UNIVERSITY OF NAIROBI
SCHOOL OF COMPUTING AND INFORMATICS

Combating Motor Vehicle theft using Decision Support Models: Nairobi Case Study

BY

NJIRI, Charles W.
P56/P/7492/04

Supervisor

Mr. Lawrence Mucemi

February 2010

Submitted in partial fulfillment of the requirements of the Master of Science in Information Systems

Abstract

Forecasting is an essential analytical tool in police planning and allocation of resources. Law enforcement agents have complex databases which has intelligence hidden as trends, patterns, dependencies, and relationships. Data Mining is the process of acquisition of this knowledge from databases inform of significant patterns and associations. This project focuses on machine learning tools and identifies Artificial Neural Network model that can be used for motor vehicle theft short term trend and patterns forecasting in Nairobi. The study involves several experiments using WEKA, Zaitun Time Series, Neuralab and Tiberius software to forecast motor vehicle theft. The data was prepared to forecast geographical location where theft was likely to occur. The forecasted results were trivial and hence disregarded.

The second set of experiment involved forecasting motor vehicle theft counts using time series Neural Network with WEKA software. The results were successful and able to forecast the motor vehicle theft tred for the succeeding 2 to 3 months. These results were further used to extract rules from the trained network. These rules explained future trends in motor vehicle theft.

The last study was to identify an open source time series neural solution that could be used to support decision making to combat motor vehicle theft. Three software solutions were studied and Zaituni Time series software performed extremely well.
The motor vehicle theft trend forecasted in this project confirmed the research hypothesis that artificial intelligence can be used to forecast crime trends.

Declaration

This project, as presented in this report, is my original work and has not been presented for any other University award.

Charles W. NJIRI 25/07/2010

Date

This Project has been submitted as partial fulfilment of requirements for Master of Science Information Systems of the University of Nairobi with my approval as the University supervisor.

Dr. Wanjiku Ng'ang'e 29.07.2010

FOR: **Supervisor: Mr. Lawrence MUCEMI** **Date**

Dedication

This project is dedicated to my loving wife Mary and my lovely daughters Emmah and Joy.

Acknowledgements

I would like to thank all the faculty members of School of Computing and Informatics of the University of Nairobi for helping me to develop the necessary skills to carry out this research. I owe special gratitude to Mr. Lawrence Mucemi for his individual support and motivation when I needed it most. The efforts of Dr. Wanjiku Nganga cannot go un mentioned. She gave me positive guidance that helped me to gain deep understanding of Artificial Intelligence and compilation of the final report.

Special thanks go to the Commissioner, Kenya Police Force without whose support and permission this research would not have been possible. I wish to thank all officers who contributed to this study, known and known, for their commitment and time.

I wish to acknowledge moral support from my family and my parents for instilling in me the thirst for education and who provided me with the opportunities to achieve. Their love, along with encouragements from relatives, classmates, friends, colleagues at the Interpol Regional Office in Nairobi and well wishers all over the world has been my strength and I am forever grateful.

Table of Contents

List of figures

List of Tables

List of Abbreviations

ANN Artificial Neural Network

ARFF Attribute-Relation File Format

CIA Criminal Intelligence Agency

FFNN Feed Forward Neural Network

GUI Graphical User Interface

INTERPOL International Criminal Police Organization

ISS Institute of Security Studies

KDD Knowledge Discovery in Databases

LEA Law Enforcement Agencies

LGPL Lesser General Public License

MAE Mean Absolute Error

NIJ United States National Institute of Justice

NN Neural Network

RMSE Root Mean Square error

SMV Stolen Motor Vehicle

SMV Stolen motor vehicles

UK United Kingdom

WEKA Waikato Environment for Knowledge Analysis

Combating Motor Vehicle theft using Decision Support Models: Nairobi Case Study

Key Words: Artificial Neural Network, Data Mining, Decision Tree, Knowledge discovery

Chapter 1

1.1.0 Introduction

A system that would predict crime trends in Kenya would be of great use to police resources allocation strategy. Such a system would ultimately lead to a reduction in the high crime rates existing in Kenya. These crimes are concentrated around the cities of Nairobi, Mombasa and Kisumu. Property crimes against Motor vehicle are the most reported crimes. About 250 motor vehicles are reported stolen every year. In Nairobi, ten carjacking are reported daily.

Crime data is arguably the most important asset to a law enforcement agent (LEA). In Kenya, Law enforcements agents include Police officer, Immigration officer, Wildlife officer etc. There exist huge seas of data that can be utilized to inform decision making and thereby improve crime intelligence to proactively detect and prevent crime.

Police and other law enforcement agencies need short-term crime level forecasts one week or one month ahead. Currently, most LEA respond to crime as they occur and follow the seasonality of the particular crimes. With short term forecasting, police would be able to get one step ahead of criminals by anticipating and preventing crime. To achieve this, Law enforcement agencies (LEA) require crime rate prediction and forecasting models to target patrols, surveillance, direct resources and ensure successful operations. Over time, LEAs have acquired skills that today are used to combat crimes. For instance, hotspots identification is in use world over.

Different techniques are today being used to forecast crime world over. In recent years the concept of neural networks has emerged as one of them.

Crime forecasting and police decision making problems are based on planning horizons: short term (tactical deployment), medium-term (resource allocation) and long-term (Strategic planning). More so, there exists enormous need to forecast large crime increases (or decreases) for tactical deployment of resources and police manpower.

This is a study aimed at forecasting motor vehicle theft trend in Nairobi using artificial neural networks.

Chapter 1 describes the problem and offers a credible justification and scope for the research. A study of previous research work in areas of motor vehicle theft, crime, data mining algorithms and machine learning are described in chapter 2. A lot of emphasis however is given to neural networks. A number of experiments were performed during this study and their methodologies given in chapter 3. Chapter 4 is dedicated to the results and discussions from the experiments. The conclusion, research contributions and future research areas are described in chapter 5.

1.2.0 Problem statements

Every year, Kenya loses more than 1.5 billion Kenya Shillings in motor vehicle theft. However, Kenyan authorities have limited modem capacity to deter and prevent the crime.

Motor vehicle crime forecasting is today based on the tradition of using the previous year's experience. The other most efficient means is 'buying information' from 'members of the public'. The latter in most cases are the police informers who could be members of the theft gang or insiders who understand their 'Modus operandi'.

Crime data contains a lot of noise from the environment and therefore linear prediction methodologies was found not be the best method to predict crime. The non linearity of crime prediction will formed the basis of this study. However, linear prediction was used as benchmark for success.

1.3.0 Justifications

Every year Kenya loses more than 1.5 billion Kenya Shillings due to motor vehicle theft. However, Kenyan authorities have limited modem capacity to deter and investigate the crime. Therefore, the science of 'informers' and 'Seasons' continues to be used to fight this crime.

Several studies have been done on crimes in Kenya. However, none has attempted to investigate the possibility of using Artificial Intelligence and neural networks to predict crime. This study was to fill this gap and predict motor vehicle theft using neural networks.

Use of small and light weapons, involvement of expensive motor vehicles and sophisticated and enlightened robbers, lack of enough law enforcement agents to fight this crime may have encouraged the proliferation of this crime.

Armed with crime trend forecasting model, the Kenya police can target patrols, direct surveillance and conduct operations to prevent crimes and enforce laws. A successful model need not predict every incident. It only needs to outperform any current model in use in Nairobi and forecast crime trend.

Over the years ANN has become popular due to its ability to learn complex and nonlinear relationships that are difficult to model with conventional techniques. The incentive is therefore to try non traditional techniques to predict this social menace.

1.4.0 Objectives

The following are the objectives of this study:

1. Demonstrate how Weka machine learning algorithms can be applied to discover knowledge from motor vehicle theft data.

2. Use Neural Network to predict uni-variate time series crime data and attempt to extract knowledge from the trained network.

2. Explore and evaluate Neural Network simulations and identify an open source solution to predict motor vehicle theft trend in Nairobi.

1.5.0 Hypothesis

The study was developed with the following hypothesis:

1. Neural Network is a practical data mining tool for anti-motor vehicle theft law enforcement agent

2. Neural Networks can be used to forecast motor vehicles crime patterns and trends.

1.6.0 Scope, Limitation and Assumption

Crime forecasting is very wide and limitless. The research was limited to motor vehicle theft, WEKA machine learning algorithms and three neural network simulation forecasting tools. However, it did not attempt to investigate how the motor vehicle crimes are committed, but attempted to develop a neural network model to forecast motor vehicle theft patterns and trends to inform decision makers to deter and combat the crime. The study was carried out using vehicle theft data from Nairobi during the period 2003-2009.

1.7.0 Proposed Solution

Artificial Intelligence is a powerful paradigm that can be used to predict crime. The proposed model was aimed to support decision making. This is based on the knowledge that the focus of crime prevention and law enforcement is on places where the crime frequently occur. The data was mined for patterns and rules. The frequency counts of crimes were used to forecast the future trend (Gorr, Olligschlaeger, & Thompson 2003)

Chapter 2

Literature Review

2.1.0 Motor vehicle Crime in Kenya

In 1998 there were 549,913 registered motor vehicles in Kenyan roads. This figure almost doubled by 2008 when there were about 1,016,766 vehicles in Kenya as reported by Jane (2009).

There exist high crime rates in Kenya. The regions identified to have above average rates include the cities of Nairobi, Mombasa and Kisumu (http://travel.state.gov/travel/cis_pa_tw/cis/cis_1151.html Retrieved on 23 Jan 2010). The most common crime in Kenya is property crimes against Motor vehicle. The vehicles are either stolen from car parks or violently robbed from their owners (Carjacking). It is estimated that about ten carjacking take place per day in Nairobi alone. In some of these carjacking, the motor vehicles are used to commit other crimes and are soon recovered (https://www.osac.gov/Reports/report.cfm?contentID=116435 Retrieved on 23 Jan 2010).

Kegoro G. (2002) in his research observed that on average 250 motor vehicles are stolen in Kenya every year. Most of these thefts are violent where guns and other small arms and light weapons are used. The total annual economic loss was estimated at 1.56 billion Kenya Shillings in 2002.

The Kenya Police annual crime reports (2009) show that in 2005, 293 vehicles were stolen while in 2006, 237 thefts were reported. This report corroborates Kegoro's year 2002 research. It is clear that the level of this crime has peaked at around 250 reported cases each year. Moreover, CIA world report describes Kenyan authorities as one that has limited capacity to deter and combat such acts of crimes against properties. This research demonstrates that internal capacity can be developed to forecast this crime.

Motor vehicle theft is defined by the US Department of Justice as the criminal act of 'stealing or attempting to steal a motor vehicle' including automobile, truck, bus,

motorcycle, snowmobile, trailer or any other speciali
(http://www.fbi.gov/ucr/cius2008/offenses/property_crime/motor_vehi
Retrieved on October 2009)

Crime data is arguably the most important asset to a Police offic officer, Wildlife officer etc. Currently there exists a huge sea of data and utilized to inform and support decision making. Research show data doubles every 2 years. Out of this, only less than 5% is electro and hence help in decision making (Bigus 1996).

Neural Networks have been used successfully in forecasting in differ fields. Moreover, in the last 10 years, computers capabilities have im immensely in speed and memory. The results of this crime forecast by police to target tactical operations. Besides, Neural Networks hav of attention from practitioners and academics. Kajitani Y., Hipel K. A. I. (2005) found that neural networks are viable contenders to vari non-linear time series models.

Haykin S. (1999) describes neural network as a simulated biologi parallel distributed processor that has a propensity for storing expe and making it available for use. Neural networks resembles the br Knowledge is acquired by the network through a learning proce connection strengths known as synaptic weights are used to store th The neural network approach helps reduce the problem as conventional forecasting methods as it has approximation abi mapping and generalization. It has the advantages of learning dire data.

2.2.0 International Motor vehicle theft

INTERPOL, the world largest police organization, recognizes highly organized criminal activity affecting the whole world. It the crime is often linked to organized crime (http://www.interpol.int/Public/Vehicle/Default.asp Retrieved vehicles are not only stolen for their own sake; sometimes th motorcycle, snowmobile, trailer or any other specialized vehicle' (http://www.fbi.gov/ucr/cius2008/offenses/property_crime/motor_vehicle_theft.html . Retrieved on October 2009)

Crime data is arguably the most important asset to a Police officer, Immigration officer, Wildlife officer etc. Currently there exists a huge sea of data that can be used and utilized to inform and support decision making. Research shows that electronic data doubles every 2 years. Out of this, only less than 5% is electronically processed and hence help in decision making (Bigus 1996).

Neural Networks have been used successfully in forecasting in different academic fields. Moreover, in the last 10 years, computers capabilities have improved immensely in speed and memory. The results of this crime forecast trends to be used by police to target tactical operations. Besides, Neural Networks have received a lot of attention from practitioners and academics. Kajitani Y., Hipel K. W. and McLeod A. I. (2005) found that neural networks are viable contenders to various linear and non-linear time series models.

