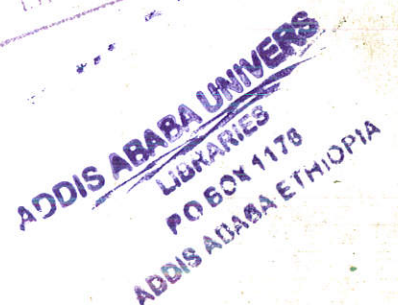
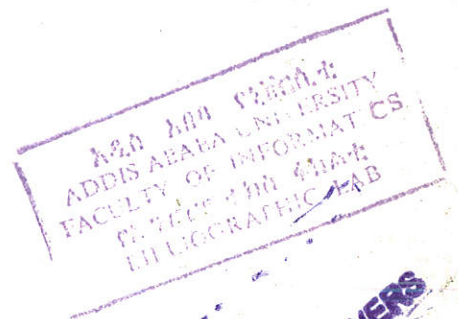


**ADDIS ABABA UNIVERSITY
SCHOOL OF GRADUATE STUDIES
FACULTY OF INFORMATICS
DEPARTMENT OF INFORMATION SCIENCE**

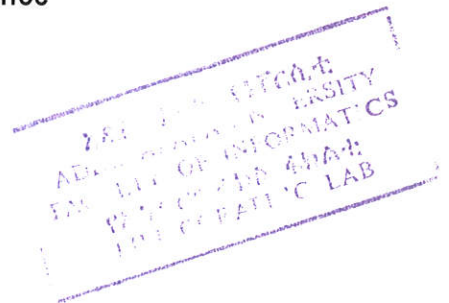
**APPLICATION OF DATA MINING TECHNOLOGY TO
SUPPORT ROAD TRAFFIC ACCIDENT SEVERITY
ANALYSIS AT ADDIS ABABA TRAFFIC OFFICE**

A THESIS SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES OF ADDIS
ABABA UNIVERSITY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF SCIENCE IN INFORMATION SCIENCE

BY
TIBEBE BESHAH TESEMA
JUNE 2005



ADDIS ABABA UNIVERSITY
SCHOOL OF GRADUATE STUDIES
Faculty of Informatics
Department of Information Science



APPLICATION OF DATA MINING TECHNOLOGY TO SUPPORT ROAD TRAFFIC
ACCIDENT ANALYSIS AT ADDIS ABABA TRAFFIC OFFICE

BY
TIBEBE BESHAH

Name and Signature of Member of the Examining Board

Ato Getahcew Jemanhe , Chairman, Examining Board

Ato Getahcew Jemanhe

Dr. Kumudha Raimond, Examiner

Dr. Kumudha Raimond

Dr. B.L. Desai, Advisor

Dr. B.L. Desai

Chairman, Faculty

Ato Getahcew Jemanhe

Signature

Date

12/07/05

Chairman, Graduate Council

Signature

Date

Dedication

*I dedicate this thesis to my parents, Zenebech Zeleke and Kes Beshaha
Tesema, who costed everything they had to me.*

Acknowledgment

First and for most I would like to thank my advisor Dr. B. L. Desai for his constructive comments and for refining my work. My special thanks also goes to Dr. Ajith Abraham from Korea, Chung-Ang University, School of Computer Science and Engineering for his useful support during my research.

I would like to thank my brothers and sister Fekade Beshah, Alemayehu G/Mariam and Etfwork Eshete for their support and encouragement during my study. Their council and encouragement was always driving me forward.

I am very grateful to the management and staff of Addis Ababa Traffic Office, especially to Commander Tegoda, Ato Daniel, and statistics department staffs, for their constant support in accessing the data and in providing relevant information to my research work.

My special thanks goes to my beloved family for their unreserved backing and encouragement. I am also very much indebted to Sisay, Aberash and Awel who stood beside me at this critical time in my life. They were there whenever I needed someone. My dear Zedy deserves special thanks for her constant prop, understanding and faith on me.

I am also grateful to my class mates and staff members of Faculty of Informatics. Special thanks to my friends Yihun, Endale and G/Meskel for their unreserved help; they often had a solution to the problems I faced.

Above all, I thank Almighty God who surrounded me with so many wonderful people.

Tibebe Beshah

Table of Contents

Dedication	iii
Acknowledgment	iv
Table of contents	v
List of figures	vii
List of tables.....	viii
Abstract	iv
 Chapter One: Introduction	 ...1
1.1. Background.....	...1
1.2. Statement of the problem and justification of the study.....	...4
1.3. Objectives of the study.....	...5
1.3.1. General objectives.....	...5
1.3.2. Specific objectives.....	...5
1.4. Methodology Adopted.....	...6
1.4.1. Literature review	...6
1.4.2. Identifying the data source: data collection.....	...6
1.4.3. Data making/preparation.....	...7
1.4.4. Model building/Data analysis.....	...7
1.5. Scope and limitation of the study.....	...7
1.6. Research contribution.....	...8
1.7. Thesis organization.....	...8
Chapter Two: Data Mining.....	10
2.1 Overview.....	10
2.2 Data Mining and the KDD process	11
2.3 Data Mining and other statistical tools	13
2.4 Data warehousing and OLAP technology for Data Mining.....	14
2.5 Data Mining Technologies.....	15
2.5.1 Predictive modelling.....	16
2.5.2 Descriptive modelling	17
2.6 Data Mining Techniques.....	18
2.6.1 Decision trees	18
2.6.2 Artificial neural networks.....	20
2.6.3 Other data mining techniques	22
2.7 Application of data mining technologies.....	23
2.7.1 General applications.....	23
2.7.2 Application of Data Mining in Road Transport and Accidents Analysis....	25
Chapter Three: Road Traffic Accident and Injury Analysis	27
3.1 Introduction	27
3.2 Road Traffic accidents	28
3.3 Road safety and accident injury severity	30
3.4 Road traffic accidents at Addis Ababa	32
3.4.1 Overview	32
3.4.2 Accident data analysis at Addis Ababa	33
Chapter Four: Preparing Data for Analysis	36

4.1	Overview	36
4.2	Data collection	36
4.2.1	Accident data set: data understanding	38
4.2.2	Defining data mining function	40
4.3	Data pre-processing	41
4.3.1	Data cleaning	41
4.3.2	Attribute/Variable Selection	43
4.3.3	Data transformation and aggregation	44
Chapter Five: Model building		46
5.1	Overview	46
5.2	Selection of the modelling technique	47
5.3	The experiment	48
5.3.1	Model Building	49
5.4	Evaluation	63
Chapter Six: Conclusion and Recommendation		65
6.1	Conclusions.....	65
6.2	Recommendations.....	66
Reference.....		69
Appendices.....		73
Appendix 1: Rules Generated.....		73
Appendix 2: Accident data recording form.....		81
Declaration		82

List of Figures

Figure 2.1: Projected changes in the ranking of the 15 leading causes of death and disease.....	28
Figure 5.1: The knowledge studio data set partition dialog	48
Figure 5.2: Overview report of the training dataset	49
Figure 5.3: Overview of the tree attributes editor	50
Figure 5.4: Output of the first Decision Tree	51
Figure 5.5: Prediction statistics of the first experiment.....	52
Figure 5.6: Output of the second Decision Tree.....	53
Figure 5.7: Validation results from the second decision tree	54
Figure 5.8: 'Best' decision tree with seven variables.....	55
Figure 5.9: Validation results from the 'best' decision tree	56
Figure 5.10: Accuracy result of the second experiment on predictive model.....	61
Figure 5.11: Accuracy result of the second experiment on predictive model.....	61
Figure 5.12: Accuracy result of the third experiment on predictive model	62

List of Tables

Table 4.1: Attributes with their description and data type.....	39
Table 4.2: Attributes with missing values	42
Table 4.3: Selected attributes with their data type and description.....	44
Table 4.4: Derived attributes with their base attributes and values	45
Table 5.1: Results of the Tests Conducted	54
Table 5.2: Experiment accuracy comparison table	62

Abstract

Road traffic accidents are among the top leading causes of deaths and injuries of various levels. Ethiopia is one of the countries of the world experiencing highest rate of such accidents resulting in fatalities and various levels of injuries. Addis Ababa, the capital city of Ethiopia, takes the lion's share of the risk having higher number of vehicles and traffic. And the cost of these fatalities and injuries due to such road traffic accidents has a great impact on the socio-economic development of a society.

This thesis reports on the findings of a research that had the objective to build a decision support system to handle road traffic accident analysis, for Addis Ababa City Traffic Office. The study focused on injury severity levels resulting from an accident. In doing so the aim of this research was to assess the potential applicability of data mining technology specifically decision tree technique to help traffic accident data analysis in decision-making process at the traffic office.

In the thesis, the process of building a model through knowledge discovery and data mining techniques on historical accident record data is described. Different tools and techniques are also used for the purpose of data analysis. The methodology adopted had three basic steps namely data collection, data preparation, and model building and validation. The required data was selected and extracted from Addis Ababa Traffic Office. Then, data preparation tasks (such as data transformation, deriving of new attributes, and handling of missing values) were undertaken. The final step was model building and validation using the selected tools and techniques.

The decision tree KnowledgeSEEKER algorithm is used in the study. The particular tool used for the model building was the decision tree incorporated in KnowledgeSTUDIO. After successive experiments, a model that can classify accidents well with a better accuracy as fatal, serious, and slight or property-damage was selected and evaluated. Experiment results reveal that the use of decision tree is helpful in detecting dangerous accidents through identifying behavioral and roadway accident patterns. The reported findings are promising, making the proposed model a useful tool in the decision making process. And the whole research process can be a good input for further in-depth research.

Chapter One

Introduction

1.1 Background

Analyzing, interpreting and making maximum use of the data has been difficult and resource demanding due to the exponential growth of many business, governmental and scientific databases. It is estimated that the amount of data stored in the world's database grows every twenty months at a rate of 100% (Whitten and Frank, 2000). 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 (Mao, 2001).

In today's world, where the accumulation of data is increasing in an alarming rate, understanding interesting patterns of data is an important issue to be considered to adjust strategies, to make maximum use of it, and find new opportunities. Organizations keeping data on their domain area takes every record as an opportunity in learning facts. But the simple gathering of data is not enough to get maximum knowledge out of it. Thus, for an effective learning, data from many sources must first be gathered and organized in a consistent and useful manner, data warehousing. Data warehousing allows the enterprise to recognize what it has noticed about its domain area. The data must also be analyzed, understood, and turned into actionable information. This is the point where the application of data mining is needed.

