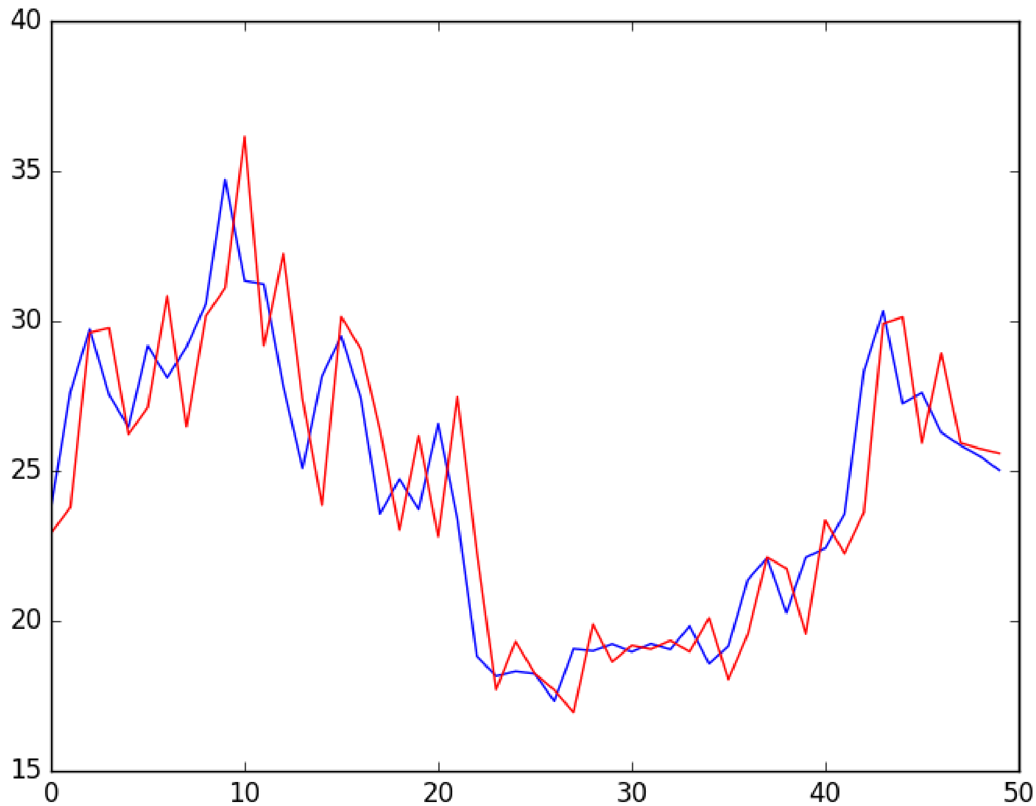
Table 4.2: Forecasting performance of ANN and ARIMA models for various horizons.

MODEL	1 month ahead		2 months ahead		3 months ahead		4 months ahead	
	RMSE	MAD	RMSE	MAD	RMSE	MAD	RMSE	MAD
ANN	2.8	1.9	3.06	2.15	3.10	2.18	3.10	2.2
ARIMA	4.691	5.159	4.807	5.159	4.918	5.159	4.931	5.159

Table 2.3: Forecasting results of ANN models for 4 months predictions.

Month-Year	Actual Price	Predicted Price	Error
Mar-12	Kshs 23.66	Kshs 23.21	0.45
Apr-12	Kshs 24.26	Kshs 22.53	1.73
May-12	Kshs 28.5	Kshs 22.64	5.86
Jun-12	Kshs 30.6	Kshs 26.34	4.26
July-12	Kshs 28.67	Kshs 28.1	0.57
Aug-12	Kshs 28.45	Kshs 27.12	1.33
Sep-12	Kshs 28.91	Kshs 26.58	2.33
Oct-12	Kshs 30.34	Kshs 26.18	4.16
Nov-12	Kshs 30.17	Kshs 27.65	2.52
Dec-12	Kshs 26.87	Kshs 27.57	0.7
Jan-13	Kshs 23.73	Kshs 25.14	1.41
Feb-13	Kshs 24.1	Kshs 22.63	1.47
Mar-13	Kshs 24.3	Kshs 22.52	1.78
Apr-13	Kshs 23.65	Kshs 22.71	0.94
May-13	Kshs 23.93	Kshs 22.87	1.06
Jun-13	Kshs 24.37	Kshs 23.07	1.3
July-13	Kshs 23.73	Kshs 23.23	0.5
Aug-13	Kshs 22.2	Kshs 22.91	0.71
Sep-13	Kshs 24.13	Kshs 21.85	2.28

Figure 4.9.5 – ARIMA prediction results from Jan-2006 to Apr -2018.



Source: Author

4.8. Chapter Summary.

Chapter 4 provides the results and then reviews, interprets and analyzes these results. Results obtained from the model in its univariate and multivariate forms are compared. The model is also evaluated against other models that can do predictions (ARIMA).

CHAPTER 5

CONCLUSION AND RECOMMENDATION.

5.1. Conclusion.

The main advantage of univariate time-series forecasting is that it requires data only of the time series in question. First, this feature is advantageous if we are to forecast a large number of price series. Second, this avoids the problem that occurs sometimes with multivariate models; for example, consider a model including import, prices and domestic production. It is possible that a consistent data on import series is available only for a shorter period of time than the other two series, restricting the time period over which the model can be estimated. Third, timeliness of data can be a problem with multivariate models.

This study has compared the ARIMA and ANN models in terms of both modelling and forecasting using monthly retail maize price data in three Kenyan counties, namely; Eldoret, Nairobi and Kisumu. The ANN model in general has provided a better forecast accuracy in terms of conventional RMSE and MAD values as compared to the ARIMA model. It has been found that the evidence of nonlinearity in a series plays a fairly good role in providing a reliable guide to post-sample forecast accuracy of ARIMA and ANN models in terms of RMSE for these price series.

Agricultural price information needs for decision- making at all levels are increasing due to globalization and market integration. This necessitates an effort towards designing a market intelligence system by integrating traditional statistical methods with soft computing techniques like neural network, fuzzy logic, etc. to provide accurate and timely price forecast by taking into account the local information to the farmers, traders and policymakers so that they may make production, marketing and policy decisions well in advance. The decision support system should provide customized advice to individual farmers in view of their local conditions.

According to impact of maize price on price of maize products and its market regulation condition, farmers and producers should take future prices of this input into account.

Considering price of maize in the future, agricultural authorities can reduce price fluctuations and consequently reduce the high risk present in maize and maize products` market and finally can increase producers and consumers` welfare. In fact, they can support maize farmers and maize product units in making the right decision by identifying and showing future price condition in this sector in different times.

5.2. Recommendation.

Since the price prediction is a key factor and it is significant to have up-to-date information, it is recommended to examine market condition in the future researches at universities and research centers and also to apply different prediction methods such as neural network method in order to reduce maize production risk in Kenyan Agricultural Sector.

Future studied needs to explore the possibility of using a hybrid model (combined linear and non- linear model) for predicting maize prices in specific Kenyan counties based on aggregated price data from all counties.

Finally, further research is needed to determine how long a trained ANN system remains valid and effective in prediction before it is found to be in need of retraining.

REFERENCES.

APPENDICES