Haykin S. (1999) describes neural network as a simulated biological cell, massively parallel distributed processor that has a propensity for storing experiential knowledge and making it available for use. Neural networks resembles the brain in two respects; Knowledge is acquired by the network through a learning process and interneuron connection strengths known as synaptic weights are used to store the knowledge.
The neural network approach helps reduce the problem associated with the conventional forecasting methods as it has approximation ability for non-linear mapping and generalization. It has the advantages of learning directly from historical data.

2.2.0 International Motor vehicle theft

INTERPOL, the world largest police organization, recognizes vehicle crime as a highly organized criminal activity affecting the whole world. It has established that the crime is often linked to organized crime and terrorism (http://www.interpol.int/Public/Vehicle/Default.asp Retrieved October 2009). The vehicles are not only stolen for their own sake; sometimes they are trafficked to finance other crimes in the underworld economies. They can also be used as bomb carriers or in the perpetration of other serious crimes.

By the end of December 2008, the INTERPOL stolen motor vehicle database held more than 4.6 million records of reported stolen motor vehicles from across the globe. The figure 2.1 below shows the growth of international stolen motor vehicle records in the INTERPOL databases.

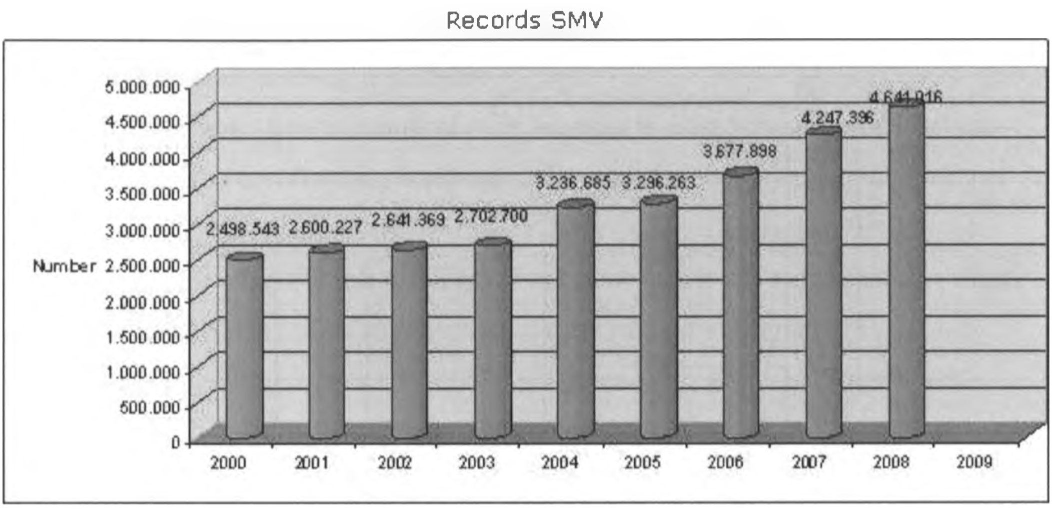


Figure 2.1. *Interpol SMV database growth (source: www.Interpol.int)*

2.3.0 Crime forecasting and its origin

US National Institute of Justice (NIJ) awarded five grants to study crime forecasting 1998. This was after the success of a crime mapping project. At about the same time, the UK's Home Office published crime forecasts for the first time. Basic needs forwarded then are the same today which includes informed decision making, budget planning and resource allocation, manpower resources redeployment, shift between law breaking prevention to law enforcement.

2.4.0 Forecasting Requirement of Police

Gorr et al (2003) classified police forecasting and decision problems based on planning horizons: short term (tactical deployment), medium-term (resource allocation) and long-term (Strategic planning). This research was based on short term forecasting and tactical deployment which can be identified as one month to three months horizon. This was motivated by the fact that police budgets are devoted to such resources as personnel and vehicles that can easily be redeployed in short-term.

2.5.0 Machine Learning

Machine learning is a scientific discipline that is concerned with the design and development of algorithms that allow computers to evolve behaviors based on empirical data, such as from sensor data or databases.

There are three main learning paradigms:

Supervised learning:- here a sample of pairs are used to infer how mapping implied by the data and the cost function is related to the mismatch between our mapping and the data.

Unsupervised learning:- in this paradigm given some data x and a cost function which is to minimized which can be any function of x and the networks output f.

Reinforcement learning: - data x is usually not given but is generated by an agent's interactions with the environment e.g. game playing.

2.6.0 Data Mining

Data Mining is the process of discovering patterns in data. This process is also called Knowledge Discovery in Databases (KDD). It is an automated process (Witten H.I & Frank E. 2004). Data Mining has also been referred to as exploratory data analysis. Jonas J. (2006) describes data mining as the process of searching data for previously unknown patterns and using those patterns to predict future outcomes.

Systematic exploration with classical statistical methods is the basis of data mining. Tools and techniques that transform stored data into knowledge include regression (normal regression for prediction and logistic regression for classification), neural networks and decision trees.

Five KDD steps were identified by Bigus P. (1996) as shown in figure 2.2. These are: Data Selection, data pre-processing and cleaning, data transformation, Data Mining and result interpretation and evaluation. The KDD process starts by understanding the problem's domain and the final objectives to be reached. The available data is arranged into an organized group, the search's target. The data-cleaning step comes next, by means of data preprocessing, integrating heterogeneous data, eliminating incomplete data and others.

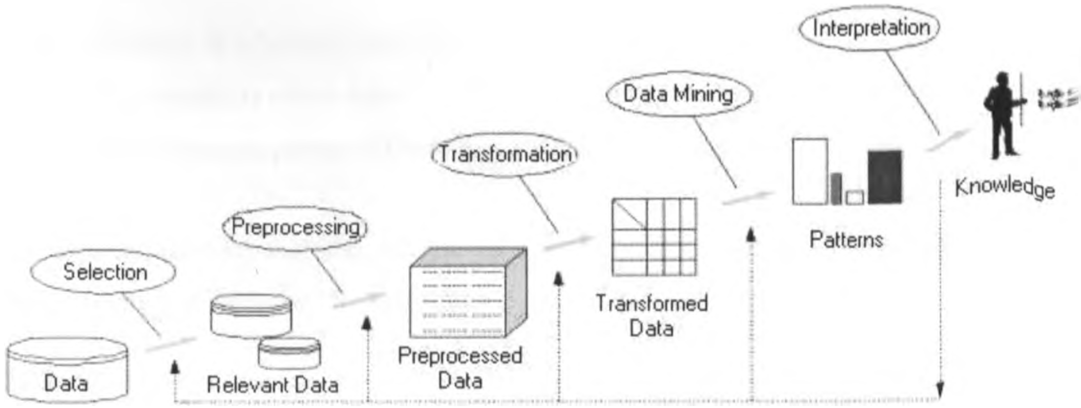


Figure 2.2: *Data Mining Knowledge Discovery Process*

In this work, these preliminary steps involved data selection, cleaning and coding (in two of the four simulations).Thereafter, data mining commences with a choice of algorithms to be used. This choice depends fundamentally on the KDD process' objective which may be: classification, grouping or association. In general, the algorithms used during the Data Mining step look for patterns in the data.

Several distinct tools, such as NNs, decision trees, systems based on rules, statistical programs and others, isolated or combined one with each other, can then be applied to the problem.

In general, search processing is interactive so as to allow analysts to review results, form a new set of questions in order to refine searches with respect to a certain aspect of results, and feed the system back with new parameters. By the end of the process, a discovery report is produced, which is then interpreted by the mining analysts and the knowledge is discovered.

Among the Data Mining techniques that are used in classification problems and forecasting we point out the Nearest Neighbors, j48 classifiers and NNs: they build internal representations of models or patterns they detect in the data, but these representations are not presented explicitly to the users.

Anbananthen (2007) observed that the rule extraction algorithms can be categorized as:

- Schemes of extracting rules from ANN
- The portability of the rules extraction techniques
- The expressive power of the rule extracted.

He further observed that the schema of extracting rules can be categorized into decomposition or pedagogical algorithms

2.7.0 Time Series

A time series is a 'set of observations, results, or other data obtained over a period of time'. Time series analysis comprises methods that attempt to understand such time series. Time series forecasting is the use of a model to forecast future events based on known past events (Bigus J.P 1996).

Time Series prediction was identified as the most appropriate method to be applied for short term Motor Vehicle theft forecasting model. Searching for systematic and recurrent relationships in the historical data and making predictions on the future based on this relationship is the key features of time series methods.

Short-term forecast models are of two primary kinds:
1). Uni-variate, extrapolative forecast models- used to extrapolate existing crime patterns. This research is based on Uni-variate forecasting.
2). Multivariate, leading indicator forecast models- can forecast new crime patterns not yet observed. 'Broken window hypothesis' posits that 'soft crimes' are leading indicators of hard crimes. Law enforcement agencies need to generate their own leading indicator data.

Haydari Z. Kavehnia F., Askari M & Ganbariyan M. (2007) identified the following components in time series:

- Trend Component
- Cyclical Component
- Seasonal Component
- Irregular Component

Trend Component: - this is gradual shift of the time series usually due to long term factors such as population changes, shift in technology, political dynamics, improvement in infrastructure and other factors which produce steady and gradual change over time.

Cyclical components:- these are regular pattern of sequences of points above and below the trend line lasting more than one year. It represents multiyear cyclical movements in the economy.

Seasonal Component:- This caused by natural seasonal phenomenon like drought, rain etc.

Irregular Component:- This is the residue of the time series if the trend, cyclic and seasonal components are removed.

2.7.1 Linear Trend

Linear trend is a simple function described as a straight line along several points of time series value in time series graph. Linear trend has a common pattern:

$$T_t = a + b.Y_t \text{..........} \quad \textit{Equation (1)}$$

Where T_t = Trend value of period t

a = Constant of trend value at base period

b = Coefficient of trend line direction

Y_t = an independent variable, represents time variable, usually assumed to have integer value 1, 2, 3,... as in the sequence of time series data. There are several methods that can be used to find the linear trend equation of a time series. Most commonly used is least squares method. This method finds the coefficient values of the trend equation (a and b) by minimizing mean of squared error (MSE). The formula is:

$$b = \frac{n\sum Y_t T_t - \sum Y_t \sum T_t}{n\sum Y_t^2 - \sum (Y_t^2)} \text{..........} \quad \textit{Equation (2)}$$

$$a = \bar{Y} - b\bar{T}_t \text{..........} \quad \textit{Equation (3)}$$

Linear trend formed the baseline of this study

2.8.0 Artificial Neural Network

The method of forecasting and prediction using artificial neural network (ANN) has been extensively studied by academics and researchers. Haykin S. (1999) described a neural network as a massively parallel distributed processor made up of simple processing units, which has a natural propensity for storing experiential knowledge and making it available for use.

ANN is a mathematical model or computational model based on biological neural networks (Fig 2.3). It consists of an interconnected group of artificial neurons that processes information using a connectionist approach to computation.

Modern software implementation of artificial neural networks is more than ever inspired by statistics than in biological cells. They are used in large systems that combine both adaptive and non-adaptive elements.

Neural networks are composed of computing units (artificial neurons) interconnected so that each neuron can send and receive signals to or from each other. Neural networks are examples of distributed nonlinear systems models. These flexible non-linear models are capable of discovering hidden patterns adaptively from data. They acquire knowledge through learning as they are designed to work like the human brain.

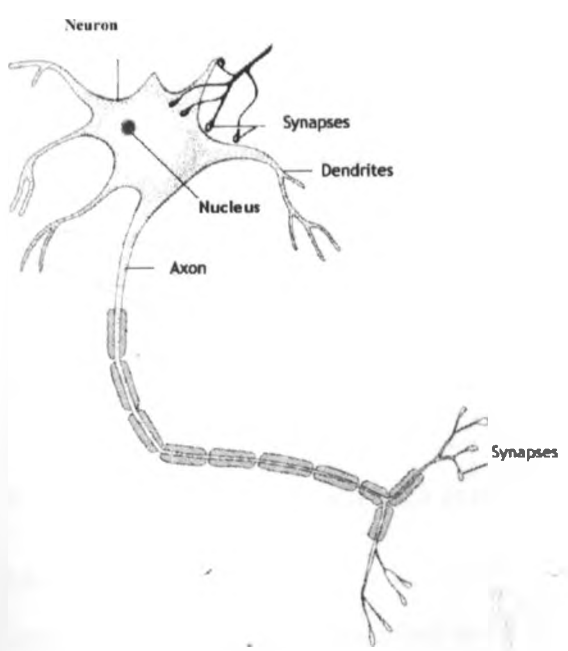


Figure 2.3. *Biological Neuron*

2.8.1 Models

Neural network (fig 2.4) models define a function

$f : U \rightarrow Z$ ....*Equation (4)*

Each type of ANN model corresponds to a class of such functions.

