FACULTY OF COMPUTING & INFORMATION MANAGEMENT

RESEARCH PROJECT

ON

USING INFORMATION GAIN TO EVALUATE WEIGH-IN-MOTION AXLE LOAD MANAGEMENT INFORMATION SYSTEM

BY

GEORGE O. ODONGO

A RESEARCH PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER OF SCIENCE IN INFORMATION SYSTEMS MANAGEMENT AT KCA UNIVERSITY

OCTOBER 2017

DECLARATION

I declare that this Research Project is my original work and has not been previously published or submitted elsewhere for award of a degree. I also declare that this work contains no material written or published by other people except where due reference is made and author acknowledged.

Student Name: Odongo George Otieno   Reg/No: 14/04325

Signature……………………………………………Date…………………………………………….

This project has been presented for examination with my approval as the appointed supervisor.

Dr. Henry Mwangi (Supervisor)

Faculty of Computing & Information Management KCA University

Signed ……………………………………………..   Date ………………………..

Dr. Richard Rimiru (Supervisor)

Faculty of Computing & Information Management KCA University

Signed ……………………………………….……….. Date

………………………...

ABSTRACT

Delays and other operational inefficiencies at weighbridges in Kenya has remained persistent over the years. This delays have led to astronomical rise in transport costs to transporters. These costs are mostly passed on to consumers of the transported goods. Causes of the delays at weighbridges stem from inspection of legal requirements such as axle load limits, validity of load permits, validity of driving license, drunk drivers and possession of fire extinguisher equipment. Not all trucks screened at the weighbridge violate these laws and regulations. The research therefore sought to establish a pattern using data mining techniques and tools from data generated from the weighbridges. Data on origins, destinations and compliance o these requirements was obtained from weighbridge data at Webuye station. Out of tens of thousands of data at the database, two thousand five hundred and fifty data items were randomly selected for this study. The selected sample was divided into training data set for building the model and testing data set for testing the model. WEKA data mining software was used to perform the data mining. J 48 classification algorithm in WEKA was employed to build the model by establishing a patter that linked origins, destinations and the likelihood of committing the offences. Performance metrics of the model indicated that there was a strong predictive accuracy by the model using the F-measure 79.9%, True Positive Rate 82.7%, Recall 82.7% and Precision at 81.1%. from the results of this study decision makers and policy formulators at the weighbridges may use them to institute policy that will address vehicle screening process based on their destinations and origins with an aim of implementing operational efficiency at the weighbridges.

Key Words: Data mining, algorithm, classification techniques, weigh-in-motion, axle load control and

ACKNOWLEDGEMENT

My special thanks go to my supervisors Dr. Richard Rimiru and Dr. Henry Mwangi for their guidance, patience, and dedication without which I would not have not successfully concluded this study. Am greatly indebted to my employer Kenya National Highways Authority (KeNHA) for having accorded me the opportunity to pursue this course. I also give special thanks to KCA University Faculty of Computing and Information Management for the delivery of course content rich management and problem solving skills. To all my colleagues at work who encouraged me to pursue this area of study, I say thank you so much and may the Almighty God bless you abundantly. Finally, but not the least, I give special gratitude to Almighty God for His enduring mercies that saw me through to the end of the course.

TABLE OF CONTENTS

DEDICATION

I would like to dedicate this Research Project to my dear wife Eunice Orina and daughters Kaisy Esther, Candy Otieno and Clancy Hope for their prayers, encouragement and support during the period of study, especially when I couldn't accompany them to church service and instead attended classes on Saturday (Sabbath Day). On several occasions, they urged me on with empowering goodwill messages and wishes. I also dedicate this project to my late parents, my mother Esther Akello and father Charles Odongo whose wise counsel, love and care saw me through my education and made me what I am today. I also appreciate support and words of encouragement from my brothers & sisters, especially my elder brother Joseph Odongo for his vision and determination that saw all his siblings pursue education against all odds.

LIST OF FIGURES

List of Tables

ABBREVIATIONS & ACRONYMS

ANPR	-	Automatic Number Plate Recognition systems
CCTV	-	Closed Circuit Television
CSV	-	Comma Separated Value
GoK	-	Government of Kenya
GVW	-	Gross Vehicle Weight
HSWIM	-	High Speed Weigh in Motion
IBRD	-	International Bank for Reconstruction and Development
KDD	-	Knowledge Discovery from Databases
KeNHA	-	Kenya National Highways Authority
KeRRA	-	Kenya Rural Roads Authority
KURA	-	Kenya Urban Roads Authority
LSWIM	-	Low Speed Weigh-in-Motion
MIS	-	Management Information Systems
ML	-	Machine Learning
RFID	-	Radio Frequency Identification System
SSADM	-	Structured Systems Analysis and Design Model
SVM	-	Support Vector Machine
WB	-	World Bank

WEKA	-	Waikato Environment for Knowledge Analysis
WIM	-	Weigh-In-Motion

OPERATIONAL DEFINITION OF TERMS

Weigh-in-motion (WIM):

Is a system equipped with an ability to measure the axle loads of vehicles, while trucks pass over installed sensors.

Weighbridge:

Is the road pavement where the vehicle weighing scale is installed/mounted for purposes of weighing vehicle by passing over it.

Axle load:

Is the load exerted by a motor vehicle through the wheels connected by the same axle.

Electronic screening:

e-screening sorts and diverts commercial vehicles that have a high likelihood of being overweight into a weigh station using weigh-in-motion technology.

Information Technology:

As defined by the Information Technology Association of America (ITAA) is the study, design, development, implementation, support or management of computer-based information systems, particularly software applications and computer hardware.

Information Systems:

Any written electronic or graphical method of communicating information.

KenWei System:

Is a weighbridge management system used on Kenyan weighbridge stations for the weighing operations.

Entropy

The entropy is what characterizes the (im)purity of an arbitrary collection of examples.

Information Gain is

Information Gain is the expected reduction in entropy caused by partitioning the examples according to a given attribute.

High Speed Weigh-In-Motion:

Is a weighing scale that is capable of capturing vehicle weigh while the vehicle is moving at a high speed of up to 80 km/hr.

CHAPTER ONE: INTRODUCTION

1.1 Background of the Study

Road transport is one of the mostly used mode of transport in Kenya and more so in the transportation of goods across the country and outside its borders. To ensure an orderly road transport, there is need to ensure adherence to road and traffic laws and regulations that govern users of the roads. Provision of efficient road transport is one of the top Government priority investments areas in Kenya as demonstrated by the annual budget allocations to road investment initiatives. In the financial year 2016/2017, the Government of Kenya allocated Kshs. 117.6 billion to roads sector. (GoK Budget statement, 2016). To realize the vision of efficient road network, it is imperative to maintain and manage activities on the roads as efficiently as possible. One such threat to efficient management of activities on the roads in many countries is violation of axle load control laws and regulations (Pinard, 2011). A lot of time is spent at the weighbridges in the process of weighing motor vehicles trying to ascertain compliance with legal requirements by truck owners. When trucks overload beyond the allowable axle load limits, the roads tend to deteriorate faster than anticipated, hence not lasting to their full lifespan. In Namibia, the average age of the bitumen road network based on the date of the first upgrade is 25.8 years, but improves to 23.1 years when the major rehabilitation interventions done over the period are taken into account. On average, the bitumen road network is therefore serving beyond its design life and major rehabilitation will be required in the short to medium term. (Pinard 2011).

To control overload and other the offenses by trucks on the roads, weighbridges have been installed at strategic locations along the road network. The weighbridges are meant to control overloading and detect compliance to other traffic offenses by screening overloaded vehicles for legal action in addition to enforcing compliance to legal requirements such as valid permits, drink driving, possession of valid driving license, fire extinguishers and functional first aid kits. (Kenya Traffic Amendment Act, 2012). The process of ascertain the legal requirements mentioned contribute to the overall time taken to clear a truck from a weighbridge station. This has led to long queues witnessed at weighbridges translating into time wastage and higher transportation costs to transporters. This research sought to propose an alternative way to ascertain the legal compliance requirements by trucks in a more efficient manner that saves time and transporter costs. Despite the installed weighbridges in Kenya's road network, overloading and the other traffic offences has continued unabated. Alongside the weighing scales at weighbridges an axle load management information system (MIS) has been used to detect, record and store information about overloading trucks for effective weight control. One important development towards this end is the use of Information Technology at designated points on the roads to weigh vehicles for purposes of determining compliance with the set axle-load limits. Use of MIS is still not properly used on Kenyan weighbridges. The application/system currently in use in Kenya called ''KenWei'' system has not yielded desired results in controlling overloading and the other offences committed by transporters. Perhaps, loopholes in the KenWei has provided an opportunity for transporters to overload and still escape from prosecution by government authorities. ''Trucks exceeding legal weight limits increase the risks of traffic accidents and damage to the infrastructure. They also result into unfair competition between transport modes and companies. It is therefore important to ensure that trucks' compliance to weight regulations and obedience to traffic rules and other legal requirements is enforced in a more efficient manner that does not result into loss of time at the weighbridges. Data mining technique was employed in this research to establish pattern between the offenses, origin and destinations of the trucks to aid in policy formulation and decision making by persons responsible for enforcement of the traffic laws and regulations. The rate at which the weighbridge systems generate data is enormous and thousands of data is generated yearly at the weighbridges. (Tiseme T.B 2011) explains that it is estimated that the amount of data stored in the world's database grows every twenty months at a rate of 100%. This fact shows that we are getting more and more exploded by data/information and yet ravenous for knowledge. Data mining therefore appears as a useful tool to address the need for sifting useful information such as hidden patterns from databases. From established patter of traffic rules offenders, there is need to formulate policies based on the vehicle weighing data to determine vehicles which are prone to overloading and committing the other traffic offences by marking vehicle destinations, origins and other attributes which can be used for policy formulation.

Roads in Kenya are classified into classes, namely A, B, C, D & E where class A are national roads (highways) linking Kenya to neighboring countries, class B are national roads linking major urban centers while class C are roads serving inter-district/sub-county linkages and D & E are rural roads within the county Governments (Kenya Roads Act, 2007). Out of these roads, classes A & B are the most used by transporters of heavy trucks, hence are monitored for axle load control. To control overloading and violation of other traffic laws by trucks on the highways, the Government has put in place measures by way of legislations, use of motor vehicle static weighing scales and law enforcement. However, in the process of implementing the measures, a problem of long trucks' queues at the weighing points became a menace which resulted in to delay in transporting goods within and across Kenyan borders, this translated into higher transport costs to transporters, created corruption opportunity for officers manning the weighing points and continued damage to the road pavements due to overloading. The weighbridges comprise weigh-in-motion (WIM) weighing scales, multi-deck static weighing scales and mobile weighing scales. Weigh-in-motion weighing scales are those scales which are capable of capturing truck weights while in motion, between 20 km/h to 80 km/h. The vehicles do not have to stop to be weighed, but would pass over the scales at specified speeds and still get weighed. Weigh-In-Motion (WIM) vehicle scales offer a cost-effective means of measuring truck axle and gross weights without affecting the flow of traffic. They are routinely used in commercial weight enforcement to screen trucks entering a weigh station returning legally loaded vehicles back to the main roadway while directing over weight trucks to the multi-deck static scales.

Other countries across the world have employed penalties as deterrent to the offences on the roads. (Nordegen, undated) observed that in South Africa, as a result of a minimum level of enforcement in many areas, low levels of fines, limited success with prosecution in the courts and varying levels of corruption at weighbridges from country to country, many transport operators seem to have adopted a policy of deliberate overloading. He further observed that in order to address the problem of lack of driver discipline, a different approach to vehicle screening has been adopted at a number of weighbridges in the province of KwaZulu-Natal, which were being upgraded to Grade A standard. The WIM sensors were located on the highway in the left-hand (slow) lane rather than in a dedicated screening lane. These permanent sensors have a dual purpose of both screening heavy vehicles which are overloaded and capturing traffic information (including vehicle speed and length, axle loads and spacing) of all vehicles on a 24-hour basis. In Europe, different countries applied different laws to regulate road transport in their countries. This posed great challenge in the enforcement of applicable laws on transporters who traversed across more than one country. (Barrot, 2005), says that in the field of abnormal road transports there are big differences between the rules and procedures currently applied in the Member States.

This research sought to establish a pattern between violation of axle load regulations, destinations and origin of the trucks. Classification algorithm in Waikato Environment for Knowledge Analysis (WEKA) was used to analyze weighbridge data generated from the weigh-in-motion system to develop a predictive pattern that can be used to formulate policies in the management of the weighbridges. Classification as a technique has been used to solve similar problems in other areas such health, agriculture, education, transport, engineering among others. This particular application of classification technique delves in analyzing data generated from a weighbridge management information system called weigh-in-motion (WIM) axle control management system. The data obtained from WIM was on vehicles which were weighed and screened at weigh-in-motion at Webuye weighbridge. The data attributes comprised weights and other legal compliance requirements such as valid load permits, drunk driving, and valid driving license.

1.2 Statement of the Problem

One major problem faced by transporters in Kenya is the operational inefficiency at weighbridge stations. Weighbridges in Kenya operate using Management Information System (MIS) called ‘‘KenWei’’ system. However, use of KenWei system has not yielded desired results in control of overloading and violation of other traffic laws and regulations such as drunken driving, use of invalid permits, lack of fire extinguishers among other offences. The inefficiencies at the weighbridge stations have led to a lot of time spent by trucks at the weighbridges translating into high transport costs resulting from the delays. (Pinard, 2010) observed that heavy vehicle overloading is a serious problem across much of Sub-Saharan

Africa. Such overloading not only significantly accelerates the rate of deterioration of road pavements but, when coupled with inadequate funding for road maintenance, it contributes significantly to poor road conditions and high transport costs. The indicative cost of overloading in East and Southern Africa has been estimated at more than US$ 4 billion per annum. This exceeds the amounts being spent on road rehabilitation. Time spent at the weighbridges by transporters to undergo the truck weighing is inordinately long. This translates into high transport costs. Some of the reasons why it takes long at the weighbridges is due to manual inspection of documents such as load permit, driving licenses, fire extinguishers among others. From the data, it was demonstrated that not all vehicles are prone to committing the offenses. In this research, information gain concept was used to evaluate data in the MIS at weighbridges on Kenyan highways. A pattern was established that links origins and destinations of the trucks to their likelihood to committing the offenses. The mined data therefore established a pattern of those who were likely to commit the offenses based on their origin and destinations. (Supee and Gael, 2009) explained that trucking companies in East Africa illustrate the opportunity cost of delays (at borders, weighbridges, and port).

For many years, trucks have been waiting between one to two days at Malaba, the main border post between Kenya and Uganda. Further, they indicated that in addition to border delays, weighbridge operations (for instance at Mariakani, Kenya, or at Mombasa port) also contribute to long delays along the northern corridor. In their study of transport costs along the northern corridor they observed that if all these delays (port, weighbridges, border) could be significantly reduced, vehicle yearly mileage should improve by at least 20,000 kilometers, which would help increase the ratio of a vehicle’s capital utilization, thus reducing average yearly operating costs per vehicle and perhaps leading to transport price reductions. This research developed a model which can be used to formulate a policy and administrative procedures on vehicle inspection at weighbridges based on their origins and destinations when screening them for certain traffic offences.