Although it is difficult to define precisely and delimit the range and limits of such scientific disciplines, many scholars try to indicate the basic tasks of data mining. In line with this, Hand, (2005) defines Data mining as the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner.

Data mining can also be seen as a combination of tools, techniques and processes in knowledge discovery. In other words, it uses a variety of tools ranging from classical statistical methods to neural networks and other new techniques originating from machine learning and artificial intelligence in improving database promotion and process optimization.

According to Berry and Linoff (2000) the six basic functions or activities of data mining are classified in to directed and undirected data mining. Specifically classification, estimation and prediction are directed, where the available data is used to build a model that describes one particular variable of interest in terms of the rest of the available data. Affinity grouping or association rules, clustering, description and visualization on the other hand are undirected data mining where the goal is to establish some relationship among all variables. And among the techniques, the most widely used once are, Artificial Neural Networks, Decision Tree, Nearest Neighbor Method, and Rule Induction (Thearling, 2003).

Up to recent time, the only analysis made on data to get meaning out of it, is simple statistical manipulation that has no power to show all the necessary information content of a given data. But data mining technology, on the other hand has the greatest potential in identifying various interesting patterns for enabling organizations to control data resources for strategic planning and decision-making in their domain area.

Road Traffic Control Systems

Traffic control system is one of the various areas, where critical data about the well-being of the society is recorded and kept. Various aspects of a traffic system like vehicle accidents, traffic volumes and concentration are recorded at different levels. In connection to this injury severities resulted from road traffic accident are one of the areas of concern.

Traffic accidents are a major cause of severe injuries in most countries. A report from World Health Organization indicates that the leading injury-related cause of death among people aged 15-44 years is traffic injuries. Of the 5.8 million people who died of injuries in 1998, 1,170,694 died as a direct result of injuries sustained in a motor vehicle accident. Worldwide, the organization reports that, 38,848,625 injuries (serious and slight) were received by people involved in motor vehicle accidents in 1998. And in relation to this, on average in the industrialized countries, and also in many developing countries, an accident victim occupies one hospital bed in ten. As indicated in various literatures, Ethiopia is one of the countries experiencing large number of road traffic accidents in the world. Specifically according to the World health Organization, Ethiopia has the highest rate of fatalities per vehicle in the world. Uganda ranks second in road fatality rates in the world behind Ethiopia. Moreover emergency medical systems are often poor and injury prevention programs are rarely available (SafeCarGuide, 2004).

Out of all accidents registered in Ethiopia, Addis Ababa holds about 60% on average. This is partly because the city has great contact through its all gates with different regions every day. In addition to this, of the registered motor vehicles in Ethiopia, the city takes about 77% of it. All these facts reveal that Addis Ababa, having a great deal of concentration of vehicles and traffic, takes the lion's share in car accidents also (Abebe, 2004). In connection to this statistical data from the office shows that Addis Ababa is experiencing around 700 accidents per month. And the costs of such fatalities and injuries due to traffic accidents have a great impact on various aspects of the society.

Being one of the various administrative regions of Ethiopia, Addis Ababa City has different offices in handling deferent socio-economic and political issues. One such office is the Traffic Control Division of Addis Ababa Police Commission.

The primary functions of the division are: -

- ◆ Ensure the safety to the citizen by promoting the safe and orderly flow of traffic on the city's street and highways.
- ◆ Enforce all laws and ordinances as they relate to each of the different forms of streets and highway traffic.
- ◆ Reduce the number of vehicular and pedestrian accidents.
- ◆ Develop and implement strategies that will improve the flow of traffic, remove obstacles to traffic movement and expedite motor vehicle traffic about the city.
- ◆ Reducing crime by improving the quality of life which is part of the goal of the police commission

In managing and controlling the city's traffic system, the Addis Ababa traffic office is structurally organized under three major departments namely Administrative support, Accident investigation, Security and Control. The office has a total of 216 staff including the department head and traffic police officers.

In discharging its task, the office collects and analyzes data on different aspects of road traffic. But most of the data manipulations in the office are more of paper based. In connection with this, Mekitew (2000) tried to propose an automated traffic information system for Addis Ababa Region Traffic Office with the aim of helping the office in

information handling. It has been seen that data especially in some regions where the traffic and number of vehicles are huge, do not get enough attention to use it as a base for decision-making. Identifying and knowing a given pattern of data in a given traffic office will help the decision makers in deciding on the specific future activities.

Thus, through this research work an attempt has been made to apply data mining tools and techniques in analyzing and determining interesting patterns especially with respect to injury severity, on road accidents data at Addis Ababa Region Traffic Control System.

1.2 Statement of the Problem and Justification of the Study

The underlying research problem that necessitated this research is the existence of high vehicle accident at the national level in general and at the city of Addis Ababa in particular. Specifically in the region, although data on vehicle accidents is periodically gathered by the traffic office, due to lack of appropriate data analysis tools, this data is not practically used to alleviate the problems faced by the people, traffic police and policy makers to identify major determinants of car accident, to predict the injury severity, to establish the most important factors influencing the severity of injury.

Thus, in order to plan and implement effective strategies in reducing the severity of the injury and vehicle accident at large in Ethiopia, there is a need for actionable information which is obviously a result of a research work.

So, in the effort of alleviating the current problem of vehicle accidents, identifying factors leading to accidents through developing a capacity to design and implement an effective traffic information system that can provide timely and accurate traffic information is very crucial. In this connection, timely and reliable data collected about vehicle accidents can be used to identify major determinants and risk factors for vehicle accidents, severe injury and fatalities and to take preventive measures so that the effort of improving the quality of life will be enhanced. It has been seen that, there have been some research made; however all those previous researches were conducted by using small proportion of the accumulated data. Besides, in those researches data analysis was conducted by using simple statistical methods.

Since the analysis made by using traditional methods focus on problems with much more manageable number of variables and cases than may be encountered in real world, they have

limited capacity to discover new and unanticipated patterns and relationships that are hidden in conventional databases (Plate et al, 1997)

The absence of significant attempt that has been made so far to this level in identifying the major determinants of car accidents and establishing the most important factors influencing the severity of an injury in Addis Ababa Region justify the importance of this research. This research work will be groundwork for the effort of reducing vehicle accident in particular and improving the quality of life in general. Moreover it will also be an input for researches in the same area.

Thus, in this study, attempt has been made to investigate how technical, social, economical behavioral, environmental and case related factors affect traffic system and cause vehicle accidents and determine the severity nature of injuries at the Addis Ababa City by applying computerized methods of data mining technology.

1.3 Objectives of the Study

The general and specific objectives of the proposed study are the following: -

1.3.1 General Objective

The general objective of the research was to investigate the potential applicability of data mining technology in developing a model that can support road traffic accident severity analysis in the effort of preventing and controlling vehicle accident at the city of Addis Ababa.

1.3.2 Specific objectives

In order to achieve the general objective indicated above, the research will have the following specific objectives

- ◆ To understand the nature of vehicle accidents and traffic control system
- ◆ To assess the potential of data mining techniques in identifying vehicle accident patterns

- ◆ To explore and choose among the various data mining software supporting different data mining techniques and functionalities that are more appropriate to the problem domain.
- ◆ To build and validate models using the selected tools and technologies.
- ◆ To report the result and forward recommendations for further studies.

1.4 Methodology Adopted

The techniques used by Two Crows Corporation (1999), which considers different data mining steps that need to be carried out to develop a decision support system, was applied for this research. Accordingly, the following methods were employed in conducting the study. That is how the target data set is going to be created, analyzed, and the model will be build and evaluation process will be done.

1.4.1 Literature Review

Since having a thorough understanding about a given subject and clearly notice its application is a critical issue to consider, literature review was made to get more insight in to the concept of data mining and its application. This also includes having multidimensional view in understanding the business problem at hand. Attempt has also been made to understand the data set under investigation through detail and intensive discussion with staff and management of the Traffic Office under study.

1.4.2 Identifying the Data Source: Data Collection

The major task at this stage was identifying and collecting the relevant dataset for the purpose of the research work at hand. Since the specific steps and tasks at this stage will determine the final result of the research work, much effort was made in creating the right dataset.

The data at the Addis Ababa Traffic Control Office were not combined and kept in a centrally organized manner. Hence it was necessary to collect data from different sources namely databases, flat files, and manual formats. The raw data to be used for this research work was

basically on various aspects of an accident, such as, road condition, weather condition, timing and other issues related to that specific accident. Apart from accident information, details of the car and the driver involved will also be considered.

1.4.3 Data Making/ Preparation

This was a step where the collected data was arranged into a form that will be suitable for the particular data-mining tool selected, KnowledgeSTUDIO. In other words, it is just to prepare data for analysis by collecting it to a new database. At this stage, pre-processing tasks like handling noisy data, unknown values, missing values, deriving new fields from the existing ones, and summarization of data was done by taking into account the selected decision tree model building techniques. This helps a great deal in effectively applying data mining tools and algorithms on the data.

1.4.4 Model Building/Data Analysis

Out of the major data mining technologies, decision tree classification modeling technique was used for this research work. Accordingly different algorithms and measures supported by the selected data mining tool have been examined by taking into consideration their application to the problem domain. Using such tools and techniques model building and training was the next important task done. Evaluation or assessing and interpreting of the results followed the model building task. While evaluating a model, its accuracy, the types of errors and the cost associated with it has been considered. One of the various approaches that employed to deal with such issues was confusion matrix. Finally, based on assessment and evaluation of such models, the best model that finds an interesting pattern was selected. And this provides a good ground for result interpretation and recommendation.

1.5 Scope and Limitation of the Study

The scope of the research is limited to assessing the possible application of data mining technology at Addis Ababa Traffic Office; it is limited to injury severities associated with road accidents to examine the potentials of data mining techniques in developing

classification model by the use of decision tree technique in support of traffic control activities.

The following are some of the limitations noticed during this research undertaking. The absence of relevant attributes especially on the early data, forces the researcher to focus on relatively recent data which has its own impact on the size of the data set. In addition to the above limitation, getting access to resources such as literature related to the study especially on Ethiopian case was another limitation.