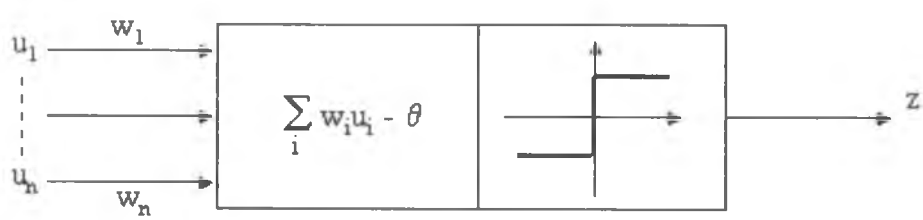


Figure 2.4. *The artificial neuron with a threshold function*

A neural network is so called because the function $f(u)$ is defined as a composition of other functions $g(u)$, which can be defined as a composition of other functions.

Consider a non linear weighted sum,

$z = f(u) = K(\sum_i w_i g_i(u))$.......................................*Equation (5)*

Where from fig 2.4

$g_i(x) = u_i$..*Equation (6)*

K is some predefined function, such as the hyperbolic tangent or sigma.

2.8.3 Learning

Learning in practice can be described as follows: given a task to solve, and a class of functions F, learning means using a set of observations, in order to find $f^* \in F$ which solved the task in an optimal sense.

This has a cost function and therefore learning algorithms search through the solutions space in order to find a function that has the smallest possible cost.

2.8.4 Advantages of Neural Networks

A neural network has several advantages. Among them include the ability to learn from data and thus potential to produce an acceptable output for previously unseen input data. This ability holds even when presented with noisy data. Neural networks are also non-linear mappers. This property allows neural network to be an ideal candidate for solving many non-linear problems.

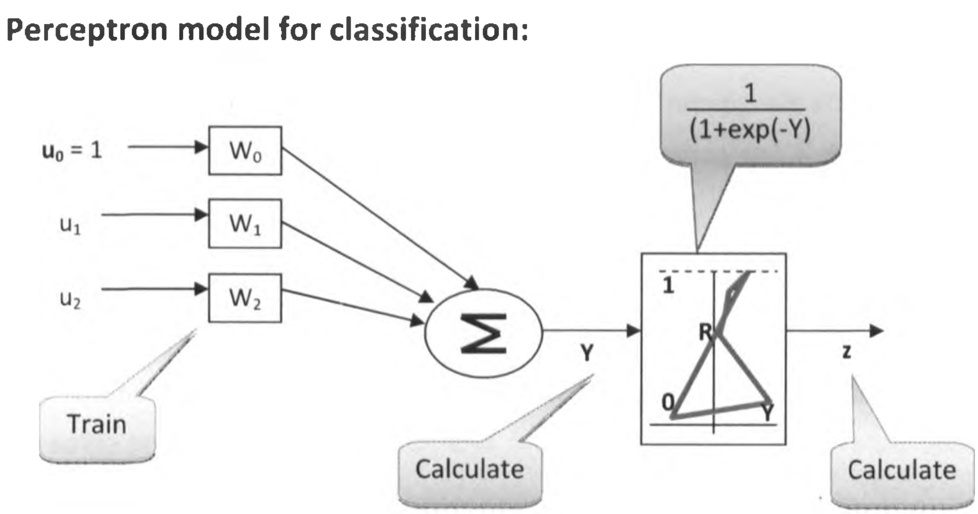


Figure 2.5. *Basic perceptron model.*

The goal of neural network learning is to find the weights W as shown in fig 2.5.

Neural networks are extremely flexible to changes in environment. This is because no program coding is required to learn the rules. The system requires to be trained to adapt to the new environment.

However, neural networks have their down sides which include the black box property. This arises from the fact that it is difficult to understand the internal decisions made by the system. However, recent research has developed new techniques to extract rules from the system.

Noisy data reinforce negative implications of incorrect causalities. Overtraining or over fitting also harm generalization. Finally, the domain knowledge is necessary to avoid trivial and irrelevant inputs and hence output.

2.8.5 Neural Network models and Architectures

The combination of topology, learning paradigm and learning algorithm define a neural network model. In data mining, the back propagation and Kohonen feature map are the most popular.

2.8.6 Feed-Forward Neural Network Models in Time Series Prediction

Many Artificial neural network architectures have been examined for addressing the time series problem. These architectures include: Multilayer Feedforward neural network (FFNN), recurrent networks and radial basis function (RBF).

There is substantial motivation for using FFNN for predicting time series. Kajitani K (2005) observed that FFNN can outperform chaos model in forecasting but both were better than traditional random walk model. He further observed that FFNN seem to be suitable for time series forecasting with small signal to noise ratios if we have enough data and use appropriate data transformation techniques..

2.8.7 Back Propagation Networks

Back propagation network applies a feed forward topology, supervised learning and back propagation algorithm. It is a powerful but expensive model in terms of computational training requirement.

Bigus P. (1996) observed that back propagation network and its variation, with a single hidden layer of processing elements can model any continuous function to any degree of accuracy. It works for a wide range of problems. The hidden layer provides the internal knowledge storage as shown in fig 2.5.

A back propagation network without a hidden layer can easily model a linear regression model relating to multiple inputs to multiple outputs. Meanwhile, addition of hidden layer transforms the network to nonlinear one capable of performing multivariate logistic regression.

Time series forecasting can be accomplished with back propagation network through the "Sliding window" technique as recommended by Bigus P (1996). Here a set period of time can be presented to the neural network, and the desired output is the function at the next time period.

Learn rate and momentum are used to control the training process of a back propagation network. Learn rate specify whether the neural network is going to make major adjustments after each learning trial. Momentum is used to control possible oscillations in the weights (Bigus P.J. 1996).

2.9.0 Knowledge Extraction from Neural Networks

In the past, neural networks were considered as 'black boxes' as they were not able to explain the knowledge acquired in the weights after the training process. The goal of knowledge extraction is to find the knowledge stored in the network's weights in symbolic form. One main concern is the fidelity of the extraction process, i.e. how accurately the extracted knowledge corresponds to the knowledge stored in the network.

Chandra R. & Omlin C. W. (2007) identified two main approaches for knowledge extraction from trained neural networks:

(1) extraction of 'if-then' rules by clustering the activation values of hidden state neurons.

(2) The application of machine learning methods such as decision trees on the observation of input-output mappings of the trained network when presented with data.

In this study, decision trees were used to extract rules from trained neural networks. The extracted rules explained the increase or decrease of SMV counts in the predicted period given the counts of the previous 12 months.

2.9.1 Decision Trees

Decision trees are machine learning tools for building a tree structure from a training dataset of instances which can predict a classification given unseen instances. A decision tree learns by starting at the root node and selects the best attributes which splits the training data. The root node then grows unique child nodes using an entropy function to measure the information gain from the training data. The process continues until the tree structure is able to describe the given data set. Compared to neural networks, they can explain how neural networks arrived to a particular solution.

Chapter 3

Methodology

3.0.0 Introduction

In this chapter, the methodologies of the following experiments are described:

1. Using machine learning techniques to mine rules from SMV data
2. Using FFNN to predict stochastic uni-variate Time Series motor vehicle crime trend and extracting forecasting knowledge.
3. Identify an open source tool to predict and forecast time series crime trend that can be used by law enforcement agents.

3.1.0 Data

Data for this study was obtained from the competent department of the Kenya police. This is the department that maintains the data regarding motor vehicle theft in the Kenya. The records contains information related to; Report date, Theft occurrence, Road/ Area Committed, Report Station, Car Make, Car Model, Color, Year of Manufacture, Driver Gender, Description of Circumstances, Description, Recovery Status.

3.2.0 Tools

The following tools were used during the experiment:

- WEKA (www.cs.waikato.ac.nz/ml/weka)
- Zaituni Time Series
- Tiberious Networks
- NeuralLabs Solutions
- Microsoft Excel
- Computer System

3.3.0 Data Preparation

The data received was prepared for two different experiments.

1. Using machine learning techniques to mine for possible rules from SMV data.

2. Time Series Neural Network Forecasting for stolen motor vehicle

3.4.0 Basic Statistical Data Analysis

The data was analyzed using Microsoft Excel and Zaituni Time series with the aim of identifying a general trend in crime. It also gives an insight into motor vehicle crime seasonality trends. The data was prepared on month to month for each of the years 2003-2009.

3.5.1 Using machine learning techniques to mine rules from SMV data

The data used in this experiment included 554 instances of reported stolen motor vehicles. From this data, 6 attributes were identified as follows: Day, Location, Vehicle Model, Color, Relative age of the car.

- Day are days of the week: Monday, Tuesday, Wednesday, Thursday, Friday, Saturday and Sunday.
- Location are the 9 police divisions in Nairobi as follows: Langata, Gigiri, Kasarani, Embakasi, Kilimani, Central, Kayole. Buruburu and Ngong. Another division called 'Others' was created to represent all other theft reported outside Nairobi.
- Vehicle Model:- Most of the vehicles reported were found to be Toyotas and therefore there was need to distinguish the instances by the models. The models identified for this study and formed this attribute are as follows: Sunny(Nissans), Corolla (Toyota), SUV (Toyota Landcurisers, RAV4s,Surfs,Harriers, Mitsubishi Pajero, Lexus etc), Other_Toyota_Models (Toyota models like Caldina, Platz, Prius etc), Pickup_Truck(Mitsubishi l200, Canter, Isuzu TFR, Toyota Hilux etc), Datsun (120y) and Vans (Nissan Caravan, Toyota Shark, Homy etc)
- Color: The main color reported was White, Red, Pearl and Black.
- Owner: three types of owners were reported: Males (M), Females (F) and Corporate (C).

3.5.2 Missing Attributes

Missing crucial attributes like Location, Vehicle models etc. were rampant. These being unique attributes the data was left as unknown.

The data was pre-processed and converted to ARFF format to be used by WEKA Machine learning tools.

3.5.3 Machine learning techniques

Among the 10 methods for obtaining classification rules that exist in the WEKA software, JRip, NNge, MLP and J48 Methods were examined for accuracy. A total of 554 instances were used.

Figure 3.1: *Weka Classifier Prediction Model*

Figure 3.1 shows the Weka classifiers prediction model complete with variable attributes and predicted attribute.

3.6.0 Using FFNN to predict stochastic uni-variate Time Series motor vehicle crime trend

3.6.1 Methodology

Predictive mining is a task that performs inference on the current data in order to make a prediction. A monthly crime data is stochastic and can be grouped as a time-series set because it consists of sequences of values in time. A time series data can be simplified as:

$y = f(t)$ .... *Equation (7)*

Where y can be any single valued variable which develops in time t In this work, y is monthly stolen motor vehicle values. To forecast time series data it involves knowing the past history of f and extrapolating it to the future. The main characteristic of the forecasting model is a non-linear system so that back propagation neural network can be applied in time-series predictions.

3.6.2 Neural Network development

The following steps were identified and followed to develop SMV Neural Network forecasting model.

- Basic Analysis of the data.
- Preliminary analysis of the data using other statistical methods.
- Neural network model design
- Network optimization
- Empirical studies on NN Solutions
- Multiple tests run on the selected model.
- Interpretation of the input/output mapping and sensitivity analysis

3.6.3 Basic Statistical Data Analysis

The data was analyzed as explained in section 3.4.0.

3.6.4 Preliminary analysis of the data using other statistical methods.

Uni-variate time series analysis of the data was performed using linear regression. A linear regression equation was modeled to predict crime trend using Weka function as shown in fig 3.2 below.

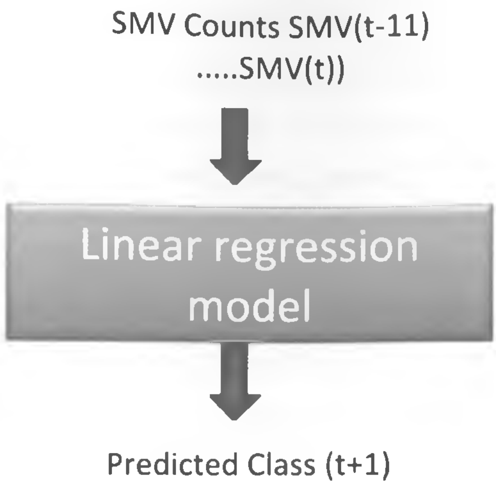


Figure 3.2: *Linear Regression Model*

3.6.5 Attributes and data preparation of Neural Network Model

In this, study 60 counts of motor vehicle thefts recorded each month for five years were used. However, not all the counts were available. Therefore the average of the available specific month was used. For instance, if Jan 2006 was missing, the mean of Jan 2003, Jan 2004, Jan 2005 and Jan 2007 would be calculated. This mean would fill for Jan 2006.

The data was arranged into an organized group of 13 variables: SMV(t-11), SMV(t-10)....SMV(t). This represented a time lag of 12 months. The predicted variable is Predict_Class is SMV(t+1) was coded as shown in table 3.1 below.