1.3 Objectives of the Study

1.3.1 Main Objective

The main objective of this study was to establish a pattern and a predictive correlation between origin and destination of trucks weighed at weighbridges and the likelihood of not complying with axle load limits and other legal requirements such as valid load permits, drink driving, lack of fire extinguishers and validity of drivers' licenses with a view to formulating a policy to deal with the problem of delays occasioned by weighing and inspections at the weigh bridges.

1.3.2 Specific objectives were;

i). To identify the relevant data generated and stored at weighbridge databases.

ii). To analyze the data obtained from Weighbridges database using relevant classification algorithm.

iii). To build a model from the analyzed data.

iv). To test and validate the model.

1.4 Significance of the Study

Delays at Kenyan weighbridges is a problem that affects businesses in virtually all sectors of the economy within the country and beyond. The delays translate into high costs of transport and overall business costs to the transporters and consumers of goods transported along the Kenya transport corridors. In East and Southern Africa, transport costs are severely affected by the opportunity cost of delays (at border crossings, weighbridges, and ports) and long custom procedures (Supee & Gael, 2009). Solutions offered to solve the problem of truck weighing delays at weighbridges will save transport costs and the country expenditures incurred on road maintenance. This study was therefore to evaluate the data generated at the weighbridges for policy making.

1.5 Motivation of the Study

The increased use of road transport in Kenya in the recently requires that all stakeholders in the transport sector must work towards efficient delivery of services to the transporters of goods within Kenya and outside its borders. My work place presented an opportunity to interact with managers, systems, operations and activities performed at Kenyan weighbridges across the country. In addition, reports on inefficiencies at the weighbridges, especially on delays experienced in weighing and inspecting trucks at the weighbridges motivated this study. The available data from the weigh-in-motion system and the readily available free source data mining tools such as Waikato Environment for Knowledge Analysis (WEKA) made this study feasible.

1.6 Scope of the Study

This study involved evaluation of data generated from the WIM system used at weighbridges in Kenya for weighing and monitoring adherence to axle load control and other legal requirements on the roads. It studied practices in other countries especially those that have fully automated their weighbridge operations and data. Data generated in Kenya from one weighbridge station located at Webuye weighbridge station on road A 104 that leads to the Kenya-Uganda boarder was used in this study. This study might be limited by the following;

i). I did not obtain comparable data from countries that have implemented WIM systems in the management of their weighbridges owing to logistical implications. Data generated from systems in those countries was therefore not obtained for analysis.

ii). Data from other weighbridges within Kenya was not obtained because officials in charge of the data were reluctant to provide such information. This might have been because of the fear of exposing malpractices in the management of Kenyan weighbridges, even though that was not the motivation of this study.

CHAPTER TWO: LITERATURE REVIEW

2.1 Introduction

This chapter provides the theoretical background and the underpinning theories used in data mining. Reference has also been made to the studies on axle load control management in other countries in Africa and across the world. Use of Management Information Systems (MIS) in axle load control management and operations of weighbridges has been widely studied and a plethora of publications exist on the same. This has mostly been done in the developed countries as opposed to developing countries where weighbridges are largely manual or semi-automated in their operations. Kenya uses a mix of automated and manual weighbridge operations in all its nine (9) weighbridge stations (KeNHA Weighbridge Manual 2011). Despite the partial automation of the weighbridges operations, the results have not met expectations with problems of corruption, long traffic queues and delays still persist. KenWei application was introduced for data capture, storage and reporting in 2011. Data mining tools and techniques available for mining such data to create patterns or knowledge for policy formulation are discussed and the theories underlying the study.

2.2 Theoretical Review

There are a number of established theories on decision making and data mining. They include, decision theory, information theory and decision tree theory.

2.2.1 Decision Theory

Decision theory is theory about decisions (Harrison, 1994). Decisions made by individuals, businesses and even governments depend something they do not know. For instance, one may want to make a decision whether to make an investment in a business stream or portfolio. The decision will depend on some information available before such a decision is arrived at. Harrison says that in the situations treated by decision theorists, there are options to choose between, and we choose in a non-random way. Our choices, in these situations, are goal-directed activities. Hence, decision theory is concerned with goal-directed behavior in the presence of options.

2.2.2 Normative and Descriptive decision theories

Normative decision theory is about how decisions should be made and descriptive theory is a theory about how decisions are actually made. This study therefore falls within the description decision theory.

2.2.3 The decision-making process

Decision making starts from identification of the problem, followed by obtaining the necessary and relevant information. The next stage is production of possible solutions, which are then evaluated to determine their efficacy. The last stage is then on selection of the strategy for the performance and implementation of the selected strategy.

2.3 Data Mining Techniques and Algorithms

Data mining has become popular in recent the times for performing analysis and pattern discovery. According to Ozer Patrick, Data Mining is a process that consists of applying data analysis and discovery algorithms that, under acceptable computational efficiency limitations, produce a particular enumeration of patterns (or models) over the data (Ozer, 2008). Ian written and Eibe Frank observed that seeking of pattern is not new, people have been seeking patters from when human life began. The only development in seeking patterns are the tools available for the exercise. They argue that hunters seek patterns in animal migration behavior, farmers seek patterns in crop growth, politicians seek patterns in voter opinion, and lovers seek patterns in their partners’ responses. The only difference is that in data mining data is stored electronically and the search is automated. Data mining is therefore about solving problems.

(Megha Gupta and Naveen Aggarwal 2010) opined that ‘‘Data mining is the discovery of knowledge and useful information from the large amounts of data stored in databases. It is referred to as knowledge discovery from databases (KDD), is the automated or convenient extraction of patterns representing knowledge implicitly stored in large databases. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. Data mining tools can answer business questions that traditionally were too time consuming to resolve. Patterns generated from data mining can be used to make predictions of certain behaviors in variables in a data set. This is useful when predictions on behavior of another set of data is to be made. The data provide a fertile base for performing analyses for decision making by business leaders and decision makers within organizations. Data mining tasks are generally categorized as clustering, association, classification and prediction (Chien & Chen, 2008).

2.3.1 Clustering

Clustering is the grouping of instances with the same attributes together. It works well when there ae not clear classes and instead the instances are divided into natural groups. According to (Written, 2005). The classic clustering technique is called k-means. First, you specify in advance how many clusters are being sought: this is the parameter k. Then ‘‘k’’ the number of clusters is first randomly identified in advance and then all the instances are assigned to their closest cluster centre according to the ordinary Euclidean distance metric. The centres, or mean, of the instances in each cluster is computed. The centres or centroids are taken to be new centre values for their respective clusters. This process is repeated several times until the same points are assigned to each cluster in consecutive rounds, the cluster centres then stabilize and remain the same for ever.

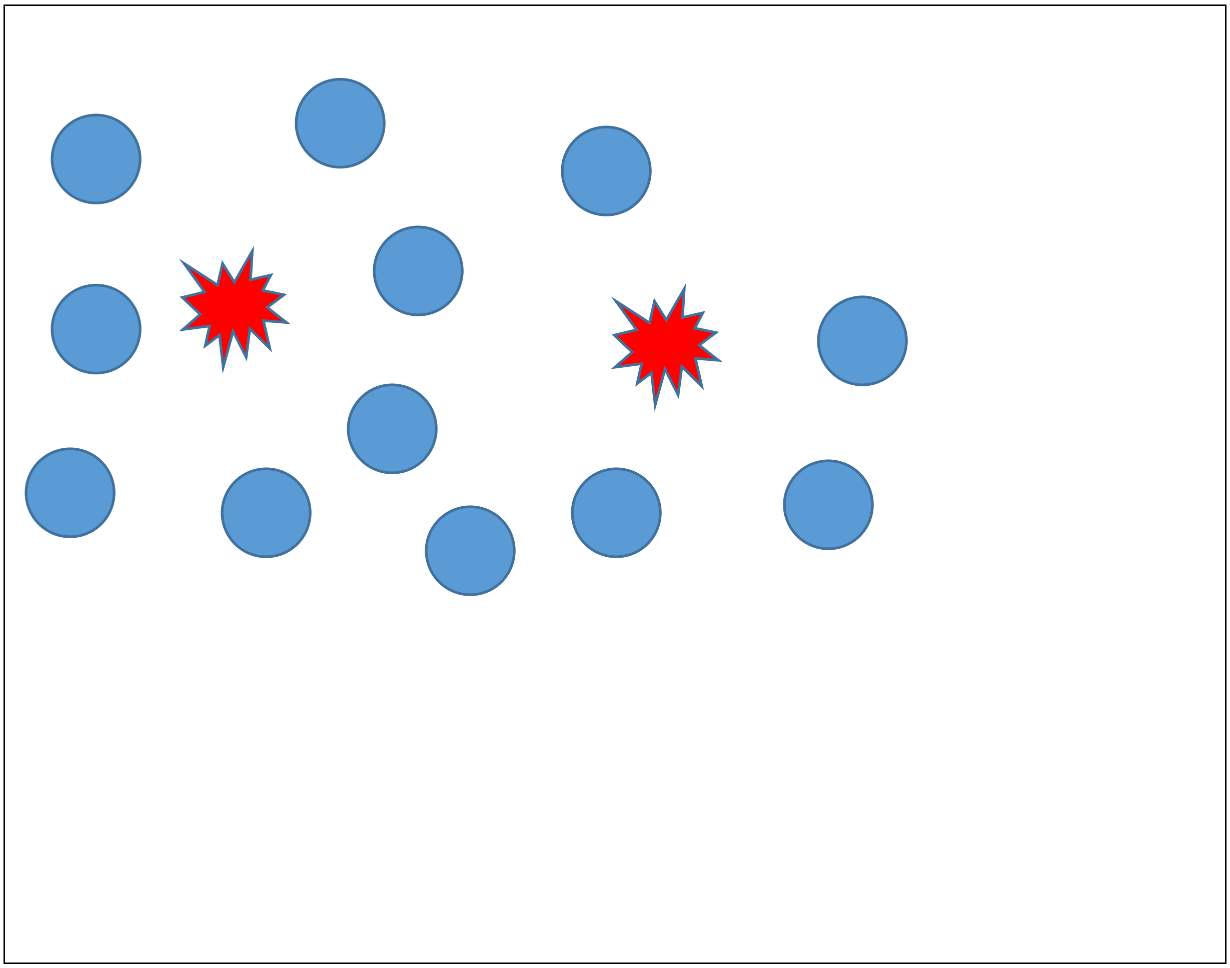
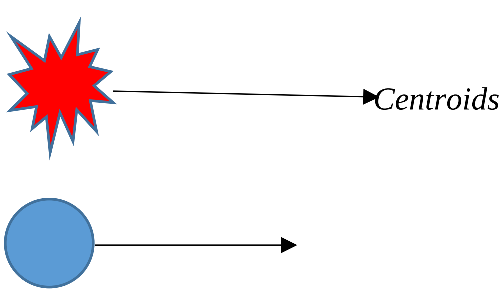


Figure 1: Clustering indicating the Euclidean distance

Legend:



Instances

2.3.2 Association

Associations rules do not represent any of causality or correlation between two or more items. Association is a technique which may be used to find groups of items that tend to occur together in transactions, typically supermarket checkout data. For instance, at a supermarket outlet, it can be automatically discerned by mining sales data that customers who buy beer also buy diaper. Such discoveries would be extremely significant to the operators of the supermarket I determining purchase orders and items arrangement within the floor of the supermarket to aid customer accessibility to the items, special discount decisions and many more business decisions. For example, weather data in the table below generated association rule

Outlook	Temperature	Humidity	Windy	Paly
Sunny	85	85	False	No
Sunny	80	90	True	No
Overcast	83	86	False	Yes
Rainy	70	96	False	Yes
Rainy	68	80	False	Yes
Rainy	65	70	True	No
Overcast	64	65	True	Yes
Sunny	72	95	False	No
Sunny	69	70	False	Yes
Rainy	75	80	False	Yes
Sunny	75	70	True	Yes
Overcast	72	90	True	Yes
Overcast	81	75	False	Yes
Rainy	71	91	True	No

Table 1: Weather data used to generate rules

The resulting rules from the above data can be explained as below;

Rule 1: If temperature = cool then humidity = normal

Rule 2: If humidity = normal and windy = false then play = yes

Rule 3: If outlook = sunny and play = no then humidity = high

Rule 4: If windy = false and play = no then outlook = sunny and humidity = high.

The rules listed above were obtained from mining the data using WEKA machine learning software.

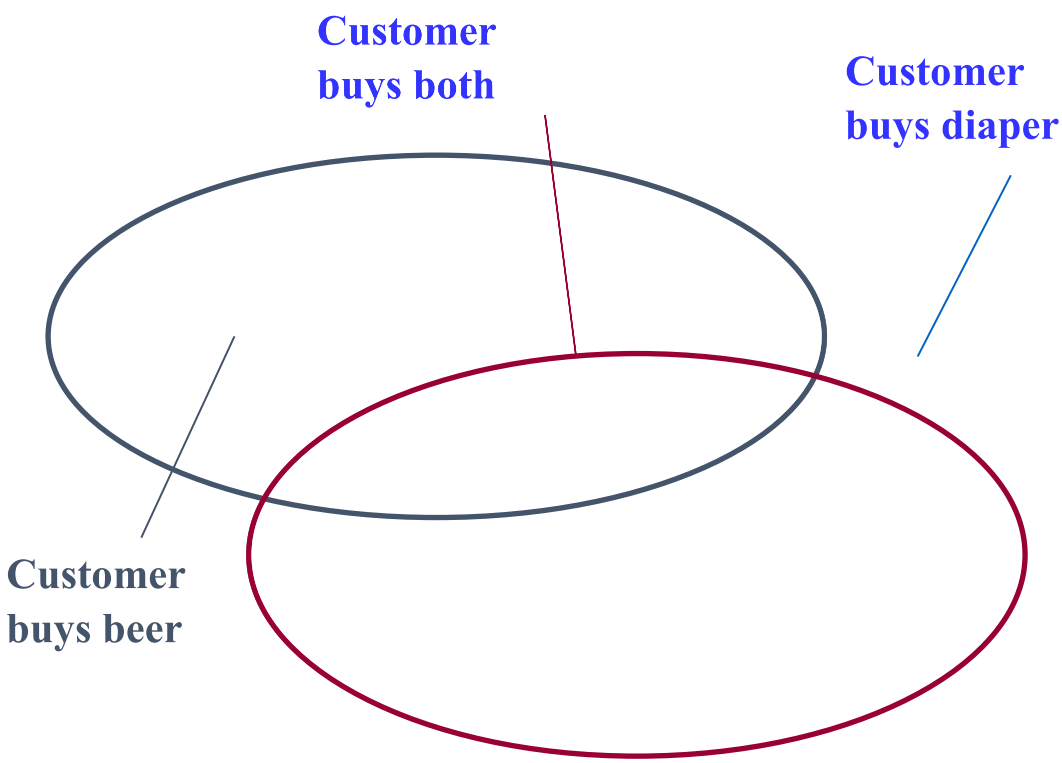


Figure 2: Association between beer and diaper

2.3.3 Classification

Classification techniques are used in data classification in various industries. It involves classifying data into various classes. (Gupta 2010) defined classification as a data mining or machine learning technique used to predict group membership for data instances. Each sample or tuple is assumed to belong to a predefined class, this is why classification is referred to as supervised learning technique. Classification process begins from developing a model using training data set. A data set comprise instances with defined attributes. A model built from the training data set is then tested using a different data set called test data. The test data essentially have similar attributes to the training data set except that the data is a separate set of data. For instance, when data extracted from a data base has one hundred instances, sixty (60) may be used to build the model and the remaining forty (40) used to test the model.