1.6 Research Contribution

In this research an attempt was made in general to find out the applicability of data mining technology in the transport sector and in particular to examine the use of decision tree techniques in assessing injury severity. The result of the study shall be used as an input for the development of full-fledged data mining application using decision trees in determining the primary determinant factors of car accident and their severity nature.

Furthermore this research will contribute to the effort of improving the quality of life in general and reduce the risk of car accident and injuries in particular. It will also pave the way to develop better parameters in all aspects of traffic control system. Specifically it will support the Traffic Control Division of the region in taking proper action, like in revising the existing traffic rules, against vehicle accidents. Although the study was aimed at addressing traffic problems in particular, the output of the study may be used as a source of methodological approach for studies dealing with the application of data mining technology on similar problem areas.

1.7 Thesis Organization

This thesis report is organized under six chapters. The first chapter deals with the general overview of the study including background, statement of the problem, objectives and methodology of the research.

The second chapter is devoted to literature review of data mining technology. Available tools and techniques in the area are reviewed with the emphasis on the tools and techniques

employed in this specific research. Assessment of the application areas of data mining in general and in the road transport sector is also part of this chapter.

Under chapter three an attempt has been made to review literatures and trends in road transport, traffic system and road safety. Injury severity levels resulting from such road traffic accidents were also assessed.

The next section, chapter four, deals with different pre-processing steps. Tasks like data collection and cleaning, attribute selection, data formatting and transformation are reported in detail.

Chapter five reports the experiment of the research. The experiment basically comprises training; building and validation of the models in addition to analysis and interpretation of the results. The last chapter presents conclusions and recommendations of the research.

Chapter Two

Data Mining

In this chapter an attempt has been made to present a review of literatures on the basic concepts of data mining technology, tools and techniques available and the application area of data mining.

2.1 Overview

The convenience and easy availability of various technologies in generating collecting, manipulating and storing data led to the today's level of data explosion. Progress in database technology and other related fields in addition to the explosive growth in data collected from applications including business and management, government administration, science and engineering, and environmental control led to an increasing demand for efficient and effective data analysis and data understanding tools (Han and Kamber, 2001). And this in turn led to another scenario. Among others, it is the increased availability of huge amounts of data, and its complex nature for simple statistical and manual analysis that necessitate the use of data mining technology.

As to the historical overview of the emerging field data mining, it has some evolutionary steps. Data collection for the purpose of retrospective, static data delivery was the primary concern in early 1960s. Late in 1980s data access with the aim of retrospective, dynamic data delivery at record level became an issue. Dynamic data delivery at multiple levels was the main feature of data warehousing and decision support in 1990s. Now day's data mining with the intent of prospective and proactive information delivery is the current trend (Thearling, 2003).

Data mining generally is extraction of interesting patterns from a given set of data to facilitate decision making in a given business area. Data mining is a process that uses a variety of data analysis tools to discover patterns and relationships in data that may be used to make valid predictions (Two Crows Corporation, 1999). Moss (2003) on his part define data mining as the analysis of data with the intent to discover gems of hidden information in the vast quantity of data that has been captured in the normal course of running the business. Data mining is a

component of the knowledge discovery in database process concerned with the algorithmic means by which patterns are extracted and enumerated from data. (Fayyad, et al, 19996)

As indicted by Berry and Linoff (2000), the three core areas of competency that are needed to be successful in data mining are data mining techniques, data and the modeling of data. Data mining techniques refer to the conceptual approach to extract information from the data. Data refers to the collection of daily transaction or any sort of information kept from day to day activities. Coming up with a precise formulation of the problem to be solved and usage of the right data are the two important keys to success in data mining. Building models is only one step in knowledge discovery. It's vital to properly collect and prepare the data and to check the models against the real world (Two Crows Corporation, 1999).

Apart from problem understanding and having right data, selecting an appropriate data mining tools is another issue to be considered. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. This is because the past can tell a lot about the future. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining tools can answer business questions that traditionally were too time-consuming to resolve. They search databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations (Thearling, 2003).

2.2 Data Mining and the KDD Process

Data mining by it self is a process of finding understanding and interpreting interesting and useful knowledge. But still its part of the wider process called knowledge discovery in database. Formalized in 1989, knowledge discovery in Data base (KDD) has some basic sub-process or steps. According to Feyyad, et al (1996) KDD is the process of identifying a valid; potentially use full and ultimately understandable structure in data.

As a wider process it involves identifying the data sources, selecting the data, pre-processing it, transforming and/or reducing it, applying a data mining component to produce a structure, evaluating and interpreting the derived structure. While knowledge discovery often refers to the process of discovering useful knowledge from a data, data mining focuses on the application of algorithms for extracting patterns from data (Chung, et al, 1998)

Two Crows Corporation, (1999) also identified the basic steps of data mining for knowledge discovery as defining business problem, build data mining database, explore data, prepare data for modeling, build model, evaluate model , deploy model and results. A short description of these steps follows.

Define the business problem: Understanding the data and the business area is a prerequisite to knowledge discovery. Without having clear statement of the objective and understanding no algorithm will help.

Build a data mining database: The data to be mined should be collected in a database. But this does not mean that a database management system must be used. Depending on the amount of the data, the complexity of the data, and the uses to which it is to be put, a flat file or even a spreadsheet may be adequate. Tasks like data collection, data description, selection, data quality assessment and data cleaning, consolidation and integration are part of a process of building a data mining database. These data preparation steps may take much of the time and effort of the entire knowledge discovery process.

Explore the data: It is focused on identifying the most important fields in predicting an outcome, and determine which derived values may be useful. In a data set with hundreds or even thousands of columns, exploring the data can be as time consuming and labor-intensive as it is illuminating.

Prepare data for modeling: This is the final data preparation step before building models which consists of, selecting variables and rows, construct new variables and transforming variables.

Data mining model building: It is an iterative process employing a certain algorithm to explore alternative models in finding the one that is most useful in solving a business problem at hand. The “best” model is often found after building models of several different types, or by trying different technologies or algorithms.

Evaluation and interpretation: validating the model both in its accuracy and consistency on a real world data is a task that needs close attention in evaluating the results of the model and interpreting their significance.

Deploy the model and results. After model building and validation, the selected model can be used in one of two main ways. The first way is for an analyst to recommend actions based on

simply viewing the model and its results. The second way is to apply the model to different data sets.

Similarly with more or less related approach, after understanding the domain area, the process of data mining can also be summarized in to four major activities. In line with this various researches such as Fayyad, et al (1996), describe the process of knowledge discovery via data mining in to four basic activities:

- ◆ **Selection:** This involves creating the target dataset, since the data may represent a number of different aspects of the domain which may not be directly related.
- ◆ **Pre-Preprocessing:** This involves preparing the dataset for analysis by the data mining software to be used. This may include resolving undesirable data characteristics such as missing data (non-complete fields), irrelevant fields, and outlying data points.
- ◆ **Data mining:** Involves subjecting the cleaned data to analysis by the data mining software in an attempt to identify hidden trends or patterns, or to test specific hypotheses.
- ◆ **Interpretation:** Involves the analysis and interpretation of the results produced

2.3 Data Mining and Other Statistical Tools

So far statistical techniques have been used for centuries to understand various aspects of the world. Techniques like regression, sampling methodologies, and experimental designs are good examples of such kind. Statistics is one of the major disciplines that have contributed to data mining. It is still an important support to the field data mining (Berry and Linoff, 2000).

Although they seem to have some commonalties, data mining is different from conventional statistical analysis. Various scholars tried to demark clearly where these two fields differ.

“The general area of exploring massive data set is referred to as data mining. The field of data mining over laps many disciplines, primarily statistics, computer science, and information systems. Data storage and retrieval is generally considered to be an information systems issue. Algorithms useful for extracting information from database are of particular

interest in computer science. Statistics is a discipline that studies the best way to analyze the data and seek meaningful relationships among sets of variables. Existing data mining techniques are useful for computer assisted sifting through the complexities interwoven among large number of variables.” (Conerly et al, 2000)

As to the difference between statistical analysis and data mining Moss (2000) stated that statistical analysis uses only numerical data while data mining tools can use different types of data not just numerical data.

Similarly Whitten and Frank (2000) tried to show the difference and similarities of the two fields-statistics and machine learning (as a data mining technique). With regard to their differences, statistics has been more concerned with testing hypothesis, whereas machine learning has been concerned with formulating the process of generalization as a search through possible hypothesis. As to their similarities, very similar schemes have been developed in parallel, in both fields. Statisticians published a book on classification and regression trees, while almost at the same time a machine learning researcher was developing a system for inferring classification trees from examples. These two independent projects produced quite similar schemes for generating trees from examples. Another similarity between the two is that the use of nearest-neighbor method for classification. These are standard statistical techniques that have been extensively adopted by machine learning researchers.

Thus data mining does not replace traditional statistical technique; it is rather an extension of statistical methods which is the result of major changes in the statistic community.

2.4 Data Warehousing and OLAP Technology for Data Mining

Organizations keep the data in a different form of data base basically for two purposes. One is to process large number of transaction very fast and the other is to support decision. As stated by Berry and Linoff (2000), decision support system contrast with the online transaction processing systems (OLTP).

“Decision support is a broad term for the entire information technology infrastructure that companies and other organization use to make informed decisions” (Berry and Linoff, 2000).

An important feature of a Decision Support System is basically the general focus area that they are designed for. In other words it is not common to look at individual instances of the database instead are designed to handle and support more general and complex queries.

A special case of a decision support database, according to Berry and Linoff (2000) is the data warehouse. A data warehouse is a large decision support database fed by many operational systems. Continuing his discussion he stated that, data warehouse is motivated by the need to view the entire enterprise from a single point of view instead of as a collation of narrowly defined “silos”. Similarly to help put the flood of data into a format that can be used, more and more data is being moved into data warehouse whose purpose is decision support. Data warehouses helps by having a common format and consistent definition for fields (Grove, 1999). But a data warehouse is only as good as the business objective that led to its creation (Mangold, 1998).

Data marts designed for on line analytical processing (OLAP) is also another common type of decision support database. Database management systems (DBMS) also provide good ground to data mining as the broader purpose of data mining is to discover knowledge from large database. The data about data (Meta data) found in the databases is also valuable for data mining as the nature of the attributes can be easily understood (Berry and Linoff, 2000)

The use of OLAP tools for the interactive analysis of multidimensional data of different parts, which facilitates effective data mining, is provided by data warehouses. In addition to this, many other data mining functionalities, like classification, prediction, association, and clustering, can be integrated with OLAP operation to enhance interactive mining of knowledge at multiple levels of abstraction (Han and Kamber, 2001).