Variable	Range Code	Meaning
<=22	1	Low thefts level
22<x<29	2	Medium thefts level
>=29	3	High Theft Levels

Table 3.1: Predicted attribute coding

The coding ensures that the results are easily understood by police officers since it is easy to appreciate levels are changing than numbers.

3.6.6 Neural Network Design

To obtain the optimum neural network performance a multilayer perceptron with 1 hidden layer was used. The number of hidden nodes was varied between 4 and 10 and RMSE recorded using learning rates of 0.3 and momentum of 0.2.

3.6.7 Neural Network model design

Motor vehicle crime data is stochastic time series data. Multilayer perceptron (MLP) is a layered feedforward neural network which is trained using Back propagation. Time series forecasting was done using a feedforward neural network through the "Sliding window" technique. Here a set 12 months period of time was presented to the neural network, and the desired output is the function at the next time period.

To avoid overtraining the 66% of the data was used for training and the rest as validation set.

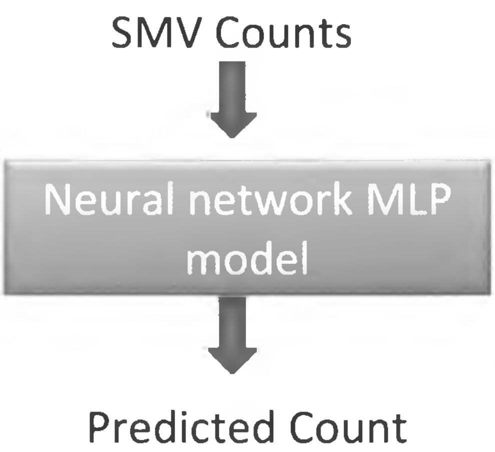


Figure 3.3: *Neural Network Model.*

One layer of hidden units was sufficient. The number of hidden units was varied to obtain the optimum performances.

3.6.8 Resources Required

The tool used during this experiment was WEKA machine learning algorithms.

3.7.0 Knowledge Extraction using Decision Trees

Weka j48 algorithm was used to extracted decision trees from trained neural networks. For the decision induction, the attributes and classification of the 66% training data and the attributes of the remaining 34% test data were used, but with their labels determined by the trained Networks. This way, decision trees became a means to knowledge extraction from neural networks. The labeling by the trained networks shows how well the network can predict on the motor vehicle theft trend given its previous history.

3.8.0 Identification of an open source tool to predict and forecast time series crime

As part of this research, it was necessary to identify a time series neural network forecasting tool that can be used by law enforcement agencies besides Weka. Three potential solutions were identified from the Internet. Table 2 below shows the initial analysis.

	GUI Friendliness	Ease of use	Computer Resources
Zaitun time Series	5	4	3
Tiberious	4	3	4
NeuralLab	3	2	4

Table 3.2: Software Analysis

In the table 3.2 above, each option considered was given a 1-5 rating in each category, with 1 being the worst case scenario and 5 the best score. These ratings were then added together and the highest scoring was considered to be the best. The chosen software was then subjected to further analysis to determine the optimal configuration.

3.8.1 Neural Network Model design

Modeling a time series involves generating a set of input vectors and corresponding output values as earlier observed. Previous research indicates that use of a single or two hidden layer is sufficient to learn any complex nonlinear function (Corcoran J. Wilson D. & Ware A. (2003).

3.8.2 Emperical NN solution studies

Zaitun Time Series is an Open Source time series forecasting application. It is used for statistical and neural forecasting. It is a simple, user friendly, easy to use and understand application available in LGPL licence.

It is a robust application which directly reads time series data directly from Microsoft Excel file and comma separated file. The application transforms the data appropriately ready for forecasting and prediction.

3.8.3 Zaitun Time Series Data preparation

The data was preprocessed into monthly SMV counts running in excel worksheet. The Zaitun application has an advanced excel file access with the ability to load the serial data with the data point frequency (annual, monthly, weekly, daily, etc) ranging from start date to end date.

3.8.4 Neurolab Neural Network Data preparation

The data was processed into 13 variables: (SMV) [t], (SMV)[t-1 to t-11] .

The input nodes correspond to sum of monthly SMV count from Nairobi for the past 12 months, (SMV)[t to t-11]. The output node is the forecasted SMV count for the current month, (SMV)[t+1] as shown in figure 3.4.

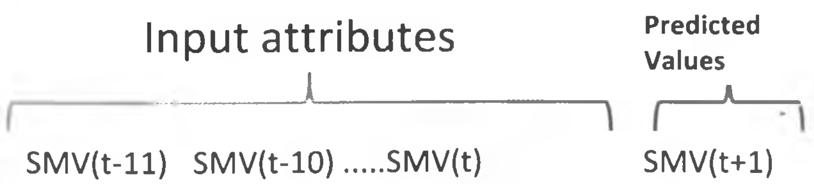


Figure 3.4: *Neurolab Prediction Model*

The second training pattern was taken from (SMV)[t-2 to SUM(SMV)[t]. And the output was (SMV)[t+1] as the desired output as shown in table 3.3. Therefore, the number of input nodes is equal to the number of input features of training examples. The training pattern for the network consisted of input attributes and a single output attribute. For example, the input features of the first input pattern consist of the sum of SMV counts for the year 2003. The output will be the SMV counts in January 2004. The input feature of the second input pattern consists of sum of SMV values for the last 12 months.

SMV(t-11)	SMV(t-10)	SMV(t-9)	SMV(t-8)	SMV(t-7)	SMV(t-6)	SMV(t-5)	SMV(t-4)	SMV(t-3)	SMV(t-2)	SMV(t-1)	SMV(t)	Predicted thefts SMV(t+1)
13	19	15	17	13	27	32	21	38	55	41	58	31
19	15	17	13	27	32	21	38	55	41	58	31	32
15	17	13	27	32	21	38	55	41	58	31	32	51

Table 3.3: Training windows for NeuralLab Solutions

3.8.5 Estimation of Accuracy

The idea behind the use of Neural Network in motor vehicle theft forecasting is that of allowing the network to map the relationship between time and actual theft.
The forecasts error for each pair of actual and forecasted theft can be given as:

$$\text{Error}=\frac{Actualtheft - predictedtheft}{Actualtheft} x100\% \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \textit{Equation} \quad (8)$$

After calculating forecast error, forecasting accuracy is given by:

Forecasting Accuracy= 100 - %age forecasting error …………………Equation (9)

Another important criterion used for the evaluation of the systems is generalization. A network is said to be generalized when the output is correct or close enough for an input which has not been included in the training set.

3.8.6 Further Zaitun Time Series Empirical Studies

Further experimental tests were conducted aimed at developing a working model to be used in forecasting. The tests included:

Determination of optimal hidden neuron

Determination of optimal transfer neuron transfer function

Determination of optimal learning rate

Chapter 4

Results and Discussion

Developments in technology have led organizations to capture and store large quantities of data. Machine learning and Data mining techniques has developed as tools that can be used to obtain patterns and trends from data warehouses.

The following discussion emanates from the following experiments:

1. Using WEKA machine learning techniques to mine rules from SMV data
2. Using WEKA FFNN to predict trends in Time Series SMV Data
3. Identification of an open source tool to predict and forecast time series crime trend.

   i. Predict SMV Crime using Neural Network Zaitun Application

   ii. Predict SMV Crime using Neural Network Tiberious

   iii. Predict SMV Crime using NeuralLab

4.1.1 Basic Statistical data Analysis

Figure 4.1, shows the SMV graphs from 2003 to 2008. A casual observation shows that motor vehicle theft in Nairobi increases between May to July every year. It was observed that year 2009 experienced less SMV counts.

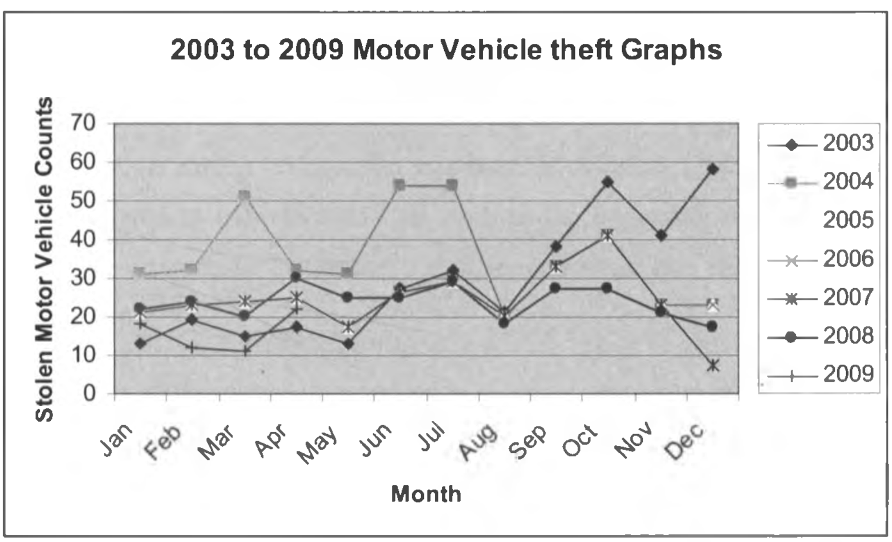


Figure 4.1: Motor vehicle theft trend

In 2006, Eastern Africa Police Chiefs Cooperation Organization (EAPCCO) with support from INTERPOL, held a successful international joint operation targeting stolen motor vehicles in East African countries of Kenya, Uganda and Tanzania. Since then, neighboring countries including Rwanda, Burundi Zambia and Malawi have made annual international joint operation against motor vehicle theft. This may have contributed significantly against the vice in the region since the exercise has disrupted the potential market and routes for stolen vehicle from Kenya to the neighboring countries.

In the years under review, it is easy to observe a common tread where there is a surge in thefts in the months of April to July. During this period, there are relatively high economic activities in Kenya and other eastern African partner states. This is especially so as with the government ministries (the highest spender) go on spending bliss to exhaust initial funding while awaiting allocations for new financial year. This is corroborated by Muchai A (2003).

Similarly, in preparation for the end of calendar year, there is a rise in economic-activities. A rise in car thefts could be associated with the anticipation by the robbers that they could steal vehicles and cash from the owners and get ready market.

Interestingly, the month of December generally has few SMV theft incidents. This could be as a result of extra vigilance from the LEA agencies during this season. The other reason could be that the perpetrators of the crime could be enjoying the 'fruit of their loot' with their families after carrying out thefts in the months of October and November. A similar observation was made by Muchai (2003) where she noted that crimes related to motor vehicle are high in the beginning of the year and reduces during the year up to December. This general trend line can be seen for the years 2006-2008.

Clearly, a trend is emerging where motor vehicle theft is high in the second and third quarter of the year. These quarters will require law enforcement agencies to raise surveillance and motor vehicle owners to take extra precautions to ensure safety of their cars.

4.2.0 Predict SMV Crime with Zaitun using trend analysis method

Results from this study were obtained after the data was run on Zaitun time series application.

Variable: Monthly counts

Observations: 72

Linear Trend Equation: Y_t=31.63-0.12152*t .... *Equation (10)*

Mean Squared Error (MSE): 99.200817

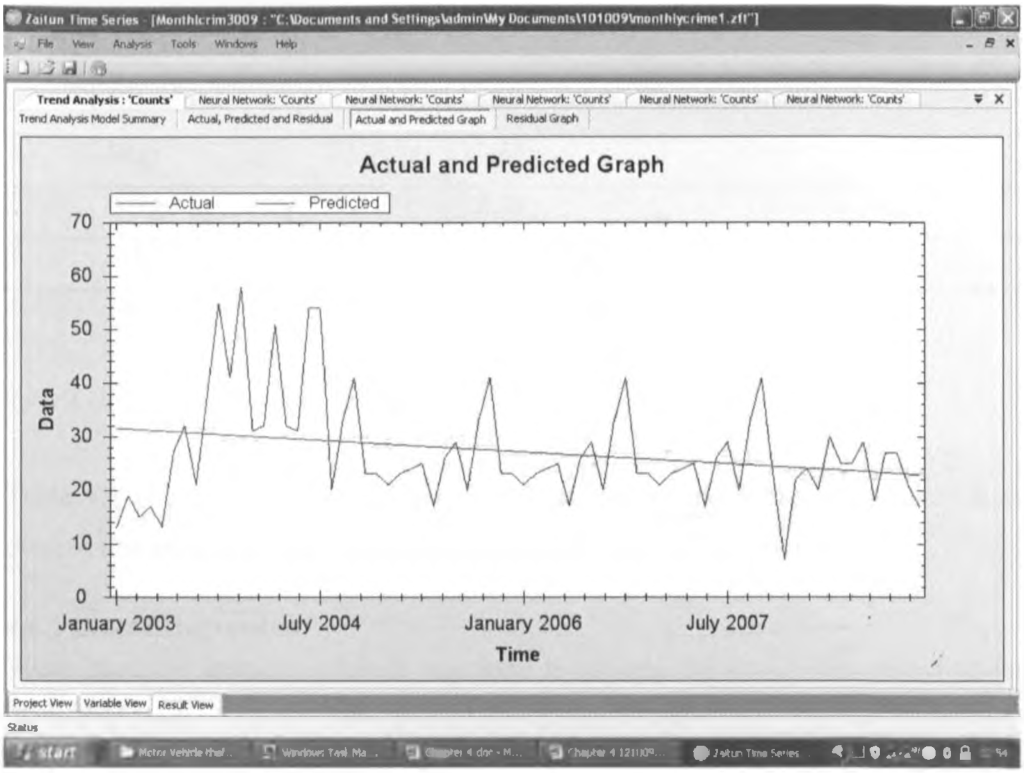


Figure 4.2: Linear Trend

It is clear from figure 4.2 that motor vehicle theft incidences have been on the downward trend since January 2003. It is also significant to note that between 2003 and 2004 the incidences were high and the economy was improving (http://imf.org/external/pubs/ft/scr/2007/cr07160.pdf Retrieved October 2009). Significant too is the huge drop of incidences during December 2007 and early 2008. This can be attributed to the post election violence experienced after the December 2007 disputed general election in Kenya.