Classification is the act of looking for a model that describes a class label in such a way that such a model can be used to predict an unknown class label. Thus, classification is usually used to predict an unknown class labels. For instance, a classification model can be used to classify bank loans as either safe or unsafe. Classification applies some methods like decision tree, Bayesian method and rule induction in building its models. Classification process involves two steps. The first step is the learning stage which involves building the models while the second stage involves using the model to predict the class labels. (Obuandike G, 2015).

Classification technique uses various algorithms to perform data mining. These algorithms are available in WEKA, they include; Rule based classifier, Decision tree induction, Nearest neighbour classifier, Bayesian classifier, Artificial neural network, Support vector machine and Regression trees.

Rule based classifier

Rule based classifiers are mostly viewed as general models as opposed to decision trees. Each path in a decision tree may be used to generate a rule. Rules created from decision trees in some instances may contradict each other. Rule based classifier was therefore not be suitable for my research.

Nearest Neighbour Classifier.

Gupta described nearest neighbor classifiers as one that uses the k-nearest neighbor's algorithm (K-NN). K-NN is a method for classifying objects based on closest training examples in the feature space. K-NN is a type of instance-based learning, or lazy learning. This method is referred to as as lazy learning, because it waits for knowledge of the test instance in order to create a locally optimized model specific to the test instance.

Support Vector Machines

Support Vector Machines (SVM) methods use linear conditions in order to separate out the classes from one another idea is to use a linear condition that separates the two classes from each other as well as possible. (Charu et al, 2015).

Decision tree induction

According to Gupta, decision tree is a flowchart like tree structures, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node holds a class label. Decision tree induction has advantages over the other classifications schemes. The advantages include;

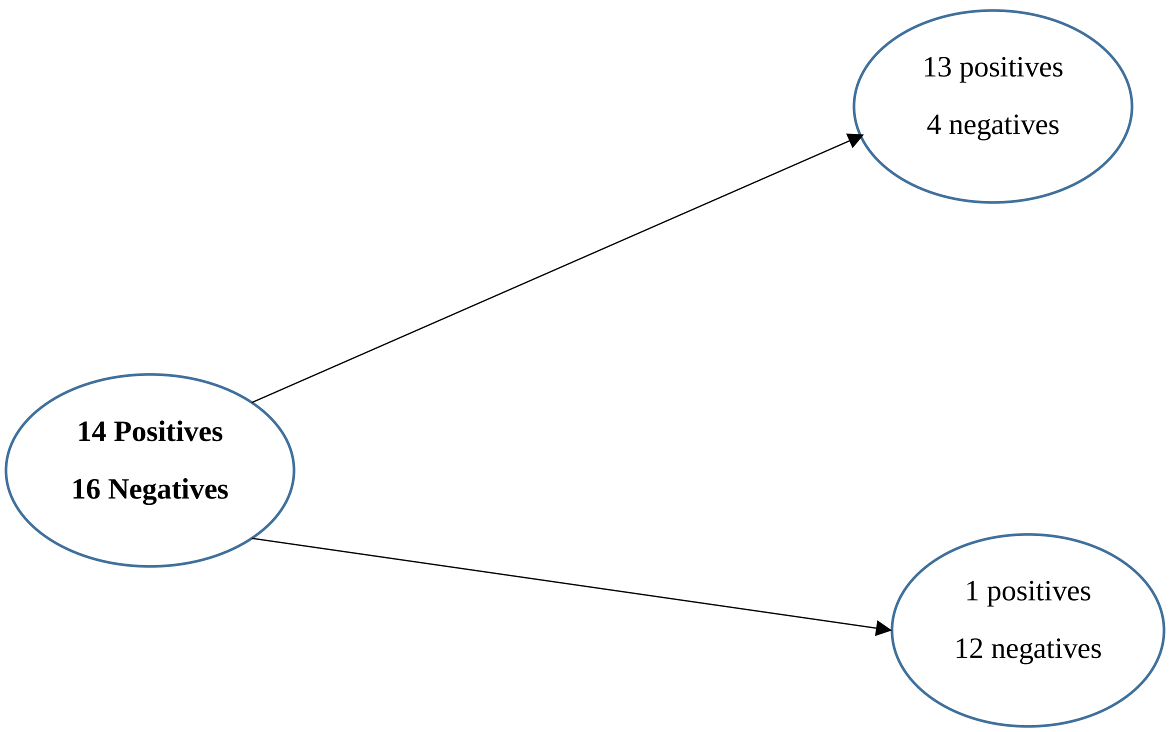
- Decision trees are simple to understand and interpret. It therefore suitable for research by all categories of students and users of classification results.
- Decision tree require little data and are able to handle both numeric and categorical data.
- It is able to perform with large data in a short time, hence suitable for quick fixes on problems that require immediate solutions.

Despite the advantages mentioned of decision trees, one major setback of decision tree is that it can create complex trees that do not generalize the data well (overfitting). To avoid the problem of overfitting, the trees can be pruned to mitigate the overfitting problem. In this research, I opted for decision tree algorithm called J 48 to mine data and build a classification model.

Information Gain and Entropy

Entropy is the measure of impurity in data set. It provides the information to measure the goodness of a split or the amount of information in an attribute. Information gain is the expected reduction in entropy caused by partitioning the data according to some selected attribute. For instance, in a data set with 30 instances, 14 positive and 16 negative. Entropy is calculated as follows;

Entropy = - 14/30 log 2 (12/30) + 16/30 log 2 (16/30); This is equal to 0.996.



Computation of Entropy and Information Gain

Parent Entropy = - {14/30. Log 14/30} – {16/30.Log2 16/30} = 0.996

Child 1 Entropy = - {13/17.Log2 13/17}- {4/17. Log 4/17} = 0.787

Child 2 Entropy = -{1/13.Log2 1/13} – {12/13.Log2 12/13} = 0.391

Weighted Average Entropy of the children = {17/30 * 0.787} + {13/30*0.391} = 0.615

Information Gain = 0.996 – 0.615 = 0.381

2.4 Current Trends in use in MIS on Axle Load Control

Many countries across the world have embraced the use of MIS in their weighbridges. Some of the most efficient functioning MIS have been deployed in the developed countries. The systems have also demonstrated high performance and efficiency. In instances where MIS have been deployed, the weighbridges have weigh-in-motion (WIM) and static scales technologies. Weigh in Motion come in two forms, the High-Speed Weigh in Motion (HSWIM) and Low Speed Weigh in Motion (LSWIM). HSWIM is where the weighed vehicles pass over the scales at a speed of 50 km/h or more while LSWIM is where the weigh vehicles move at as speed of 15 km/h or less. Weigh-in-Motion technology has been used for many years in commercial vehicle operations and is installed in hundreds of locations across the United States of America (U.S.). Klebe & Susor, (undated) observe that WIM system can be configured to meet specific user requirements. The WIM controller and user interface can be located in a building near the scale, or fiber optic connection can be used for scale to user interface distances greater than 2,500 Kilometers. WIM systems, that screen/weigh trucks at high speeds, reduce delays that would otherwise be incurred by weighing trucks on static scales. (Rakha et al, 2006).

2.4.1 MIS in Axle-Load Control in Sweden

The Swedish Government has implemented WIM based on the bending plate technology. The bending plates are installed on the road surface and measures the weights of passing vehicles. The information is stored in the road side unit and can be accessed by the built-in web-interface or from a central system. The Swedish system automatically collects information on overweight vehicles and processes fine to the offenders. In Africa, the use of weighbridges to control loads has been ineffective, several road sections with weighbridges are in no better condition than sections where there are fewer or no weighbridges. There is sufficient evidence to assume that overloading is the main cause of road deterioration and, therefore, that the weighbridges are not being utilized effectively to control overloading. (Supee & Gael, 2009).

2.4.2 MIS in Axle-Load Control in India

In India, ‘‘the government installed a video camera system to register all trucks coming to the border and to check the permissible weight for each truck using the data from a central database. An electronic weighbridge weighs the vehicle, and a computer automatically issues a fine. Drivers can use a stored value card for payments, obviating the need for them to carry large sums of cash, thereby reducing corruption at the border. The system increased tax collection from $12 million to $35 million over two years and reduced the average time required to clear a vehicle from 30 minutes to 2 minutes’’ (Drüke, 2007). Use of WIM in India therefore demonstrates how use of MIS can be applied to achieve efficiency and economy (cost savings). The Indian situation mirrors what happens in Kenya where weighbridges have been associated with corruption and inefficiency. ‘‘The government installed a video camera system to register all trucks coming to the border and to check the permissible weight for each truck using the data from a central database. An electronic weighbridge weighs the vehicle, and a computer automatically issues a fine. Drivers can use a stored value card for payments, obviating the need for them to carry large sums of cash, thereby reducing corruption at the border’’ (Shah, 2007).

2.4.3 MIS in Axle-Load Control in Finland

In Finland, the National Road Administration Agency installed WIM equipment to collect information on the extent of overloading and to publicize the names of persistent offenders. The Finland Road Agency therefore adopted use on MIS in the capture, storage, dissemination and retrieval of axle load information on vehicles that overloaded on the roads. ‘‘In the case of overloading, a picture of the vehicle is captured by a camera which is activated by the main part of system using sensors' data and analyzing axle loads based on authorized weights. The gathered data are normally transferred through a fibre optic device to the closest stop station and the vehicle is checked with high accuracy devices’’ (Abbas Mahmoudabadi, 2013).

2.4.4 MIS in Axle-Load Control in South Africa

South Africa like many other countries face the challenge overweight vehicles on the road.

‘‘The low level of overloading enforcement in certain provinces is largely due to a lack of manpower and weighing facilities and to some extent due to the perception that heavy vehicle overloading is not a serious traffic offence’’ (Nordengen, 1997). Weighbridges in South Africa consist of four independent decks on which approximately 95 percent of vehicle combinations can be weighed in a single operation (about 3 percent of vehicles weighed have 5 axle groups and about another 2 percent have uncommon axle/axle group spacing). If the vehicle is found to be legal or within the 5 percent tolerance limit, the driver can proceed back onto the highway. If the vehicle or one or more of the axles/axle units is more than 5 percent overloaded, it is directed by means of traffic signals to the holding area where prosecution is instituted. The Traffic Control Centre has a computer system which monitors the weighing/prosecution process from when the vehicle is intercepted until it leaves the center. The system also includes a process for the issuing of prosecution notices, which apart from speeding up the issuing of these documents, ensures uniformity and minimizes transcription errors. (Nordengen, 1997).

In the South African weighbridges, if vehicle is found to be legal or within the 5 percent tolerance limit, the driver is allowed to proceed back onto the highway. If the vehicle or one or more of the axles/axle units is more than 5 percent overloaded, it is directed by means of traffic signals to the holding area where prosecution is instituted. The Traffic Control Centre has a computer system which monitors the weighing/prosecution process from when the vehicle is intercepted until it leaves the weighbridge. The system also includes a process for the issuing of prosecution notices, which apart from speeding up the issuing of these documents, ensures uniformity and minimizes transcription errors. (Nordengen, 1997).

2.4.5 MIS in Axle Load Control in Namibia

One of the problems facing overload control in Namibia is the occurrence of malpractices at weighbridge facilities due to human interventions. The Roads Authority implemented overload control computer- based information called ‘‘TrafMan’’ system to network the operations of all weighbridges. The system has the capability of transmitting live weighing data to a central system at the head office for ease of access by the data manager and the Roads Management System. In addition, a fully integrated management information system at each weighbridge record processes and produces reports. (Mike Ian Pinard, 2011).

2.4.6 MIS in Axle-Load Control in Kenya

In Kenya, there are nine (9) fixed axle load enforcement weighbridges spread across the country and six (6) mobile roaming the network. (Kenyan highways serve as gateway to landlocked economies of Uganda, Rwanda, Burundi and DR Congo with a combined transshipment of over 2.2 million tonnes of load every year with an annual growing rate of twenty (20) percent. (SGS Kenya Ltd Report, 2016). The current system uses HSWIM scale alongside static scales. In some stations the static scale is multi deck while in others it single deck. Kenya use ‘‘MIS called KenWei’’ system. The systems integrate several components of the weighbridge station operations. The components include Traffic Control (TC), cameras capturing motor vehicles approaching the station at HSWIM, multi-deck, holding yard, exit from the station and those that abscond weighing by avoiding the HSWIM and static scale. Weighing application, is the application used on the HSWIM and the multi deck scale to capture, record and store the vehicle weights. They also compare the weights to the axle load legal limits as per the Traffic Act Cap 403. Vehicles found to overweight are dealt with according to the provisions of the law. In spite of the use of KenWei, overloading has persisted. This will be examined when conducting the study where I will evaluate performance of this system for purposes of improving it.

2.4.7 The WIM system used in Kenya

Application of the existing weigh in motion systems in other countries are largely similar to what is installed on the Kenyan roads. However, the administrative and policy aspect of operating the weighing systems are not the same. Main reasons being that axle load control is administered under various legal provisions of the country where they operate. Legal frameworks on axle load control differ from country to country. Further, the WIM systems take cognizance of the topographical orientation of the roads, countries do not have uniform such orientation, hence each country must design the WIM by taking cognizance of the topology of their road terrains. In addition, various technologies have been used to design the road pavements, WIM deployment considers the road pavement designs. Countries have employed varied road pavement designs depending on available technologies, levels of financing and the traffic volumes expected in the various roads. For instance, in the European Union Countries, laws governing obtaining of permits for abnormal (overload) motor vehicles take different duration under their local laws.

Figure 3: Pavement mounted Weigh-in-Motion weighbridge

Source: Jacques Barrot, 2014

2.4.8 Types of Motor Vehicle Weighing

There are two major types of weighing scales. These are static and weigh-in-motion weighing. The static weighing is the traditional weighing of trucks on the roads. Trucks step onto the weighing scale by passing each axle onto the scale, weights for each axle is obtained after the gross vehicle weight (GVW) is obtained from the axle weights. There have however been some developments in the static weighing where the whole truck steps on the scale and the motor vehicle weight on all the axles is captured as the same time.

The other weighing method is the weigh-in-motion. This is the weighing of vehicles while they are in motion. This is further divided into low speed weigh-in-motion and high speed weigh-in-motion. The low speed WIM is where the vehicle passes over the scale at a speed of not more than five (5) km per hour. The high-speed WIM is where the motor vehicle passes over the weighing scale at between ten (10) to fifty (50) km per hour.

