2017

# An Algorithm for predicting road accidents based on traffic offence data

Levice Obongo Jwan
*Faculty of Information Technology (FIT)*
*Strathmore University*

Follow this and additional works at http://su-plus.strathmore.edu/handle/11071/5608

Recommended Citation

# AN ALGORITHM FOR PREDICTING ROAD ACCIDENTS BASED ON TRAFFIC OFFENCE DATA

## JWAN, LEVICE OBONGO

**Submitted in Partial Fulfilment of the Requirements for Degree of Master of Science in Computer-Based Information Systems at Strathmore University.**

**Faculty of Information Technology**

**Strathmore University**

**Nairobi, Kenya**

**JUNE, 2017**

## Declaration

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

© No part of this thesis may be reproduced without the permission of the author and Strathmore University.

............................................................ [*Name of Candidate*]

............................................................ [*Signature*]

............................................................ [*Date*]

## Approval

The thesis of Jwan, Levice Obongo was reviewed and approved by the following:

Dr. Vitalis Ozianyi,
Senior Lecturer, Faculty of Information Technology,
Strathmore University


Dr. Joseph Orero,
Dean, Faculty of Information Technology,
Strathmore University


Prof. Ruth Kiraka,
Dean, School of Graduate Studies,
Strathmore University

## Dedication

I dedicate this work to my family. The journey has been quite long. The days I spent away from home due to research were among my hardest. I however appreciate your sacrifice and support all along. I wish to dedicate this thesis to you; my uncle Dr. Julius Jwan for your support and encouragement that motivated me all through, aunt Sella Kisaka for her unconditional love and prayers throughout the journey and my brother Ian Jwan for being supportive. Without your support, I would not have completed my research. A special debt of gratitude is due.

# Acknowledgements

# Abstract

Drivers with multiple records of road traffic violations for instance speeding, driving under influence of alcohol and using mobile phones while driving have been considered as a high risk group for possible involvement in road accidents. Studies have shown that there are links between these reckless behaviors and road accidents. It is therefore critical that such drivers be identified early in advance to eliminate that likelihood. Currently, the road traffic offence data collected by National Transport and Safety Authority for instance speeding and drunk driving data is solely used for reporting and prosecution hence not adequately utilized in ensuring road safety. Effective utilization of these data can positively impact road safety management since authorities can put in place mitigation mechanisms in order to prevent the frequent road accidents. The algorithm-based system developed in this study makes use of traffic offence data to predict the likelihood of a driver causing road accident. Data was gathered using close-ended questionnaires and interviews. The questionnaires and interviews intended to determine causes of road accidents and specific aspects about; booking an offender, relaying of traffic accident data and the need for a system among users within the transport sector in Kenya. Three categories of respondents were used; the National Transport and Safety Authority, the Kenya Police and the motorists. Similar questionnaires were given to the police and the NTSA officials while the motorists had their own set of questions. From the research, it emerged that the major causes of accidents in Kenya were; speeding, dangerous overlapping and drunk driving. Of the 37 respondents; 22 supported the algorithm-based system, indicating a 59.47% approval for the system. The implication of the research is that there will be more people booked for traffic offences and it will be possible for law enforcement to know the risk level of a driver based on the offences committed.

*Keywords: road traffic offence, predict, road accident*

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF ACRONYMS

NTSA        :                National Transport and Safety Authority

PSV         :                Public Service Vehicle

WHO         :                World Health Organization

**CHAPTER 1: INTRODUCTION**

**1.1 Background**

Road traffic crashes occur across every country world-wide. Each year, millions of people lose their lives as many more succumb to severe injuries. Increased economic development has resulted in an increase in the transport sector in the country. However, the development in the transport sector has been a dilemma in road safety by increasing the number of road accidents. Road Traffic accident is an unplanned occurrence of auto crash that may result in injuries, loss of lives and properties (Kual et al., 2005).

According to World Health Organization (2009), road traffic accidents cause approximately 1.2 million deaths, and between 20 and 50 million sustaining non-fatal injuries worldwide annually. Nevertheless, road traffic accidents have a great impact on the global economy. World Health Organization (2009) estimates that governments across the globe lose US $518 billion due to road traffic injuries. In Kenya, the cost to the economy from road traffic accidents is in excess of US$ 50 million exclusive of the actual loss of life.

Muchene (2013) reports that most of the road traffic accidents are caused by poor driving habits, that is, the total disregard for traffic laws and regulations. It is estimated that a large percentage of the accidents are contributed by such factors as speeding, drunk driving, use of mobile phones while driving, and plain recklessness. Similarly,Ipsos Synovate (2012) reported that 71% and 58% of Kenyans attributed the causes of accidents to speeding and drunk driving respectively.

In addition, based on the Accident Cause Code Classification, a report by Kenya Police reveal that 85.5% of crashes are as a result of poor driving behavior, of which driver error constitutes 44.4%, pedal cyclists 7.2% , pedestrians and passengers 33.9% (Kenya Police Service, 2016).

Nevertheless, a study by Goldenbeld, Reurings , Van Norden and Stipdonk (2013) on crash involvement of motor vehicles in correlation with the number and severity of traffic offences reveals that rise in traffic offence frequency concurs with a stronger increase in relative crash involvement.

In their study, Sun, Das, Sk and Wang (2013) suggested that drivers with recurring crash history should be targeted for specialized safety programs regularly through sensitization and

regulations. For instance, a motor vehicle registration office in a given country could work closely with the enforcement institutions to establish a driver's license reviewing program that has authority to send warnings or to suspend a driver's license, or to ask the driver undertake a mandatory safety class if the driver has had multiple crashes within a short period of time.

## 1.2 Problem Statement

The cost of fatalities and injuries due to road traffic accidents have a significant impact on the well-being of the society as well as socio-economic development. In Kenya, an estimated 3,000 deaths from road crashes occur annually despite the declining death rates in high income countries. It is reported that unless immediate action is taken, road accidents will be the leading cause of deaths in the future (National Transport and Safety Authority, 2016).

The Kenya Police and NTSA have a mandate of ensuring road safety. The two institutions thus collect huge amount of traffic data relating to speeding and drunk driving. However, these entities provide inconsistent data which cannot be used in determining patterns in road accidents. In addition, the data collected is merely used for reporting hence not effectively used in ensuring road safety as there are increasing cases of recklessness among drivers especially those in public transport.

## 1.3 Research Objectives

i. To evaluate factors contributing to road accidents in Kenya,
ii. To review related systems and models for predicting the likelihood of a driver causing an accident,
iii. To develop an algorithm for predicting the likelihood of a driver causing an accident,
iv. To test and validate the developed algorithm.

## 1.4 Research Questions

i. What are the factors contributing to road accident in Kenya?
ii. What are the existing algorithms and models for predicting the likelihood of a driver causing an accident?
iii. How can the developed algorithm be used to predict the likelihood of a driver causing an accident?
iv. How can the developed algorithm be tested and validated?

**1.5 Justification**

With the growing number of road fatalities across the globe, several measures have been put in place to ensure road safety. In addition, a number of models and systems have been proposed by various scholars to solve this global issue. However, these measures are in themselves inadequate to deal with the vice as most of them are reactive.

This research therefore aimed at providing an efficient and cost effective alternative to managing road safety by introducing use of a system that predicts the likelihood of a given driver causing an accident based on traffic offence data; hence high risk drivers can be identified early enough and appropriate actions taken to ensure road safety. This is a proactive approach to road safety.

**1.6 Scope**

The research focused on matatu drivers within the environs of Nairobi thus the findings may not be generalized to other areas. In addition, the study majorly focused on speeding, reckless overtaking and drunk-driving data collected by NTSA.

**1.7 Limitations**

Since matatu drivers were part of the population from which data was collected, the researcher was limited to use of interviews in data collection as most of them were always on the move.

## CHAPTER TWO: LITERATURE REVIEW

### 2.1Introduction

As noted from the previous chapter, road safety is a major issue not only in developing countries but globally. This Chapter explores the status of road safety in Kenya and other parts of the world by reviewing prevailing statistics, methods of collecting road traffic offence data as well as factors contributing to road accidents in Kenya and around the globe. This review will also cover related models and designs for predicting road accidents.

### 2.2 Status of road safety

Road traffic injuries remain an important public health problem at national, regional and global level. It is indeed saddening that despite the measures that have been put in place by various authorities across the globe, the number of road traffic accidents is still on the rise. The increase of accidents especially in developing countries has been attributed to increase in number of motor vehicles. World Health Organization estimates that 1.27 million people die in a year as a result of road accidents thus impact the economy of a country.

### 2.2.1 Status of road safety around the world

World Health Organization (2009) reported that over 90% of the world's fatalities on roads occur in low income and middle countries which have less than half of the world's vehicles. Shabeer and Banu (2012) notes that these accidents result from use of mobile phones while driving, which reduces concentration because of both physical and cognitive distraction. According to the report on status of road safety by WHO, approximately 62% of reported road traffic deaths happen in 10 countries including India, China, United States of America, Russia, Brazil, Iran, Mexico, Indonesia, South Africa and Egypt. In the United States for instance, 30,000 deaths were caused by accidents, and 2.5 million people remained injured. It is thus worthy to note that if this trend continues, deaths as a result of road accidents shall supersede other causes such as diseases.

### 2.1.2 Status of road safety in Kenya

Kenya has one of the highest road mortality rates in relation to vehicle ownership in the world, with an average of 7 deaths from the 35 road crashes that occur each day. An estimated 3,000 deaths resulting from road crashes occur annually in Kenya of which about 40% are pedestrians

National Transport and Safety Authority (2016). The noticeable causes of road accidents include bad roads, speeding, driving while drunk or fatigued and ignoring road signs. However these causes are dependent. For instance, a sober driver on a bad road would do quick judgment and avoid an accident compared to a drunk driver. Bachani et al., (2012) suggests that speeding and poor conditions of some public service vehicles causes most accidents. Some drivers not only ignore road signs but also they do not understand them because they never attended driving classes. Table 2.1 shows distribution of road fatalities by population in the last 11 years in Kenya

Table 2. 1: Distribution of Road Fatalities by population (Manyara, 2016)

| Year | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Population (Millions)[a]** | 35.1 | 36.1 | 37.2 | 38.3 | 39.3 | 40.3* | 41.4* | 42.5* | 43.7* | 44.9* | 47.8* |
| **Road Fatalities [b]** | 2533 | 2715 | 2921 | 3149 | 4032 | 3045 | 3302 | 3141 | 3218 | 2907 | 3057 |
| **Fatalities per 100,000 popn** | 7.42 | 8.21 | 7.93 | 8.31 | 9.54 | 8.25 | 7.97 | 6.82 | 7.22 | 6.34 | 6.4 |

**2.3 Recording road traffic offence data**

Traditionally, the Kenya Police is charged with the responsibility of collection and storage data relating to road safety management. This includes data on road accidents as well as road traffic violations. However, NTSA has been included in the process in a bid to bring sanity on the Kenyan roads. According to Kenya Police Service (2016), road traffic offence data is recorded on a standard document called the charge sheet. In addition, speed cameras have been introduced to capture data from speeding vehicles as well as breathalyzers commonly known as 'alcoblow' to help in identifying drivers driving under influence.

## 2.4 Causes of road accidents in Kenya

Based on Accident Cause Code Classification, reports by Kenya Police reveal that 85.5% of crashes are caused by poor driver behavior, of which driver error represents 44.4%, pedestrians and passengers 33.9% and pedal cyclists 7.2%. Other proximal factors include vehicle defects 5.1%, road environment 2.9%, and other factors 6.4% (Odero, Khayesi, & Heda, Road traffic injuries in Kenya: magnitude, causes and status of intervention, 2003). Table 2.2 shows the causes and percentages of traffic crashes in Kenya.

Table 2. 2*:* Causes and percentages of traffic crashes in Kenya (Odero, Khayesi, & Heda, 2003)

| Causes (specific factors) | Particulars | 1990 | 1985-90 |
|---|---|---|---|
| Human -(speed, misjudgment, improper overtaking, alcohol, traffic violation) | Drivers and motor-cyclists | 44.4 | 44.3 |
| | Pedestrians | 27.1 | 27.4 |
| | Passengers | 6.8 | 6.7 |
| | Pedal cyclists | 7.2 | 5.9 |
| **Sub-Total** | | **85.5** | **84.3** |
| Vehicle - (overload, defective breaks, tires, steering system, headlights, tire burst) | Tires or wheels | 2.5 | 2.2 |
| | Others defects | 2.6 | 3.9 |
| Sub-Total | | 5.1 | 6.1 |
| Traffic Environment - (potholed, sharp/steep bends, slippery road) | Road defects | 1.3 | 1.4 |
| | Animal | 0.7 | 0.9 |
| | Obstruction | 0.5 | 0.8 |
| | Weather | 0.4 | 1.4 |
| **Sub-Total** | | 2.9 | 4.5 |

| Other causes | | 6.4 | 6.2 |
|---|---|---|---|

Nevertheless, the contributing factors to road accidents have remained invariable over the years. Driver faults for instance speeding, improper overtaking, and misjudgment have accounted for the largest percentage of all causes associated with human error (Odero, Khayesi, & Heda, 2003). However, among the human factors, drunk driving has been established as a greater contributing factor to increased incidences of vehicle crash. In a survey conducted by Odero (1998) on alcohol related road traffic injuries, 40% and 20% of drivers and pedestrians respectively were drunk at the time of the crash. In addition, Osoro, Ng'ang'a, & Yitambe (2015) in their study on causes of road accidents, 59.6% of their respondents attributed road traffic accidents to human error, while 20.9% and 19.5% of the respondents attributed the accidents to defective vehicles and roads respectively. Similarly, the study revealed that the human errors associated with road traffic accidents included over-speeding, improper overtaking, drunk-driving and miscalculation or poor judgement. In general, drivers' errors, often accompanied by law violations are in the chain of events leading to more than 90% of all highway fatalities.

## 2.5 Road accident predictive models and techniques

### 2.5.1 Traffic accident surveillance scheme

Hu, Xiao, and Xie (2003) proposed a traffic accident surveillance scheme for predicting road accidents using 3D model based vehicle tracking. The system is composed of 3 main modules: 3D model based vehicle tracking, learning of activity patterns and prediction of traffic accidents as shown in Figure 2.1.
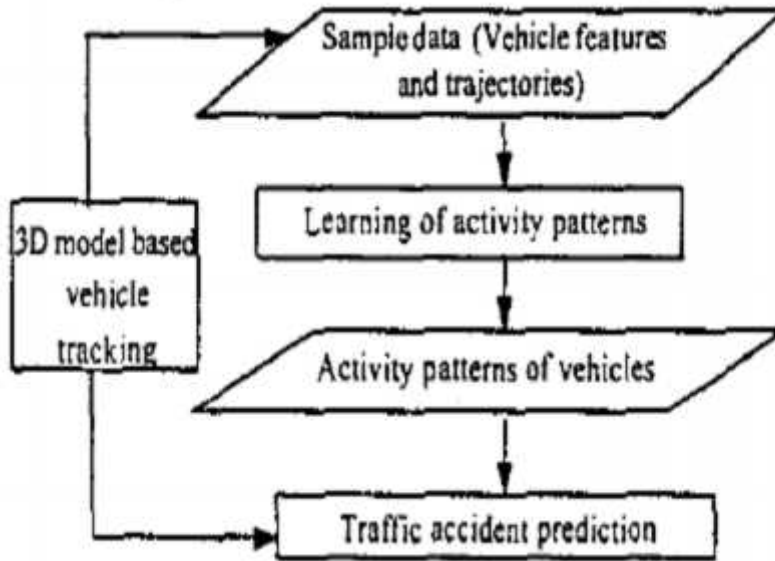
*Figure 2. 1:* Traffic accident surveillance scheme (Hu, Xiao, and Xie 2003)

3D model based vehicle tracking is effected by matching the 3D vehicle models constructed in advance with the calibrated image sequences. The outcomes of this module are the 3D trajectories of vehicles and the features of vehicles such as size. These outputs form the sample data for learning activity patterns.

After a series of observations during tracking, the distribution of vehicle activity patterns can be learnt from the observation data using a fuzzy self-organizing neural network. For the module for traffic accident prediction, partial trajectories are matched to the learned activity patterns, and the occurrence probability of an accident is inferred from a probabilistic model. Nonetheless, this model may not accurately predict an accident since it is based on trajectory and does not consider other factors such as driver's state or behavior as well as visibility.

### 2.5.2 Logistic regression method

The logistic regression model or the logit model as it is often referred to, is a special case of a generalized linear model and analyzes models where the outcome is a nominal variable. Analysis for the logistic regression model assumes the outcome variable is a categorical variable. In logistic regression, the dependent variable is binary or dichotomous, i.e. it only contains data coded as 1 (TRUE, success, pregnant, etc.) or 0 (FALSE, failure, non-pregnant, etc.) (Agresti, 2007).