2.5. Data Mining Technologies

In whatever way described, data mining tasks can be categorized into either prediction or description. In line with this, two styles of data mining are identified by Berry and Linoff, (2000). Directed data mining, which is a top-down approach used when the ultimate result is

expected. It is often also called predictive modeling. Undirected data mining, basically used to identify patterns in a given data and let the user to determine the interestingness of it. It is normally a bottom-up approach of describing the data.

Thus various literatures agree to classify data mining technology in to either predictive or descriptive modeling. These two basic techniques basically differ in their approach in solving problems of various natures. There is usually a particular type of problem to be solved by each of the applications. This is to say that there is always a specific algorithm that will be used for a problem posed by people who are going to mine the data.

Different problem types in a data mining process to ultimately solve the business problems are often also called data mining tasks. The most widely used tasks are classification, association, clustering, dependency analysis, prediction, segmentation, description (Whitten and Frank, 2000).

2.5.1. Predictive Modeling

The goal of a predictive modeling is to predict the value of one column based on the value of other columns. It refers to the process of building a model that will permits the value of one variable to be predicted from the known values of other variables (Hand, 2005). Thus it is also called, supervised modeling as a task. In supervised modeling, there is a special attribute called the “label” that you intend to predict. By encoding the relation between the label and other attributes, the model can make predictions about new data. In addition to this, by visualizing the model itself, we can gain insight into the relationship between labels and other attributes (Gerritsen, 1999).

Classification and Regression are the two most common tasks in predictive modeling. If the label is discrete (containing a fixed set of values), the task is called classification. If the label is a continuous value, the task is called regression.

Classification is the task of assigning a discrete label value to an unlabeled record. In doing so, records are divided into predefined groups. It is the process of finding a set of models that describe and distinguish data classes or concepts, for the purpose of being able to use the model to predict the class of objects whose class label is unknown (Han and Kamber, 2001). For example, as cited by Chong et al (2004) Sohn et al (2003) applied data fusion, ensemble

and clustering to improve the accuracy of individual classifiers for two categories of severity (bodily injury and property damage) of road traffic accidents. Classifiers can also predict the probability that the label will take on a specific value. For example, the probability that an accident will result in bodily damage can be seen. Classification as a task frequently employs decision tree or neural network-based algorithms (Moxon, 1996).

Regression on the other hand used to deal with non-discrete that means continuous variable. Regression is similar to classification, except that the label is not discrete. For example, predicting salary or the price of stock is a regression, whereas predicting whether the salary is in a given range or whether a stock will go up or down is a classification task. It uses existing values to forecast what other values will be. Although at a simplest senses regression uses standard statistical techniques such as linear regression, because of the complex nature of the real world problems, more complex techniques like decision tree and neural networks may be necessary to forecast future values (Two Crows Corporation, 1999).

2.5.2. Descriptive Modeling

Descriptive modeling on the other hand deals with describing all of the data and discovering patterns and segments of the data (Hand, 2005). Thus it is also called unsupervised technique. Being unsupervised task it helps us to see patterns and segments that behave similarly. The two most common descriptive modeling tasks are association and clustering.

Association refers to finding the association of two items in certain degree. Thus associations are used to find affinity groupings that discover what patterns are usually go together. It addresses a class of problems typified by a market-basket analysis (Moxon, 1996). The classic affinity grouping is market basket analysis, predicting the frequency with which certain items are purchased together.

Clustering on the other hand, analyze and segment the data into groups of records or clusters that have some similar characteristics. Thus it address segmentation problem. This approach assigns records with a large number of attributes in to a relatively small set of groups or segments. This assignment process is performed automatically by clustering algorithms that identify the distinguishing characteristics of the dataset and then partition the dimensional space defined by the dataset attributes along natural cleaving boundaries (Moxon, 1996). The best clustering should exhibit higher inter-class dissimilarity and intra-class similarity.

2.6. Data Mining Techniques

Data mining may be performed using a variety of techniques. That means in identifying interesting patterns, be it predictive or descriptive, employing some sort of technique is a must. And the type of technique to be used is determined by the goal of the data-mining task. This is because no data mining technique is right for all situations. As stated by Berson (1999), various data mining techniques exist today. And these methods can be broadly categorized in to two groups, classical techniques which include statistics, neighborhoods and clustering and next generation techniques consisting of trees, networks and rules.

In connection with this, Thearling (2003), identify six namely, Artificial Neural Networks, Decision Trees, Genetic algorithms, Nearest Neighbor method and rule induction as the most commonly used data mining techniques. Berry and Linoff (2000), on their part, identifies the three namely, Artificial Neural Networks, Decision Trees and Automatic Cluster Detection techniques implemented in a large number of commercial software applications and cover a wide range of data mining situations. Data mining is not a technology that can be applied without discretion to solve multitude of problems. It requires a thoughtful application of the right technique to the right data. As discussed by Coppock (2003) the selection of data mining techniques to be used on a given data set and the identification of most useful data patterns examined by the data mining algorithms are important points to be considered to get maximum out of the data mining technology.

2.6.1 Decision Tree

Being one of the most commonly used data mining technique, it is a technique in which tree-shaped structures that represent sets of decisions can be constructed to generate rules for the classification of a dataset. A decision tree is described as a way of representing a series of rules that lead to a class or value. It is a flow-chart like tree structure, where each internal node denotes a test on attribute, each branch represents an outcome of the test, and leaf nodes represent classes or class distribution (Han and Kamber, 2001).

A decision tree is composed of hierarchically arranged nodes, growing from the top most nodes called root node to leaf node. So it can be thought as the tree growing upside down,

splitting the data at each level to form new nodes. The resulting tree comprises of many nodes connected by branches. Nodes that are at the end of branches are called leaf nodes and play a special role when the tree is used for prediction. That means each node in the tree specifies a test of some attribute of the instance, and each branch descending from that node corresponds to one of the possible values for this specific attribute.

Through examining the data and inducing the tree and its rules decision tree models are used to make predictions. In doing so, there are various algorithms that are supported by different data mining tools and being used for decision tree construction like CHAID (Chi-squared Automatic Interaction Detection), C4.5/C5.0 , CART (classification and Regression) and many others which are less familiar (Berry and Linoff, 2000).

As stated by Two Crows Corporation (1999), there are two main types of decision trees: classification trees, and regression trees. Classification trees are used to predict categorical variables that label records and assign them to the proper class that can also provide the confidence that the classification is correct. Regression trees on the other hand are used to predict continuous variables that estimate the value of a target variable that takes on numeric values.

Decision tree learning has therefore been applied to problems such as learning to classify medical patients by their disease, equipment malfunctions by their cause and loan applicants by their likelihood of defaulting on payments (Brand and Gerritsen, 1998). They are also said to be a good choice when the data-mining task is classification of records or prediction of outcomes. The training process that creates the decision tree is usually called induction.

Constructing decision trees has basically two phases: a tree-growing (splitting) phase followed by a pruning phase (Brand and Gerritsen, 1998). The tree-growing phase is an iterative process, which involves splitting the data into progressively smaller subsets. The importance of each attribute will be tested to be used as a splitting node. Such a measure is referred to as an attribute selection measure or a measure of the goodness of split. For example an attribute with the highest information gain (or greatest entropy reduction) is chosen as the test attribute for the current node.

After growing the tree using the right split node exploring the tree is necessary to find out any nodes or sub-trees that are undesirable because of overfitting, or may contain rules that the domain expert feels are inappropriate. Overfitting is a situation when a tree that precisely

describes the data from which it was derived is unlikely to generalize well to another sample drawn from the same population. Thus pruning is a common technique used to make a tree more manageable and usable having the necessary features. Generally, it is an approach to solve overfitting problems by allowing the initial decision tree to grow quite deep and then find ways to prune off the branches that fail to generalize. Pruning could be either pre or post. Pre-pruning is when a tree is “pruned” by halting its construction early. Upon halting the nodes become leafs. Post-pruning removes branches from a “fully grown” tree. Pre-pruning and post-pruning may be interleaved for a combined approach. Post-pruning requires more computation than pre-pruning, yet generally leads to a more reliable tree. Most decision tree algorithms stop growing the tree when either the segment contains only one record, all the records in the segment have identical characteristics or the improvement is not substantial enough to warrant making the split (Brand and Gerritsen, 1998).

2.6.2 Artificial Neural Networks

They are basically non-linear predictive models that learn through training and resemble biological neural network in structure. Although, artificial neurons are much simpler than the biological neurons the basic unit of neural networks, the artificial neurons, simulates the basic functions of the natural neurons.

It is the most widely known and the least understood of the major data mining techniques. Neural network attempts to simulate within specialized hardware and sophisticated software, the multiple layers of simple processing elements called neurons. Each neuron is linked to certain of its neighbors with varying coefficients of its connectivity that represent the strengths of these connections. Learning is accomplished by adjusting these strengths to cause the overall network to output appropriate results (Klerfors, 1998).

As to the architecture of neural networks it could be either Feed-forward or Feedback networks. Feed forward artificial neural networks allow signals to travel one way only; from input to output. There is no feedback that means, the output of any layer does not affect the same layer. Feed forward neural networks tend to be straight forward networks that associate inputs with outputs. They are extensively used in pattern recognition. They work with the backpropagation algorithm. This model generalizes the network of perceptron for the architecture with hidden layers, so called multilayered perceptron (Sima, 1998)

Feedback networks on the other hand can have signals traveling in both directions by introducing loops in the network. Feedback networks are very powerful and can get extremely complicated. They are dynamic; their state is changing continuously until they reach all equilibrium point. They remain at the equilibrium point until the input changes and a new equilibrium needs to be found (Stergion and Siganos, 1996).

A neural network may have different layers. It starts with an input layer, where each node corresponds to a particular variable. These input nodes are connected to a number of nodes in a hidden layer. Each input node is connected to every node in the hidden layer. The nodes in the hidden layer may be connected to nodes in another hidden layer, or to an output layer. The output layer consists of one or more response variables. After the input layer, each node takes in a set of inputs, multiplies them by a connection weight, adds them together, applies a function (called activation function) to them, and passes the output to the node(s) in the next layer.