Weighbridges can be mounted or installed in different ways. These include the following;

2.5 Justification and study gap

Data mining tools and techniques have been used in various fields such as health, education, agriculture, engineering transport among others. However, no such study has been done on management of weighbridges in Kenya. There is however some related study in the transport industry but they addressed other aspects such as transport scheduling in cities, traffic control systems, intelligent transports systems and traffic accidents as was used to establish a pattern of accidents in Addis Ababa Ethiopia. (Tiseme T.B 2011). No study has been conducted in Kenya to link violation of traffic rules and axle load control regulations on the Kenyan roads using data mining techniques. Hamidah Jantan1, Abdul Razak Hamdan2 and Zulaiha Ali used data mining techniques to study human talent forecasting in Malaysia. The study concluded that generated classification rules can be used to predict the potential talent for the specific task in an organization and that in Human Resource Management (HRM), there are several tasks that can be solved using this approach, for example, selecting new employees, matching people to jobs, planning career paths, planning training needs for new and senior employees, predicting employee performance, and predicting future employee.

2.6 Conceptual Framework

The conceptual framework demonstrates interaction among various system's parts. It starts from a traffic system which notifies on coming trucks to slow down to a desirable speed of between 20 km/h to 80 km/h. vehicles then proceed to the weighing scale, compliant vehicles are allowed to proceed while those that don't comply to the weight limits are directed to a static weigh scales. A traffic light will be used to signal the vehicles whether they have complied or not. Green light will indicate that the vehicle has complied while red light will indicate that the vehicle should divert to static weighing scale for further confirmatory weighing. Weighing data is captured and stored in a centralized storage system from where reports are generated for decision making purposes.

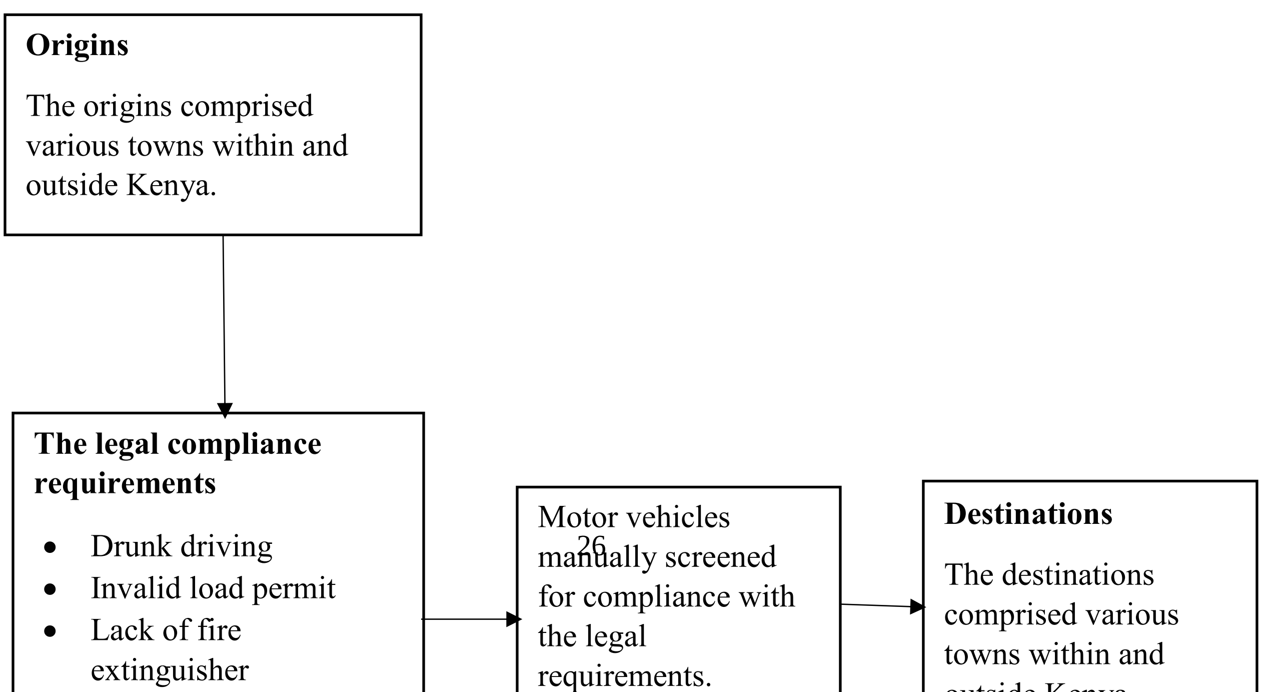


Figure 4: Conceptual Framework

CHAPTER THREE: METHODOLOGY

3.1 Introduction

This chapter is about the method used to perform that study. Research methodology refers to the techniques, tools and procedures used to collect and analyze data. It refers to frameworks and assumptions used to inform the research. There are several methodologies used in research, for instance; quantitative, qualitative, exploratory, experimental, survey, action research and case studies. It is the onus of the researcher to decide on the most appropriate methodology to employ in his/her research. (William, 2011) describes Research Methods as the tools and techniques for doing research. Research is a term used liberally for any kind of investigation that is intended to uncover interesting or new facts.

3.2 Research Design

Research design is the structure within which the research is conducted. It involves identifying source and method of obtaining the relevant data for analysis and production of the final research results. Data used in this research was secondary data obtained from data base of axle load control management information system at the Webuye weighbridge station. The data obtained from the weighbridge was then analyzed using relevant data classification technique. Data in the weighbridge database comprise thousands of data. From the available data, I randomly sampled a data set comprising two thousand five hundred and fifty (2,550) items. From the sample, I divided the data set into training and test data sets. The training set comprised 1,581 items and the test data set had 969. Data classification process proceeds in two phases. The first phase is on model construction and then use of the model in classification. This has been depicted in the diagram below.

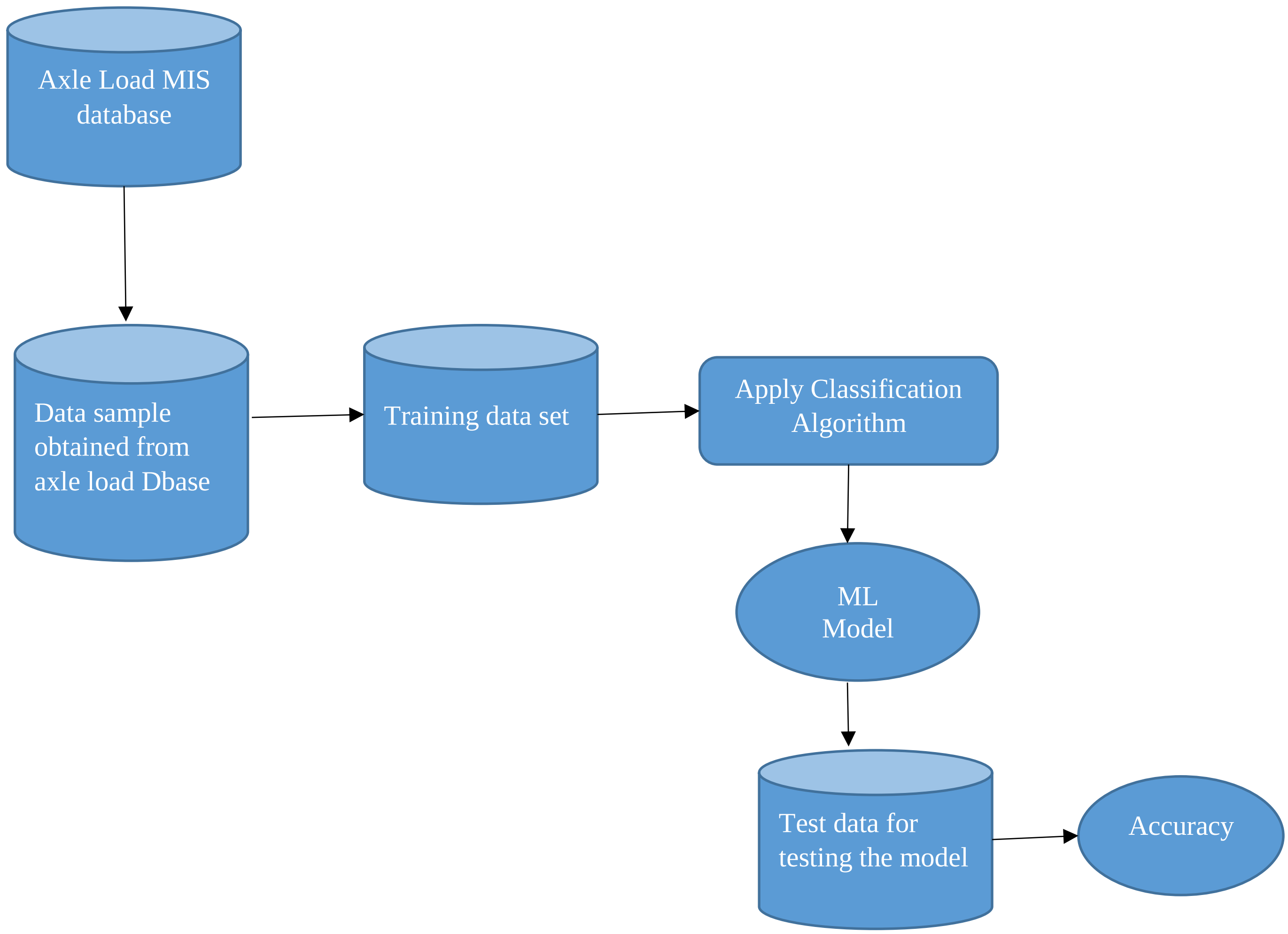


Figure 5: Research Design

3.3 Evaluation of the methods applied

There are a number of decision tree algorithms in WEKA which can be applied to build and test a classification model. The broad categories are bayes, functions, lazy, meta, rules and trees. For purposes of building this model, I evaluated four classifier algorithms to determine one that gives the best performance, the algorithms are J.48, Random Forest, IBK- (K-nearest neighbor) and NaiveBayes. The WEKA experimenter function was used in the evaluation.

J.48 Algorithm

J.48 implements a later and slightly improved version of C 4.5. Algorithm C4.5 was developed by Ross Quinlan in 1986 at the University of Sydney, Australia. It started from a version called ID 3. the algorithm has capabilities of handling missing values in a data set, visualization, numeric and nominal data. This makes it a robust tool for classification of data obtained from the weighbridge database. J 48 algorithm uses entropy and information gain.

$$Entropy(S) = \sum_{i=1}^{c} - p_i \log_2(p_i)$$

Where S = Sample of training examples

i = number of instances start from one

c = total number of instances

p = proportion of data in a given class

Gain (S, A) = Entropy (S) - ∑Sv/S Entropy (Sv), Where Sv is the sub set of S

Random Forest

(Written 2011) says that a popular algorithm for learning random forests builds a randomized decision tree in each iteration of the bagging algorithm, and often produces excellent predictors. For this reason, I selected it as one of the algorithms to be evaluated for use in my analysis. It has the capability of to handle missing values, nominal and numerical attributes.

IBK (K-Nearest Neighbor)

This algorithm is capable of handling missing values, numeric and nominal values.

NaiveBayes

It is capable of handling data with missing class values, numerical and nominal attributes. Naive Bayes uses the normal distribution to model numeric attributes. This is a classification method that is based on Bayes' theorem which is used to predict class labels. This classifier is based on probability theorem and is named after Thomas Bayes who is the founder of the theorem (Obuandike G 2015).

From the summary of performance evaluation results by the algorithms, J 48 emerged the best algorithm with the highest accuracy, precision, recall and F-Measure. I therefore used J 48 to construct the tree whose detailed explained is contained in the next chapter.

Summary of algorithms evaluation results

Classifier	True Positive Rate	Precision	Recall	F-Measure
J 48	82.7%	81.1%	82.7%	79.9%
Random Forest	59.4%	54.6%	59.4%	54.4&
Naive Bayes	55.6%	37.0%	55.7%	44.1%
K-Nearest Neighbor (IBK)	59.4%	54.5%	59.4	54.6%

Table 2: Metrics of algorithm classifiers.

True Positive (TP) Rate

This is the statistics that shows the correctly classified instances. In the case summarized above, J 48 correctly classified 82.7 % of all the instances in the data. The other classifiers had their percentages f correctly classified as indicated in the table above.

Precision

Measures the exactness of the relevant data retrieved. High precision means the model returns more relevant data than irrelevant data. On the data used in this research J48 returned 81.1%.

Recall

Measures the percentage of all relevant data that was returned by the classifier. A high recall means the model returns most of the relevant data.

F-Measure

This is a combined measure of precision and recall. It is computed as *F-Measure= 2*Precision*Recall/(Precision+Recall).*

The figures below indicate results which have been summarized in table 2. Figure 6 shows the metrics derived using naïve bayes algorithm and those for K-nearest neighbor, random forest and J,48 are shown on figures 7, 8 & 9.