The trend line also shows that Motor vehicle crime has been on the downward trend. This is consistent with reports by Kenya Police 2008 crime report and research results by Synovate K. Ltd (Formerly Steadman group) that has been releasing three-month crime statistics since 2003.

4.3.0 Using WEKA machine learning techniques to mine rules

An attempt to mine forecasting rules from raw data resulted in trivial rules which the motor vehicle crime expert disapproved. The algorithms which were tested included Weka's JRip, NNge, MLP and J48 Classifiers. The algorithms were to predict motor vehicle theft incident location.

	Classifier	% of correctly classified instances
1.	JRip	17.54
2.	NNge	12.28
3.	MLP	12.28
4.	J48	17.50

Table 4.1:Percentage of correctly classified instances

Table 4.1 shows the performance of some of the trained Weka classifiers. It is evident that the percentage of correctly classified instances is very low.

4.4.1 Linear Regression

Weka Machine learning software was used to process the data. The mathematical prediction model from the data can be described as follows:

SMV(t+1)=-0.2405 x SMV2 + 0.1641 x SMV3 + 0.391 x SMV9 +0.2819 x SMV11 + 10.6465 ....*Equation 11*

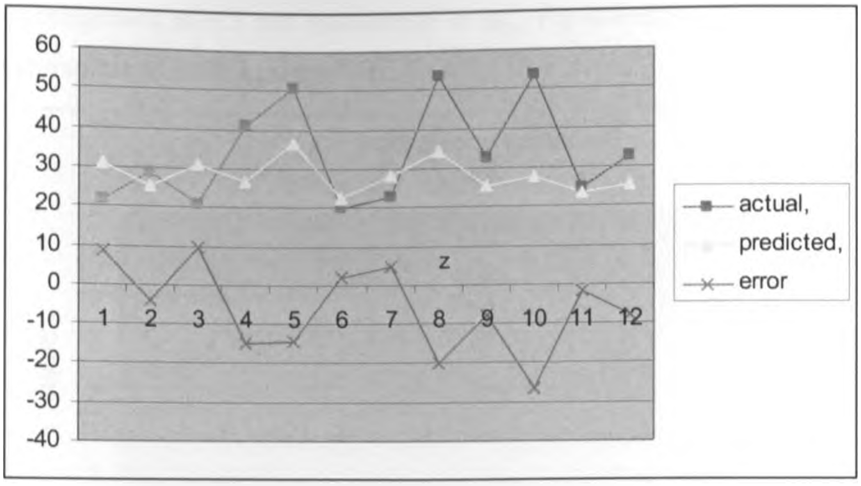


Figure 4.3: Figure of Actual and Predicted motor vehicle theft and prediction Error

Fig 4.3 shows graphical predictive performance of linear regression model.
The performance indicators of this model are shown in table 4.2 below:

Correlation coefficient	0.5253
Mean absolute error	10.1108
Root mean squared error	12.3795
Relative absolute error	91.8005%
Root relative squared error	83.6571
Total Number of Instances	12

Table 4.2: Linear Regression performance table

It is clear that this is a credible model that can be used for law enforcement decision support.

4.2.0 Using WEKA FFNN to predict trends

4.2.1 Basic Statistical data Analysis

The results of basic statistical analysis are as shown in section 4.1.0.

4.2.2 Neural Network model design

Multilayer perceptron with 1 hidden layer was used. The number of hidden nodes was varied between 1 and 11 and root means square error (RMSE) recorded using learning rates of 0.3 and momentum of 0.2. Fig 4.4 shows the RMSE variation with the number of hidden nodes

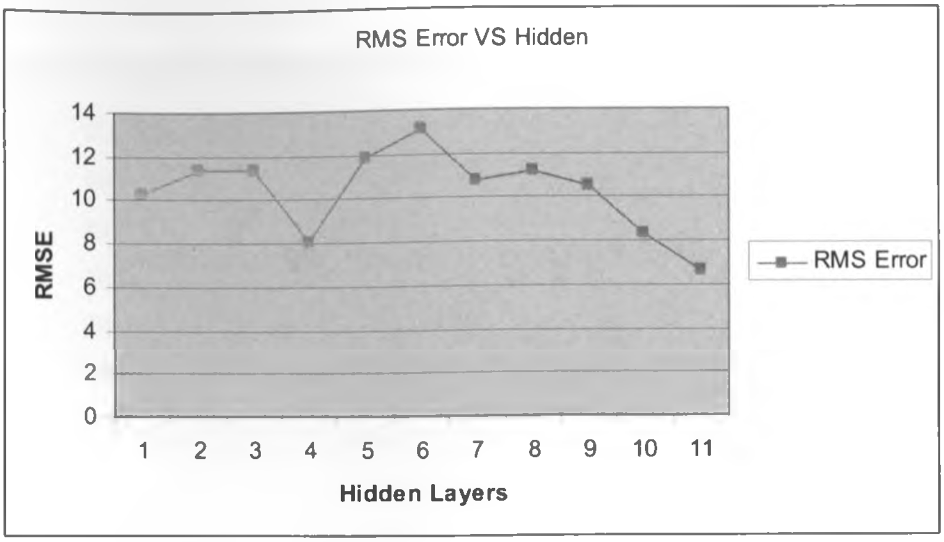


Figure 4.4: RMSE vs Hidden layers

It is observed from figure 4.4 that the lowest RMSE was obtained with 4 hidden layer nodes.

Determination of optimal Learning Rate

With 4 hidden nodes, the learning rate was varied from 0.1 to 0.8 as shown below.

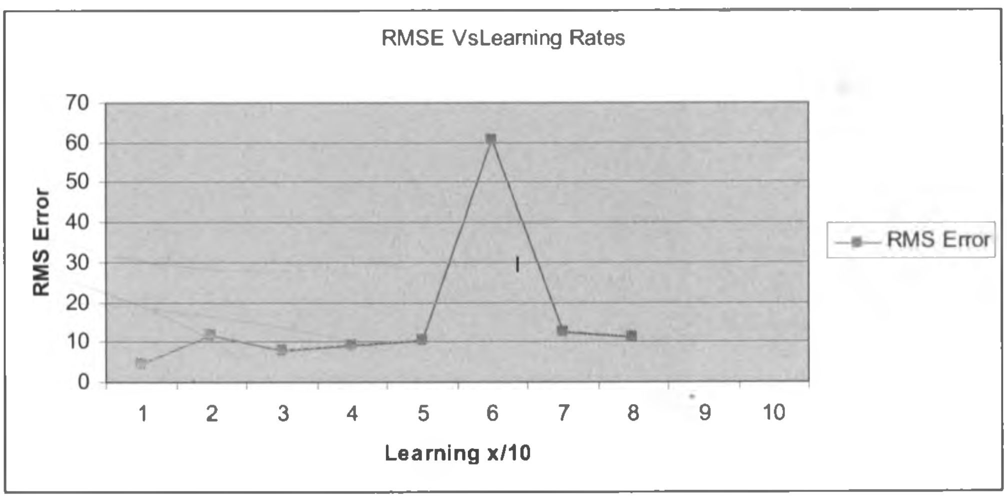


Figure 4.5: RMSE against learning rates

The results shown in figure 4.5 shows that a learning rate of 0.3 was optimal.

Therefore, the optimal Neural Network model was found to have:

12 inputs, 1 hidden layer with four nodes and 1 output as shown in figure 4.6 below:

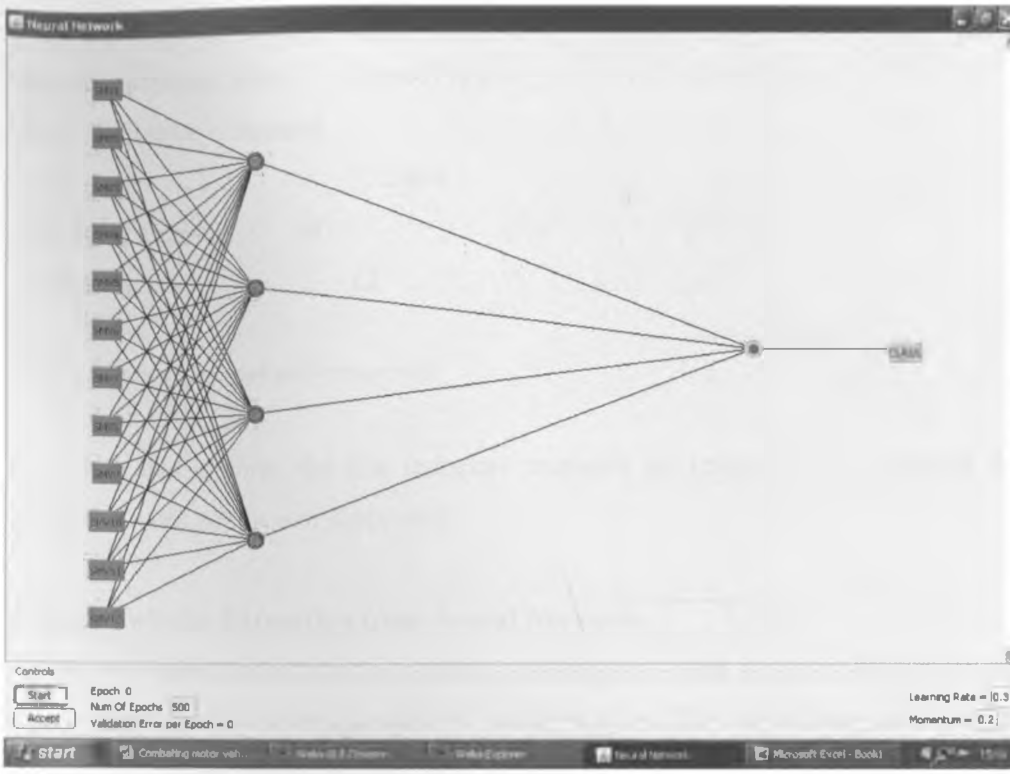


Figure 4.6: Weka Neural Prediction Model

Figure 4.7 shows the performance of the prediction model.

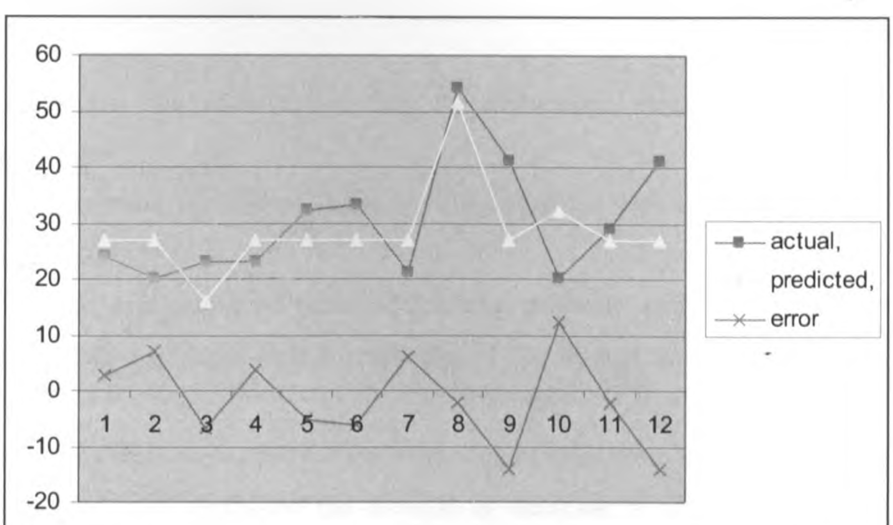


Figure 4.7: Graph of Neural Network performance.

Correlation coefficient	0.6456
Mean absolute error	6.8822
Root mean squared error	8.0564
Relative absolute error	83.4207%
Root relative squared error	73.3664
Total Number of Instances	12

Table 4.3: Neural Network performace table

It is clear that despite, the few instances available for training and validation, the Neural network performed fairly well.