8

The goal of logistic regression is to find the best fitting (yet biologically reasonable) model to describe the relationship between the dichotomous characteristic of interest (dependent variable = response or outcome variable) and a set of independent (predictor or explanatory) variables.

The logistic regression model has linear form for the logit of this probability as follows:

**logit [ (x)] = log[ (x)/1- (x)] = + x**

*where and are the regression parameters estimated by the maximum likelihood method* (Agresti, 2007).

Lu, Lixin, Dunyao and Pan (2015) used this method to develop a model for predicting traffic accident hotspot on Beijing's 10 main roads. The independent variables used in this study included: driver state, vehicle condition, visual condition, road surface condition, road safety grade, and location of the car in road intersects with the dependent variable being traffic accident hotspot. The model showed that it could provide correct predictions against traffic accident with a correct rate of 86.67%. However, this model is limited to identifying hotspot and cannot be used in identifying individual drivers who are likely to cause the accidents. Moreover, the model did not considers such factors as traffic flows hence the findings may be limited.

Similarly, in an attempt to determine likelihood of future crashes for crash-prone drivers, Das, Sun, Wang and Leboeuf (2015)used eight years' traffic crash data (2004-2011) in Louisiana to develop a model. Road crash predictors such as the driver's crash involvement, crash and road characteristics, collision type, human factors and environmental factors were considered in the model. The model could pinpoint as many as 62.40% of the crash incidence of at-fault drivers in the upcoming year.

### 2.5.3 Use of zero inflated negative binomial regression

Kim, Ramjan, & Mak (2016) in an aim to identify whether driver's characteristics and past traffic violations may predict vehicle crashes in Korea, used the zero inflated negative binomial regression to develop a model to aid in the same. This method according to Agresti (2007) is for modeling count variables with excessive zeros and it is usually for over-dispersed count outcome variables. It assumes that zero outcome is due to two different processes. During the study, a total of 500,000 drivers were randomly selected for the study with the independent variables including driver's demographic information, driving experience, car types as well as traffic

violations in the past 2 years. The study revealed that drivers involved in drunk driving as well as speed violation had a higher risk of a vehicle crash. The study concludes that driver's characteristics and past traffic violations could predict crashes in Korea.

### 2.5.4 The sleep/wake predictor model

Sleepiness is related to aspects such as the time of day, the time since awakening and the duration of prior sleep. The present study investigated whether actual road crashes could be predicted from a mathematical model based on these three factors (the Sleep/Wake Predictor—SWP). Data were derived from a population-based case-control study of serious injury crashes. Data on accident time (or control sampling time) and start and end of prior sleep were entered into the model (blind). The predicted sleepiness values were used in logistic regressions. According to a research by Akerstedt et al. (2008) the results showed a highly significant odd ratio (OR) = 1.72 (confidence interval = 1.41–2.09) for each incremental step of sleepiness on the output sleepiness scale (nine steps). There was also a significant interaction with blood alcohol level, showing high OR values for high sleepiness levels and alcohol levels above 50 mg% (0.05 g/dl). It was concluded that the model is a good predictor of road crashes beyond that of alcohol level, and that interaction between the two carries a very high risk.

### 2.5.5 Using Data Mining Algorithms to predict collision patterns

Shanthi & Ramani (2011) emphasizes the importance of Data Mining classification algorithms in predicting the vehicle collision patterns occurred in training accident data set. Their research was aimed at deriving classification rules which can be used for the prediction of manner of collision. The classification algorithms viz. C4.5, C-RT, CS-MC4, Decision List, ID3, Naïve Bayes and RndTree have been applied in predicting vehicle collision patterns. The road accident training data set obtained from the Fatality Analysis Reporting System (FARS) which is available in the University of Alabama's Critical Analysis Reporting Environment (CARE) system. The experimental results indicated that RndTree classification algorithm achieved better accuracy than other algorithms in classifying the manner of collision which increases fatality rate in road accidents.

## 2.5.6 Crash-prediction model for multilane roads

According to Caliendo, et al (2007), considerable research has been carried out in recent years to establish relationships between crashes and traffic flow, geometric infrastructure characteristics and environmental factors for two-lane rural roads. Crash-prediction models focused on multilane rural roads, however, have rarely been investigated. In addition, most research has paid but little attention to the safety effects of variables such as stopping sight distance and pavement surface characteristics. Moreover, the statistical approaches have generally included Poisson and Negative Binomial regression models, whilst Negative Multinomial regression model has been used to a lesser extent. Finally, as far as the authors are aware, prediction models involving all the above-mentioned factors have still not been developed in Italy for multilane roads, such as motorways. Thus, in this paper crash-prediction models for a four-lane median-divided Italian motorway were set up on the basis of accident data observed during a 5-year monitoring period extending between 1999 and 2003. The Poisson, Negative Binomial and Negative Multinomial regression models, applied separately to tangents and curves, were used to model the frequency of accident occurrence. Model parameters were estimated by the Maximum Likelihood Method, and the Generalized Likelihood Ratio Test was applied to detect the significant variables to be included in the model equation. Goodness-of-fit was measured by means of both the explained fraction of total variation and the explained fraction of systematic variation (Caliendo, et al., 2007).

## 2.5.7 System Dynamics approach to road accident prediction

Partheeban, Arunbabu, & Hemamalini (2008) propose a dynamics approach to accident prediction. This is done by developing a model for road accident through evaluation of various factors involved in assessment of road accidents in Kenya. To build an accident model, various factors causing the road accident and cost were identified. This model is capable of calculating the accident rate and its costs for the future. In this study the accident caused by bus alone is considered. The cost model is dealt more in this study as it requires more complex assessment. The accident model is built on the year 2000 data and predicted the accidents up to 2020 for every 5 year interval. The accident model is evaluated by comparing the predicted and actual accident data for the year 2005. Three scenarios were studied by changing the income growth rate and discount rate. Finally, best scenario is suggested for implementation. The outcome of the

study is highly useful for the planners, administrators and police to make their decisions effectively for road safety investment projects.

## 2.6 The proposed Fast correlation- based filter algorithm

Fast Correlation Based Filter algorithm was designed for high dimensional data and has been shown effective in removing both irrelevant features and redundant features. The algorithm is practical for feature selection for classification of high dimensional data. It can efficiently achieve high degree of dimensionality reduction and enhance classification accuracy with predominant features (Yu & Liu, 2003). In order to get traffic offence data from a possible list of 10 million drivers on our roads, it would suffice to use a data mining algorithm that will capture all the information efficiently and ensure that there is minimal searching and a lot of referencing to get the exact driver details when needed. The algorithm is expected to correlated data from a set of files within a directory in order to specifically identify the specific identity searched. For instance, in a database with thousands of records such as the NTSA database, the algorithm is supposed to reference close-knit word sequences to identify a specific driver. The logical layout of the algorithm is as shown in the pseudocode below;

Begin

      Create an array micFC [i..k]

      Create an array Subset [i..k]

          numSubset = 1

          for I = 1 to k

                micFC[i] = MIC(F[i], C)

                if micFC >= r

                      Subset [num, Subset] = i

                      numSubset = numSubset + 1

                endif

          endfor

      rank the items in Subset [1..numSubset] in the descending order by micFC[Subset[i]]

          for e=1 to numSubset

                q = e+1

                while q<=numSubset

```
                        if MIC(F[Subset[e]], F[Subset[q]]) >= micFC [Subset [q]]

                                       for I = q to (numSubset -1)

                                       Subset [i] = Subset[I +1]

                                       numSubset = numSubset -1

                                       endfor

                              else q = q+1

                              end if

               end while

               e = e +1

        end for

   return
```

## 2.7 Conceptual Framework

Existing and new traffic offence data such as speeding, drunk driving, inappropriate overtaking and use of mobile phones while driving is sorted out and recorded. The data is then structured in form of tables and normalized for easy retrieval, and processing in the database as shown in Figure 2.2.The Prediction algorithm uses the data stored in the database for computation and analysis in order to determine the probability of a given driver causing an accident. The algorithm applies set rules (parameters) in computing the probability. Finally, a prediction report is produced revealing driver with a higher probability of causing an accident shown in Figure 2.2.
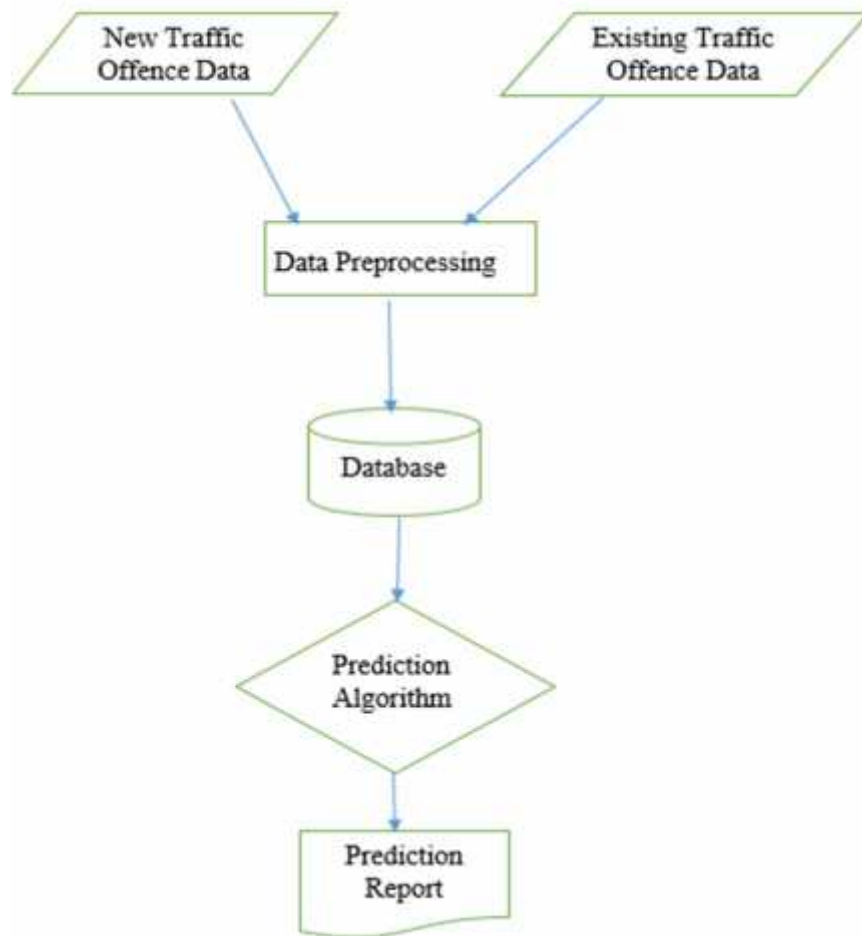
*Figure 2. 2: Conceptual Framework*

# CHAPTER THREE: RESEARCH METHODOLOGY

## 3.1 Overview

This chapter discusses the research methodology applied in this study and gives a general framework for this research. The chapter outlines details of the research design, population targeted, sample and sampling procedures, research tools to be used, data analysis and ethical considerations while conducting the study.

## 3.2 Research Design

A research design is used to structure the research, to show how all the major parts of the project work together to try to address the central research questions. It is the conceptual structure within which research is conducted; it constitutes the blueprint for the collection, measurement and analysis of data (Kothari, 2004). For this study, exploratory research was used because it facilitates in-depth investigation to unravel unclear areas relating to this study (Creswell, 2014). For instance, road traffic offences associated with road crashes in Kenya, how data relating to these offences is recorded, current mechanisms of determining crash prone drivers as well as means of quantifying an offence based on its magnitude.

In designing the system, the research implemented waterfall development methodology. This methodology allowed the researcher to determine system requirements long before programming so as to minimize changes as development progressed.

## 3.3 Development Methodology

In this particular research, the technique employed is the Software Development Life cycle (SDLC). The software development life cycle is the process that documents software evolution slowly, from the time the idea of the software is conceived to the realization of the final product. This involves; specification of the software's expected inputs, processes and outputs; a design layout of the software's major components; a validation tool for the software developed; and an evolution mechanism to track new changes or versions of the software.

This is based on the waterfall model which is a sequential design process often used in software development process. The development is seen as owing downwards steadily. In the model, one phase has to be finalized before the process can progress to the next phase.

*Figure 3. 1:* System Development Life Cycle

## 3.4 Population and Sampling

This section describes the population to be used in the study and the process utilized in selecting a sample. According to Best and Khan (2006), a population refers to any group of individuals that has one or more characteristics in common and that are of interest to the researcher. The target population for this study comprised of Kenya Police officers in the Traffic Department, NTSA road safety officers and matatu drivers.

Since it was quite expensive and impractical to collect data from all potential target units covered in the research problem, sampling was done in order to come up with a manageable group. According to Ogula (2005), sampling is a procedure, process or technique of choosing a sub-group from a population to participate in the study.

This study adopted a purposive sampling design. In this type of sampling, items for the sample are selected deliberately by the researcher; his choice concerning the items remains supreme (Kothari, 2004).

Nevertheless, in determining the sample size for this study, the following sample size formula was used:

$$S = \frac{X^2NP(1-P)}{d^2(N-1)+X^2P(1-P)}$$

*Figure 3. 2: Sample Size Formula for Finite Population (Krejcie & Morgan, 1970)*

Where:

S       =    Required Sample size

X       =  Z value (e.g. 1.96 for 95% confidence level)

N       =  Population Size

P       =    Population proportion (expressed as decimal) (assumed to be 0.5 (50%)

d       =  Degree of accuracy (5%), expressed as a proportion (.05); It is margin of error

## 3.5 Data Collection Methods

In order to achieve the objectives of the study, both primary and secondary data was used. Secondary data used in this study included road traffic offence records and accident records from Kenya Police and NTSA, as well as road traffic offence manuals and other documentations relevant to the study.

The researcher used both questionnaires (for NTSA officials and traffic police officers) and interviews (for matatu drivers) in collecting primary data. This data was useful in understanding the underlying factors associated with road crashes, determining violations considered high determinants to road accidents, feasibility of the system, as well as testing whether the system meets the users' desired functionalities.

## 3.6 Data Analysis

Data obtained from the field in raw form is difficult to interpret; such data must be cleaned, coded, keypunched into computer and analyzed. It is from the results of such analysis that researchers are able to make sense of the data (Mugenda & Mugenda, 1999).

During data analysis, relationships or differences supporting or conflicting with original or new hypotheses should be subjected to statistical tests of significance to determine with what validity data can be said to indicate any conclusions (Kothari, 2004). This research utilized both qualitative and quantitative methods of data analysis. SPSS-a statistical package and excel sheets were used in data analysis.

Reliability of the data was measured against past research data on accident records as well as database records for integrity measurement. The viability of the data is guaranteed by the validation rules in the system. There are input masks to ensure that the data taken for different tests is only the desired set of information needed.

### 3.7 Research Quality

In order to ensure quality, the research was conducted in an ethical manner and line with the regulations of the university as well as other set standards. Vale (2010) noted that data quality is generally understood to be the degree to which data, including research processes such as data collection and statistical accuracy, meet the needs of users.

### 3.7.1 Validity

The sample population used in the study was calculated using globally accepted sample size formulas in order to ensure trustworthiness of the study. In addition the questions asked during data collection was be pre-tested on subject outside the sample population to determine their validity.

### 3.7.2 Reliability

The sample population used in the study comprised of traffic police officers, NTSA officials in charge of road safety and matatu drivers. The traffic police officers targeted were randomly selected from various police stations within Nairobi so as to ensure consistency. Moreover, the researcher interviewed matatu drivers operating in different routes in Nairobi.

### 3.7.3 Objectivity

In order to ensure objectivity, this study was not based on personal bias and conclusion was based on the data collected and analyzed. In addition, the tools used in data collection, were developed with the objectives in mind.

### 3.8 Ethical Considerations

The researcher explained to the respondents about the research and that the study was for academic purposes only. It was made clear that participation is voluntary and that the respondents were free to decline or withdraw any time during the research period. Moreover, respondents were not coerced into participating in the study. The participants made informed consent to make the choice to participate or not. Consequently, the privacy of the respondents was guaranteed by strict standard of **anonymity.** In addition, the data collected was not used for monetary gain without the knowledge of the persons involved.

Nevertheless, materials used in this study from various sources that are not original by the researcher, were acknowledged in the reference. Also, access to the data collected was restricted to only authorized persons.

# CHAPTER 4: FINDINGS AND ANALYSIS

## 4.1 Introduction

Data analysis is important after every data collection process. It is the process that ensures all data collected from a research activity is fully represented and tabulated in a way that can be understood by other readers or researchers. For this particular research, the analysis was from two major respondents; the traffic police and the members of the matatu industry (drivers and touts). Questionnaires were issued to the traffic police as they were likely to find time to answer the questionnaires. For the drivers, this was not feasible as most of them were always on the road. The choice of data collection from the drivers and touts was thus through interviews. In order to get the right sample to interview, there was a significant number of sampling techniques considered. However, the sample formula in *Figure 3.1* was used. From the formula, an effective sample of 30 policemen and NTSA officials were considered for the questionnaires. The interviews would then target 15 drivers plying one of the busy routes within the city center.