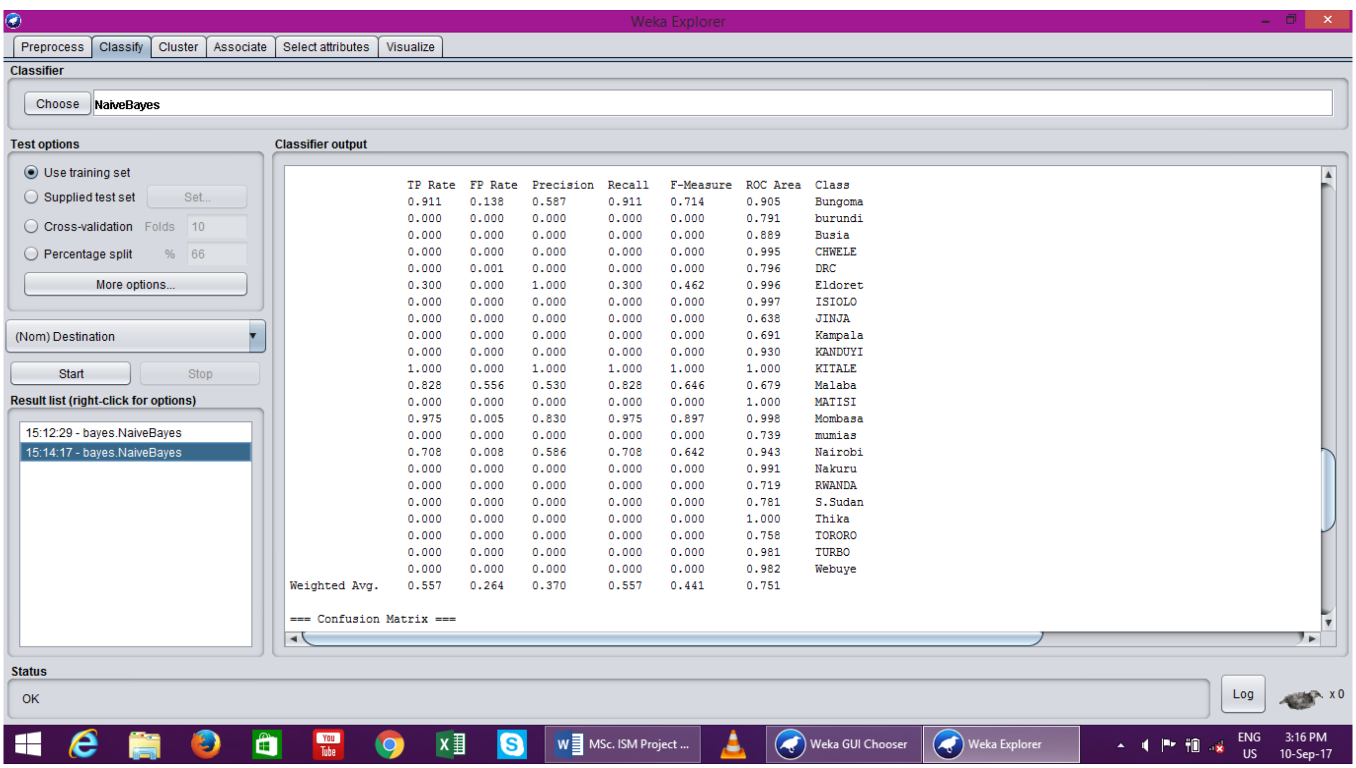


Figure 6: Naive Bayes Evaluation Results in WEKA

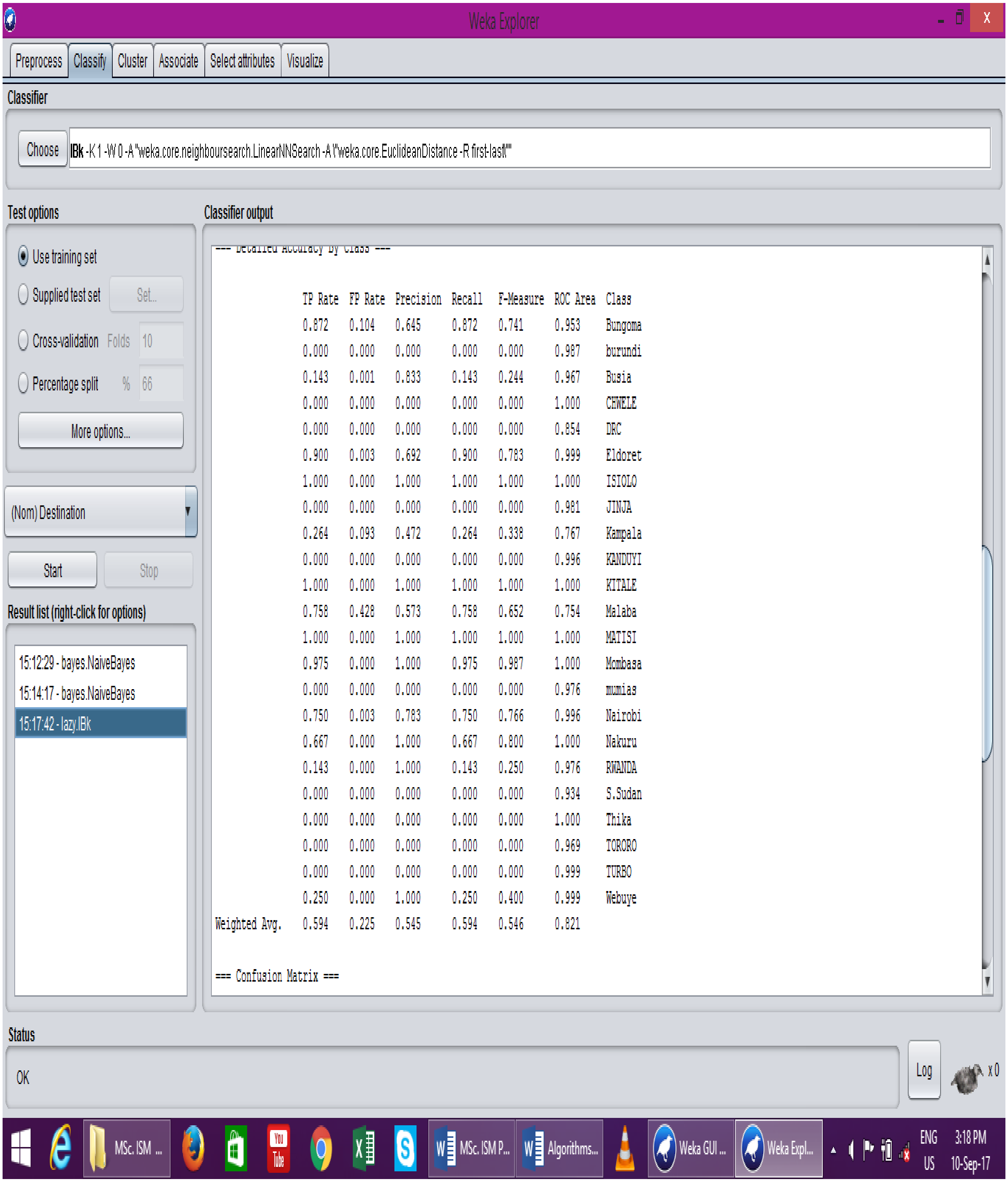


Figure 7: Nearest Neighbor Evaluation Results in WEKA

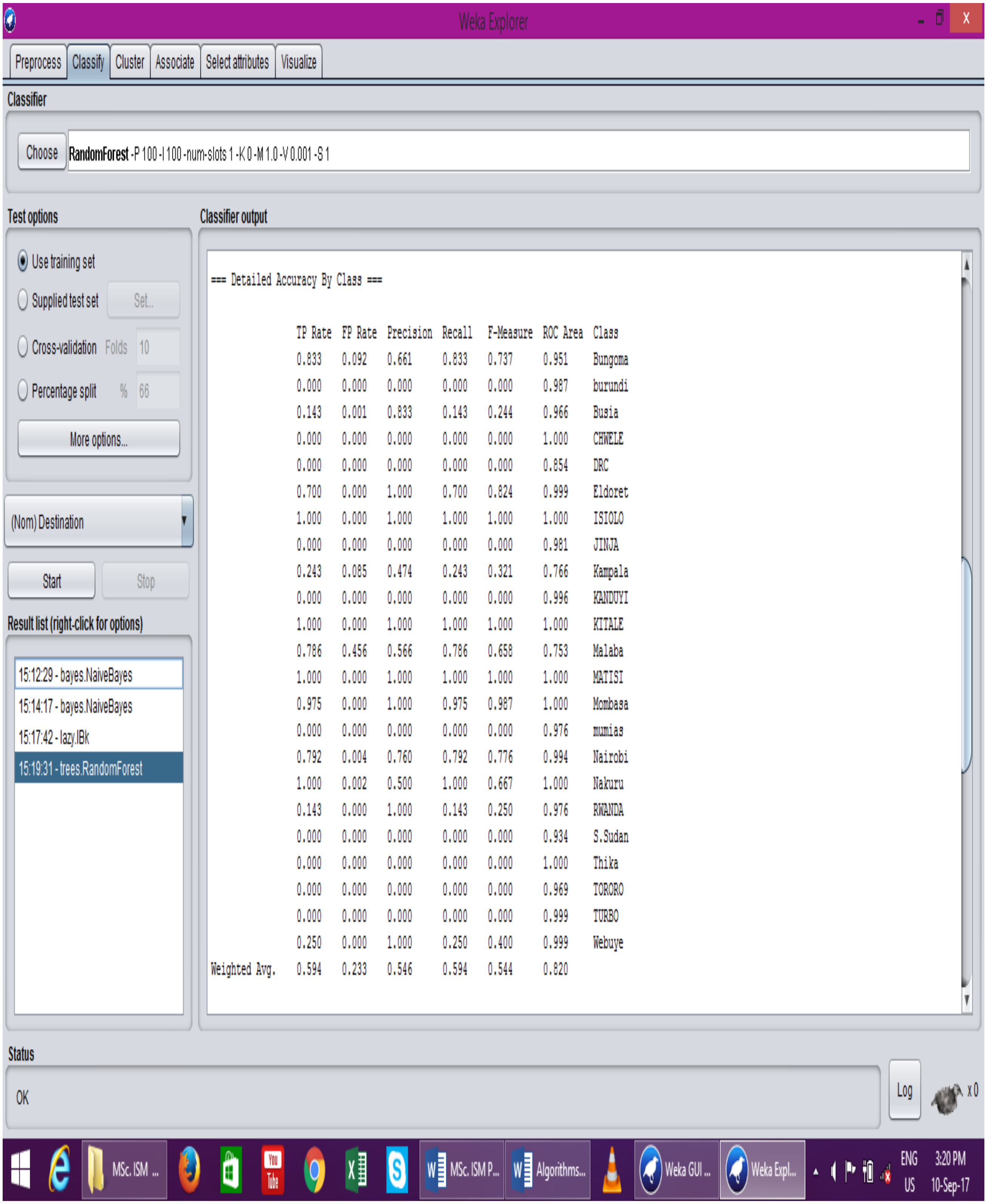


Figure 8: Random Forest Evaluation Results in WEKA

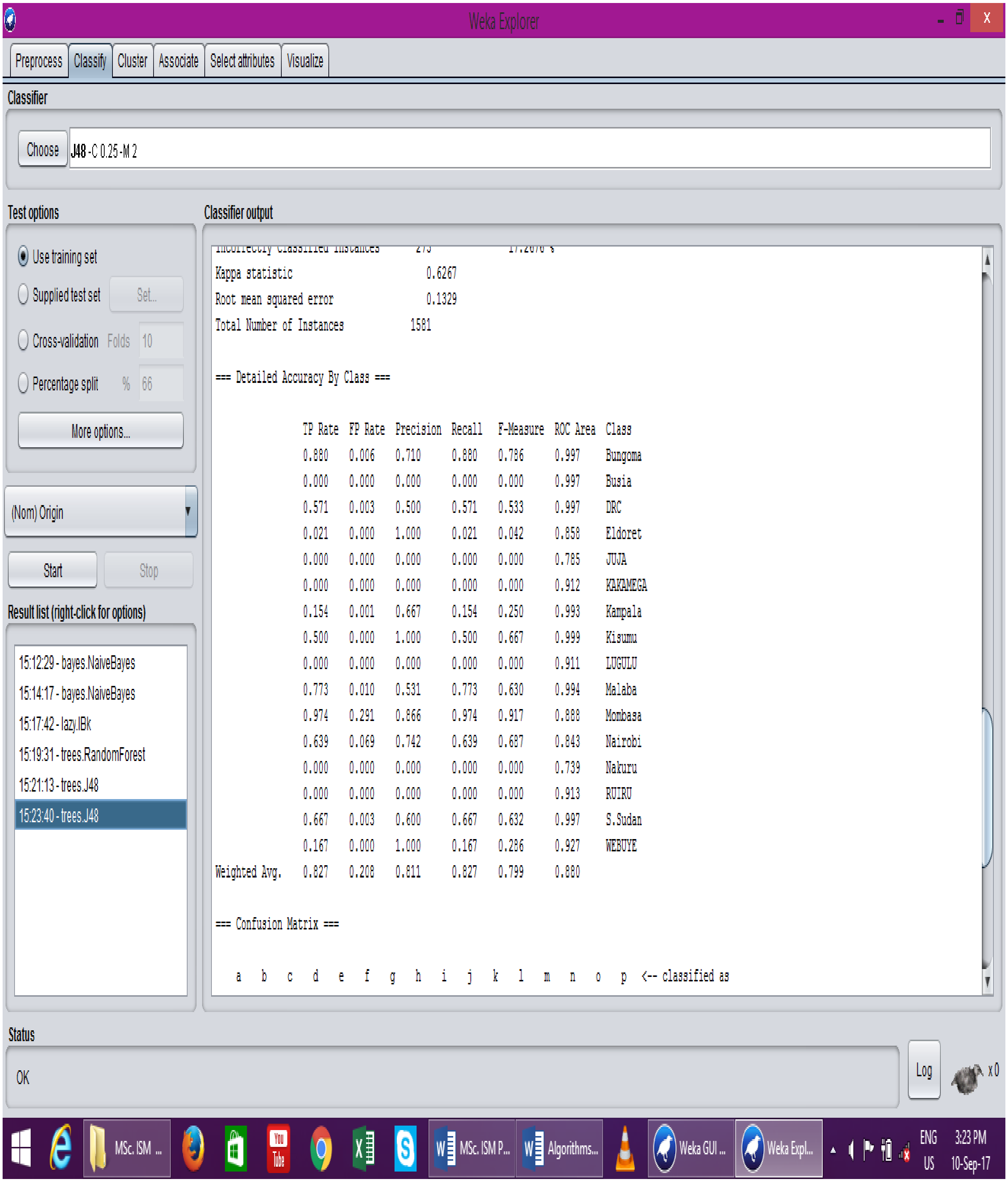


Figure 9: J 48 Evaluation Results in WEKA

CHAPTER FOUR: DATA ANALYSIS, FINDINGS AND DISCUSSION

4.1 Introduction

This chapter delves on the analysis of data obtained from the weighbridge database, examine the results and provide explanation to the model build from the data. It also covers testing of the model to determine well it predicts the classes in the analyzed data. The classes in the data are destinations of the trucks whose records were captured at the weighbridge at stored in the weighbridge MIS.

4.2 Descriptive statistics

Data from the weighbridge comprised a total of 2,550 items. Out of which 1,581 was used to build the model. The data was raw in an MS excel format. The data was therefore cleaned and pre-processed before analysis to conform to formats supported by WEKA, it supports comma separated version (CSV), attribute relational file format (ARFF).