The success of a neural network will basically depend on its training. Training a neural network is the process of setting the weights on the inputs of each of the units in such a way that the network best approximates the underlying function, or does the job of predicting the target variable. Neural networks are sometimes called machine learning algorithms, because changing of its connection weights causes the network to learn the solution to a problem. This resembles them with the learning process of the brain. The strength of connection between the neurons are stored as a weight-value for the specific connection. The system learns new knowledge by adjusting these connection weights (Klerfors, 1998)

Most software packages for building and training neural network models use some variation of the algorithm known as backpropagation algorithm which has basically three steps namely calculating an output for a training instance using the existing weights in the network, calculating the error by taking the difference between the calculated result and the expected result and using the error adjust the weights (Stergion and Siganos, 1996).

As stated by Lloyd-williams (1997), the main techniques employed by neural networks are known as supervised learning and unsupervised learning. Supervised learning is essentially a two stage process; firstly training the neural network to recognize different classes of data by exposing it to a series of examples, and secondly, testing how well it has learned from these examples by supplying it with a previously unseen set of data. On the other hand, no initial information regarding the correct classification of the data it is presented with is required in

the case of unsupervised learning. Thus neural networks using this technique are able to identify their classification schemes based on structure of the data provided, thus reducing its dimensionality.

As to its application, neural networks are good choices for most classification and prediction tasks when the results of the model are more important than understanding how the model works. Neural Networks actually represent complex mathematical equations with lots of summation, exponential functions, and many parameters. Neural networks do not work well if there are many hundreds or thousands of input features. Large numbers of features make it more difficult for the network to find patterns and can result in long training phases that never converge to a good solution (Berry and Linoff, 2000).

Neural networks have broad applicability to real world business problems. They have already been successfully applied in many industries. Since neural networks are best at identifying patterns or trends in data, they are well suited for prediction or forecasting needs including; sales forecasting, customer research, data validation, risk management, and target marketing (Stergion and Siganos, 1996).

2.6.3 Other Data Mining Techniques

A genetic algorithm is an optimization technique that use processes such as genetic combination, mutation, and natural selection in a design based on the concepts of evolution. It employs the selection, crossover, and mutation operators to evolve successive generation of solutions (Berry and Linoff, 2000). Another technique that classifies each instance in a database based on the classes of the k records which are most similar to it in a historical dataset is referred to as nearest neighbor method (Thearling, 2003). On the basis of the number of records (k) it is often known as k -nearest neighbors. It is an instance-based learning in which a distance function is used to determine the class of a new instance (Witten and Frank, 2000). Rule induction, which is basically an extraction of useful if-then rules from database based on statistical significance, is also another technique possibly applicable in some domain areas (Thearling, 2003).

2.7. Application of Data Mining Technologies

Under this section application areas of data mining technology will be presented. This will help to better understand the scope or delimitation of the field which will in turn facilitate its effective implementation.

2.7.1. General Applications

Technological advancement in data generation, collection, organization and storage necessitate a more advanced way of extracting meaning full information. One of such advanced technologies in identifying patterns from cumulated data is data mining. Its successful application in wider areas make data mining very popular and many organizations are benefiting from it. Various literatures showed us that data mining is applicable in areas like business data analysis and decision support, fraud detection, text mining and web mining. But successful application of data mining calls for large well organized and integrated data in addition to enough understanding of the domain knowledge where data mining is to be applied.

Many organizations are using data mining to help manage their specific tasks. Identifying customers who are most likely interested in a new credit product is one instance of data mining application. Detecting fraudulent use of services using data mining technologies is also another area that interests a credit card company (Thearling, 2003). The application of data mining in biomedical and DNA data analysis is also discussed by Han and Kamber (2001). In line with this discovering large-scale sequencing pattern, in identifying and study human gene for the development of new pharmaceuticals and cancer in cancer therapies is a recent focus of data mining.

Data mining offers value across a broad spectrum of industries, telecommunications and credit card companies to detect fraudulent use of their services, insurance companies and stock exchanges to reduce fraud, medical applications to predict the effectiveness of surgical procedures, medical tests or medications. Companies active in the financial markets use also data mining to determine market and industry characteristics as well as to predict individual company and stock performance. Retailers are making more use of data mining to decide which products to stock in particular stores (and even how to place them within a store), as well as to assess the effectiveness of promotions and coupons. Pharmaceutical firms are mining large databases of chemical compounds and of genetic material to discover substances

that might be candidates for development as agents for the treatments of disease (Two Crows Corporation, 1999).

According to Hong (2000), data mining in general and predictive modeling in particular, which is perhaps the most-used subfield of data mining, has got great attention in insurance, fraud detection and text categorization. In addition, the power of data mining is being used for many other purposes, such as analyzing Supreme Court decisions, discovering pattern in health care, pulling stories about competitors from newswires, resolving bottlenecks in production process, and analyzing sequences in human genetic makeup. There really is no limit to the type of business or area of study where data mining can be beneficial (Tracey Dishman, 2000).

Similarly Mangold, (1998) stated that data mining has proves itself immensely useful in a number of fields, including detecting auto insurance fraud, diagnosing failures in manufacturing processes, market- basket analysis, market segmentation and assessing risk patterns in consumer credit loans. Health specialist, doctors, may also use the discovery of unrecognized patterns in their database to predict which patients are good candidates for a surgical procedure disease (Grove, 1999).

As indicated above a number of researches have been done to show the possible applications of data mining techniques in many areas and organizations in different parts of the world. This also holds true in Ethiopia. With this regard, a number of researches have also been conducted at Addis Ababa University, Faculty of Informatics. Just to mention a few of them, Henok (2002), Tesfaye (2002) and Shegaw (2002) have tried to show its application in different areas in Ethiopian context. More over Leul (2003) has also assessed the potential applicability of data mining technology in crime prevention. It is the researcher's belief that data mining techniques are also applicable to the selected traffic control system, where very large data is accumulated, in assessing vehicle accident patterns.

Thus from the above discussion it is apparent that the application areas of data mining as a field of study is growing time to time. And it is getting more and more acceptance in areas where the need for important decision is crucial and data is available.

2.7.2. Application of Data mining in Road Transport and Accident analysis

As it is apparent from the previous discussion, the success of data mining application lies basically on the availability of enough amounts of data that can speak about itself. In addition, identifying, interesting pattern in a given set of data is a key in decision making. With respect to transport in general, and road transport in particular, data of various kinds are recorded every day. Road traffic accident data is one, among others, which is recorded and kept about individual accidents on the public highway.

A data mining application always starts with a problem. We can consider the following cases 'Sadly, there are traffic accidents. On some places there are more than others. What makes part of a road dangerous?' (Siebes, 2000)

In view of Chong et al (2004), the application of data mining in the area of road transport can be classified in to three basic categories namely;

- ◆ Traffic density analysis; measurement and investigation of traffic accident volumes.
- ◆ Traffic accident analysis; identifying determinant factors in road traffic accidents and other related issues
- ◆ Injury severity analysis; modeling and predicting the severity of injuries resulting from traffic accidents.

With this regard, various researchers try to employ different data mining techniques in their effort of traffic density, accident and injury severity analysis. In line with this, as cited by Chong, et al (2004), Bedard, et al (2002), studied the independent contribution of driver, crash and vehicle characteristics to driver's fatality risk. And it was found that the increase use of seatbelt and reducing speed might prevent fatalities. Similarly, Shankar, et al (1996) on his part tried to study the accident severity likelihood conditioned on the occurrence of accident. It was found that there is a greater probability of evident injury or disabling injury (fatality) relative to no evident injury if at least one driver did not use a restraint.

2.7.2. Application of Data mining in Road Transport and Accident analysis

As it is apparent from the previous discussion, the success of data mining application lies basically on the availability of enough amounts of data that can speak about itself. In addition, identifying, interesting pattern in a given set of data is a key in decision making. With respect to transport in general, and road transport in particular, data of various kinds are recorded every day. Road traffic accident data is one, among others, which is recorded and kept about individual accidents on the public highway.

A data mining application always starts with a problem. We can consider the following cases ‘Sadly, there are traffic accidents. On some places there are more than others. What makes part of a road dangerous?’ (Siebes, 2000)

In view of Chong et al (2004), the application of data mining in the area of road transport can be classified in to three basic categories namely;

- ◆ Traffic density analysis; measurement and investigation of traffic accident volumes.
- ◆ Traffic accident analysis; identifying determinant factors in road traffic accidents and other related issues
- ◆ Injury severity analysis; modeling and predicting the severity of injuries resulting from traffic accidents.

With this regard, various researchers try to employ different data mining techniques in their effort of traffic density, accident and injury severity analysis. In line with this, as cited by Chong, et al (2004), Bedard, et al (2002), studied the independent contribution of driver, crash and vehicle characteristics to driver’s fatality risk. And it was found that the increase use of seatbelt and reducing speed might prevent fatalities. Similarly, Shankar, et al (1996) on his part tried to study the accident severity likelihood conditioned on the occurrence of accident. It was found that there is a greater probability of evident injury or disabling injury (fatality) relative to no evident injury if at least one driver did not use a restraint.

Yang et al (1999) also in his part used neural network approach to detect safe driving patterns that have less chance of causing death and injury when a car crash occurs. The research indicates that by controlling a single variable (such as the driving speed, or the light conditions) they potentially could reduce fatalities and injuries by up to 40%.

Conerly, et al (2000) examines the Alabama Accident Database using different data mining and visualization modeling techniques. The aim of the research was to find useful information that might lead to improved highway safety. Particularly they tried to identify important determinants of injury and fatalities in traffic accidents using multivariate data exploration tools like classification and regression trees (CART) and mosaic plots supported by SAS software package. Unfortunately due to numerous data entry errors and variable definition problems in the Alabama Accident Databases, the research founded out that it was not feasible to build a model for the prediction of injuries and fatalities.

It is apparent that the result of such researches may vary from time to time and /or place to place depending on the socio-economic and infrastructure facilities available at that specific situation. As to the knowledge of the researcher no research has conducted so far in Ethiopia employing data mining technology in support of the decision making in this area.

So far in this chapter the basic tools and techniques of data mining and its application area were discussed. And it revealed that theses data mining tools and techniques can also be employed in the specific domain area of this research, road traffic system in Ethiopian context. The application of data mining in such area in Ethiopian case can also be in the analysis of traffic density, accident factors and injury severity, from which the researcher picked up the road traffic accident severity analysis. This will be the focus of the subsequent chapters.