4.3.0 Knowledge Extraction from Neural Networks

A few years ago, neural networks were considered as 'black boxes' as they were not able to explain the knowledge acquired in the weights after the training process. The goal of knowledge extraction is to find the knowledge stored in the network's weights in a symbolic form (fig.2.5). One main concern is the fidelity of the extraction process, i.e. how accurately the extracted knowledge corresponds to the knowledge stored in the network.

There are two main approaches for knowledge extraction from trained neural networks:

(1) extraction of 'if-then' rules by clustering the activation values of hidden state neurons.

(2) the application of machine learning methods such as decision trees on the observation of input-output mappings of the trained network when presented with data.

Decision trees were used for the extraction of rules from trained neural networks. The extracted rules explained the increase or decrease in SMV counts in the predicted period given the counts of the previous 12 months.

4.3.1 Decision Trees

Decision trees are machine learning tools for building a tree structure from a training dataset of instances which can predict a classification given unseen instances. A decision tree learns by starting at the root node and selects the best attributes which splits the training data. The root node then grows unique child nodes using an entropy function to measure the information gain from the training data. The process continues until the tree structure is able to describe the given data set. Compared to neural networks, they can explain how they arrive to a particular solution.

In this research, decision trees were used to extract rules from the trained neural networks. A dataset was obtained by presenting instances and record the generalization made by the output of the trained network. The decision trees was extracted from trained neural networks using the Weka j48 algorithm. For the decision induction, attributes and classification of the 66% training data and the attributes of the remaining 34% test data was used, but with their labels were determined by the trained networks.

A decision tree extracted from experiment is shown in Figure 4.8. It was observed that the rules defined by training knowledge-based neural networks are much simple. Clearly, neural network training has successfully refined and simplified the rules experts use for predicting the increase or decrease of motor vehicle loss.

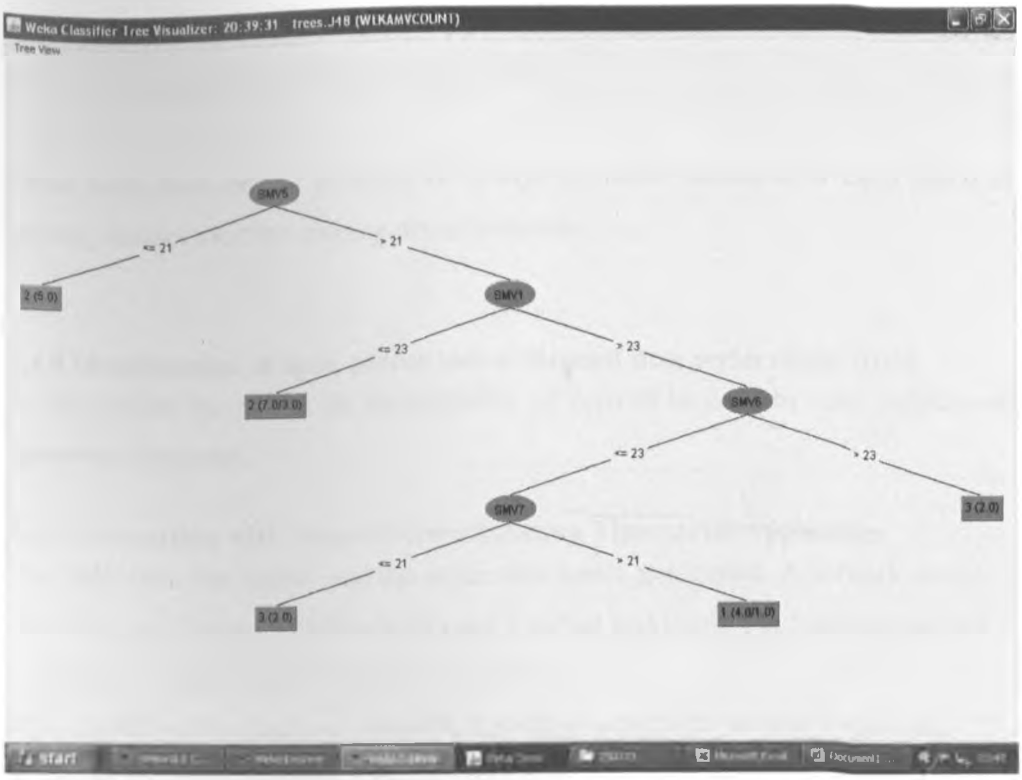


***Figure 4.8**: Weka J48 classification of trained NN output*

The knowledge obtained from the trained network about the increase or decrease of motor vehicle theft counts is summarized in the following rules

a) If SMV 5 is low then it is predicted that coming month will have average SMV count loss.

b) If SMV5 and SMV6 count is average then predicted month will experience high level of SMV crime.

c) If SMV5 and SMV1 is average then it is predicted that the coming months to have moderate crime rates of between 23 and 29 losses.

d) If SMV 5, SMV1, SMV7 and SMV6 counts are average then it is predicted that low vehicle thefts counts in coming months.

e) If SMV 5, SMV1, SMV6 counts are average then it is predicted that high vehicle thefts counts in coming months.

These rules describe the presence of change in motor vehicle theft trend based on existing dataset used for training neural networks.

4.4.0 Identification of open source tool to forecast time series crime trend

In this section the results on identification of tools to be used by Law enforcement agents are discussed.

4.4.1 Forecasting with Neural Network Zaitun Time series Application

The SMV data was loaded into the zaitun time series application. A network model consisting of 12 input 4 hidden layers and 1 output was used. The learning rate and momentum were set at 0.05 and 0.5 respective.

More models were used to evaluate the forecasting suitability of this application.

Some of the results from this study are as follows in figures 4.9 and 4.10.

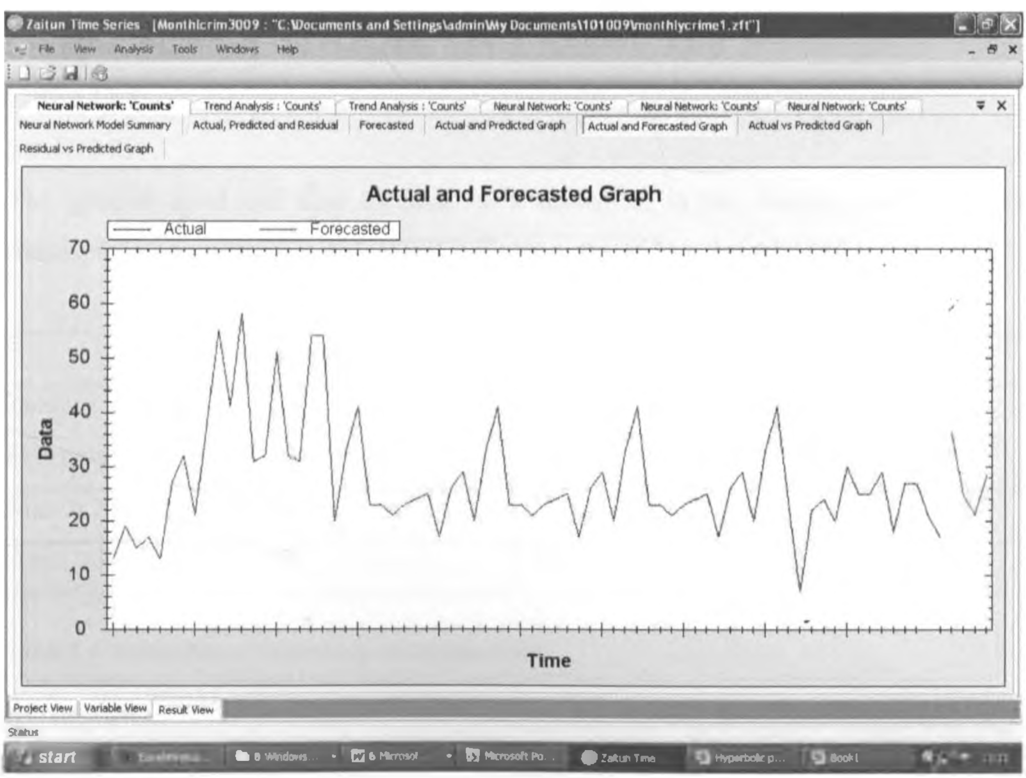


Figure 4.9. Forecasts for January 2009 to April 2009

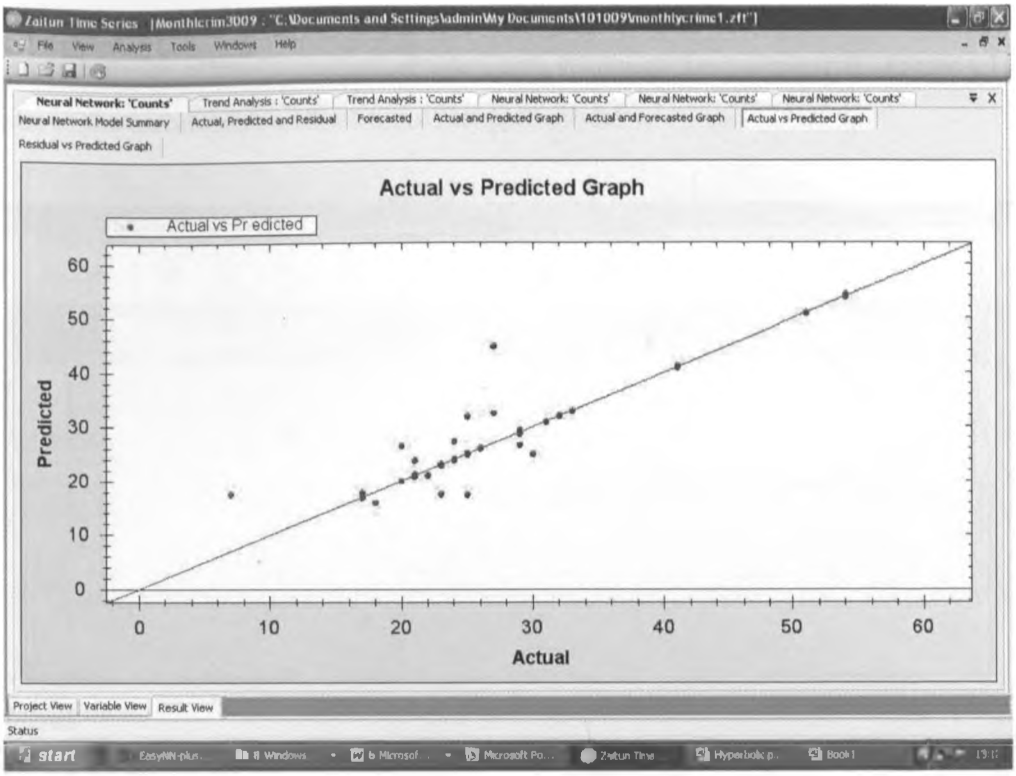


Figure 4.10. Graph of prediced and Actual SMV theft trend for Year 2009

The forecasts were almost the actual theft in January to April.

MSE : 11.8

MAE: 1.59

The general trend can also be seen as a downturn in the January and February followed by an upsurge in SMV theft in the months of March and April.

	Actual Theft	Predicted	Error
January 2009	18	36	100
February 2009	12	24	100
March 2009	11	21	10
April 2009	22	28	0

Table 4.4: Zaitun Neural forecasting percentage Error

4.4.2 Predict SMV Crime using Tiberious Neural Network

As observed earlier, this is commercial software used for data mining. However, the demo software allows some limited use which is enough for our research. Figure 4.11 shows the architecture of the Neural Network model.