Origin	Destination	Number of trucks
Bungoma	Eldoret	2
	Nairobi	5
	Eldoret	1
	Nairobi	1
	Webuye	4
	Nairobi	1
	Nairobi	2
	Eldoret	1
	Webuye	6
	TURBO	1
	Webuye	1
	Eldoret	7
	Nairobi	1
	Nakuru	9
Eldoret	Bungoma	31
	Chwele	1
	Kampala	3
	Malaba	5
Mombasa	Bungoma	22
	burundi	3

Origin	**Destination**	**Number of trucks**
	Busia	1
	DRC	41
	JINJA	1
	Kampala	395
	KANDUYI	1
	Malaba	497
	Nairobi	2
	RWANDA	13
	S. Sudan	16
	TORORO	1
Nairobi	Bungoma	212
	Busia	57
	Kampala	16
	Malaba	82
	Mumias	4
Nakuru	Bungoma	8
	Busia	1
	Malaba	33
Malaba	Eldoret	2
	Mombasa	20
	Nairobi	22
	Nakuru	2
	Thika	1
Kampala	Nairobi	21
	Mombasa	4
DRC	Mombasa	5
Busia	Eldoret	12
Webuye	Bungoma	4
Total		**1581**

Table 3: Origin-Destination distribution of trucks in the training data.

```
@RELATION WEIGHBRIDGE

@attribute InspectorName{Edwin,Eric,Reuben}
@attribute
Origin{Bungoma,Busia,DRC,Eldoret,JUJA,KAKAMEGA,Kampala,Kisumu,LUGULU,Malaba,Mombasa,Nairo
bi,Nakuru,RUIRU,S.Sudan,WEBUYE}
@attributeDestination{Bungoma,burundi,Busia,CHWELE,DRC,Eldoret,ISIOLO,JINJA,Kampala,KANDUYI,KI
TALE,Malaba,MATISI,Mombasa,mumias,Nairobi,Nakuru,RWANDA,S.Sudan,Thika,TORORO,TURBO,Webu
```

```
ye},
@attribute FirstAidKit{No,Yes}
@attribute FireExt{No,Yes}
@attribute DrunkenDriving{No,Yes}
@attribute DriversLicence{Invalid,None,Valid}
@attribute LoadPermit{Invalid,None,Valid}

@DATA
Reuben,Mombasa,Kampala,No,Yes,Yes,Valid,Invalid
Reuben,Mombasa,Kampala,Yes,Yes,No,None,Valid
Reuben,Mombasa,Kampala,No,Yes,No,Valid,Valid
Reuben,Mombasa,Kampala,Yes,Yes,Yes,Valid,Invalid
Reuben,Mombasa,Kampala,No,No,No,None,Valid
Reuben,Mombasa,Kampala,No,Yes,Yes,Invalid,Valid
Reuben,Mombasa,DRC,Yes,Yes,No,Invalid,Valid
Reuben,Mombasa,Kampala,No,Yes,Yes,Invalid,Valid
Reuben,Nairobi,Bungoma,Yes,No,Yes,Invalid,Invalid
Reuben,Nairobi,Malaba,No,Yes,No,Valid,Valid
Reuben,Mombasa,Kampala,Yes,No,Yes,None,Invalid
Reuben,Nairobi,Bungoma,Yes,Yes,No,Invalid,Valid
Reuben,Mombasa,Kampala,Yes,No,Yes,Invalid,Invalid
Reuben,Nairobi,Bungoma,No,Yes,No,None,Valid
Reuben,Nairobi,Malaba,Yes,Yes,Yes,Invalid,Invalid
Reuben,Mombasa,Kampala,No,No,No,Invalid,Valid
Reuben,Mombasa,Kampala,Yes,Yes,Yes,Invalid,Valid
Reuben,DRC,Mombasa,No,Yes,No,Valid,None
Reuben,Mombasa,Kampala,No,No,No,None,Valid
Reuben,Nairobi,Malaba,No,Yes,Yes,Invalid,Invalid
Reuben,Nairobi,Malaba,No,Yes,Yes,Invalid,Invalid
Reuben,Nairobi,Malaba,No,No,No,Valid,Invalid
Reuben,Mombasa,Kampala,No,Yes,Yes,None,Valid
Reuben,Mombasa,Kampala,No,Yes,No,Invalid,Valid
Reuben,Mombasa,Kampala,Yes,No,Yes,Valid,Invalid
Reuben,Mombasa,Kampala,No,Yes,No,Invalid,Valid
Reuben,Mombasa,Kampala,Yes,No,No,Valid,Invalid
Edwin,Mombasa,Malaba,Yes,Yes,Yes,Invalid,Valid
Edwin,Mombasa,Malaba,No,Yes,Yes,Invalid,Invalid
Edwin,Mombasa,Malaba,Yes,Yes,Yes,None,Valid
Edwin,Mombasa,Malaba,Yes,Yes,No,Invalid,Invalid
Edwin,Nairobi,Bungoma,Yes,Yes,Yes,Valid,Valid
Edwin,Nairobi,Bungoma,Yes,No,No,None,Invalid
Edwin,Mombasa,Malaba,No,Yes,Yes,Valid,Valid
Edwin,Mombasa,Malaba,Yes,Yes,No,Invalid,Invalid
Edwin,Mombasa,Malaba,No,Yes,Yes,None,Valid
Edwin,Malaba,Nakuru,No,No,No,Valid,Invalid
Edwin,Mombasa,Malaba,Yes,No,Yes,Valid,Valid
Edwin,Mombasa,Malaba,Yes,No,No,None,Invalid
Edwin,Nairobi,Malaba,Yes,Yes,Yes,Invalid,Valid
```

Edwin,Mombasa,Malaba,Yes,Yes,No,Invalid,Invalid
Edwin,Nairobi,Bungoma,Yes,No,No,Valid,Invalid
Edwin,Mombasa,Malaba,No,No,Yes,None,Valid
Edwin,Mombasa,Malaba,Yes,Yes,No,Valid,Invalid
Edwin,Mombasa,Malaba,Yes,Yes,Yes,Invalid,Valid
Edwin,Mombasa,Malaba,Yes,No,Yes,Valid,Invalid
Edwin,Mombasa,Malaba,No,Yes,No,Valid,Valid
Edwin,Mombasa,Malaba,Yes,Yes,Yes,None,Invalid
Edwin,Mombasa,Malaba,Yes,No,No,Invalid,Valid
Edwin,Mombasa,KANDUYI,Yes,Yes,Yes,Invalid,Invalid
Edwin,Mombasa,Malaba,No,No,No,Valid,Valid
Edwin,Mombasa,Malaba,Yes,No,Yes,Valid,Invalid
Edwin,Mombasa,Malaba,No,Yes,No,None,Invalid
Edwin,Mombasa,Malaba,No,Yes,No,Valid,Valid

Table 4: Part of the data extracted from weighbridge database

Table 3 shows part of the data that was analyzed in WEKA software. The data comprised seven attributes, namely; origin, destination, first aid kit, fire Ext, Drunken Driving, Drivers License, and Load Permit. Total data for the training data set had 1581 instances. Figure 10 shows a preprocess tab in WEKA. The tab includes a bar chart presentation of the attributes using various and difference colors for each attribute. Figure 11 shows all the attributes on one window with all the numbers of instances per attribute.

Figure 10 is the WEKA window showing all the eight attributes in the analyzed data. The tabs on the upper bar are preprocess, classify, cluster, associate, select attributes and visualize. All are for the purpose of performing the data mining activities. Figure 11 is a visualization of all the attributes in bar chart format. It shows instances for all the attributes when a particular attribute is selected. The displayed attributes in bar charts are indicated at the top of each segment of the chart, they include, origin, destination, first aid kit, fire extinguisher, drunken driving, drivers license and load permit.

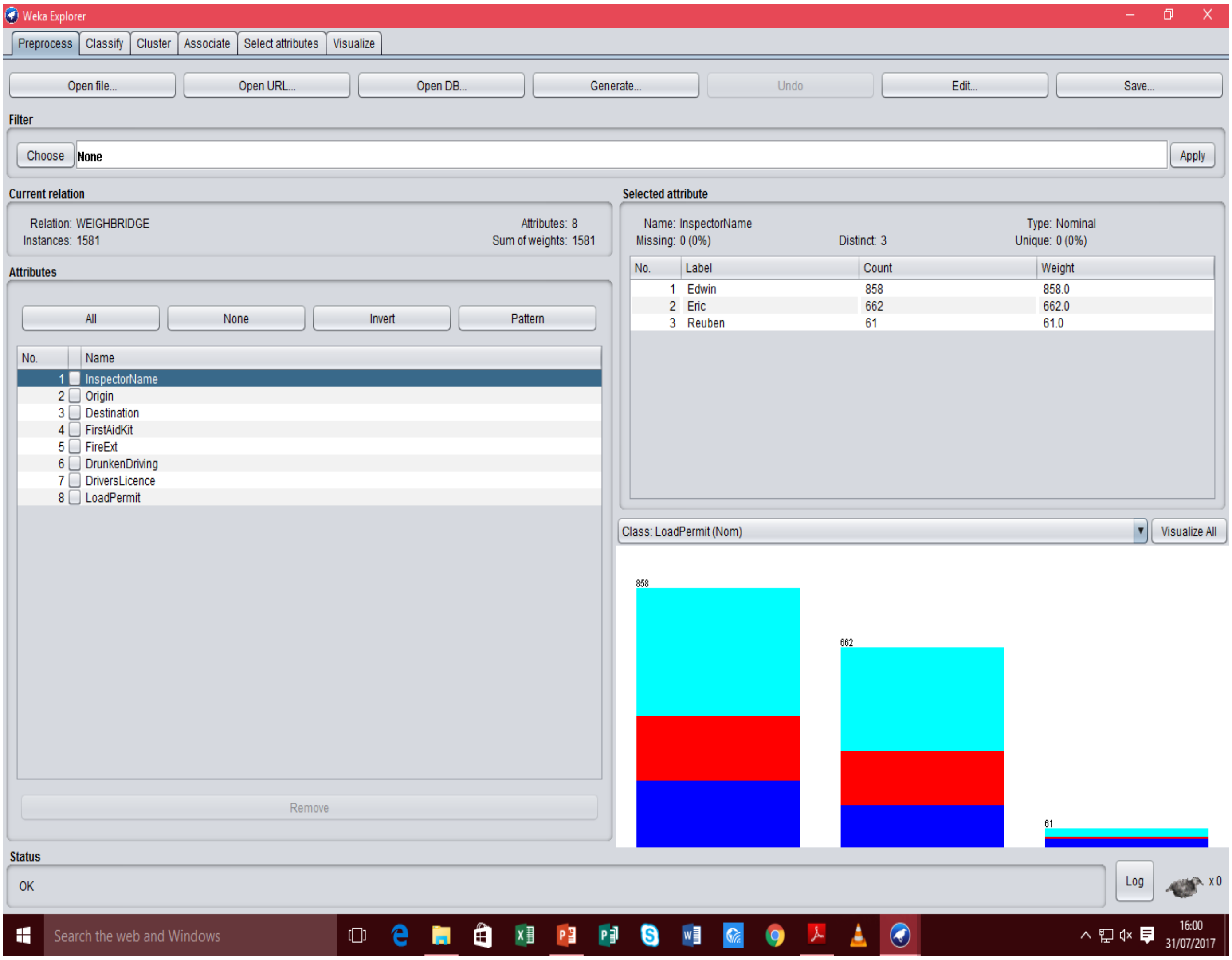


Figure 10: Training Data Set Loaded in WEKA (1,581 items).

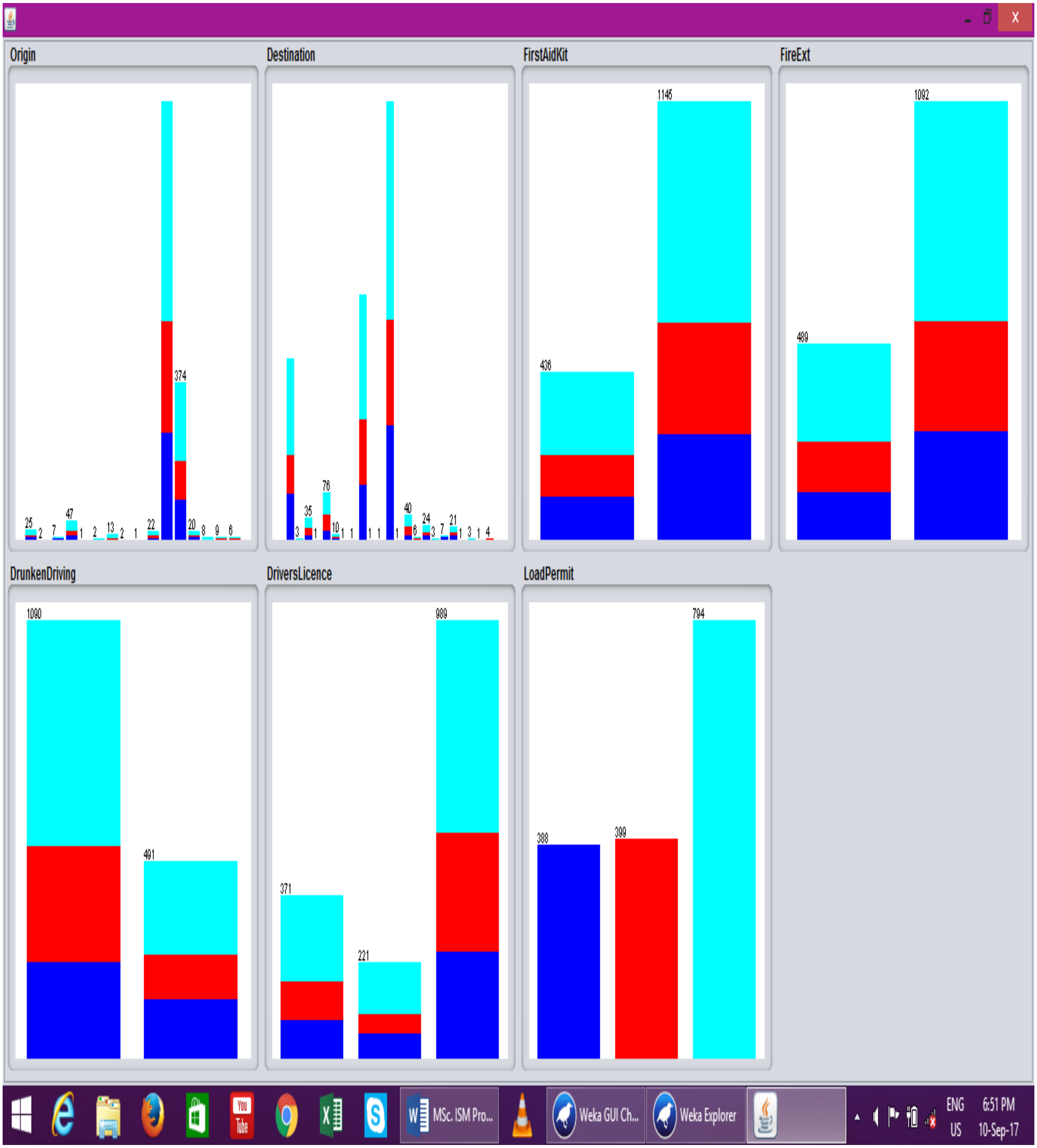


Figure 11: Data in WEKA showing all the attributes

4.3 Data Analysis and Results

A model resulting from the weighbridge data run in WEKA is as below. It comprised seven attributes with 1,581 data items in total. The tree had 34 leaves and 42 nodes. A visualization of the tree is also presented s a decision tree.