Yang et al (1999) also in his part used neural network approach to detect safe driving patterns that have less chance of causing death and injury when a car crash occurs. The research indicates that by controlling a single variable (such as the driving speed, or the light conditions) they potentially could reduce fatalities and injuries by up to 40%.

Conerly, et al (2000) examines the Alabama Accident Database using different data mining and visualization modeling techniques. The aim of the research was to find useful information that might lead to improved highway safety. Particularly they tried to identify important determinants of injury and fatalities in traffic accidents using multivariate data exploration tools like classification and regression trees (CART) and mosaic plots supported by SAS software package. Unfortunately due to numerous data entry errors and variable definition problems in the Alabama Accident Databases, the research founded out that it was not feasible to build a model for the prediction of injuries and fatalities.

It is apparent that the result of such researches may vary from time to time and /or place to place depending on the socio-economic and infrastructure facilities available at that specific situation. As to the knowledge of the researcher no research has conducted so far in Ethiopia employing data mining technology in support of the decision making in this area.

So far in this chapter the basic tools and techniques of data mining and its application area were discussed. And it revealed that theses data mining tools and techniques can also be employed in the specific domain area of this research, road traffic system in Ethiopian context. The application of data mining in such area in Ethiopian case can also be in the analysis of traffic density, accident factors and injury severity, from which the researcher picked up the road traffic accident severity analysis. This will be the focus of the subsequent chapters.

Chapter Three

Road Traffic Accident and Injury Analysis

Under this chapter attempt has been made to present review of literatures and trends in road transport, traffic system, road safety and accident injury severity. This will help a great deal in understanding the domain area where data mining technology is going to be used.

3.1 Introduction

In the history of mankind human beings use transport of different kinds for a variety of reasons. Citizens use transport to travel for their essential tasks like work, schooling etc and/or for optional purpose like pursuing leisure interests. Transport is an important sector for facilitating different socio-economic activities in the national economy. In early days, horses, mules and carts were used as a means of transport. But with technological development, different kinds of modern transportation; road, air, rail and marine, came into existence.

Along with the development of modern transportation system, traffic accidents are becoming point of attention. And of the different modern transportation system, road transport which involves peoples with different types of vehicles is the most widely used and the most dangerous.

Of all the systems that people have to deal with on a daily basis, road transport is the most complex and the most dangerous. World wide, the number of people killed in road traffic crashes each year is estimated at almost 1.2 million, while the number of injured could be as high as 50 million- the combined population of five of the world's large cities. The tragedy behind these figures regularly attracts less media attention than other, less frequent but more unusual type of tragedy (WHO, 2004)

The term 'traffic' literally refers to any movement of people and vehicles along with roads and streets, or aircrafts in the sky (Addis Ababa Transport Authority, 2004). In general, the movement of people along with transportation means is referred to as traffic (Meketew, 2000). For the purpose of this research, the term traffic is used to refer to the movements by road or highway users; people, vehicles and animals along with the road.

3.2 Road Traffic Accidents

In perusing day to day activities human beings face various types of accidents. Among others, road traffic accidents also called vehicle accident cause considerable damage to both human life and property. A road traffic accident is defined as any vehicle accident occurring on a public highway. It includes collisions between vehicles and animals, vehicles and pedestrians, or vehicles and fixed obstacles. Single vehicle accidents, which involve a single vehicle, that means without other road user, are also included (Safecarguide, 2000).

At all levels, that is nationally or internationally road traffic accidents continue to be a growing problem. In connection with this, according to a World Health organization/World Bank Report, deaths from non-communicable diseases are expected to grow from 28.1 million a year in 1990 to 49.7 million by 2020, which is an increase in absolute number of 77%. Traffic accidents are the main cause of this rise. Road traffic injures are expected to take higher place in the rank order of disease burden in the near future (Safecarguide, 2004). The magnitude of the situation can be seen from the following figure 2.1 projected by World Health Organization.

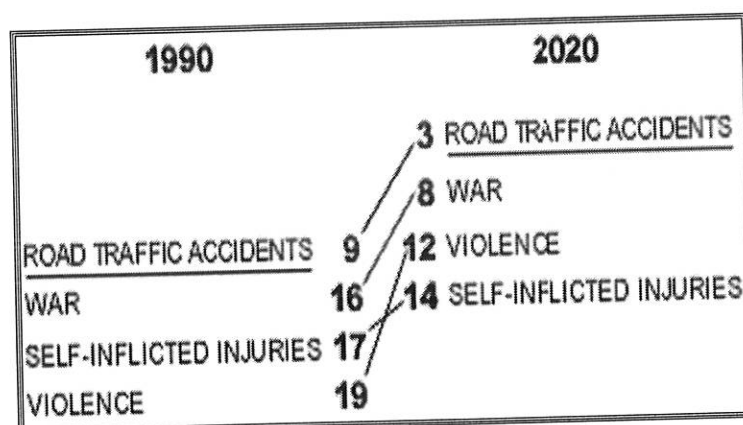


Figure 2.1: Projected change in the ranking of the 15 leading causes of death and disease

It is easy to learn from the figure that, road traffic accidents, the 9th leading cause of death and disease in 1990, will be the third most dangerous cause of death and disease in 2020. Again according to the WHO, the leading injury-related cause of death among productive age group is traffic injuries. Of the 5.8 million people who died of injuries in 1998, 1,170,694 died as a direct result of injuries sustained in a motor vehicle accident. World wide the WHO

reports that 38,848,625 injuries were received by people involved in motor vehicle accidents in 1998.

The tragedy is more or less similar in Ethiopia, Addis Ababa. The rate of traffic accidents in Addis Ababa goes up together with the increase of motor vehicles and population size. The rise in automobile ownership together with the poor condition of the roads has resulted in the high level of traffic safety and congestion problems.

At a country level, above 1,800 people died while above 7,000 were crippled or injured in 2003. Moreover the death rate is 136 per 10,000 vehicles and Ethiopia is losing over 400 million birr yearly as a result of road traffic accidents (Abebe, 2004). The share of Addis Ababa city in the total number of accidents was 60 percent in 1989 with annual average traffic accident growth 31.4 percent (Addis Ababa Transport Authority, AATA, 2004). Nowadays Addis Ababa is experiencing around 700 accidents per month resulting in various level of injury severity

As cited by Meketew (2000), Institution of Highways and Transportations (IHT) attributes road traffic accidents to numerous factors such as weather, light conditions, faulty design and /or inadequate maintenance of the road infrastructure development, as well as some vehicles with mechanical defects and human error.

As to the cause of road traffic accidents in Ethiopia, the first four leading causes as identified by the Road Transport Authority, RTA (2004) are, not respecting speed limit, driver characteristics, not giving priority for pedestrian, and vehicle defects. Especially with respect to the vehicle defect although there is an annual program for technical investigation of vehicles, it is not enough when compared to the magnitude of the problem. And consequently conducting occasional technical investigation have got due attention now days.

In connection with this the contribution of respecting speed limit in reducing vehicle accident is immense. Different rules and regulations regarding the speed limits of a given vehicle were set up so far in Ethiopia. But due to low level of awareness of drivers and lack of enough control, it has been one of the few important factors for accident (RTA, 2004).

And it is crystal true that loss of human life and various level injuries along with the extensive property damage creates huge burden on the effort of socio-economic development of a given nation. Obviously the situation is more severe in developing countries.

3.3 Road safety and Accident Injuries Severity

As it can be seen from various successive reports of WHO and Traffic Control Office at Addis Ababa, without increased efforts and new initiatives, the total number of road traffic deaths and injuries at all levels are growing at an alarming rate. It should get due attention and scale of resources that is currently paid to other health issues, if the problem with its devastating human impact and large economic cost to society are to be averted.

In road traffic, severity is normally a function of three sets of elements. The first set is factors to be considered before the crash, pre-crash preventions. The second set of elements is injury prevention during the crash and the last one is post crash activities of life sustaining. As cited by the WHO, Rothe (2002), propose a set of activities in a form of strategy which includes interventions to reduce exposure to risk; to prevent road traffic crashes from occurring; to reduce the severity of injury in the event of a crash and to reduce the consequence of injury through improved post-collision care among others. Some of the important points of attentions in the new strategy in understanding and dealing with road safety is summarized as follows:

- ◆ Crash injury is largely predictable preventable. It is a problem amenable to rational analysis and remedy.
- ◆ Road safety policy must be based on a sound analysis and interpretation of data, rather than on anecdote.
- ◆ Road safety is a public health issue that intimately involves a range of sectors including that of health. All have their responsibilities and all need to be fully engaged in injury prevention.
- ◆ Since human error in complex traffic systems cannot be eliminated entirely, environmental solutions (including the design of roads and of vehicles) must help in making road traffic systems safer.
- ◆ The vulnerability of the human body should be a limiting design factor for traffic systems, i.e. for vehicle and road design, and for setting speed limits (WHO, 2004).

Thus it is crystal true that the formidable challenge of reducing the level of human loss on the roads requires good deal of cooperation among different sectors increased commitment and capacity for research, policy making and creating conducive environment. Having said this, it is clear that not all accidents have equal consequences. As reported by WHO injuries sustained by victims of a road traffic crash vary in type and severity. While some accidents resulted in fatal injuries, serious or slight others end with non-injury damages. Serious injuries refer to any one or combination of problems like fracture, concussion, internal crushing, severe cut and laceration, severe general shock requiring medical treatment. Slight injuries on the other hand refer to secondary injuries such as sprain, bruises and cuts not judged to be severe.

“The ability of predicting a fatal and non-fatal injury is very important since driver’s fatality has the highest cost to society economically and socially” (Chong, et al, 2004).

Chong et al (2002) tries to model the severity of injury resulting from traffic accidents. They have applied artificial neural networks and decision trees on an actual data set obtained from the National Automotive Sampling System (NASS) General Estimate System (GES). They finally came up with the result of identifying the three most important factors in fatal injury as driver’s seat belt usage, light condition of the road way and driver’s alcohol usage. The result of the research revealed also that in all the cases the decision tree out performs the neural network.