## 4.2 Response Rate

### 4.2.1 Response to Questionnaires

Thirty questionnaires were used for the research. Of these, fifteen were given to different traffic police bases within the city while the other 15 were handed to the National Transport and Safety Authority, at their headquarters, within the city. The respondents were given a duration of two weeks to come up with the appropriate responses before the questionnaires would be picked from them. From the set of thirty questionnaires, 24 were returned. The rest were either tampered with or lost. The analysis thus depended on the 24 questionnaires that were efficiently filled. Table 4.1 shows the tabulated categorization of returned questionnaires;

*Table 4. 1: Questionnaires returned by NTSA and Traffic police respondents*

| Category | Questionnaires Offered | Questionnaires Returned |
|----------|------------------------|-------------------------|
| NTSA Officials | 15 | 11 |
| Traffic Police | 15 | 13 |
| Totals | 30 | 24 |

**4.2.2 Response to interviews**

The interviews used presented significant data that was crucial to this research. This was because; it was easy to determine the validity of responses by inquiring more from the drivers about the safety of roads in the country, as well as their experience and history with traffic offences. The drivers were assured that they were not being recorded. They were also informed of their right to abdicate the interview at any point necessary. A consent form was signed between the researcher and each driver to ensure that the information given by the driver was in confidence, and could not be used for any other purposes other than to meet the objectives of this particular research. Although requests to interview 15 drivers/touts were made, 2 declined to be interviewed. The analysis is thus based on data from 13 drivers/touts.

**4.3 Summary responses from the National Transport and Safety Authority (NTSA)**

**4.3.1 Information required when booking a traffic offender**

The different responses gathered from the NTSA officials were assessed and evaluated. These officers were mostly related with booking traffic offenders, controlling traffic where road signs were inefficient and monitoring speeding vehicles along major highway outlets within the city. The responses are as represented in Figure 4.1
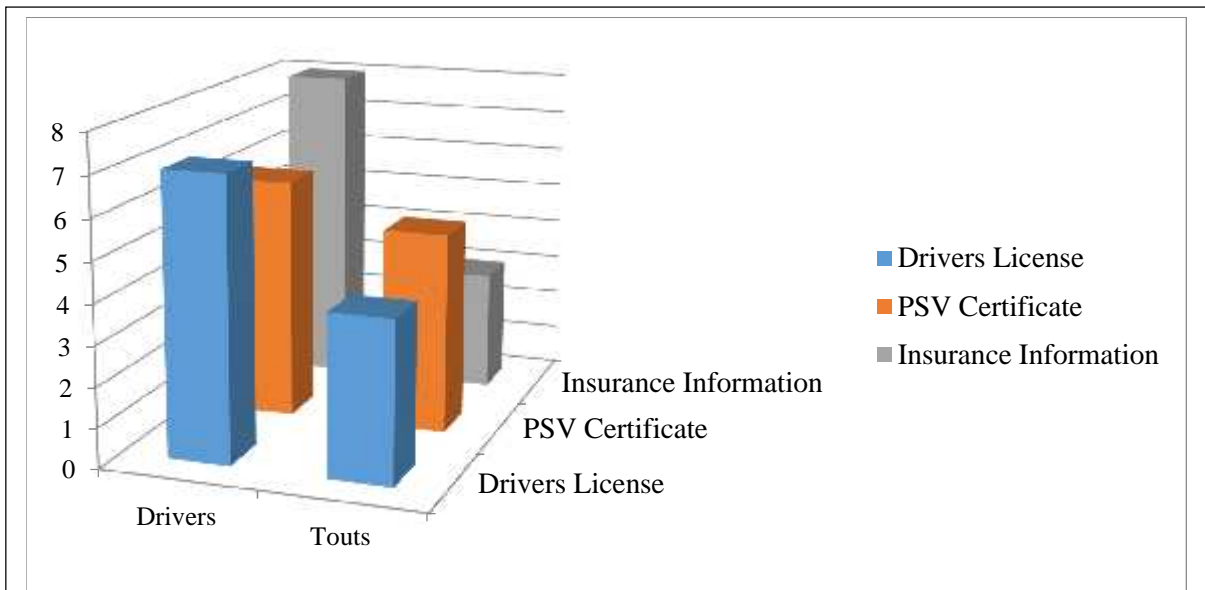


*Figure 4. 1: Information required when booking a traffic offender*

The responses gathered from the NTSA indicate when a PSV vehicle is involved in a traffic offence, most of the time, the driver's license from the driver is required more than the tout. The insurance information is however paramount and should be presented by either of the two. PSV certificates are also demanded of people in the public transport sector. The information needed is this generally of better offered by the driver, and is used to book the PSV vehicle and its operators.

**4.3.2 Main Causes of Road Accidents in Nairobi**

The NTSA officers were also asked to mention the main causes of road accidents in Nairobi. A list of options was offered in a questionnaire, which they were to fill. Within this list, there was data from major reports on road accidents, especially data published by the organization. The data was used to determine which traffic offences were majorly the leading causes of road accidents. From such data, the study sought to offer the major offences as inputs to the algorithm-predictive system for tracking and punishing reckless road users. Likert-scale answers were desired of the respondents. These answers were decoded in Table 4.2;

*Table 4. 2: Likert scale answers as decoded*

| Strongly Agree | Agree | Neutral | Disagree | Strongly Disagree |
|---|---|---|---|---|
| 5 | 4 | 3 | 2 | 1 |

The analysis of their responses is as shown below;



**Major causes of road accidents in Kenya**

- Overspeeding
- Reckless overtaking
- Overlapping
- Drunk driving
- Use of mobile phones

*Figure 4. 2: Major causes of road accidents in Kenya*

The data indicated that over speeding and reckless overtaking were the major causes of road accidents in Kenya. However, drunk driving was a significant contributor as well. The least recorded cause of road accident was overlapping. Use of mobile phones contributed about 10 percent of the road accidents in the country. From such information, the study determines to ensure that over speeding, and reckless overtaking be major inputs in the design of the proposed algorithm based system. This would drastically reduce the number of road accidents in Kenya, if the algorithm is well utilized by the respective authorities.

### 4.3.3 Relationship between traffic offences and road accidents

The NTSA respondents offered their views on whether the susceptibility of drivers to road accidents was a result of previous history of traffic offences. The algorithm that would be developed for traffic accident monitoring and management of traffic offences required an efficient way of linking these two variables. It was thus necessary that the answers offered by the respondents be effective and ample. This information is as indicated in Figure 4.3
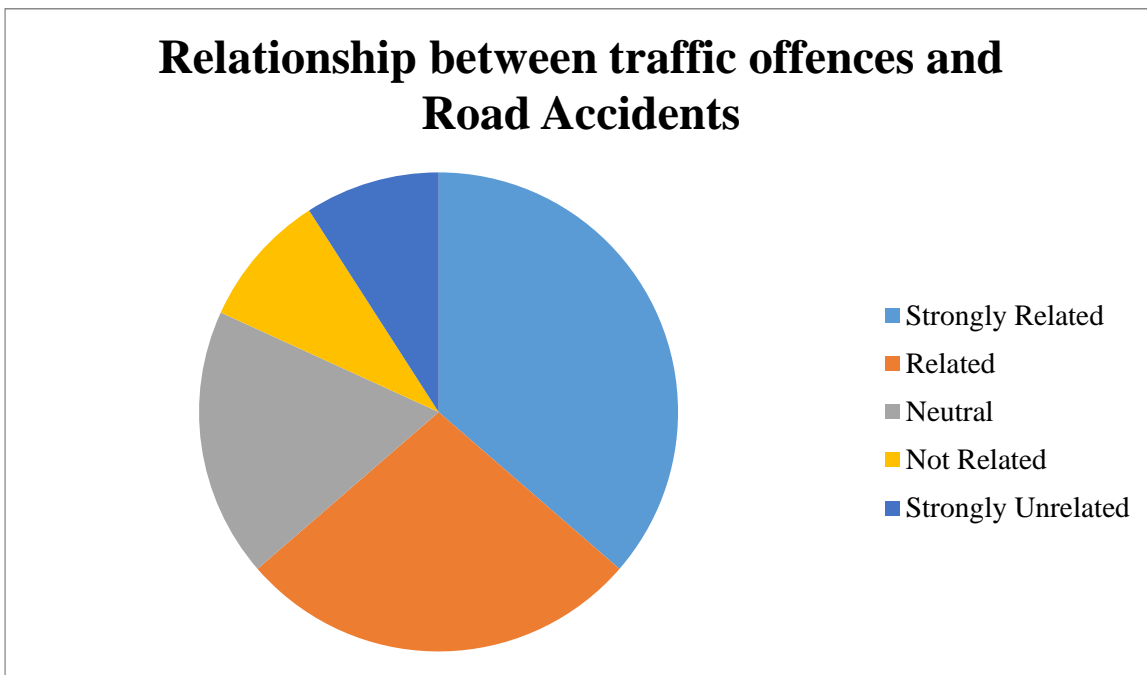


*Figure 4. 3: Relationship between traffic offences and road accidents*

Of the 11 NTSA officers that responded, 4 indicated that indeed, traffic offences and road accidents were strongly related. 3 of them rated the relationship at 'related' while two were

neutral on the relation. Two others indicated that there was no significant relationship. This however represented a minority of the population surveyed.

### 4.3.4 Efficiency of the current system for tracking road offenders

The NTSA officers were asked about the existing system used to track traffic offenders. They all indicated that there is an existing database, which is often presented when a traffic offender is charged. However, they mentioned several faults with the existing system. These faults include; inability to access the system remotely, challenges in referencing the database when filing a police case, errors with the records in the database and the lack of efficient data mining tools within this database. This can be represented in Figure 4.4
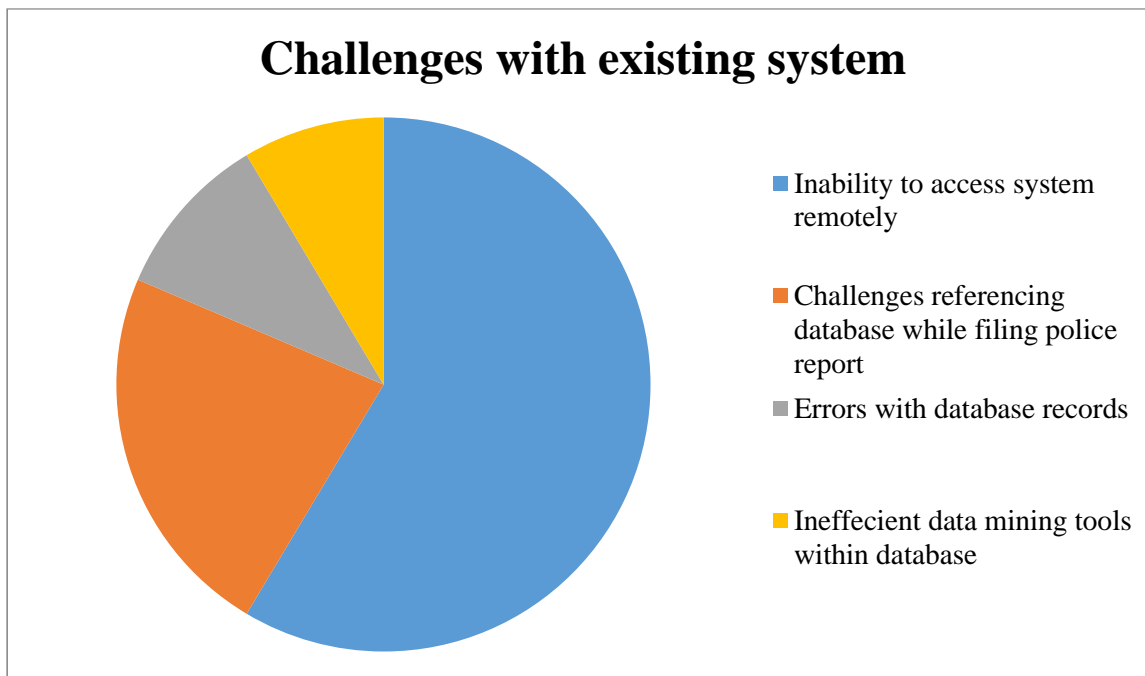


*Figure 4. 4: Challenges with existing system for tracking frequent offenders*

### 4.3.5 Willingness to adopt algorithm-based system

The respondents were then quizzed over their willingness to adopt an algorithm-centric system for tracking frequent road offenders. The details of the system were not indicated but the research question indicated that the system would be computerized. This is an aspect that has not been implemented in the current systems used in the country. However, it was anticipated that there would still be some minor resistance from the respondents. Their responses are as shown in Figure 4.5
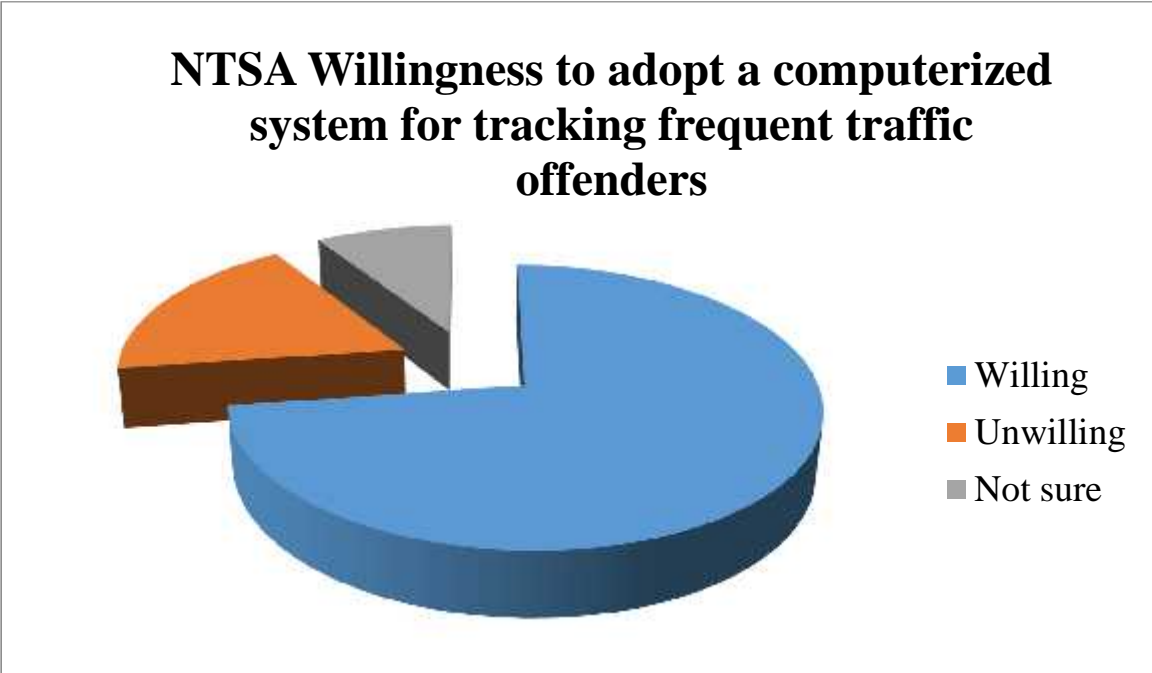
*Figure 4. 5: Willingness to adopt proposed system for tracking frequent traffic offenders*

The responses were gathered from the NTSA officers based on their indication of; willingness to support the new system, or unwillingness to support the new system. No reasons were required for the choice given but of the 11 respondents, 8 were willing to adopt the system. 2 of the respondents were not willing to adopt the system while 1 respondent needed to test the new system first before deciding on whether or not to adopt it.