The Model

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: WEIGHBRIDGE-weka.filters.unsupervised.attribute.Remove-R1

Instances: 1581

Attributes: 7

Origin

Destination

FirstAidKit

FireExt

DrunkenDriving

DriversLicence

LoadPermit

Test mode: evaluate on training data

=== Classifier model (full training set) ===

J48 pruned tree

Destination = Bungoma: Nairobi (281.0/77.0)

Destination = burundi: Mombasa (3.0)

Destination = Busia: Nairobi (35.0/4.0)

Destination = CHWELE: Eldoret (1.0)

Destination = DRC: Mombasa (76.0)

Destination = Eldoret: Bungoma (10.0/5.0)

Destination = ISIOLO: Malaba (1.0)

Destination = JINJA: Mombasa (1.0)

Destination = Kampala: Mombasa (379.0/27.0)

Destination = KANDUYI: Mombasa (1.0)

Destination = KITALE: Kisumu (1.0)

Destination = Malaba: Mombasa (681.0/130.0)

Destination = MATISI: WEBUYE (1.0)

Destination = Mombasa

| FireExt = No: Malaba (16.0/9.0)

| FireExt = Yes

| | DrunkenDriving = No

| | | FirstAidKit = No

| | | | LoadPermit = Invalid: Kampala (0.0)

| | | | LoadPermit = None: Kampala (3.0/1.0)

| | | | LoadPermit = Valid: S.Sudan (5.0/2.0)

| | | FirstAidKit = Yes

| | | | LoadPermit = Invalid: DRC (3.0/1.0)

| | | | LoadPermit = None: S.Sudan (5.0/2.0)

| | | | LoadPermit = Valid: DRC (5.0/3.0)

| | DrunkenDriving = Yes: Malaba (3.0/1.0)

Destination = mumias: Nairobi (6.0/2.0)

Destination = Nairobi

| DriversLicence = Invalid

| | LoadPermit = Invalid: Bungoma (0.0)

| | LoadPermit = None: Malaba (2.0/1.0)

| | LoadPermit = Valid: Bungoma (3.0)

| DriversLicence = None: Malaba (6.0/3.0)

| DriversLicence = Valid: Bungoma (13.0/3.0)

Destination = Nakuru: Malaba (3.0/1.0)

Destination = RWANDA: Mombasa (7.0)

Destination = S.Sudan: Mombasa (21.0)

Destination = Thika: Malaba (1.0)

Destination = TORORO: Mombasa (3.0)

Destination = TURBO: Bungoma (1.0)

Destination = Webuye: Bungoma (4.0/1.0)

Number of Leaves : 34

Size of the tree : 42

Time taken to build model: 0.03 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0.11 seconds

=== Summary ===

Correctly Classified Instances	1308	82.7324 %
Incorrectly Classified Instances	273	17.2676 %
Kappa statistic	0.6267	
Mean absolute error	0.0354	
Root mean squared error	0.1329	
Relative absolute error	55.2024 %	
Root relative squared error	74.6091 %	
Total Number of Instances	1581	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.880	0.006	0.710	0.880	0.786	0.787	0.997	0.797	Bungoma
0.000	0.000	0.000	0.000	0.000	0.000	0.997	0.200	Busia
0.571	0.003	0.500	0.571	0.533	0.532	0.997	0.472	DRC
0.021	0.000	1.000	0.021	0.042	0.144	0.858	0.136	Eldoret
0.000	0.000	0.000	0.000	0.000	0.000	0.785	0.001	JUJA
0.000	0.000	0.000	0.000	0.000	0.000	0.912	0.007	KAKAMEGA
0.154	0.001	0.667	0.154	0.250	0.318	0.993	0.427	Kampala
0.500	0.000	1.000	0.500	0.667	0.707	0.999	0.591	Kisumu
0.000	0.000	0.000	0.000	0.000	0.000	0.911	0.004	LUGULU
0.773	0.010	0.531	0.773	0.630	0.635	0.994	0.599	Malaba
0.974	0.291	0.866	0.974	0.917	0.739	0.888	0.904	Mombasa
0.639	0.069	0.742	0.639	0.687	0.602	0.843	0.608	Nairobi

	0.000	0.000	0.000	0.000	0.000	0.000	0.739	0.031	Nakuru
	0.000	0.000	0.000	0.000	0.000	0.000	0.913	0.028	RUIRU
	0.667	0.003	0.600	0.667	0.632	0.630	0.997	0.539	S.Sudan
	0.167	0.000	1.000	0.167	0.286	0.408	0.927	0.184	WEBUYE
Weighted Avg.	**0.827**	**0.208**	**0.811**	**0.827**	**0.799**	**0.666**	**0.880**	**0.775**	

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	<-- classified as
22	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	a = Bungoma
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	b = Busia
0	0	4	0	0	0	1	0	0	2	0	0	0	0	0	0	c = DRC
0	0	0	1	0	0	0	0	0	0	9	37	0	0	0	0	d = Eldoret
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	e = JUJA
0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	f = KAKAMEGA
1	0	1	0	0	0	2	0	0	6	0	0	0	0	3	0	g = Kampala
1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	h = Kisumu
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	i = LUGULU
3	0	1	0	0	0	0	0	0	17	0	0	0	0	1	0	j = Malaba
1	0	0	0	0	0	0	0	0	3	1015	23	0	0	0	0	k = Mombasa
1	0	0	0	0	0	0	0	0	0	134	239	0	0	0	0	l = Nairobi
0	0	0	0	0	0	0	0	0	0	13	7	0	0	0	0	m = Nakuru
0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	n = RUIRU
0	0	2	0	0	0	0	0	0	1	0	0	0	0	6	0	o = S.Sudan
0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	1	p = WEBUYE

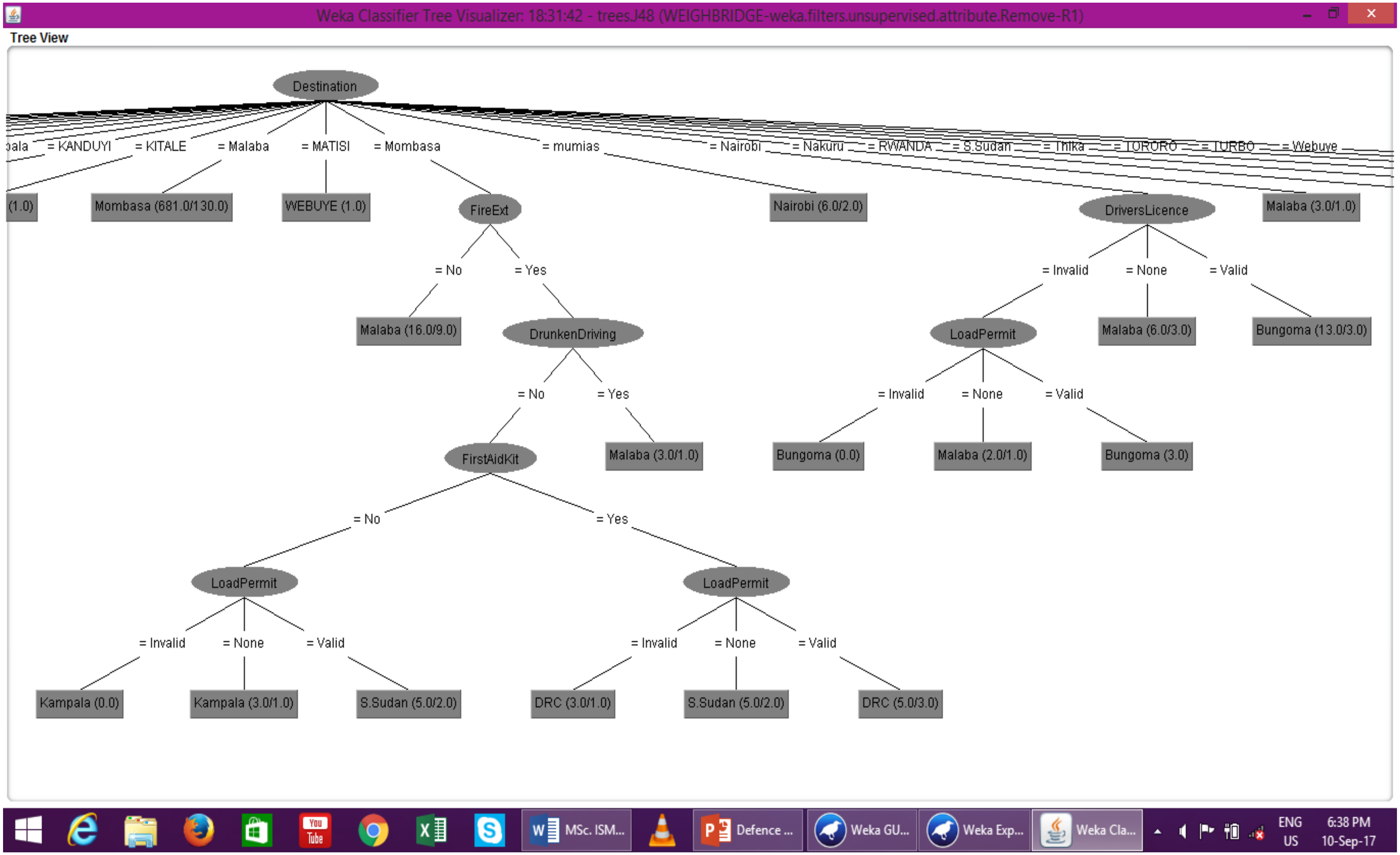


Figure 12: Tree visualization of the model build from training data set.

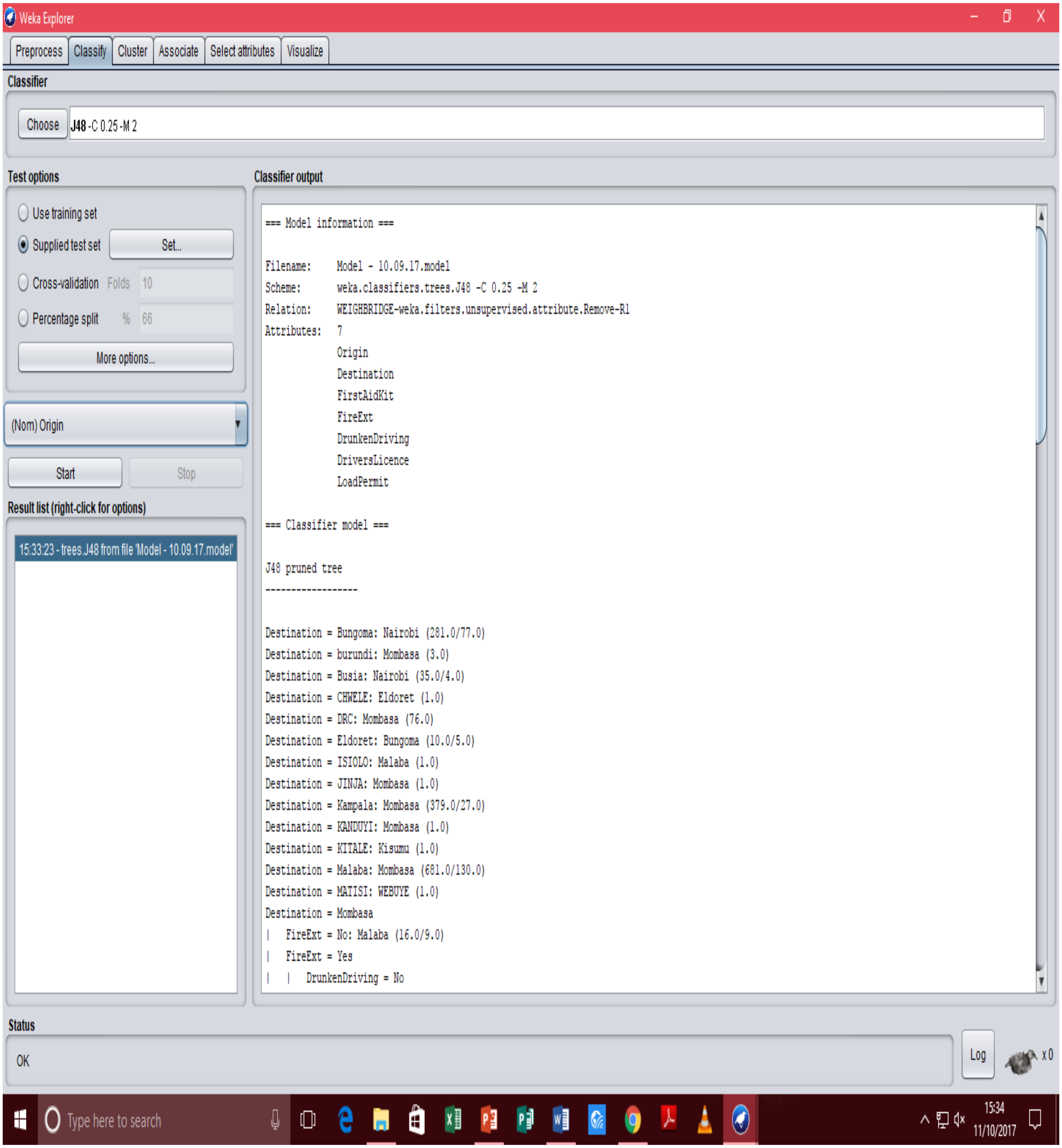


Figure 13: In text visualization of the Tree

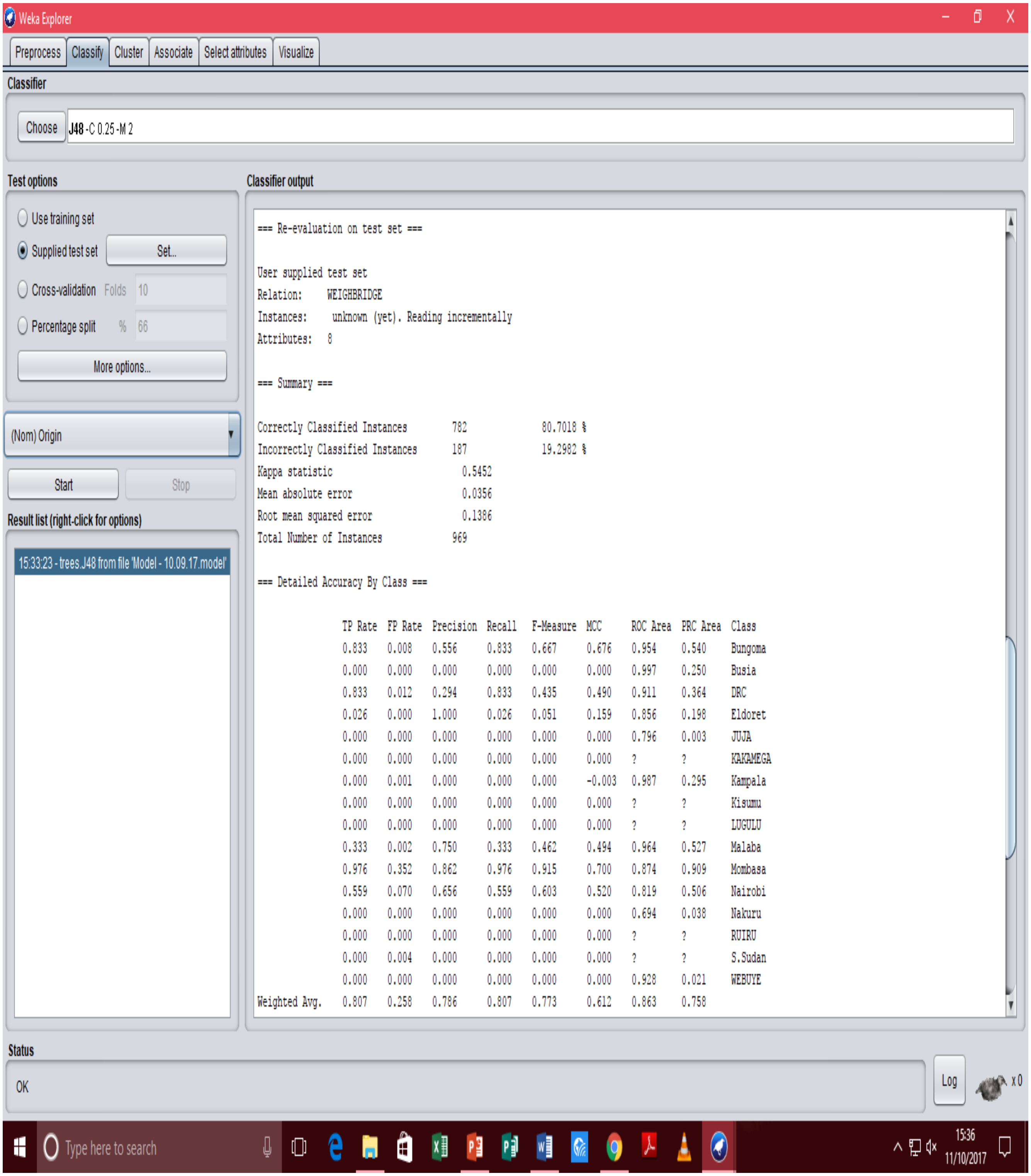


Figure 14: Model test in text form

In figure 12, is a decision tree showing how trucks from certain origin and headed to some destinations comply with the legal requirements. The tree was interpreted to come up with decision rules in the results. Figure 13 is a display of of WEKA window with the supplied test data loaded in the software ready for testing, this is continued in figure 14 where the test result is shown and indicates 80.7% accuracy.