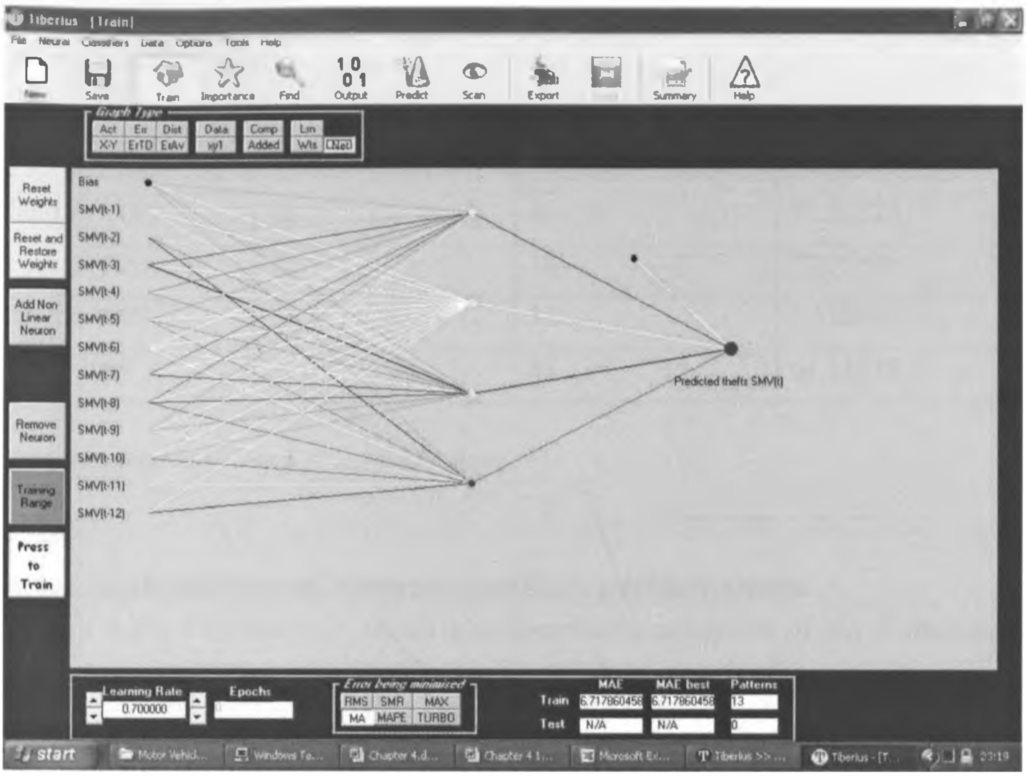


Figure 4.11: Tiberious Neural Network Model

	Actual Theft	Predicted	Error %
January 2009	18	19	5.56
February 2009	12	31	158.33
March 2009	11	33	200
April 2009	22	27	22.7

***Table 4.5**: Tiberious Neural forecasting % Error*

It is easy to see from table 4.5 that the error demonstrated by Tiberious Neural Network is relatively high. But the general trend in Crime trend is forecasted. It is expected that the April motor vehicle theft will go up.

4.4.3 SMV Crime Prediction with NeuralLab

NeuralLab is a free neural network software used for research simulation. NeuralLab analysis produced a Mean square error of 0.000307819 is incredible. However the learning is based on annealing simulation technique. Table 4.6 shows the performance of NeuralLab software.

	Actual Theft	Predicted	Error %
January 2009	18	5	72.22222
February 2009	12	18	33.3333
March 2009	11	12	5.55556
April 2009	22	11	61.11111

Table 4.6:. NeuralLab Neural forecasting % Error

4.4.4 Artificial Neural Network solution performances

Figures 4.13, 4.14 and 4.15 shows a performance comparison of the 3 simulation networks. It is clear that Zaitun Neural Network and NeuralLab outperforms Tiberious Neural network model.

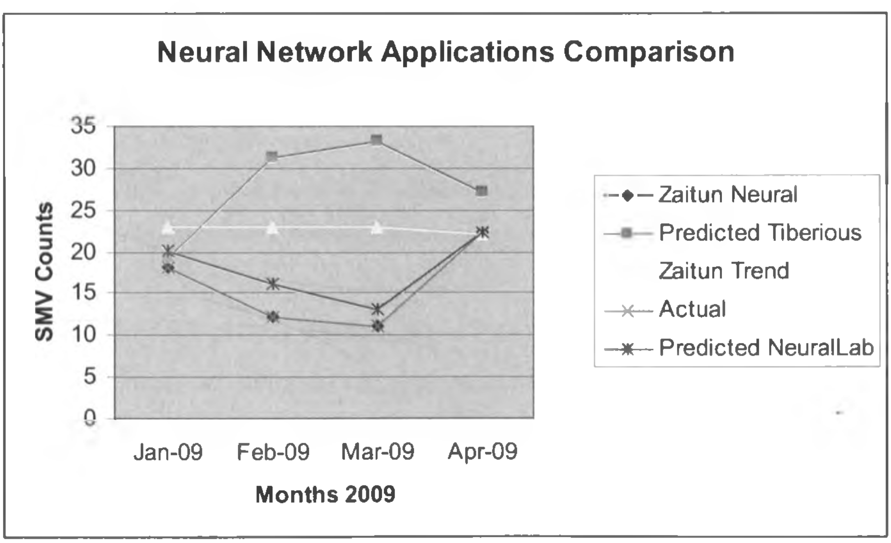
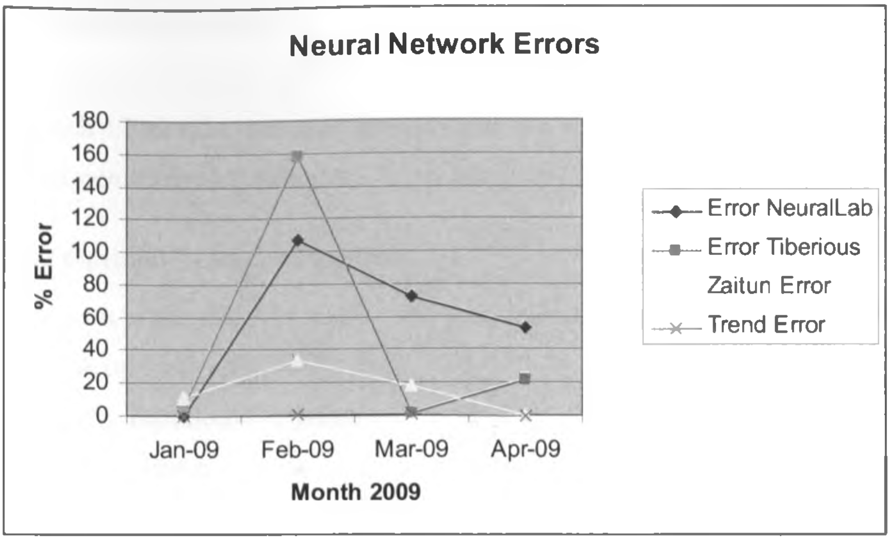


Figure 4.12: A comparison of forecasted SMV counts for 2009



4Figure 4.13: A comparison of forecasted SMV errors for year 2009

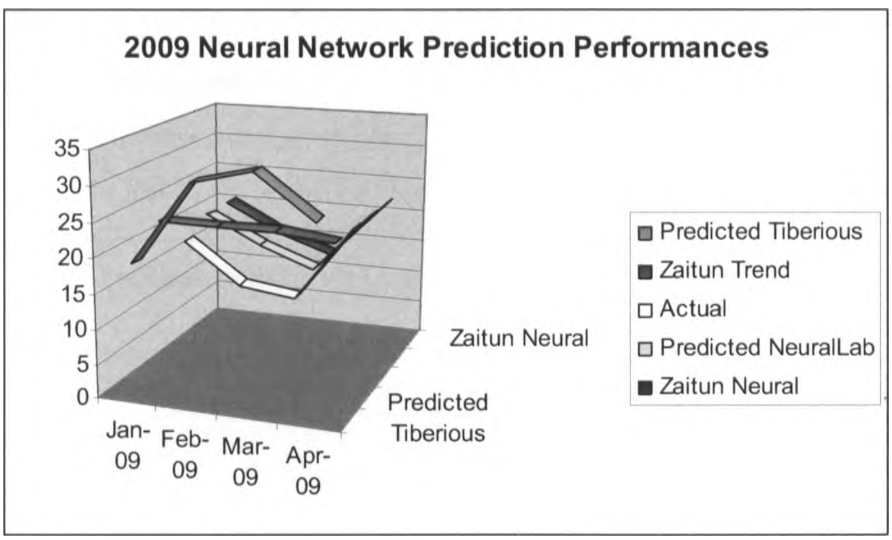


Figure 4.14: Neural Network forecast performance in relation to the trend line.

4.5.0 Further Empirical Study Zaitun Neural Network Application

After the study to identify an application that can be used by law enforcement agencies, it was found critical to identify the most optimal neural settings for the Zaituni neural network.

4.5.1 Input Layer size

The number of input neurons depends on the size of the sliding window. In this research 12 months was used to determine the window size and consequently predict the 13month (see 4.2.2).

4.5.2 Optimal Transfer Function

As shown in the table 4.6 below, the best transfer function Bipolar Sigma was found to be the optimal.

Transfer function	MAE	MSE
Bipolar Sigmoid	4.1	27.32
Sigmoid	5.04	43.93
Hyperbolic Tangent	4.286	28.99
Semi Linear	5.52	59.74

Table 4.7: Performance Errors against the number of Zaitun Neural nodeTransfer function

4.5.3 Optimum Learning Rate

To determine the optimum learning rate the learning rate was varied from 0.01 to 0.12. The results are shown on figure 4.17.

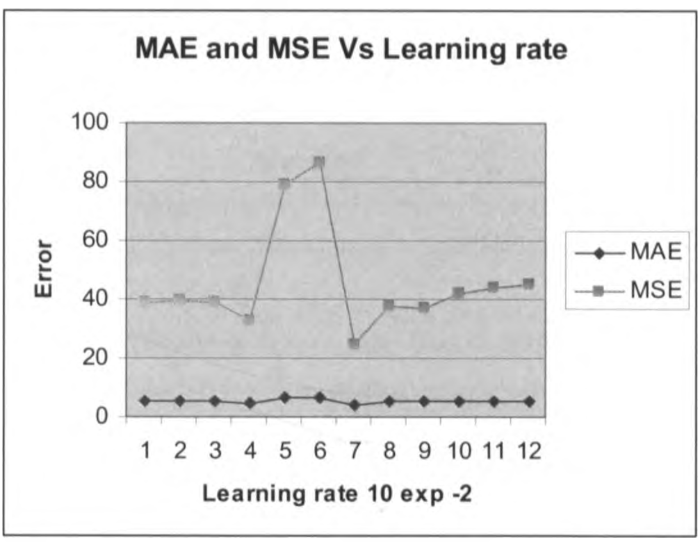


Figure 4.15 Performance Errors against the number of NeuralLab Neural Nodes

The rate of 0.04 was found to be the optimal learning rate.

4.5.4 Other parameters

Other parameters settings found appropriate are as given on table 4.7.

Network Parameters	Optimal value
Learning rate	0.005
Momentum	0.02
Maximum epoch	1000

Table 4.8: Other Performance parameters

4.6.0 Zaitun Forecasting Performances Tests

This is the performance of Zaitun time series in actual forecast of motor vehicle theft in months of January, February, March and April 2009.

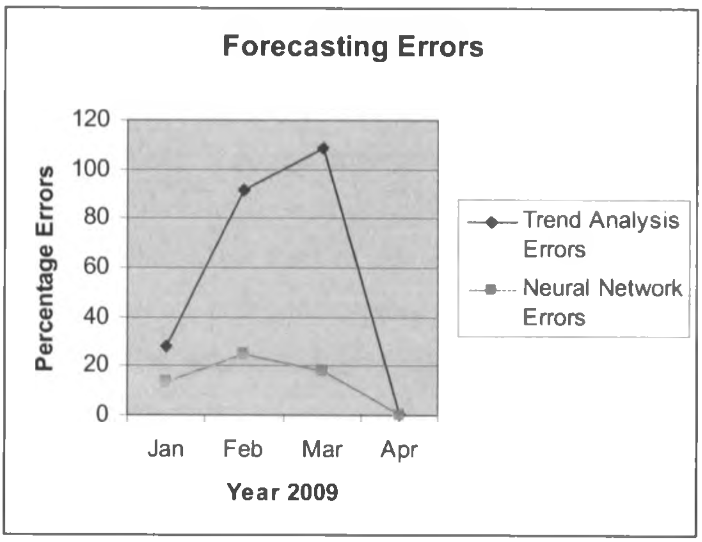


***Figure 4.16**: Optimum Zaitun Neural network Model forecasting errors for year 2009*

Figure 4.16 above shows that Neural Network has the least errors. The Neural Network Model demonstrated an accuracy of 85.9%.

Factors that could contribute to the high level of accuracy include:

1. Zaitun Time analysis application is an open source software. It is therefore open and available to many programmers who contribute daily to ensure stability.
2. Zaitun Time Analysis application is developed using Microsoft C sharp and .Net Framework.
3. Zaitun Time Analysis application uses Encog-CS neural network engine.

4.7.0 Summary of results

Each of the Neural models was able to predict the general trend in crime for the month of January, February, March and April for the year 2009.

There are several possible reasons for the divergences between actual and expected crime levels. The reasons advanced include that the models applied is on recorded crime. There are possibilities that some motor vehicles were recovered after theft reports were made. However, after recovery the records were not amended.

The security awareness practiced by the owners and stressed by the Insurance companies require the installation of Anti-motor vehicle theft devices including car alarms, fuel cut system etc could also explain the down trend of the number of stolen motor vehicle. It is also important to note that new motor vehicles are preinstalled with state of the art security gadgets.

Even as this is not part of this research, it is possible that motor vehicle theft perpetrators are now concentrating on other type of crimes. As a matter of fact, it is possible that perpetrators have shifted to new crime areas like cyber crime and kidnappings.

From the results of this study, it can also be concluded that recent government and international interventions has lead to gradual successful reduction in motor vehicle crime.

The multiple experiments on different Neural network implementations show it is not possible to use an optimized model from one solution to the other. For instance an optimized model from WEKA cannot be used by Zaitun Time Series application.