## 4.4 Responses gathered from the Traffic police officers

Thirteen traffic police officers respondent to the questionnaires given. Two were not returned. The indications from the questionnaire were used to form an analysis based on the questions asked of the police officers. From the questions asked, the responses had an almost similar bearing to those offered by the NTSA officers. This was however anticipated as most of these officers worked alongside the NTSA in handling traffic offences on Kenyan roads. The responses given can be categorized as follows;

## 4.4.1 Information required when booking a traffic offender

The same details mentioned in the NTSA questionnaire analysis were mentioned by the traffic police officers. However, the traffic police officers had additional details they mentioned as

being critical data given by traffic offenders to the police. These details can be summed up as shown in Figure 4.6
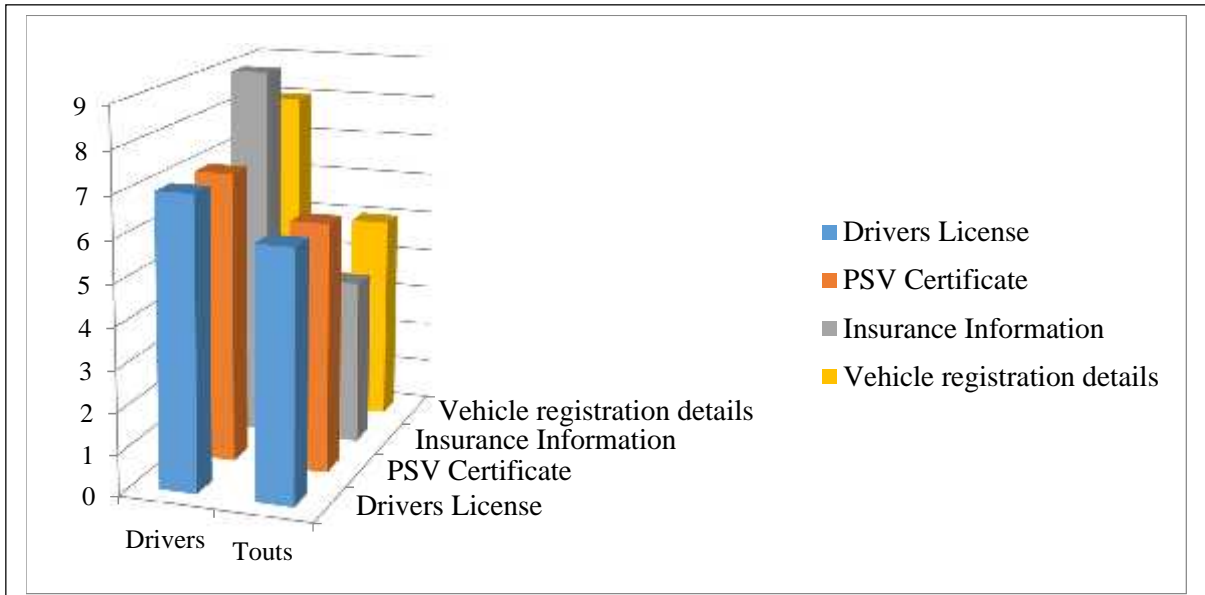


*Figure 4. 6: Information required when booking a traffic offender*

The responses gathered from the traffic police officers in the country indicate that traffic offenders are asked to provide their drivers' license, their PSV registration certificates, insurance information and the vehicle registration details. The PSV registration certificate is the authorization to drive a PSV matatu or van and is offered by the NTSA upon testing of the driver's road skills. It is subject to renewal annually. This could be the controlling factor in the development of the algorithm.

### 4.4.2 Main Causes of Road Accidents in Nairobi

The traffic police officers were also asked to categorize the main causes of road accidents in Nairobi. This was based on various options indicated in a Likert-type list; similar to the questions asked of the NTSA officers. From the responses, a Likert chart was developed for analysis. This chart is as shown in Table 4.2. The analysis of responses from the Kenya police traffic officers is as shown in Figure 4.7

The analysis of their responses is as shown in Figure 4.7;
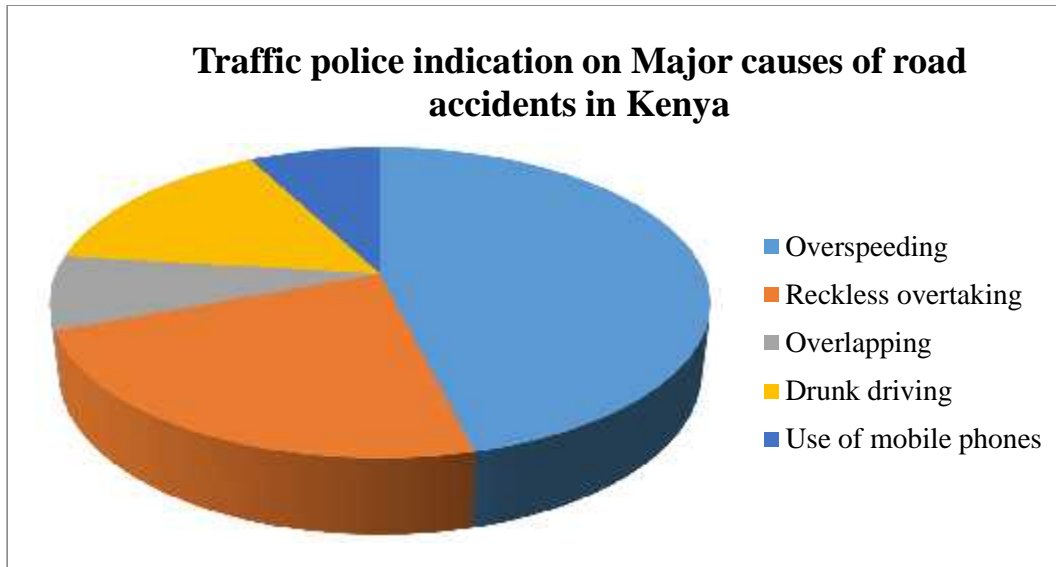
*Figure 4. 7: Traffic police indication on major causes of road accidents in Kenya*

The data indicated that over speeding and reckless overtaking were the major causes of road accidents in Kenya, similar to what had been indicated by the National Transport and Safety Authority respondents. This response was based on similar data gathered on Kenyan roads, as represented in the literature section of the document. The research was thus based on the past information on road accident records. The police officers and NTSA officials involved in this study thus confirmed what had been previously documented. It also became an emergent issue that the use of cell phones while driving either for texting or calling was a major distraction to drivers. This factor would be considered in the development of the algorithm that would be used to track frequent traffic offenders and propose punitive measures on them.

### 4.4.3 Relationship between traffic offences and road accidents

The traffic police respondents were also expected to give their view on whether or not there was a relationship between traffic offences and road accidents in Kenya. This data would be applied in the correlation analysis that would be done in the course of the research, using appropriate tools in the Statistical Package for Social Science (SPSS) program. The responses were gathered from all the 13 officers as follows; 8 believed that frequent traffic offenders were majorly involved in road accidents in the country, 3 could not relate the two variables while 2 of the respondents thought there was no direct relationship between the two variables. This information is as indicated in Figure 4.8

**Relationship between traffic offences and Road Accidents as determined by traffic police officers**

- Related
- Not Related
- No Direct Relationship
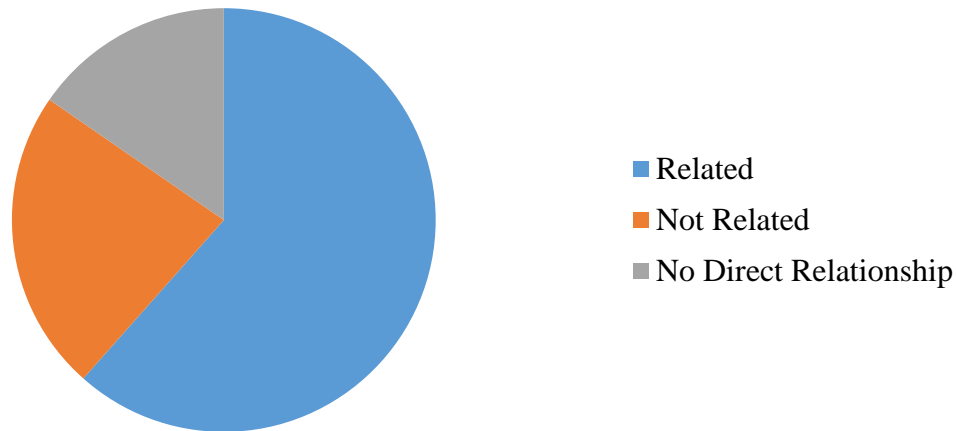
*Figure 4. 8: Relationship between traffic offences and road accidents*

**4.4.4 Efficiency of the current system for tracking road offenders**

The traffic police officers were asked about the existing system used to track traffic offenders. The current system for tracking road offenders was also not preferred by the traffic police officers. They cited challenges with the system relating to how they booked traffic offenders. It was essentially a concern since most of these officers needed previous records to file police cases on accident reports but could not get the ample information they required. Part of the existing challenges mentioned by police officers included; existing manual records on traffic occurrence reports, widespread records on traffic accidents by a single offender, lack of sufficient computer systems to track vehicles on roads and the need for digital records such as drivers' licenses and PSV information. This can be represented in Figure 4.9

**Indications from Traffic police officers about need for traffic offence detailing systems**

- Manual records on traffic occurrence
- Widespread records by a single offender
- Lack of sufficient computer systems for vehicle tracking
- Digital records on PSV operators and road users
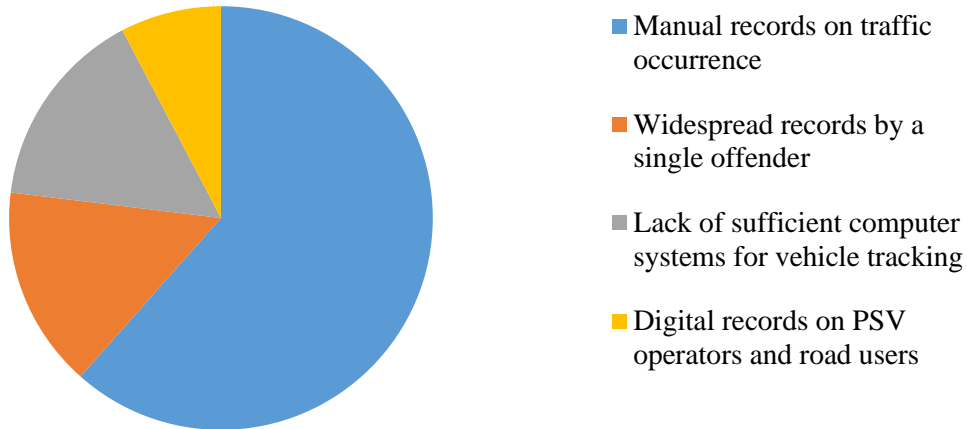
*Figure 4. 9: Challenges with existing system for tracking frequent offenders*

### 4.4.5 Willingness to adopt algorithm-based system

The traffic police officers were asked to indicate their willingness to adopt a system that would computerize their records and make it easier for them to monitor traffic offenders wherever they may be within the country. The system would come in handy especially to deal with frequent offences committed by persons in different parts of the country who would on major occasions escape arrest for lack of precedence data. This information would also be considered important to the police as it would estimate better how to deal with imminent traffic offenders. Their responses are as shown in Figure 4.10
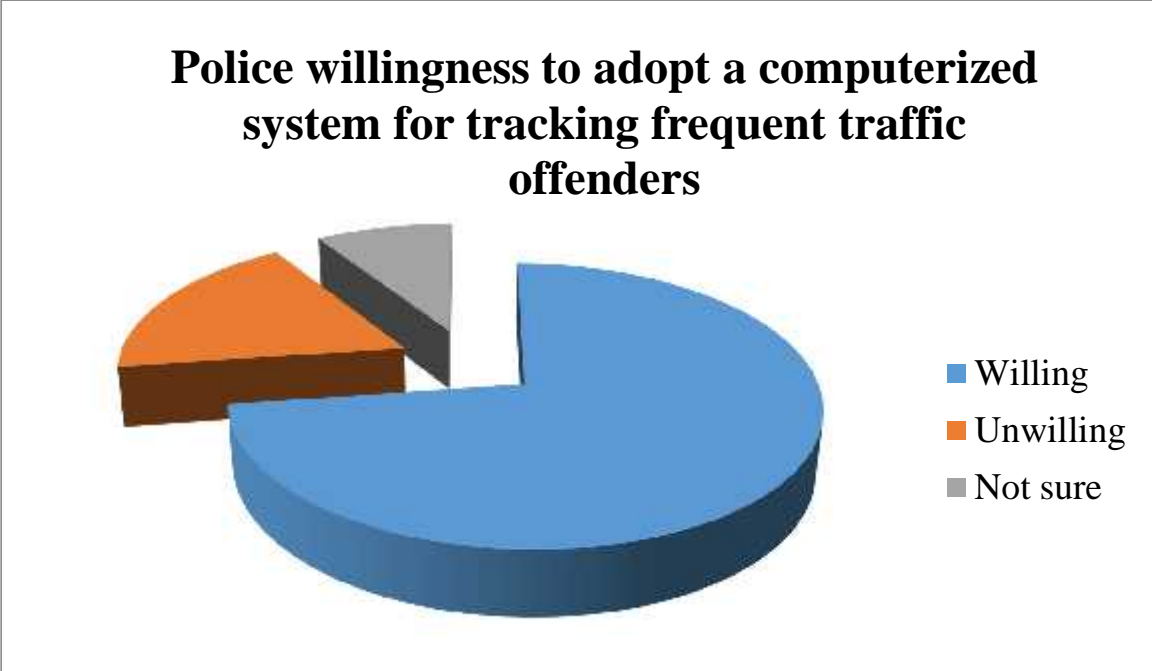
*Figure 4. 10: Willingness to adopt proposed system for tracking frequent traffic offenders*

Of the thirteen respondents surveyed, 8 indicated that they were willing to adopt the computerized system for tracking frequent traffic offenders. Three of the respondents were not willing to adopt the system while 2 of the respondents were not sure about the system. To convince them further, it was necessary to develop a prototype for presentation in order to prove the functionalities of the system.

## 4.5 Interview responses from PSV drivers

Thirteen PSV drivers were interviewed. This is because; of the fifteen surveyed, two were failed to show up for the interview process. The questions asked to these drivers indicated their driving prowess as well as their ability to be vigilant on the road and the offences they had committed before. The responses would be significant to the algorithm-based system they would minimize road traffic offences in Kenya. The responses can be analyzed as follows;

## 4.5.1 Experience in the transport business

The study sought to find out the experience these drivers had in the transport business. All the thirteen drivers indicated different experience levels in the transport business. This information was classified in three different categories; 1-5 years, 5-10 years and more than 10 years. The

responses were analyzed and a graph developed for the data. These details can be summed up as shown in Figure 4.11

# Experience of drivers on Kenyan Roads



■ 5-10 years experience

■ More than 10 years experience

■ 1-5 years experience
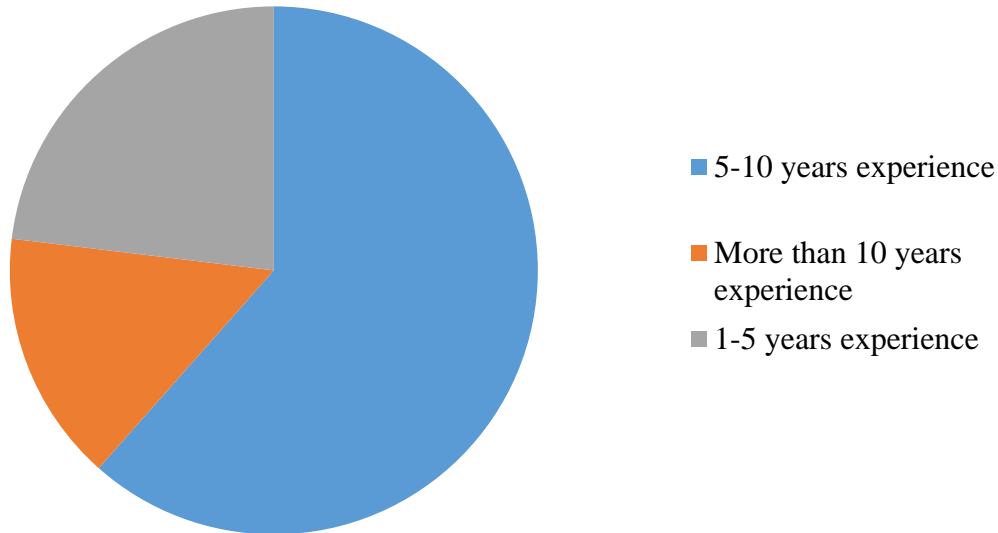
*Figure 4. 11: Driving experience by Kenyan drivers*

From the responses gathered, it was clear that most of the respondent had significant experience, with regard to the driving of public service vehicles in Kenya. Indeed, of the 13 interviewed, only three mentioned of having less than 5 years driving experience on Kenyan roads. The sample was thus vastly experienced and could offer ample data on the issues concerning the traffic offence monitoring systems, as well as the proposed algorithm-based system for tracking traffic offenders in Kenya.