Indicating the focus of previous researches as a binary classification of accident results into non-injury and injury including fatality, Chong, et al (2004), in their Traffic Accident Analysis extend their research to include new categories: passable injury, non-incapacitating injury, incapacitating injury and fatal injury. They also indicated that, the light condition of the road, driver’s alcohol usage and driver’s seat belt usage were most important factors in fatal injury.

Hallberg (2004) used time dependent modeling in order to analyze traffic accident statistics in Nordic countries. The objective of the research was to predict traffic accident levels based on an estimate of the future car population. The study revealed that measure reductions in accident rates relate to the implementation of major traffic safety. In addition to this the study also indicated that the net increase of the car population is a main factor to consider when predicting the number of people that will be killed in future traffic.

3.4 Road Traffic Accident at Addis Ababa

3.4.1 Overview

In developing African nations, an increasingly large proportion of financial resources are being diverted to the importation of vehicles and new road construction. The health impact of the transition has not been well documented in Africa and motor vehicle injuries have not figured prominently in the past as a priority public health problem (Taddele and Larson, 1991).

Apart from being serious health problem, road traffic accidents also create various socio-economic problems specifically on developing nations like Ethiopia. Especially if these accidents are accompanied by a greater volume of traffic system the problem associated with them will be more serious. In connection to this, volume and composition of traffic on the Ethiopian roads was reviewed by Ethiopian Road Authority, ERA (2002). The study provides analysis of traffic information of the year which is a primary input to the maintenance, construction, improvement, planning and management needs of Ethiopian Road Authority. The report revealed that the need for better traffic system management and maintenance capacity strengthening is evident from simple consideration of volume and composition of vehicles on the road network of the country.

The concept of traffic control system in Addis Ababa is dated back around 1900 E.C with the introduction of motor vehicles. As the case is in any other areas, regulations and laws also govern road traffic. In connection with this, the first cited traffic control rule having 18 articles was in action in 1918 E.C. This rule was aimed at facilitating the traffic system, which involves movement of pedestrians, animals and motor vehicles. It was in 1935 E.C that formal and legal driving license become in action and the present driving license issuance rule is formally implemented since 1960 E.C. With respect to comprehensiveness, it is the Transport Act numbered 361 /1961 E.C, which is more comprehensive in having most of the major regulations commonly used. (Addis Transport, 2004). The road traffic safety regulation number 5/1998 and number 4/2004 are also the recent rules and regulation in use at the City of Addis Ababa.

3.4.2 Accident Data Analysis at Addis Ababa Traffic Office

As cited by Mekitew (2000) IHT (1990) lists the basic data items that should be included in an accident reporting. They are basic accident description, road type, environmental features, driver features, causality details and traffic characteristics related to time and location.

Similarly, according to the discussion with the traffic officers, upon receiving a notification about an accident, Addis Ababa Traffic Police Department assigns an investigator to collect the necessary details about a given accident. Notifications are normally reported by the drivers or any party being involved or have interest on it because the law requires doing so. On site investigation and recording is done with the aim of finding detailed and accurate information as to its cause, determine whether or not there has been violation of the law and ultimately to prevent the re-occurrence of further accidents. But some times as reported by the officers, due to time gap between the accident and the arrival of traffic officers, some details like the severity level and cause of an accident may not be identified effectively.

This accident record is basically used for various purposes in the office and for other stakeholder. National and regional transport offices use the data in directing their focus of attention in decision and policymaking's with regard to road safety. Different health offices and non-governmental organizations working in this area use the data in determining and managing health problem in society.

However as observed by the researcher, with absence of efficient and effective data management system and sufficient research in the area the effort of using this data for different problem solving is not that much fruitful.

Apart from keying some basic attributes in an Excel flat file form, at the office level what actually the staff in a statistics section of the office does is classifying accident records by accident time and date, age and sex of drivers, education level and experience of the driver, road and light conditions and by possible cause of accident summary of reports.

Recognizing the complexity and magnitude of the vehicle accident problem, developed nations take necessary measures like putting resource in studying the various aspects of these accidents and use the result in decisions making. But this is not the case in developing countries like Ethiopia. In connection with this, safer no comprehensive research work has been done specifically with respect to road accident severity in Addis Ababa.

3.4.2 Accident Data Analysis at Addis Ababa Traffic Office

As cited by Mekitew (2000) IHT (1990) lists the basic data items that should be included in an accident reporting. They are basic accident description, road type, environmental features, driver features, causality details and traffic characteristics related to time and location.

Similarly, according to the discussion with the traffic officers, upon receiving a notification about an accident, Addis Ababa Traffic Police Department assigns an investigator to collect the necessary details about a given accident. Notifications are normally reported by the drivers or any party being involved or have interest on it because the law requires doing so. On site investigation and recording is done with the aim of finding detailed and accurate information as to its cause, determine whether or not there has been violation of the law and ultimately to prevent the re-occurrence of further accidents. But some times as reported by the officers, due to time gap between the accident and the arrival of traffic officers, some details like the severity level and cause of an accident may not be identified effectively.

This accident record is basically used for various purposes in the office and for other stakeholder. National and regional transport offices use the data in directing their focus of attention in decision and policymaking's with regard to road safety. Different health offices and non-governmental organizations working in this area use the data in determining and managing health problem in society.

However as observed by the researcher, with absence of efficient and effective data management system and sufficient research in the area the effort of using this data for different problem solving is not that much fruitful.

Apart from keying some basic attributes in an Excel flat file form, at the office level what actually the staff in a statistics section of the office dose is classifying accident records by accident time and date, age and sex of drivers, education level and experience of the driver, road and light conditions and by possible cause of accident summary of reports.

Recognizing the complexity and magnitude of the vehicle accident problem, developed nations take necessary measures like putting resource in studying the various aspects of these accidents and use the result in decisions making. But this is not the case in developing countries like Ethiopia. In connection with this, safer no comprehensive research work has been done specifically with respect to road accident severity in Addis Ababa.

However, recently Abebe (2004), tried to assess points that needs immediate attention with respect to road safety. He stated that 81% of the accident all over the county is due to drivers fault and the other is due to vehicle, pedestrian and road faults. He also listed out that the main road safety problems are:

- ◆ Drivers not respecting pedestrian priority
- ◆ Over speeding
- ◆ Unsafe utilization of freight vehicles for passenger transportation.
- ◆ Poor skill and undisciplined behavior of drivers
- ◆ Less engineering effort in road design to consider safety
- ◆ Poor vehicle conditions
- ◆ Pedestrian not taking proper precautions
- ◆ Weak traffic law enforcement
- ◆ Lack of proper emergency medical services

Finally he conclude that, road safety publicity, targeted traffic law enforcement, hazardous location identification, pedestrian awareness, upgrading drivers skill and behavior both technically and with respect to keeping rules should get due consideration.

Wondwossen(1999) had also tried to study correlates of car accidents in Addis Ababa. He used chi-square test and logistic regression analysis in identifying correlates of car accidents. The research indicates that in addition to other variables like light condition of the road, 'not giving priority to pedestrians' has the highest association with physical damage.

Taddele and Larson (1991) on their part studied the occurrence and driver characteristics associated with in Addis Ababa. The purpose of the study was to determine the incidence density of hospital treated motor vehicle injuries and to identify driver and vehicle characteristics placing them at increased risk of inflecting injuries. And he comes up with the result of finding the incidence density of motor vehicle injuries rate of 279 per 100,000 person-years exposure. In addition to this he point out that drivers involved in a motor vehicle injuries are more likely to be male, young and less experienced. With respect to vehicle

characteristics associated with involvement in motor vehicle injuries, elevated odds ratios were found for newer and government owned vehicles and for taxis and buses.

As it is apparent from the above discussion the unavailability of comprehensive research, the availability of the accident data and the magnitude of the problem call for some advance technology in drawing meaning full information and pattern that will be help full in making sound and informed decision in the effort of saving life.

After clear understanding of the domain area and the data mining technology to be employed the next chapters focuses on the actual work of the research in attaining the objectives stated earlier.

Chapter Four

Preparing Data for Analysis

This chapter details the different steps in data mining that are carried out in preparing data for analysis in this particular research. The first part of the chapter explains the general over view followed by target data set creation. The next part explains how the collected data is cleaned, reduced and transformed to a required format.

4.1 Overview

It is crystal true that there is no data mining without data. And in order to get maximum out of the data using data mining tools and techniques, the data need to be cleaned and organized. Among others, data collection, data cleaning, attribute selection, data formatting and transformation, and dimensionality reduction are the most important activities under data preparation, which finally resulted in creating target data set.

Thus right after data collection, data pre-processing is a step where the data to be used in finding interesting pattern and useful data is cleaned, organized, reduced and transformed to a suitable format for the data mining tool.

Since the data set created/selected will determine the final result of the research work, it needs much attention and due consideration. Accordingly, being part of a wider process of data mining activity, pre-processing takes much of the researcher's time and effort.

4.2 Data Collection

Collecting, analyzing and understanding the content and structure of the data available is one of the most important tasks that need close attention. With respect to this specific research the sole source of data about an accident is the daily accident report form to be filled and reported by the traffic police officers. It consists of full details about a given accident. Through successive update, the office keeps this data in excel flat file format.

As indicated in the previous, sections a corporate data warehouse is a good source of data for a given data mining task (Berry & Linoff, 2000). It is just because data in a data warehouse has a common format with consistent definition for fields and their values. But unfortunately, some of the available data for this specific research were in a manual file format and some on a computer based flat file format. In line with this, the raw data was initially collected regarding vehicle accidents from the office's flat database and from daily accident record file. After keying the manual format of data, which took considerable time, careful analysis of the data and its structure was done together with the domain experts by evaluating the relationship of the data with the problem at hand and the particular data mining task to be performed.

This helps a great deal in understanding the data and viewing it from different angle in exploring its potential. In the following section, an attempt has been made to describe the nature of the data and its structure.

As indicated above, this study used data from the Addis Ababa Traffic Office (AATO). The initial data source for the study contained traffic accident records from September, 1995 E.C upto March 1997 E.C a total number of 25,560 cases. The first attempt of the office to electronically record the accident data in September, 1995 E.C. was limited to some attributes like labels of date and time, accident id , driver's name, driver's age, driver's gender, driver's license status, relation of the driver and vehicle, driver's experience, possession of the vehicle, vehicle type, accident area, accident road name, road segment separation, road direction, road surface type, roadway surface condition, vehicle maneuver, accident type, total vehicles involved, total number of victims, accident victims category, victims profession, victims health condition, pedestrian maneuver, vehicle plate number, cost estimate of the damage and injury severity. Through time specifically starting from September 2005, attempt has been made to incorporate some more attributes like, driver's educational level, vehicle defect, vehicle age, weather condition, light condition, cause for accident. The accident record's database AATO is still under revision to incorporate as many attribute as possible so that to create a more comprehensive database.