Model Testing

The model was re-evaluated using test data that had 969 items and the result was as below;

```
=== Model information ===
Filename:     Model - 10.09.17. model
Scheme:       weka. classifiers. trees -C 0.25 -M 2
Relation:     WEIGHBRIDGE-weka. filters. unsupervised. attribute. Remove-R1
Attributes:   7
              Origin
              Destination
              FirstAidKit
              FireExt
              DrunkenDriving
              DriversLicence
              LoadPermit

=== Classifier model ===
J48 pruned tree
------------------
Destination = Bungoma: Nairobi (281.0/77.0)
Destination = burundi: Mombasa (3.0)
Destination = Busia: Nairobi (35.0/4.0)
Destination = CHWELE: Eldoret (1.0)
Destination = DRC: Mombasa (76.0)
Destination = Eldoret: Bungoma (10.0/5.0)
Destination = ISIOLO: Malaba (1.0)
Destination = JINJA: Mombasa (1.0)
Destination = Kampala: Mombasa (379.0/27.0)
Destination = KANDUYI: Mombasa (1.0)
Destination = KITALE: Kisumu (1.0)
Destination = Malaba: Mombasa (681.0/130.0)
Destination = MATISI: WEBUYE (1.0)
Destination = Mombasa
|  FireExt = No: Malaba (16.0/9.0)
|  FireExt = Yes
|  |  DrunkenDriving = No
|  |  |  FirstAidKit = No
|  |  |  |  LoadPermit = Invalid: Kampala (0.0)
|  |  |  |  LoadPermit = None: Kampala (3.0/1.0)
|  |  |  |  LoadPermit = Valid: S.Sudan (5.0/2.0)
```

| | | FirstAidKit = Yes
| | | | LoadPermit = Invalid: DRC (3.0/1.0)
| | | | LoadPermit = None: S.Sudan (5.0/2.0)
| | | | LoadPermit = Valid: DRC (5.0/3.0)
| | DrunkenDriving = Yes: Malaba (3.0/1.0)
Destination = mumias: Nairobi (6.0/2.0)
Destination = Nairobi
| DriversLicence = Invalid
| | LoadPermit = Invalid: Bungoma (0.0)
| | LoadPermit = None: Malaba (2.0/1.0)
| | LoadPermit = Valid: Bungoma (3.0)
| DriversLicence = None: Malaba (6.0/3.0)
| DriversLicence = Valid: Bungoma (13.0/3.0)
Destination = Nakuru: Malaba (3.0/1.0)
Destination = RWANDA: Mombasa (7.0)
Destination = S.Sudan: Mombasa (21.0)
Destination = Thika: Malaba (1.0)
Destination = TORORO: Mombasa (3.0)
Destination = TURBO: Bungoma (1.0)
Destination = Webuye: Bungoma (4.0/1.0)

Number of Leaves : 34
Size of the tree : 42

=== Re-evaluation on test set ===
User supplied test set
Relation: WEIGHBRIDGE
Instances: unknown (yet). Reading incrementally
Attributes: 8

=== Summary ===

Correctly Classified Instances	782	80.7018 %
Incorrectly Classified Instances	187	19.2982 %
Kappa statistic	0.5452	
Mean absolute error	0.0356	
Root mean squared error	0.1386	
Total Number of Instances	969	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.833	0.008	0.556	0.833	0.667	0.676	0.954	0.540	Bungoma
0.000	0.000	0.000	0.000	0.000	0.000	0.997	0.250	Busia
0.833	0.012	0.294	0.833	0.435	0.490	0.911	0.364	DRC
0.026	0.000	1.000	0.026	0.051	0.159	0.856	0.198	Eldoret
0.000	0.000	0.000	0.000	0.000	0.000	0.796	0.003	JUJA
0.000	0.000	0.000	0.000	0.000	0.000	?	?	KAKAMEGA
0.000	0.001	0.000	0.000	0.000	-0.003	0.987	0.295	Kampala
0.000	0.000	0.000	0.000	0.000	0.000	?	?	Kisumu
0.000	0.000	0.000	0.000	0.000	0.000	?	?	LUGULU
0.333	0.002	0.750	0.333	0.462	0.494	0.964	0.527	Malaba
0.976	0.352	0.862	0.976	0.915	0.700	0.874	0.909	Mombasa
0.559	0.070	0.656	0.559	0.603	0.520	0.819	0.506	Nairobi

	0.000	0.000	0.000	0.000	0.000	0.000	0.694	0.038	Nakuru
	0.000	0.000	0.000	0.000	0.000	0.000	?	?	RUIRU
	0.000	0.004	0.000	0.000	0.000	0.000	?	?	S.Sudan
	0.000	0.000	0.000	0.000	0.000	0.000	0.928	0.021	WEBUYE
Weighted Avg.	**0.807**	**0.258**	**0.786**	**0.807**	**0.773**	**0.612**	**0.863**	**0.758**	

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	<-- classified as
10	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	a = Bungoma
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	b = Busia
0	0	5	0	0	0	0	0	0	1	0	0	0	0	0	0	c = DRC
0	0	0	1	0	0	0	0	0	0	8	29	0	0	0	0	d = Eldoret
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	e = JUJA
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	f = KAKAMEGA
1	0	4	0	0	0	0	0	0	0	0	0	0	0	3	0	g = Kampala
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	h = Kisumu
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	i = LUGULU
3	0	7	0	0	0	1	0	0	6	0	0	0	0	1	0	j = Malaba
2	0	0	0	0	0	0	0	0	0	655	14	0	0	0	0	k = Mombasa
0	0	0	0	0	0	0	0	0	0	83	105	0	0	0	0	l = Nairobi
0	0	0	0	0	0	0	0	0	0	13	9	0	0	0	0	m = Nakuru
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	n = RUIRU
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	o = S.Sudan
0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	p = WEBUYE

4.4 Interpretation of Results

From the data mining results in the model, the following rules emerge. Rules are described from the decision tree diagram and model built from the data that was analyzed.

a) Trucks from Nairobi to Malaba did not have fire extinguisher.

b) Trucks from Mombasa to Kampala were likely to involve in drunk driving, lacked first aid kit and did not have load permit.

c) Trucks from Mombasa to South Sudan involved in drunk driving, did not have fire extinguisher but had valid load permit.

d) Trucks from Nairobi to Malaba had invalid driver's licence and had no load permits.

e) Trucks from Nairobi to Malaba had invalid driver's licence and had no load permits.

The listed rules above were confirmed when the model was tested. The test model returned accuracy of 80.7%. This level of accuracy of certainly high, hence the model provides a reliable prediction tool.

4.5 Discussion of Results

There were a total of eleven (11) origin points and twenty (20) destination points in the data that used. The trucks origins were matched to a destination point for each truck contained in the data. Similar studies, which have focused on traffic origins and destinations, aimed at investigating certain patterns in the behaviors.

A study using data mining techniques to build a Classification Model for Predicting Employees Performance by Qasem A. Al-Radaideh and Eman Al Nagi and Amman, Jordan (2012). In the study data mining techniques were utilized to build a classification model to predict the performance of employees. Decision tree was the main data mining tool used to build the classification model, where several classification rules were generated. To validate the generated model, several experiments were conducted using real data collected from several companies. The model was intended to be used for predicting new applicants' performance.

Study on Human Talent Forecasting Data Mining Classification Techniques for Human Talent Forecasting by Hamidah Jantan1, Abdul Razak Hamdan2 and Zulaiha Ali Othman in Malaysia. This study described the significance of using data mining for talent management especially for classification and prediction. This generated classification rules can be used to predict the potential talent for the specific task in an organization.

CHAPTER FIVE: CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction

This chapter outlines the conclusions reached from the study results. The conclusions are derived from the data mining analysis done using J.48 classification algorithm. The variables in the analysis comprised destination as dependent variable one hand and validity of load permits, fire extinguisher, drunk driving, valid driving licenses on the other side. The rules derived from the model explains the predictive pattern created using the variables.

5.2 Contributions to Knowledge

In conclusion, it has been established that there is a pattern between the destinations of the trucks and the likelihood of them not complying with legal requirements and axle load control regulations. From this study, management and policy makers at weighbridges should formulate policies which determine strict inspection of motor vehicles which head o the destinations that mostly break the laws and axle load regulations. Those that head to destinations that mostly comply with the regulations should receive less inspection time to save on the duration they spend the weighbridges. This will ease the delays currently witnessed at the weighbridges hence, saving on transportation costs.

5.3 Conclusions and Recommendations for Future Works

This research provides a basis to carry out a deeper inquiry into the causes of the problem of non-compliance with axle load regulations and laws by trucks that head to some identified destinations. Future research may then seek to establish why the vehicles fail to comply as required. Additionally, a research may also be performed on ways to completely automate the inspection process at the weighbridge so that it takes less time and avoid the long queues irrespective of where the vehicles head. The data obtained of this research related to the years 2015 and 2016.

REFERENCES

1. Abbas Mahmoudabadi & Seyed Mohammad Seyedhosseini (2012): *Improving the efficiency of weigh in motion systems through optimized allocating truck checking oriented procedure.*
2. Abbas Mahmoudabadi (2013); *Improving the efficiency of weigh in motion systems through optimized allocating truck checking oriented procedure.*
3. Anwar Shah, (2007); *Performance Accountability and Combating Corruption; A publication of the World Bank.*
4. Barrot Jacques (2005) *European Best Practice Guidelines for Abnormal Road Transports*
5. Benard Jacob (Paris, France 2008): *Weigh-in-motion for road safety, Enforcement and Infrastructure.*
6. Charu C. Aggarwa et al (2015): *Data Classification Algorithms and Applications: Chapman & Hall/CRC Data Mining and Knowledge Discovery Series*
7. Chien, C. F., & Chen, L. F. (2008). *Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry. .*
8. David J.C. MacKay 2005; *Information Theory, Inference, and Learning Algorithms*
9. Felix Klebe & Bob Susor (undated): *Using Weigh in Motion to Improve Operational Productivity.*
10. Government of Kenya Printers, (2008); *Kenya Roads Act, 2007*
11. Hamidah Jantan, Abdul Razak Hamdan and Zulaiha Ali Othman (2011). *Data Mining Classification Techniques for Human Talent Forecasting, Knowledge-Oriented Applications in Data Mining, Prof. Kimito Funatsu (Ed.),*
12. Helmut Drüke (2007): *Can E-Government Make Public Governance More Accountable? The International Bank for Reconstruction and Development / The World Bank; Performance Accountability and Combating corruption*.

13. Hesham Rakha, Bryan Katz & Ahmed Al-Kaisy (2006): *Field Evaluation of Truck Weigh Station Operations, Journal of Intelligent Transportation Systems.*

14. Hua Z, Nasim U, Xudong S, Ping Z & Chengjun T (2015): *Field-calibrated influence lines for improved axle weight identification with a bridge weigh-in-motion system, Structure and Infrastructure Engineering.*

15. Hua Zhao, Nasim Uddin, Xudong Shao, Ping Zhu & Chengjun Tan (2015): *Field-calibrated influence lines for improved axle weight identification with a bridge weigh-in-motion system, Structure and Infrastructure Engineering.*

16. Ian G. Heggie & Piers Vickers (1998): *Commercial Management and Financing of Roads WORLD BANK TECHNICAL PAPER NO. 409 The World Bank Washington, D.C.*

17. Ian Written and Eibe Frank (2005): *Data Mining Practical Machine Learning Tools and Techniques Morgan Kaufmann Publishers is an imprint of Elsevier*.

18. Jinwoo Lee & Garland Chow (2011): *Operation Analysis of the Electronic Screening System at a Commercial Vehicle Weigh Station, Journal of Intelligent Transportation Systems.*

19. Kenya National Highways Authority Weighbridge Operations Manual, (2011)

20. Kothari C. R, 2004: *Research Methodology. Methods and Techniques- 2nd Edition*

21. Masinga Gisore Walter (2013): *A smart Traffic Management System, KCA University Project in submitted for MSc. Data Communication*.

22. Megha Gupta, Naveen Aggarwal (2010): *Classification Techniques analysis, UIET Punjab University*

23. Michael Ian Pinard (March 2010): *Guidelines on Vehicle Overload Control in Eastern and Southern Africa. A publication of International Bank for Reconstruction and Development/World Bank*.

24. Nicholas William, (Taylor & Francis e-Library, 2011): *Research Methods, The Basics*

25. Nordengen P. A (Not dated) pp 34-50: *New Era of Overloading Control in South Africa*.

26. Obuandike Georgina N. et al 2015: *Analytical Study of Some Selected Classification Algorithms in WEKA Using Real Crime Data*

27. Ozer Patrick: (2008) *Data Mining Algorithms for Classification Radboud University Nijmegen,*

28. Patrick J. Szary (2009): *Implementation of Weigh-in-Motion (WIM) Systems*.

29. Pinard Mike Ian (2011): *Emerging Good Practice in Over Load Control in Eastern and Southern Africa – Selected Case Studies.*

30. SGS Kenya Limited (2016): *Report on Management Contract for Webuye Weighbridge Station and Adjacent Road Network*.

31. Silvester M. Liech (2013): *Framework for Axle Load Management Information System for Kenya. University of Nairobi project submitted for MSc. Information Management System, School of Computing and Informatics.*

32. Supee Teravaninthorn Gaël Raballand (2009): *Transport Prices and Costs in Africa: A Review of the Main International Corridors*

33. Sven Ove Hansson 1994; *Decision Theory – A brief Introduction*

34. Tesema Tibebe Beshah, Ajith Abraham and Crina Grosan, 2011: *Rule mining and classification of road traffic accidents using adaptive regression trees.*

35. The National Treasury (2016): *Government of Kenya Budget Statement in the Financial Year 2016/2017*.