## 4.5.2 Involvement in PSV accidents

The drivers were then asked to recall whether they had been in accidents that had been recorded with the traffic police before. All the thirteen drivers were at first hesitant to offer this information willingly. However, they were given a certainty that the information offered would be kept confidential by the research team. The information on whether or not they had been in accidents before was interlinked with their driving experience. This data was then analyzed and the findings presented in the graph shown in Figure 4.12

## Drivers' Indication on their involvement in Road accidents

- ■ Experienced and had been in road accidents before
- ■ Experienced but had never been in road accidents before
- ■ Inexperienced but had been in road accidents before
- ■ Inexperienced but had not been in road accidents before
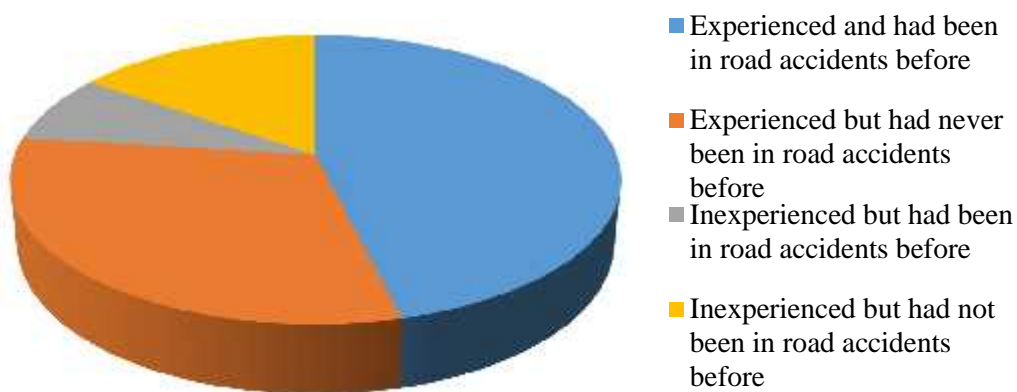
*Figure 4. 12: Drivers' indication on involvement in road accidents*

The data gathered from the PSV drivers was quite revealing. Of the thirteen drivers interviewed, 10 were considered to be experienced; having indicated to have driven for more than 5 years as represented in Figure 4.11. Of these, six indicated that they had been in road accidents before. This represents more than 50 per cent of the experienced population. Among the inexperienced drivers (with less than 5 years of driving experience), only one indicated that they had not been involved in an accident before. Generally, eight of the thirteen drivers had been involved in PSV accidents on Kenyan roads.

### 4.5.3 Main determinants to road accidents in Kenya

The drivers were asked to mention the factors they though could be used to predict road accidents in Kenya. In this question, they were guided to offer the aspects of driving and the issues on the road that made drivers more susceptible to road accidents. Although the research anticipated certain responses, these responses were not offered to the drivers as choices. Indeed, the drivers were expected to come up with their own factors that they thought could be better utilized by the relevant authorities in the prediction of accidents on Kenyan roads. The information gathered is as indicated in Figure 4.13

# Factors contributing to road carnage among PSV operators on Kenyan roads



Legend:
- Ineffective offender booking systems
- Indiscipline of PSV operators
- Corruption on Kenyan roads
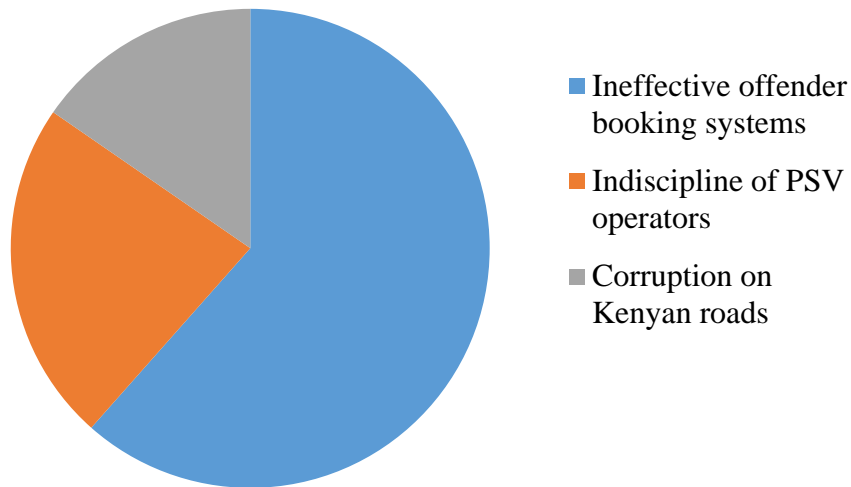
*Figure 4. 13: Relationship between traffic offences and road accidents*

Most of the drivers felt that there were no significant punitive measures for PSV drivers on Kenyan roads. This was because; the booking systems used to capture traffic offenders did not effectively capture those who committed offences. It was determined by a fraction of the respondents that indiscipline of PSV operators and corruption on Kenyan roads were key contributors to continued carnage in Kenya.

## 4.5.4 Drivers' view on the relationship between traffic offences and accidents in Nairobi

Drivers in Nairobi (sample interviewed) were asked to answer with a simple 'yes' or 'no' as to whether there was a relationship between frequency of traffic offences and the susceptibility of drivers to road accidents in Kenya. Of the thirteen respondents, only two thought that there was no relationship between the two variables (traffic offence and accidents). This is as represented in Figure 4.14

# Indication by PSV drivers on the relationship between traffic offences and accidents in Kenya



■ There is a direct relationship between traffic offences and road accidents in Kenya

■ There is no direct relationship between traffic offences and road accidents in Kenya
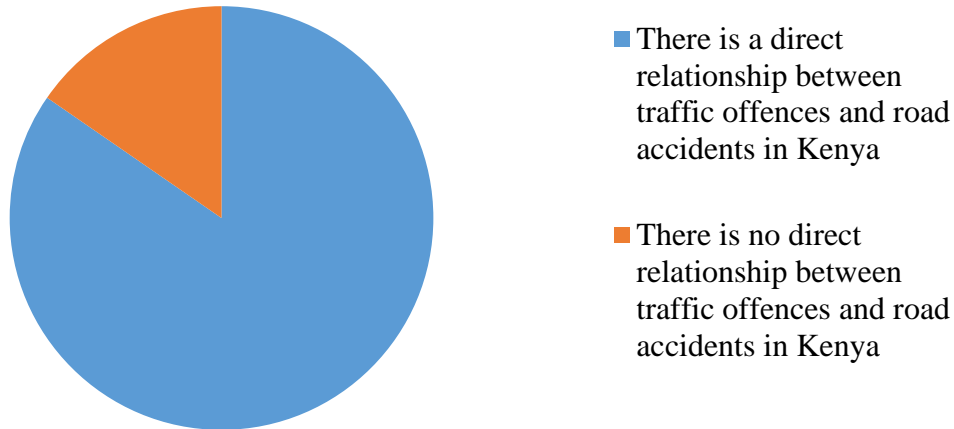
*Figure 4. 14: Drivers' indication on relationship between traffic offences and road accidents*

**4.5.5 Willingness to support algorithm-based system**

In order to gauge the willingness of drivers to support the proposed algorithm-based model for predicting road accidents based on traffic offence occurrence data, the final two questions of the interview were clumped into one, during the analysis. This entailed; first determining the number of drivers within the sample population that would be subjected to vetting and monitoring of traffic offences committed frequently. The study then sought their opinion on whether or not they would support an algorithm based system for accident prediction based on traffic offence occurrence data. Their responses are as shown in Figure 4.15

**Drivers' willingness to support an algorithm-based system for tracking frequent traffic offenders**
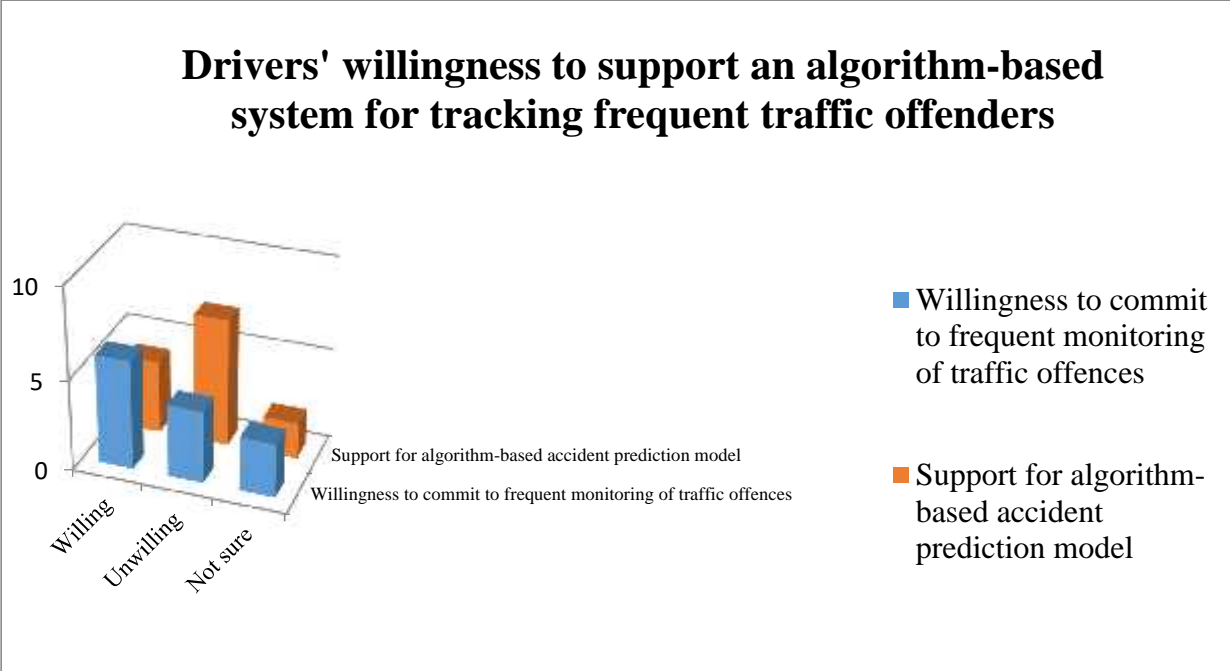
*Figure 4. 15: Willingness to support proposed system for tracking frequent traffic offenders*

For this particular question, the respondents were asked to further state their reasons as to why they made the decisions made. Many of the drivers who were unwilling to be subjected to frequent monitoring and tracking of their traffic offence data felt that they would be exposed to their employers and probably lose their jobs. Drivers who supported and expressed willingness to be monitored felt that they drove carefully and did not have any traffic information to hide. On the issue of support for an algorithm-based accident prediction model, most of the drivers were unwilling. They cited disruptive technologies in the transport business such as; speed governors, alcohol breathalyzers and speed guns as some of the most abused technologies on Kenyan roads. However, of the 13 interviewed, 5 were willing to adopt a new system nonetheless.

**4.6Analysis using SPSS**

SPSS was used to analyze the findings to weight their statistical viability and possibility to make conclusive assertions from the data. The version used for the Software is IBM Statistics version 24. The data was analyzed as follows;

### 4.6.1 Convergent Test

A convergent test is performed to establish whether data from the three variables had any relationship; and whether an infinite set of respondents would give an indication of trend in data. The analysis was done using SPSS and the output presented as shown in Figure 4.16

## Convergence Information[a,b]

| | |
|---|---|
| Maximum Number of Iterations | 20 |
| Converge Tolerance | .00100 |
| Final Maximum Absolute Difference | 1.59534E-5[c] |
| Final Maximum Relative Difference | 9.95056E-5 |
| Number of Iterations | 3 |

a. Model: Poisson

b. Design: Constant + Respondent_ID

c. The iteration converged because the maximum absolute changes of parameter estimates is less than the specified convergence criterion.

*Figure 4. 16: Convergent test using SPSS*

### 4.6.2 Univariate Analysis of Variance

Variance was determined using analysis of the offence_accident_relationship, against the category of data members used in the research. The results are as indicated in Figure 4.17.

## Univariate Analysis of Variance

### Between-Subjects Factors

| | | N |
|---|---|---|
| Category | NTSA | 11 |
| | POLICE | 13 |
| | PSV OPERATORS | 13 |
| Offence_Accident_Relatio nship | NEUTRAL | 4 |
| | RELATED | 12 |
| | STRONGLY RELATED | 14 |
| | STRONGLY UNRELATED | 1 |
| | UNRELATED | 6 |

*Figure 4. 17: Univariate Analysis of Offence_Accident relationship*

From the data, it is indicative that the values, 'strongly related' and 'related' appear twenty six of the thirty seven times. This is a strong indicator for the support for the hypothesis; that accidents in Kenya were directly related to the occurrence of traffic offences on Kenyan roads. This large number can be attributed to the general agreement in views among all three sets of respondents.

# CHAPTER 5: SYSTEM ARCHITECTURE AND DESIGN

## 5.1 Overview

Upon ascertaining that the research hypothesis is proven, the design of the system was the next step. The research was angled around realizing an effective solution for the challenge of recurrent traffic offenders being allowed on Kenyan roads. Through a digitalized register for the recording of PSV driver information as well as any data held by the National Transport and Safety Authority on Kenyan drivers, the system interlink drivers to their recurrent traffic offences through an algorithm. The algorithm basically gain inputs from every booking, manned by traffic police officers and the National Transport and Safety Authority database. Upon reaching a critical level of traffic offences, the driver shall be listed as a frequent offender and punitive action such as license suspension or payment of fines is levied on them in order to encourage them to be more prudent in observing traffic regulations. This chapter details the design of the system.

## 5.2 Traffic monitoring system Architecture

The architectural design of the proposed accident prediction system is as shown on Figure 5.1. The system requires input from a policeman, with an interface to the system. Upon filling in of driver data, the system checks for the offences committed by the driver and determines whether or not to book the driver, based on frequency of offences. Each time the driver commits an offence, they are booked. The policeman can thus check the status of a traffic offender and propose punitive measures against them.

The system requires that the policeman (system user) access an interface, either through a cell phone or a personal digital assistant such as a tablet computer that can query data from a database, save data into the database and book the driver where necessary. The policeman enters the necessary information into the system, upon which the information initiate a server request for a session. The server then allows the policeman to access the database for data storage and retrieval. Upon an offender committing more than the allowable number of traffic infringements, they are flagged for action, regardless of their location in the country.
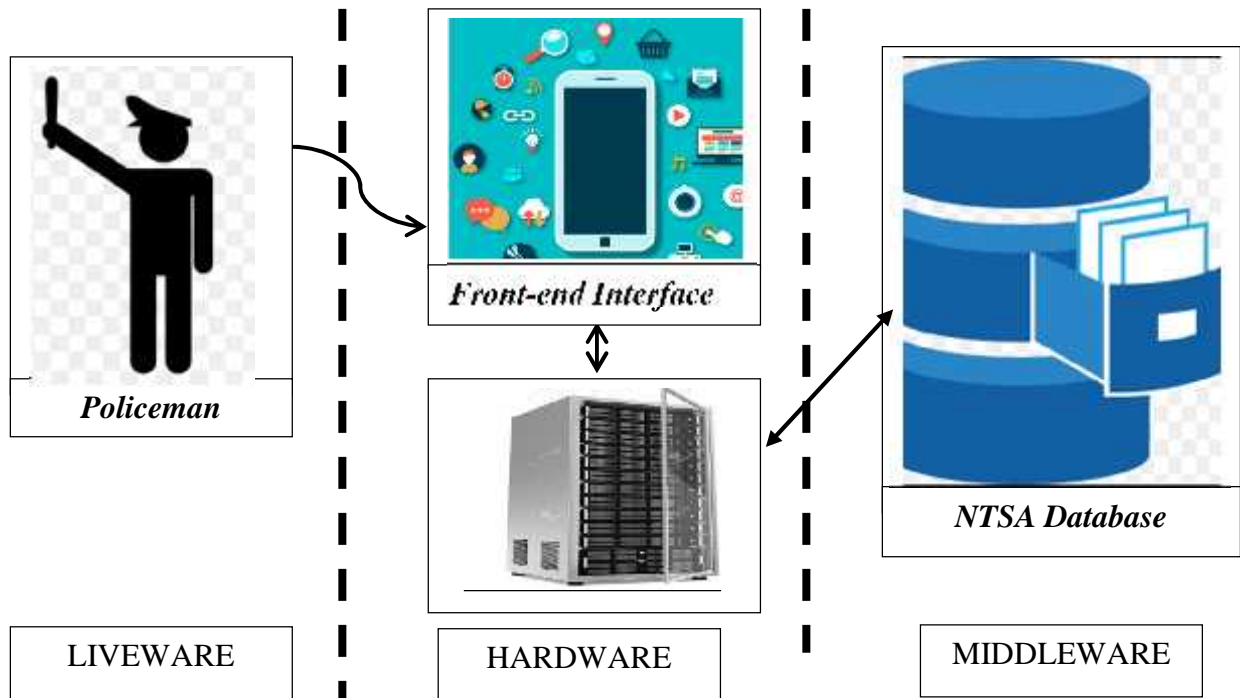
*Figure 5. 1: System Architecture Diagram*

## 5.3 System Design

The design of the traffic prediction system was done using the Unified Modeling Language (UML). The unified modeling language is universally accepted for object oriented languages. For this case, the object oriented language proposed for use shall be the JavaScript and PHP. The goal was to get the policeman to access an abstract system through running remote scripts from their gadgets, hence accessing the server, with the embedded PHP scripts. These scripts allow the policeman to access the SQL database that stores and manipulates data on frequent road user traffic offences. Here are design diagrams that demonstrate the layout of the system;