Chapter 5

Conclusion

Recent advances in the use of artificial neural networks have been exciting. Artificial neural network provides a level of flexibility and adaptability which has not been exploited this far. This thesis has developed a uni-variate crime forecasting model based on artificial neural network. The results confirm that ANN can learn hidden patterns in crime data and forecast future patterns and trends as demonstrated by the performance of SMV data.

It further identifies an open source solution that potentially can be used to forecast time series crime events by law enforcement agencies. However, it is also noted that neural applications has different implementation that require users to identify optimum models for forecasting purposes.

This study also demonstrates that artificial neural network has matured enough to be applied in crime forecasting in developing countries. It is therefore recommended that law enforcement consider applying neural network technology to proactively fighting crime. To achieve this however, there is need for the country to embrace the information technology before practical forecasting can be done.

Neural networks are successful in predicting trend in motor vehicle theft based on information on time series historical data. The prediction accuracy can be increased by having more training samples in the dataset.
Decision trees have been useful in knowledge extraction from trained neural networks. We have obtained rules for forecasting future motor vehicle theft trends; these rules can help experts in better understanding the domain knowledge about the motor vehicle theft patterns. This knowledge discovery paradigm can hence be applied to other domains in combating crimes.

5.1.0 Research Achievements, contributions and limitations

The research undertaken identifies Neural Network tools that can be used by crime analyst to identify patterns, predict and forecasts trends in crime. This knowledge can be used to support sound decision-making process in combating Motor vehicle thefts. Neural Networks are used in this research for short term prediction on motor vehicle theft as they are natural candidates for forecasting models. The quality of trend forecasting observed confirm the research hypothesis that artificial intelligence can be used to forecast motor vehicle crime trends.

The success of this research is about how good the model was able to predict the motor vehicle theft trend projection. It does not attempt to predict the number of crimes as there are many factors that affect crime. This is a limitation of uni-variate crime forecasting. The factors that have significant influence on crime include economic growth, criminal justice policies and penalties, parenting, crime prevention measures and programmes. It is unlikely that all these can be captured in a single model.

The dynamic prediction of theft of neural network has one main drawback: the amount of data required to properly train the neural network is low. This is so because inactive police data is destroyed after 5 years. This affects the number of patterns available to train the neural network model. It is worth mentioning too that law enforcement in Kenya have not embraced computers as a working tool. In recent times, projects aimed at supplying computer and promote computer use in Kenya Police to fight crime has been dogged with claims of corruption. While donor sponsored projects are sometimes stopped mid term as they are attached to donor conditions like good governance, human rights and transparency.

5.2.0 Future Research Areas

Future research areas include modeling of scenarios to facilitate prediction and forecasting other crimes using multivariate inputs and leading indicator models including political, professional intervention, land use, infrastructure, population literacy etc.

References

Afolabi A., Olatunji B. & Ajayi A. (2008). *Electricity load forecasting using artificial Neural Networks*. Journal of Engineering And Applied Sciences 3(2):149-158.

Akumu Jane (2009). Patnership for clean fuels and vehicles: africa activities. December 2009. Retrieved from http://www.unep.org/pcfv/PDF/MotorVehicleTraining_UNEP.pdf

Author Unknown (Last accessed on 10 Jan 2010) http://www.aganoconsulting.com/businessnews/wp-content/uploads/2009/03/12_africa_report09_kenya.pdf Country Report for Kenya

Bigus JP (1996). *Data mining with neural networks solving business problems from Application development to decision support*. McGraw-Hill Companies, US

Chandra R.& Omlin C. W. (2007). *Knowledge discovery using artificial neural networks for a conservation biology domain*, Proceedings of the 2007 International Conference on Data Mining, Las Vegas, USA, In Press, 2007.

Chantasut N., Charoenjit C., Tanprasert C. (2004), *predictive mining of rainfall:l Predictions using artificial neural networks for chao phraya river*, AFITA/WCCA Joint Congress on IT in Agriculture. Bangkok, Thailand. August 9-12 2004. (Electronic Version) Retrieved from http://www.nectec.or.th/NTJ/No15/papers/11.pdf

Cohen, J., W. L. Gorr, and A. Olligschlaeger (2007). L*eading indicators and spatial Interactions: a crime forecasting model for proactive police deployment Geographical Analysis*, January 2007, Vol. 39, Issue1, pp. 105-127.

Corcoran J. Wilson D. & Ware A. (2003). *Predicting the geo-temporal variations of crime and disorder*. International journal of forecasting 19 623-634.

Felson, M., & Poulsen, E. (2003). *Simple indicators of crime by time of day.*(Electronic Version) International Journal of Forecasting 19 623-634

Gorr W. & Harries R. (2003) *Introduction to crime forecasting*. International Journal of Forecasting 19 551-555.

G orr, W. L., Olligschlaeger, A. M., & Thompson, Y. (2003). *Short-term time series forecasting of crime* (Electronic Version). International Journal of Forecasting 19 623-634, Gorr W. Harries R. (2003). *Introduction to crime forecasting*. International journal of forecasting 19 551-555

Gorr W. L Olligschlaeger A.M, Szcypul J. & Thompson Y. (1999). Forecasting Crime.

Graupe D. (2007). *Principles of artificial neural networks (*2nd Edition). Advanced Series on Circuits and Systems – Vol. 6 University of Illinois, Chicago, USA

Haydari, Z., Kavehnia, F., Askari, M.& Ganbariyan, M. (2007). *Time series load modelling and load forecasting using neuro-fuzzy techniques*. In 9th International Conference - Electrical Power Quality and Utilization, Espanha, Barcelona, Outubro 2007.

Haydari Z. Kavehnia F. ,Askari M & Ganbariyan M (2007.) *Time-series load modelling and load forecasting using neuro-fuzzy techniques.* 9th International Confrence. Electrical Power Quality and utilization. Barcelona 9-11 October 2007

Haykin, S. (1999). *Neural networks a comparative foundation*. 2nd edition, Prentice Hall, Upper Saddle River, NJ, USA.

Hudson B, Whitley D, Browne A & Forda M. (2003). *Extraction of Comprehensible Logical Rules from Neural Networks. Application of TREPAN in Bio and Chemoinformatic.* Centre for Molecular Design, Institute of Biomedical and Biomolecular Sciences, University of Portsmouth,Portsmouth, Hants; PO1 2DY, UK

Ian H. Witten & Eibe Frank, *Data mining: practical machine learning tools and techniques with Java implementations*, Morgan Kaufmann Publishers Inc., San Francisco, CA, 2000

Jonas, J., Harper, J. (2006). *Effective counterterrorism and the limited role of predictive data mining*, *Policy Analysis*, CATO Institute; Washington, DC, No. 584, pp.1-11.

Kajitani Y, K. W. Hipel and A. I. McLeod (2005) *Forecasting nonlinear time series with feed-forward neural networks: a case study of canadian lynx data.* Journal of Forecasting J. Forecast. 24, 105–117 (2005)

Kamruzzaman S, Islam M (2005). Extraction of symbolic rules from artificial Neural networks. December 2009. World Academy of Science, Engineering and Technology. Retreived from http://www.waset.org/journals/waset/v10/v10-51.pdf

Kegoro G. (2002). *Laundering Patterns in Kenya. Institute for Security Studies Money* Retrieved October 2009, from http://www.iss.co.za/pubs/Monographs/No90/Chap5.html

Kenya Police (2008) The. Annual crime report for the year 2008 (last accessed on November 2009). www.kenyapolice.go.ke/resources/2008%20Crime%20Statistics.pdf

Mohamed M. (2005). A *new approach for extracting fuzzy rules using artificial neural networks. A Msc Thesis*. Cairo University Institute Of Statistical Studies And Research Department Of Computer And Information Science

Olson, David L.; Delen, Dursun (2008) *advanced data mining techniques* Springer.

Rangaswamy N, Chaiboonsri C. *Time series forecasting*: International Tourist Arrivals to Thailand (Working paper No. 6/2006)

Shabri A.N (2001). C*omparison of time series forecasting methods using neural networks and box-jenkins model*

Steiner, M. T. A.; Steiner Neto, P. J.; Soma, N. Y.; Shimizu, T.; Nievola, J. C. (2006). Using neural network rule extraction for credit-risk evaluation. International Journal of Computer Science and Network Security, 6(5A). Brazil.

Synovate K. Ltd. (2009). SPEC Barometer Results QUARTER 3: OCTOBER 2009 (Last Accessed October 2009) http://www.google.co.ke/url?sa=t&source=web&ct=res&cd=1&ved=0CBcQFjAA&url=http%3A%2F%2Fwww.nation.co.ke%2Fblob%2Fview%2F-%2F674224%2Fdata%2F108357%2F-%2Fk1nq8v%2F-%2FSynovate%2Bpolls.ppt&rct=j&q=Synovate+polls+kenya+crime+2009&ei=X9n2S8WZLY3w0gTgz4XqBw&usg=AFQjCNEvQOp6OOqvY5aVBWsfP7GWWElo9Q

Ye N. (2003). *The handbook of data mining* (Rev. ed.). Lawrence Erlbaum Associates, Inc., Publishers

Zhang, G. P. (2007). A neural network ensemble method with jittered training data for time series forecasting. Information Sciences. vol. 177, pp. 5329-5346 (2007).

Appendix A

A.0 WEKA Neural Network Model

=== Run information ===

Scheme:	weka.classifiers.functions.MultilayerPerceptron	Learning Rate	0.3	Momentum	0.2	
Relation:	WEKAMVCOUNT					
Instances:	60					
Attributes:	13					
	SMV1					
	SMV2					
	SMV3					
	SMV4					
	SMV5					
	SMV6					
	SMV7					
	SMV8					
	SMV9					
	SMV10					
	SMV11					
	SMV12					
	CLASS					
Test	mode:	split	80.00%	train,	remainder	

=== Classifier model (full training set)

Linear	Node	0	
	Inputs	Weights	
	Threshold	23.01851	
	Node	1	-1.73872
	Node	2	-2.18188
	Node	3	1.130068
	Node	4	5.750679
Sigmoid	Node	1	
	Inputs	Weights	
	Threshold	-7.41285	
	Attrib	SMV1	11.13704
	Attrib	SMV2	-9.60957

	Attrib	SMV3	-7.36493
	Attrib	SMV4	23.88437
	Attrib	SMV5	32.22963
	Attrib	SMV6	26.73318
	Attrib	SMV7	9.333896
	Attrib	SMV8	14.31987
	Attrib	SMV9	15.12976
	Attrib	SMV10	10.83386
	Attrib	SMV11	6.003921
	Attrib	SMV12	16.70887
Sigmoid	Node	2	
	Inputs	Weights	
	Threshold	-8.68222	
	Attrib	SMV1	11.77255
	Attrib	SMV2	-8.54907
	Attrib	SMV3	-6.55193
	Attrib	SMV4	25.50603
	Attrib	SMV5	32.64726
	Attrib	SMV6	26.35765
	Attrib	SMV7	10.36915
	Attrib	SMV8	14.53316
	Attrib	SMV9	13.48142
	Attrib	SMV10	10.83494
	Attrib	SMV11	3.70386
	Attrib	SMV12	16.91844
Sigmoid	Node	3	
	Inputs	Weights	
	Threshold	-15.9917	
	Attrib	SMV1	15.80227
	Attrib	SMV2	-6.55015
	Attrib	SMV3	-5.11283
	Attrib	SMV4	25.46283
	Attrib	SMV5	35.02158
	Attrib	SMV6	29.32209
	Attrib	SMV7	8.495254
	Attrib	SMV8	11.76023
	Attrib	SMV9	13.58355
	Attrib	SMV10	6.328661
	Attrib	SMV11	7.770283
	Attrib	SMV12	17.11825
Sigmoid	Node	4	

Inputs	Weights	
Threshold	-0.23243	
Attrib	SMV1	5.699444
Attrib	SMV2	-15.8868
Attrib	SMV3	-15.6296
Attrib	SMV4	19.47558
Attrib	SMV5	32.27388
Attrib	SMV6	24.52564
Attrib	SMV7	8.876426
Attrib	SMV8	19.95669
Attrib	SMV9	20.41355
Attrib	SMV10	18.59831
Attrib	SMV11	11.28952
Attrib	SMV12	18.49166

Class

Input

Node 0

Time taken to build model: 0

=== Predictions on test split ===

inst#,	actual,	predicted,	error
1	24	26.738	2.738
2	20	26.738	6.738
3	23	16.207	-6.793
4	23	26.738	3.738
5	32	26.776	-5.224
6	33	26.738	-6.262
7	21	26.738	5.738
8	54	51.632	-2.368
9	41	26.738	-14.262
10	20	32.202	12.202
11	29	26.738	-2.262
12	41	26.738	-14.262

=== Evaluation on test split ===

=== Summary ===

Correlation coefficient 0.6456

Mean	absolute	error	6.8822		
Root	mean	squared	error	8.0564	
Relative	absolute	error	83.4207 %		
Root	relative	squared	error	73.3664	%
Total	Number	of	Instances	12	