### 5.3.1 Use Case Diagram

The system has use cases to demonstrate the flow of data and relationship between actors in the system. Figure 5.2 demonstrates the use case diagram for the system
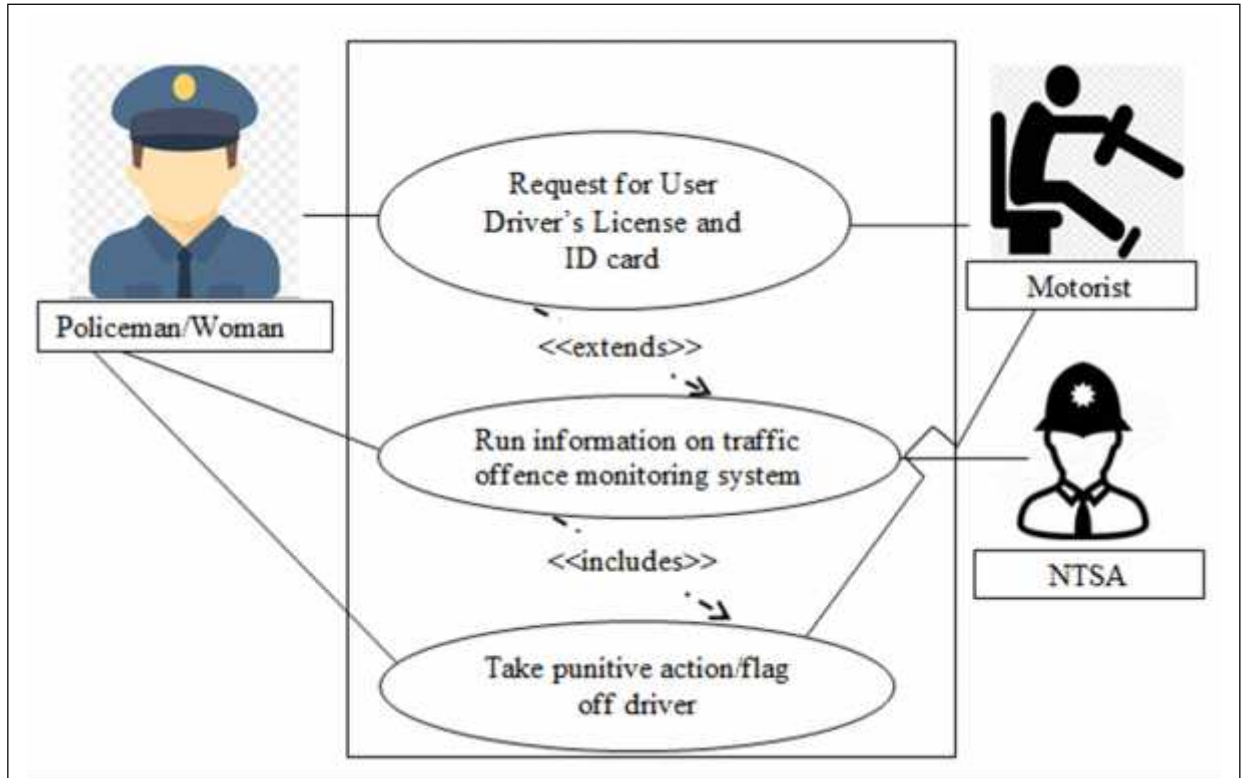


*Figure 5. 2: Use case diagram*

*Table 5. 1: Summary of the use case diagram*

| Actor | Use Case |
|---|---|
| Police Man/police woman | Requests for driver's ID and license for data entry into system |
| Motorist | Gives license information to policeman or woman |
| NTSA | Maintains information database |

Fig 5.2 shows the design of the use cases the system has. The basic use cases include; a policeman getting license information from the driver, filling the information in the system, and determining the frequency of an offence. The policeman runs the information on the accident

prediction system. This case requires that the NTSA and the Police be involved. The actor; NTSA offers system support in case any challenges exist with the system. Upon verifying the user data, the police officer then chooses to book the motorist for an offence or simply let the motorist drive off.

## 5.3.2 Sequence Diagram

In order to actualize the functionalities of the system, a sequence ensued, for communication between the front-end application, the server and the database (which is an abstract server layer). This required consistent functions to be called from the user and system server side. The policeman will first log into the system, make requests to a server, feed driver information into the system and update an offence by booking the offender. This can be represented as shown in Figure 5.3.
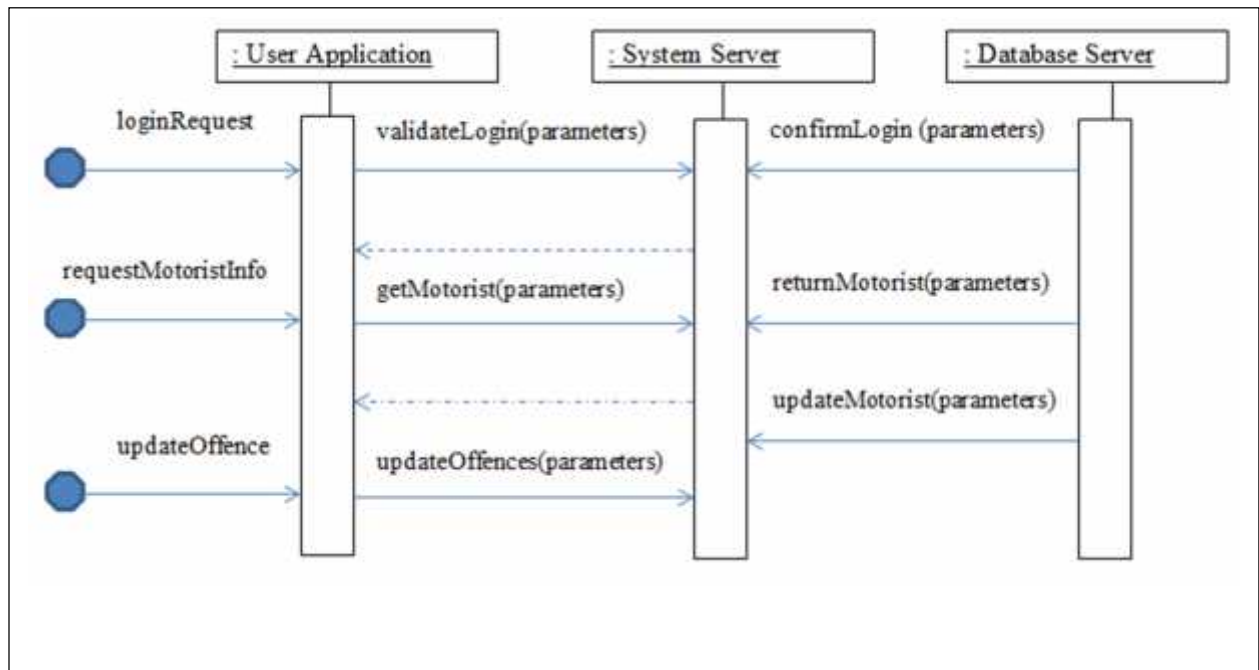


*Figure 5. 3: Sequence Diagram*

The sequence diagram demonstrates communication sessions between the user, the system server and the database server, all of which are critical elements within the system architecture. The sequence diagram is very important in determining how the system handle sessions.

### 5.3.3 Class Diagram

Classes are important tools for data encapsulation in any system. A class is a tool that contains the basic parameters of a module and is very significant to the system. The requirement for an effective class diagram is paramount. The system requires several classes to be fully functional. This is as shown in Figure 5.4.
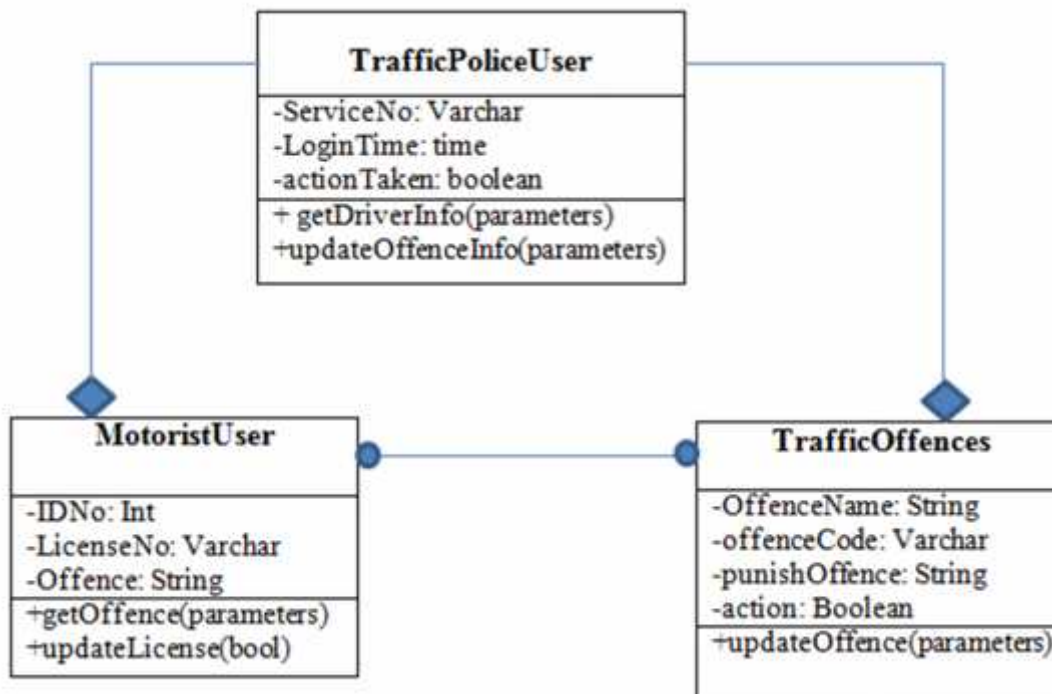


*Figure 5. 4: Class Diagram*

The class diagram in Figure 5.4 illustrates the main classes that are used in the system. Classes developed are the traffic police user, traffic offences and motorists. Each of the classes represent data saved into different tables in the database. This data is utilized in booking an offender and determining further action using the accident prediction system.

### 5.4 Context level diagram

The context level diagram indicates the interaction between the various system states as well as the entities, processes and data storage units. The context in which the system is used depends on whether the policeman wants to book the offender, whether the motorist wants to view information from the system or whether the NTSA wants to offer their support to the system. Each context have different players. The dataflow diagram is as shown;
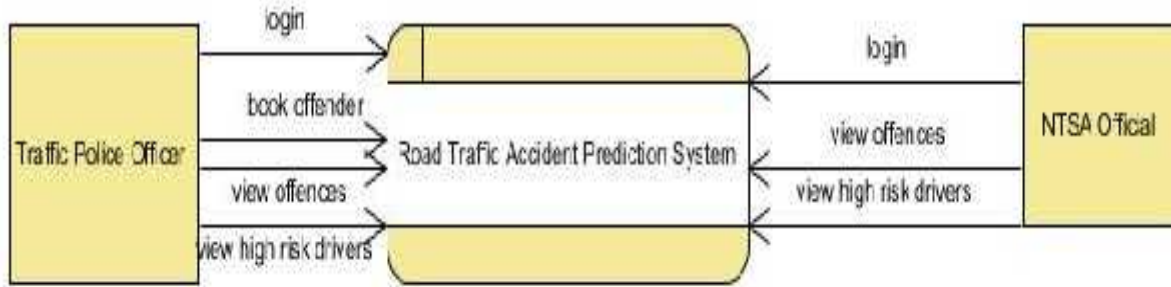
*Figure 5. 5: Context level diagram*

## 5.5 Level 0 Diagram

The level 0 diagram indicates the system's processes and activities. This includes all the activities that are either part of the system or key players within the system's environment. This includes all the inputs getting into the system from external players, as well as system users; and the output given out by the system. This is as shown in Figure 5.6
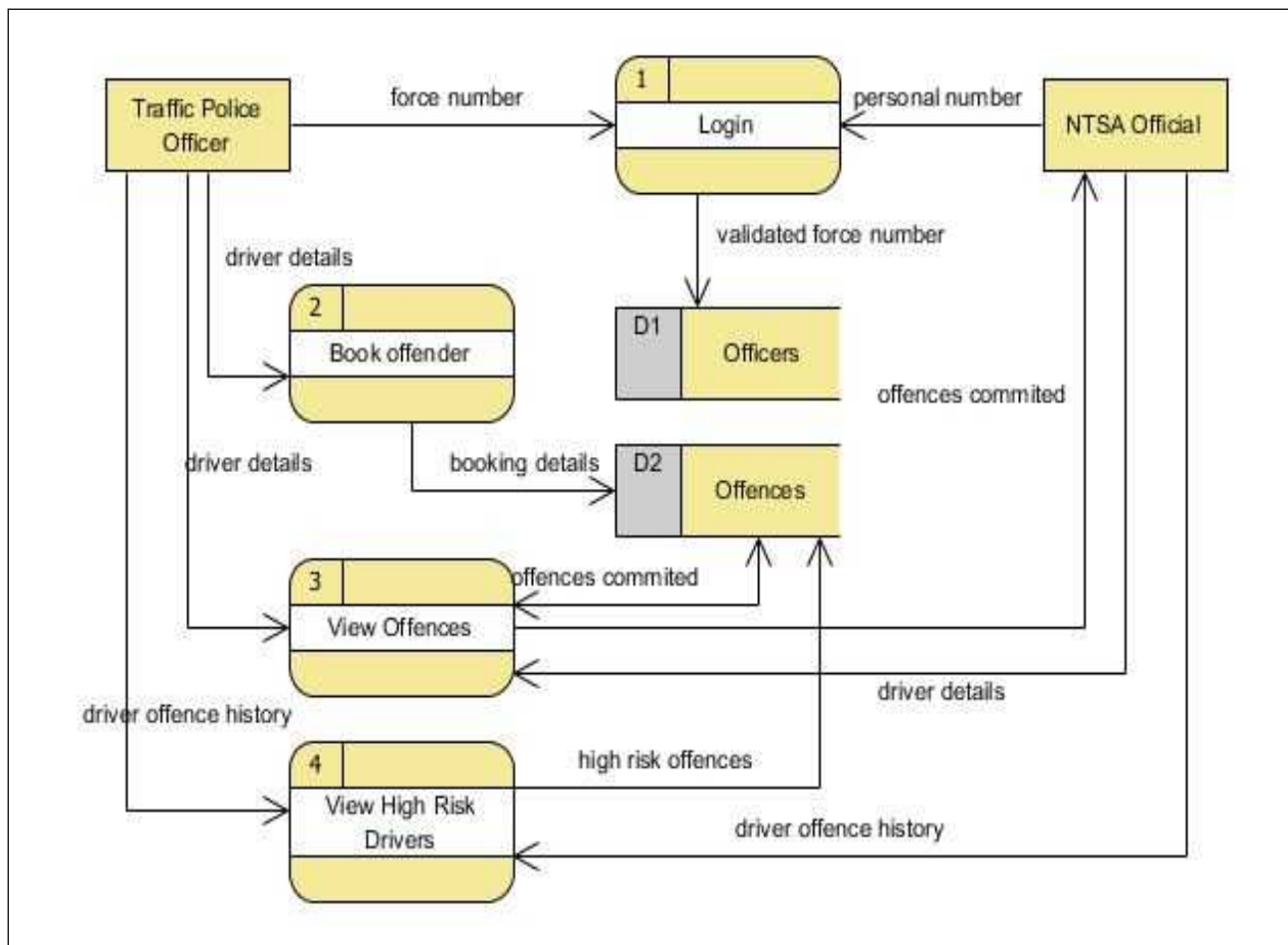


*Figure 5. 6: Level 0 Diagram*

## 5.5 Database Design

The database used in the traffic offence management system was developed using structured query language. Since the system was developed in an open-source environment, it was majorly meant to enforce cross-platform compatibility. The system was designed to utilize a number of interfaces, all with an equal access to the system's database. Various tables were modeled from the existing classes. These were structured into a relational database. This is as shown in Figure 5.7
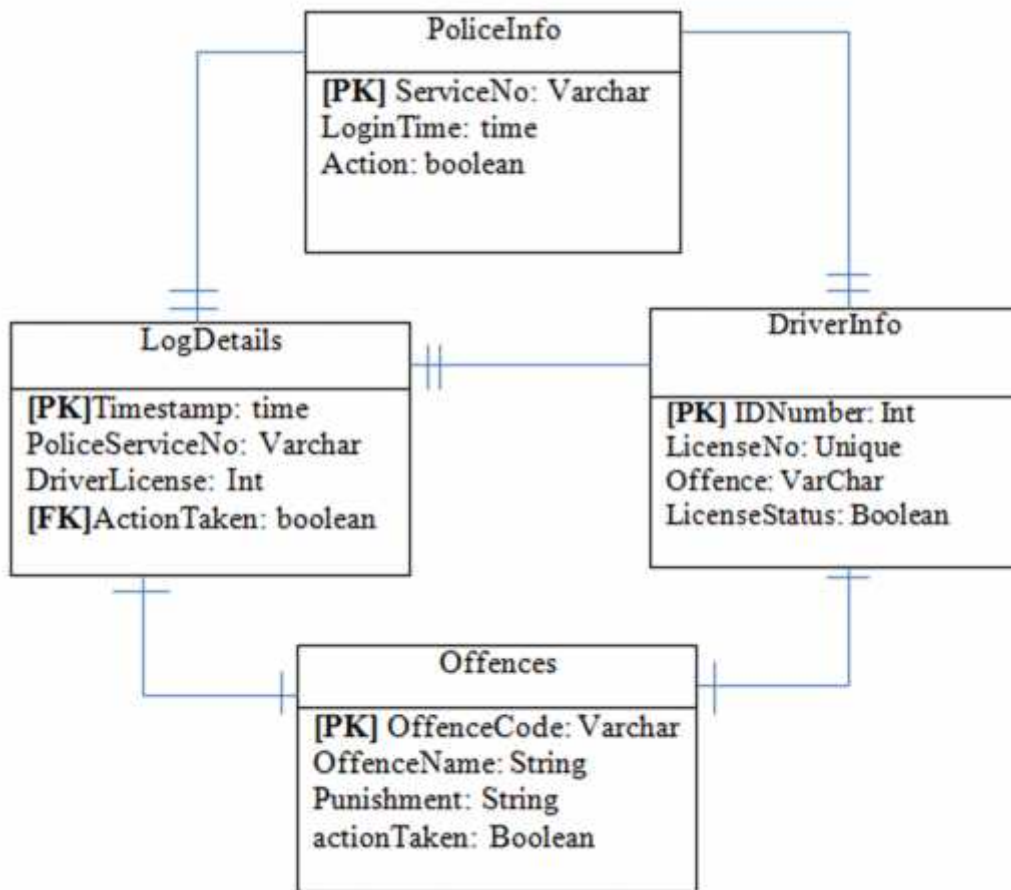


*Figure 5. 7: System Database Entity Relations*

The entity relationship diagram on Figure 5.7 demonstrates the layout of the database tables and the relationships between these tables. The tables will be interlinked using database queries. The queries will select data for various desired outputs. However, there will be an implemented algorithm for the analysis of the action to be taken by a policeman, depending on the offence

committed by the driver. This algorithm will also use previous user history from the tables and determine the culpability of the user based on history of traffic offences.

## 5.5 User Interface Design

The user interface determines the usability of a system to a great extent. Usable systems tend to be interactive and easy to understand, with little effort to capture data and perform other significant functionalities. The user often depends on the usability of a system to interact with it effectively. For the accident prediction system, the interface generally target the policeman, within a busy road. They thus not need to input a lot of data, in order to make the system not only usable but ease congestion while booking drivers on roads. The interface on Figure 5.8 demonstrates the login interface for the policemen on the roads.
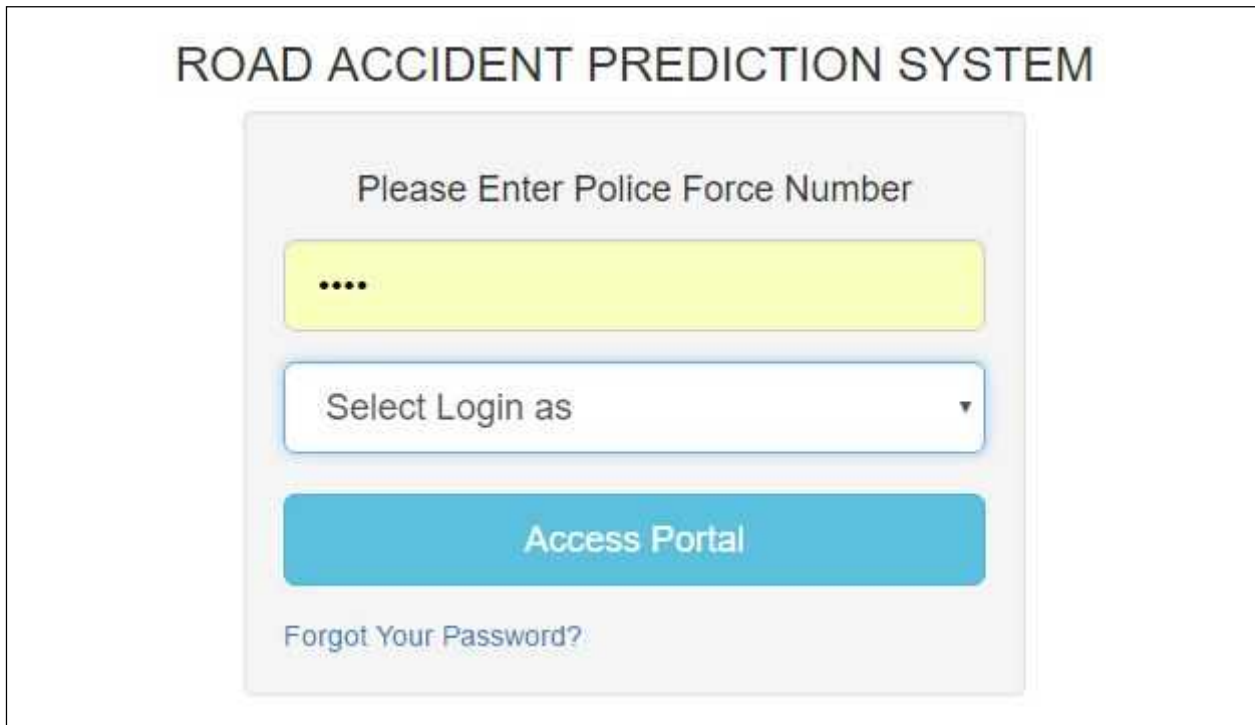


*Figure 5. 8: System Interface*

Upon Entry of driver's license number in the field of Driver offence, the driver's details is automatically displayed together with the offences committed and rating. More interfaces are shown in Appendix C.

## CHAPTER SIX: SYSTEM IMPLEMENTATION, TESTING AND DISCUSSIONS

### 6.1 System Requirements

### 6.1.1 Software Requirements

 - Windows XP service pack 2 OS or above versions

 - Net beans 7.0 and above

 - JDK 7

 - Wamp server

 - Microsoft Office 2007 suite/better

 - Internet resources

### 6.1.2 Hardware requirements

- Core i3 processor or a processor above that specification limit

 - 2 GB RAM

- Portable memory device (flash disk-8GB)

- 250 GB Hard disk

### 6.2 Input Requirements

To input data into the system, there is need for a graphical user interface; either a smartphone interface or a PC. A keyboard or keypad shall enter data into the system and thus allow the policeman to process driver road offences data. The interface also includes various command buttons that require a mouse or touch pad to manipulate.

### 6.3 Output Requirements

The system is able to perform querying from the database from all the details the system stores and display them for the police user. The system predicts the probability of a driver committing

an accident based on the prediction factors used by the system. These include; the frequency of times the user is found to have committed an offence.

## 6.4 Implementing the System

Once the hardware and software requirements of the system are met, implementation can be done. It involves the proper transitioning and changeover strategies if a similar system was in place. In this case however, no such system is presumed since the goal is to initiate a new mechanism to handle data and documents in general. Testing process of executing software in a controlled manner and examining it with data was carried out to check whether the output was as expected. The process involved coming up with test data and its expected output then trying the system with the data to examine its outputs. Unit and system and integration testing were carried out on the system. During these tests, errors were realized and corrected.

The changeover strategy for the system shall be the phased changeover technique. Policemen in different parts of the country will have the application installed on their cell phones or personal digital assistants. The policemen shall use existing data from the NTSA servers, on driver offences and accidents. Once a driver is flagged, they shall be asked for their driving licence number, upon which the policeman shall be able to tell the offences committed by the driver and tell whether or not the driver is likely to commit an accident.

## 6.5 System Testing

Testing was done by asking the respondents involved in the research to use the system and given their opinions regarding the use of the system. The major components reviewed on the system's interfaces shown in Appendix C include; predictability of road accidents, proper management of traffic data and accurate determination of drivers to prosecute on roads. The three aspects were tested independently among the 36 respondents.

### 6.5.1 Results for predictability of road accidents

Of the thirty-six respondents assessed, twenty-two were from the NTSA. The rest of the respondents were policemen and women. The findings from the clients can be represented in the graph on figure 6.1.
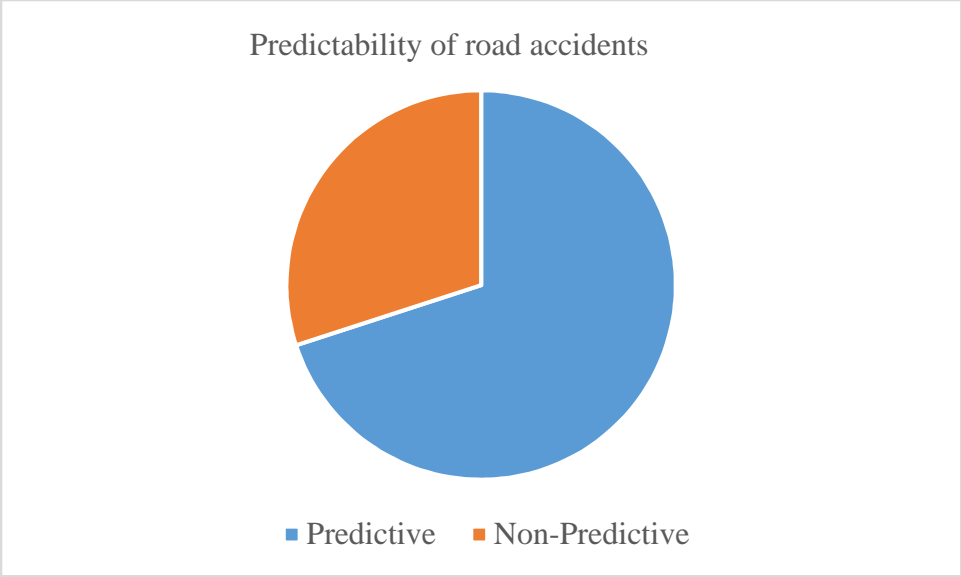
*Figure 6. 1: Predictability of road accidents*

**6.5.2 Proper management of traffic data**

With regard to the proper management of traffic data, the research indicated that 33 of the 36 respondents had confidence in the system's ability to properly manage traffic data using a database. This is as shown in the graph on figure 6.2
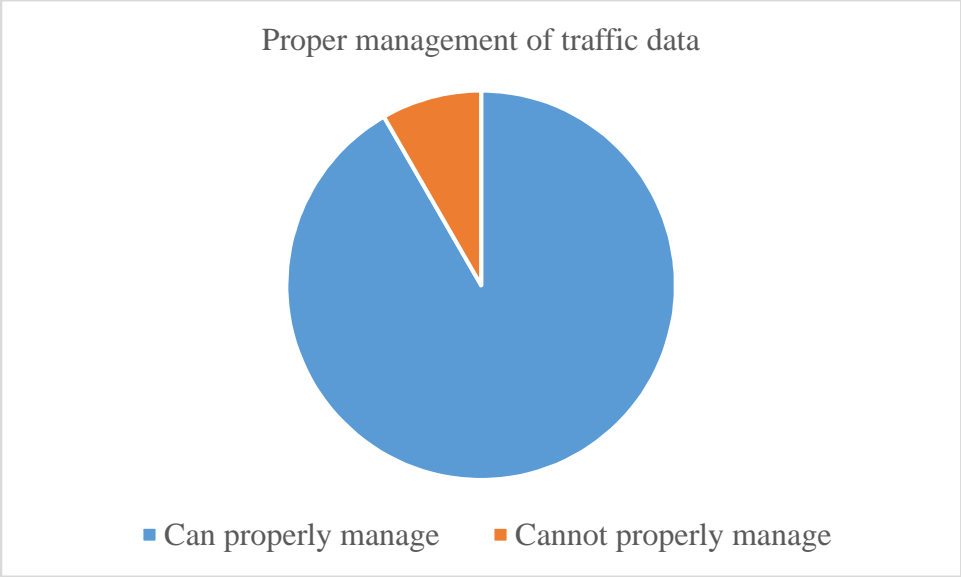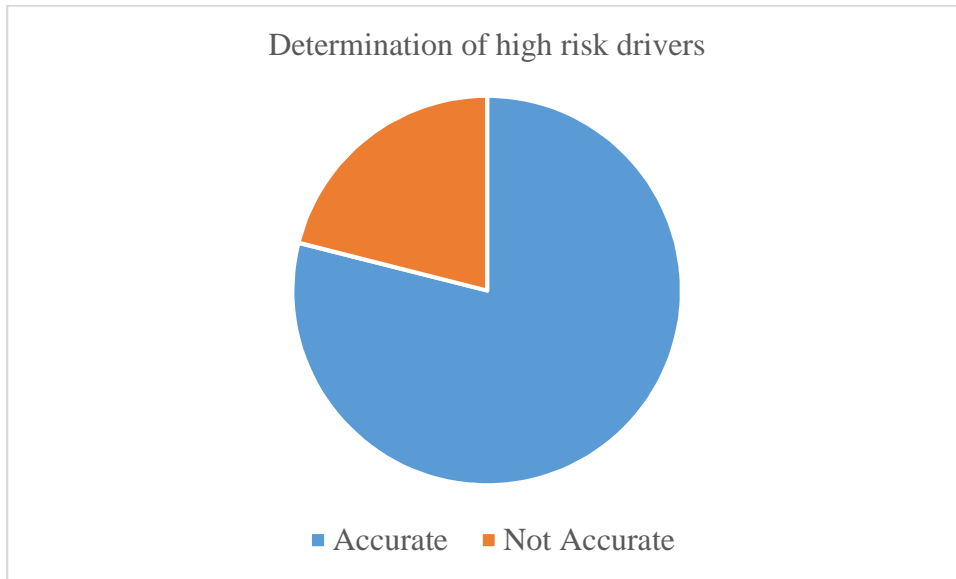


*Figure 6. 2: Proper management of traffic data*

### 6.5.3 Accurate determination of drivers to prosecute

Accurate determination of drivers to prosecute was another important aspect of the system. This research question had a significant deviation from the responses given as compared to the previous two aspects of the system tested. This is as shown in figure 6.3



*Figure 6. 3: Determination of high risk drivers*

### 6.6 Discussions

The research done in this project encompassed the use of both primary and secondary data. Primary data was sourced from different stakeholders in the Kenyan transport and safety Authority as well as the traffic police officers within the Kenya Police Service; who have interacted with road offenders and have witnessed motor vehicle accidents in the past. The research also sought input from public service vehicle drivers. From their findings, there was need for an effective system to monitor road use in Kenya, and minimize traffic accidents that were all too common in the country. The input from these respondents was critical in actualizing the design and development of the accident prediction system. However, there was need for more study on how best to implement the system in a holistic manner, covering the public transport infrastructure, minor offences that were committed by motorists, cyclists and pedestrians on Kenyan roads. The research revealed a lot of enthusiasm from the stakeholders in the transport sector on the need to arrest the accident problem in Kenya, minimizing consistent accidents on Kenyan roads.

A lot of secondary data and information was acquired from past literature work as well. Such information detailed the continued loss of lives on Kenyan roads due to road accidents that could be avoided. The research from external research work revealed details such as the number of times traffic police officers in the country were unable to flag down offenders and the number of repeat offenders who had caused fatal accidents. Repeat offenders were majorly involved in accidents such as over speeding-related car crashes, dangerous overtaking issues and even the use of road unworthy vehicles on Kenyan roads. Many Kenyans involved in these accidents were also not aware of the previous histories these drivers had on roads. Drivers were also consistently flagged for offences that could be avoided. However, they were often able to escape arrest and punitive measures by changing routes, hiring different persons to take charge of the vehicles and even bribing the police. This was confirmed from both the primary and secondary research work. The research thus proved that there was need for a system to track accident-likely vehicles in Kenya, flag their drivers and book them for punitive action. This would reduce the casualties on Kenyan roads.

# CHAPTER SEVEN: CONCLUSIONS AND RECOMMENDATIONS

## 7.1Conclusions

The research determined that dangerous driving was the main contributor to accidents in Kenya. Dangerous driving in this case involved; over speeding, dangerous overtaking of vehicles, overlapping, drunk driving and the use of mobile phones while driving. These factors made it possible for drivers in the city to put pedestrians and passengers at risk. The research may have majorly focused on public service vehicles but the indication was; while these vehicles were contributors to a majority of the accidents; there were other vehicles that contributed to some of the accidents, which were from personal vehicle owners. Accidents were also associated with weather conditions and the state of roads in Kenya.

The research revealed that traffic accident surveillance schemes, logistic regression methods and zero inflated negative binomial regression methods are the major available accident-prediction systems in use. These systems have been used consistently to help many countries across the world to determine risky drivers, vehicles and roads in their traffic systems. Drivers have thus been cautioned from infringing on some of the traffic laws in the country, and always becoming complacent to the use of roads. Systems for traffic monitoring and management have also been assisted by the fact that many of the stakeholders in the transport sector in these countries have been directly involved in the road safety campaigns, thus willing to support the existing systems.

The ultimate goal of the research was to develop an algorithm for predicting accidents in Kenya. This algorithm would be designed and implemented in a system, developed from PHP, JavaScript and HTML languages. The research work eventually bore fruit and realized a system that could use motorist data to come up with an ultimate system for accident prediction in Kenya. The system would also determine how best to handle issues such as consistent infringement of traffic laws and support policemen on roads to monitor offences committed by drivers. The system was designed in a manner that helped the existing setup of traffic monitoring; involving NTSA officers and traffic policemen to consistently monitor drivers and come up with an effective way to monitor accidents in Kenya.

Validation of the algorithm-based system for accident prediction was done through the use of sufficient road monitoring systems. It was especially done using test data from previous research. Such data indicated that the trend of accidents in Kenya increased as the Kenyan road network expanded. Testing of the system was done using test done, based on anticipated results from a traffic stop. The test cases were the NTSA officials and traffic policemen that had been involved in the research. Through their efforts, the test was effectively conclusive and could establish a conclusive finding. The data fed into the system returned results that could predict the possibility of an accident on Kenyan roads. These tests affirmed that by having the system developed and implemented in full, there would be fewer accidents on Kenyan roads. Road accidents would also be predicted right before they happened.

The research proved that indeed, an accident prediction system can be designed and implemented in Kenya. This system was designed using the parameters used by a traffic police officer to note down an offender and flag them for offences committed. These parameters were then used to effectively develop the algorithm-based system for accident prediction in Kenya. Since the input parameters used in booking a traffic offender were similar to the expected inputs by the system, the design of the system was quite flexible and appealing. It did not require too many details to develop the system interface. The study population involved in the development of the system was also the same population that had assisted in system research. This meant that the system did not really require further development research, until it had been optimally used in the Kenyan transport sector.

## 7.2 Recommendations

These recommendations shall improve the nature of the management of traffic data in Kenya, thus minimizing road accidents in the future.

a. All traffic offenders should be listed in a database that is updated regularly
b. Traffic offence data should be referenced even in Kenyan courts of law
c. There is need for the government departments responsible for managing traffic such as the National Transport and Safety Authority and the Ministry of Transport to consider developing this research further into an electronic system for real-time traffic offence booking and management

d. There is need for a web-based database where drivers and other road users can check their traffic offence status and get the necessary recommendations to improve their road discipline, including possible license revocation.

e. If these measures are adopted, there is need for phased changeover strategies do as to manage the infrastructure costs that come with the system

## 7.3 Suggestions for further study

Among the areas noted for further study research include;

i. The research and improvement of the process of flagging vehicles on Kenyan roads; including possible flagging bays.

ii. There is also need to have an interactive traffic control system on roads; handling the challenges people in Nairobi face such as traffic jams and unnecessary roundabouts.

iii. There is need for research on the algorithms that link vehicle condition to traffic accidents. These algorithms should be based on a study of the mechanical diagnosis of vehicles in a bid to understand when to stop driving a car, how best to deal with certain vehicle damages and the need for proper electrical and electronic systems on cars.

iv. Scientific research is necessary to establish the degree to which conditions on roads and vehicles can render the two accident-prone. Such research would greatly advance the available research on accidents in Kenya, and minimize the extent to which drivers in Kenya are exposed to accidents.

# REFERENCES

Agresti, A. (2007). *An Introduction to Categorical Data Analysis.* Wiley.

Åkerstedt, T., Ingre, M., Broman, J. E., & Kecklund, G. (2008). Disturbed sleep in shift workers, day workers, and insomniacs. *Chronobiology international*, *25*(2-3), 333-348.

Bachani, A., Koradia, P., Herbert, H., Mogere, S., & Akungah, D. (2012). Road traffic injuries in Kenya: the health burden and risk factors in two districts. *Traffic Injury Prevention*, 24-30.

Best, J., & Khan, J. (2006). *Research in education.* New Jersey: Pearson Education Inc.

Caliendo, C., Guida, M., & Parisi, A. (2007). A crash-prediction model for multilane roads. *Accident Analysis & Prevention*, *39*(4), 657-670.

Creswell, J. (2014). *Research design : qualitative, quantitative, and mixed methods approaches* (4th ed.). Sage Publications.

Das, S., Sun, X., Wang, F., & Leboeuf, C. (2015). Estimating likelihood of future crashes for crash-prone drivers. *Traffic and Transportation Engineering*, 145-157.

Goldenbeld, C., Reurings , M., Van Norden, Y., & Stipdonk, H. (2013). Crash involvement of motor vehicles in relationship to the number and severity of traffic offenses. An exploratory analysis of Dutch traffic offenses and crash data. *Traffic Injury Prevention*, 584-591.

Hu, W., Xiao, X., & Xie, D. (2003). Traffic accident prediction using vehicle tracking and trajectory analysis. *Intelligent Transportation Systems.*

Ipsos Synovate. (2012). *Road Safety.* Retrieved May 31, 2016, from http://www.ipsos.co.ke/report_downloads/downloads.php?dir=lifestyle_issues&file=Road%20Safety_Majority%20of%20Kenyans%20do%20not%20take%20personal%20responsibility%20for%20road%20safety%20(October%202012).pdf

Kenya Police Service. (2016). *Traffic police department*. Retrieved June 1, 2016, from Kenya police service: http://www.kenyapolice.go.ke/2015-09-07-17-41-13/traffic-police-department.html

Kim, D.-H., Ramjan, L., & Mak, K.-K. (2016). Prediction of vehicle crashes by drivers' characteristics and past traffic violations in Korea using a zero-inflated negative binomial model. *Traffic Injury Prevention*, 86-90.

Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and psychological measurement*, *30*(3), 607-610.

Kothari, C. (2004). *Research methodology: methods and techniques* (2nd ed.). New Delhi: New Age International Ltd.

Kual, A., Sinha, V., Pathak, Y., Singh, A., Kopoor, A., Sharma, S., & Singh, S. (2005). Fatal Road Traffic Accidents, Study of Distribution, Nature and Type of Injury. *Journal of Indian Academy of Forensic Medicine*, 71-78.

Lu, T., Lixin, Y., Dunyao, Z., & Pan, Z. (2015). The traffic accident hotspot prediction: based on the logistic regression method. *Transportation Information and Safety.* Wuhan.

Muchene, L. (2013). *Road accidents in Kenya: a case of poor road network or human error?* Universiteit Hasselt, Diepenbeek. Retrieved May 31, 2016, from http://www.statistics.gov.hk/wsc/CPS030-P2-S.pdf

Mugenda, O., & Mugenda, A. (1999). *Research methods: quantitative & qualitative approaches.* Nairobi: Acts Press.

National Transport and Safety Authority. (2016). *Road Safety status report 2015.* Nairobi.

Odero, W. (1998). Alcohol related road traffic injuries . *East African Medical Journal*, 708-711.

Odero, W., Khayesi, M., & Heda, P. (2003). Road traffic injuries in Kenya: magnitude, causes and status of intervention. *Injury Control and Safety Promotion*, 53-61.

Ogula, P. (2005). *Research Methods.* Nairobi: CUEA Publications.

Osoro, A., Ng'ang'a, Z., & Yitambe, A. (2015). An analysis of the incidence and causes of road traffic accident in Kisii, Central district, Kenya. *IOSR Journal Of Pharmacy*, 41-49.

Partheeban, P., Arunbabu, E., & Hemamalini, R. R. (2008). Road accident cost prediction model using systems dynamics approach. *Transport*, *23*(1), 59-66.

Shabeer, A., & Banu, W. (2012). Mobile phone accidents-experience of India. *Transport and Telecommunication*, 193-208.

Shanthi, S., & Ramani, R. G. (2011). Classification of vehicle collision patterns in road accidents using data mining algorithms. *International Journal of Computer Applications, 35(12), 30-37*.

Sun, X., Das, S., Sk, R., & Wang, F. (2013). Predicting Drivers' Crash Risk Based-on Previous Crash History. *Road Safety on Four Continents.* Beijing,.

Vale, S. (2010). *Statistical data quality in the UNECE, 2010 version.*

World Health Organization. (2009). *Global status report on road safety: time for action.* Geneva.

Yu, L., & Liu, H. (2003). Feature selection for high-dimensional data: A fast correlation-based filter solution. In *ICML* (Vol. 3, pp. 856-863).

# APPENDICES

## Appendix A: Research Questionnaire for Police and NTSA

### User Requirement Questionnaire

**Researcher**: Jwan Levice Obongo

MSIS, Strathmore University

This research will be used for academic purpose only. Its main objective is to collect user requirements to develop a predictive algorithm for road accidents in Kenya. Kindly provide your honest answers to the following questions. Please note that your responses will be treated as private and confidential.

Interviewee/Designation:……………………………………… Location:…………………………

Date……………………………

1. What information is required when booking a road offender?

……………………………………………………………………………………………………..

……………………………………………………………………………………………………..

……………………………………………………………………………………………………..

**Please Tick One**

2. The following offences are the main causes of road accidents in Nairobi.

| | Strongly Agree | Agree | Neutral | Disagree | Strongly Disagree |
|---|---|---|---|---|---|
| Speeding | | | | | |
| Drunk driving | | | | | |
| Reckless overtaking | | | | | |
| Overlapping | | | | | |
| Using phone while driving | | | | | |

3. I believe that a driver with multiple counts of road offences has a higher chance of committing an accident?

☐ Strongly Agree
☐ Agree
☐ Neutral
☐ Disagree
☐ Strongly Disagree

4. The current method of identifying drivers likely to cause accidents is efficient.

☐ Strongly Agree
☐ Agree
☐ Neutral
☐ Disagree
☐ Strongly Disagree

5. It is easy to track a frequent road offender using the current system.

☐ Strongly Agree
☐ Agree
☐ Neutral
☐ Disagree
☐ Strongly Disagree

6. If a proper computer system is implemented, I believe that tracking of frequent offenders and identification of high risk drivers would be made easier.

☐ Strongly Agree
☐ Agree
☐ Neutral
☐ Disagree
☐ Strongly Disagree

## Appendix B: Interview Questions for Matatu Operators

Researcher: Jwan Levice Obongo

MSIS, Strathmore University

This research will be used for academic purpose only. Its main objective is to collect user requirements to develop a predictive algorithm for road accidents in Kenya. Kindly provide your honest opinions. Please note that your responses will be treated as private and confidential.
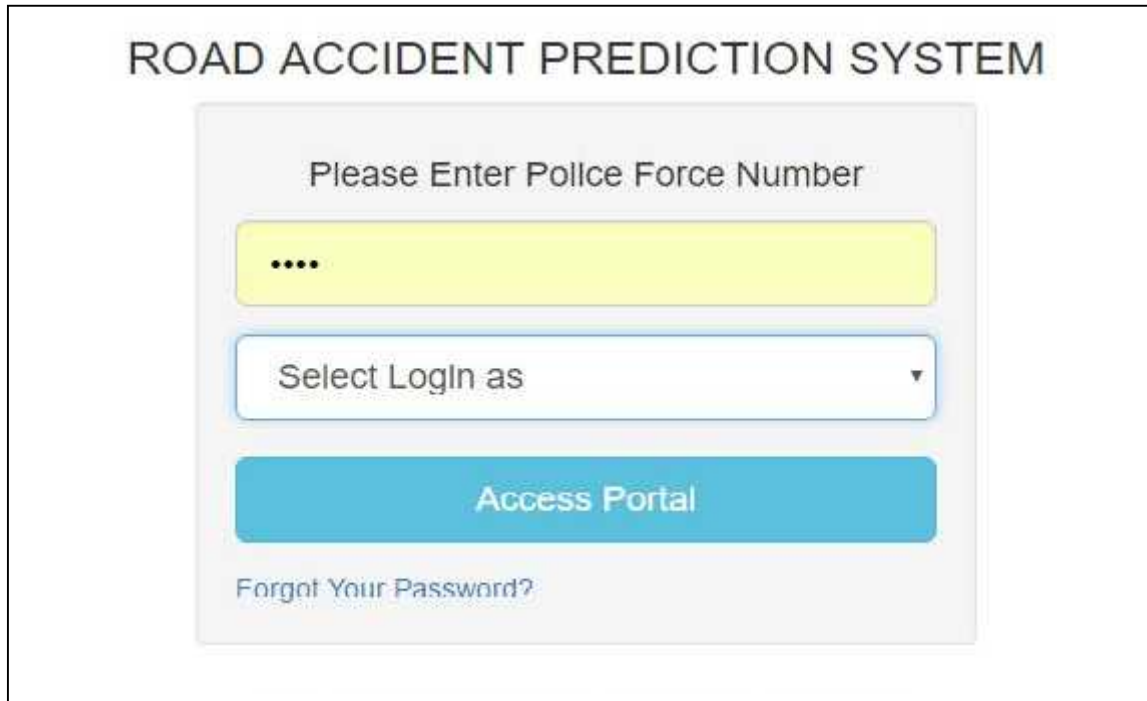Interviewee:……………………………………………. Location:…………………………
Date……………………………

1.      How long have you worked in the transport business?
…………………………………………………………………………………………………
…………………………………………………………………………………………………

2.      Have you ever been in an accident before?
………………………………………………………………………………………………..
…………………………………………………………………………………………………

3.      What factors do you consider as the main determinants to road accident in Nairobi?
………………………………………………………………………………………………..
………………………………………………………………………………………………..
………………………………………………………………………………………………..

4.      Do you believe that there is a relationship between traffic offences and accidents in Kenya? Explain
………………………………………………………………………………………………..
………………………………………………………………………………………………..
…………………………………………………………………………………………………

5.      Would you subject yourself to vetting and monitoring of traffic offences committed periodically?
………………………………………………………………………………………………..
………………………………………………………………………………………………..

6.      Would you support a system that suspended frequent traffic offenders from driving in Kenya?
………………………………………………………………………………………………..
………………………………………………………………………………………………..
………………………………………………………………………………………………..

Thank you for your time

**Appendix C: Accessing the System**

**C.1 The login session**

A login prompt indicates the first interface the user has to interact with before accessing the system. The login prompt is as shown in C.1



Figure C. 1: Login Interface

The options to log into the system include; NTSA officers and Police officers. Once the user selects the option 'NTSA Officer', the interface in Figure C.2 appears

Figure C. 2: Dashboard for NTSA officers accessing the system

If the user logs into the system as a police officer, the Interface in Figure C.3 appears;



Figure C. 3: Dashboard for police officers accessing the system

## C.2 Simulating the System

Once a driver is booked, the offence is indicated in the system for analysis and possible diagnosis of conditions. This includes determination of the driver's offence and possibility of committing an accident. Sample operationsof the system are as shown in Figure C.3



Figure C. 4: Form for booking a traffic offender

Once the traffic offender has been booked, the police officer can view the previous offences as well as the recently booked offence through the form shown in figure C.5

Figure C. 5: Form for viewing booked offences through a search algorithm

## Appendix D: Turnitin Originality Report

# An Algorithm for Predicting Road Accidents Based on Traffic Offence Data

**ORIGINALITY REPORT**

| 23% | 18% | 11% | 14% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

turnitin

| Class Portfolio | Peer Review | My Grades | Discussion | Calendar |

NOW VIEWING: HOME > THESIS

Welcome to your new class homepage! From the class homepage you can see all your assignments for your class, view additional assignment information, submit your work, and access feedback for your papers.

Hover on any item in the class homepage for more information.

### Class Homepage

This is your class homepage. To submit to an assignment click on the "Submit" button to the right of the assignment name. If the Submit button is grayed out, no submissions can be made to the assignment. If resubmissions are allowed the submit button will read "Resubmit" after you make your first submission to the assignment. To view the paper you have submitted, click the "View" button. Once the assignment's post date has passed, you will also be able to view the feedback left on your paper by clicking the "View" button.

#### Assignment Inbox: Thesis

| | Info | Dates | | Similarity | |
|---|---|---|---|---|---|
| Thesis | ⓘ | Start 25-Apr-2017<br>Due 23-Jun-2017<br>Post 23-Jun-2017 | 8:40PM<br>11:59PM<br>12:00AM | 23% ▪ | Resubmit View ⬇ |