Because of the unavailability of important attributes in the accident data of the previous years, the data set selected for this specific research covers the time from September upto March 2005, which is a total of 5,207 records. This data was in an excel file format with 36 attributes

As indicated in the previous, sections a corporate data warehouse is a good source of data for a given data mining task (Berry & Linoff, 2000). It is just because data in a data warehouse has a common format with consistent definition for fields and their values. But unfortunately, some of the available data for this specific research were in a manual file format and some on a computer based flat file format. In line with this, the raw data was initially collected regarding vehicle accidents from the office's flat database and from daily accident record file. After keying the manual format of data, which took considerable time, careful analysis of the data and its structure was done together with the domain experts by evaluating the relationship of the data with the problem at hand and the particular data mining task to be performed.

This helps a great deal in understanding the data and viewing it from different angle in exploring its potential. In the following section, an attempt has been made to describe the nature of the data and its structure.

As indicated above, this study used data from the Addis Ababa Traffic Office (AATO). The initial data source for the study contained traffic accident records from September, 1995 E.C upto March 1997 E.C a total number of 25,560 cases. The first attempt of the office to electronically record the accident data in September, 1995 E.C. was limited to some attributes like labels of date and time, accident id , driver's name, driver's age, driver's gender, driver's license status, relation of the driver and vehicle, driver's experience, possession of the vehicle, vehicle type, accident area, accident road name, road segment separation, road direction, road surface type, roadway surface condition, vehicle maneuver, accident type, total vehicles involved, total number of victims, accident victims category, victims profession, victims health condition, pedestrian maneuver, vehicle plate number, cost estimate of the damage and injury severity. Through time specifically starting from September 2005, attempt has been made to incorporate some more attributes like, driver's educational level, vehicle defect, vehicle age, weather condition, light condition, cause for accident. The accident record's database AATO is still under revision to incorporate as many attribute as possible so that to create a more comprehensive database.

Because of the unavailability of important attributes in the accident data of the previous years, the data set selected for this specific research covers the time from September upto March 2005, which is a total of 5,207 records. This data was in an excel file format with 36 attributes

to describe each record. Familiarization and understanding of this data has taken quite a good part of the research time.

4.2.1 Accident Data Set: Data Understanding

A good understanding of the data at hand leads to a better success in achieving the data mining goal. The success criterion for this data mining research is the discovery of accident severity classification rules that would find out and differentiate accidents which are serious to those which are potentially not serious in different levels. Provided that reasonable accident severity classification rules are discovered, the office could devise a means to reduce the number of fatal and serious injuries and be able to recognize the level of severity when an accident has occurred. In short, this can help decision makers to formulate better traffic safety control policies.

As indicated above the organization keeps detail information about an accident through various attributes. Types of information that the office records about an accident includes driver and vehicle characteristics, road and weather conditions, date and time of the accident, type, injury severity and possible causes of such accidents.

The specific attributes by which a given accident can be described are labels of date and time, accident id, driver's name, vehicle type, driver's age, driver's gender, driver's educational level, driver's license status, relation of the driver and vehicle, driver's experience, possession of the vehicle, vehicle defect, vehicle age, accident area, accident road name, road segment separation, road direction, road surface type, roadway surface condition, light condition, weather condition, vehicle maneuver, accident type, total vehicles involved, total number of victims, accident victims category, victims profession, victims health condition, pedestrian maneuver, vehicle plate number, cost estimate of the damage and cause for accident. In addition to the input variables mentioned above the output variable for this research that is injury severity is also another attribute of a given accident. Attributes with their description are presented as follows.

<i>S.No</i>	<i>Attribute name</i>	<i>Type</i>	<i>Description</i>
1	Accident_Date	Date/time	Date of an accident
2	Accident_Day	Text	The day of an accident in week days
3	Accident_Time	Date/time	The time of the accident
4	Accident_ID	Number	A number to identify a given accident uniquely
5	Driv_Name	Text	Name of the driver who cause an accident
6	Driv_FName	Text	Driver' last name
7	Driv_Age	Number	Age of the driver
8	Driv_Sex	Text	Gender of the driver
9	Driv_Edu_Lev	Text	Educational level of the driver
10	Driv_Lice_Lev	Text	Status of the driver's license
11	Driv_Vehic_Rel	Text	The relation of the driver and the vehicle
12	Driv_Exp	Number	Driving experience of the driver
13	Vehic_Age	Number	Service year of the vehicle
14	Vehic_Type	Text	The type of the vehicle
15	Vehic_Poss	Text	Possession of the vehicle
16	Vehic_Defect	Text	Problem or defect of the vehicle
17	Acci_Area	Text	The area where an accident occurred
18	Acci_Road_Name	Text	The name of the road where an accident occurred
19	Road_Seg_Sep	Text	The road segment separation
20	Road_Direc	Text	The direction of the road
21	Road_Surf_Type	Text	The surface type of the road
22	Road_Cond	Text	Road surface condition at the time of accident
23	Light_Cond	Text	Light condition at the time of Accident
24	Weather_Cond	Text	The weather condition at the time of accident

25	Vehic_Maneuver	Text	Vehicle maneuver during an Accident
26	Acci_Type	Text	The type of the accident
27	Tot_Vehic_Involved	Number	Total number of vehicles involved in an accident
28	Tot_Acci_Vict	Number	Total number of victims
29	Vict_Cat	Text	Whether they are drivers, passengers or pedestrians.
30	Vict_Ped_Prof	Text	Profession and/or special skill of the victims
31	Vict_Health_Cond	Text	Physical and/or mental conditions of the victims
32	Vict_Ped_Maneuver	Text	At the time of accident.
33	Acci_Cause	Text	Causes for the accident
34	Vehic_Plate_Num	Text	Plate number of the vehicle involved
35	Damage_Cost_Est	Number	Estimation of the cost of property damaged
36	Injury_Severity	Text	The injury severity level due to an accident

Table: 4.1 Attributes with their description and data type

As it clear from the above list, the office has been trying to record as much information as possible about an accident; however some attributes like vehicle load and driving speed, which will be important, are overlooked by the office. In the process of familiarization and understanding the dataset, possible values and characteristics of each attributes were also explored. Accordingly, just to mention the dependent variable, the target attribute, injury severity has four classes: fatal injury, property damage (no injury), serious injury and slight injury.

4.2.2. Defining the Data Mining Function

As researches tried to indicate in the above sections 2.5 and 2.6, there are different data mining functionalities or tasks under two categories: descriptive and predictive. The researcher's task here was to develop intelligent models that could accurately classify the severity of injuries into four categories; namely fatal , serious, slight and non-injury or

5.2. Selection of Modeling Technique

This is a step where selection of the actual modeling technique that is to be used in the model building is done CRISP-DM (2000). In data mining there are various techniques available. Based on the goal of the data mining task, the nature of data and results needed etc, one technique may prove to be more appropriate than the other. The most widely used tools include decision tree, neural networks and cluster analysis. The modeling technique to be used in this research is decision tree. In line with this, the selection of the algorithm and the measure that Knowledge tree can use to compute the best split is also another issue to consider.

An important advantage of decision tree models is that they are explainable. It is possible to build them interactively, which allows the researcher to incorporate his/her knowledge automatically. More over in either way, it is possible to view the tree and see steps to arrive at a leaf or terminal node. Decision trees are more suitable than either neural network or clustering, when disjunctive descriptions are required. Decision tree methods are a good choice when the data mining task is classification of records or prediction of outcomes. Decision trees are also a natural choice when the goal is to generate rules that can be understood (CRISP-DM, 2000).

KnowledgeSTUDIO supports two types of algorithms namely KnowledgeSEEKER and HeatSEEKER. KnowledgeSEEKER is a flexible algorithm that is especially good for exploration purposes and manual tree building and it can also handle a large amount of variables with either a continuous or discrete dependent variable. Unlike KnowledgeSEEKER, HeatSEEKER is a fast algorithm that is especially good for automatically generating a tree, and even if it can handle a very large number of records, it performs better with fewer variables. Since in the present research the number of variables considered were many (i.e 36), and there was a need for manually adjusting the tree and to explore the results obtained, the algorithm selected in model building was knowledgeSEEKER. It works first by identifying relationship between each of the predictor variables in the analysis and the dependent variable and then setting the significance of the relationship.

As to the measure used in building the tree is Adjusted-P-value Bonferroni Adjustment measure was selected because it is the most statistically sound approach given that it reduces

the possibility of producing chance results and puts all variables on a common statistical footing. The choice is also based on the fact that unordered grouping variables and variables with many values have a biased chance of being selected as branches unless adjusted tests are used.

5.3. The Experiment

As mentioned earlier the dataset was obtained in Excel flat file format. KnowledgeSTUDIO supports the import of data from a large number of data formats. Using the data import wizard the data was imported from an Excel to the KnowledgeSTUDIO. Then the dataset was partitioned into two separate datasets. Knowledge STUDIO partitioning dialog box, shown in figure 5.1 is used to partition the dataset into learning and validation sets. This was done to create separate data for training and validation. Knowledge studio has the facility to create partition in two ways. One is to partition the data randomly. The other is to construct a balanced partition using some variable of interest.

Since the dataset of this experiment has an unbalanced dependent class, injury severity, using a random sample will most likely result the minor class belonging in only one of the partitions. In order to overcome this problem the balanced partitioning option was used. To validate model, a dataset is usually divided into two partitions. One is used for training, the learn partition, and the other is held back for testing, the validation partition. Consequently the dataset was divided into the training and validation set on a 75: 25 ratio. After partitioning the training dataset contained 3493 (75% from each class) records and the validation set contained 1165 (25% from each class) records.

Partition Dialog

Defined Partitions

- Accident dataset
 - Learn
 - Validate

Add Delete

Partition Details

Name: Learn Type: ☐ Random Sample ☒ Balance Sample

Random Seed: 1

Sample Selection

Type: ☐ Record Count ☒ Percentage of Total 74.989265779

Balanced Sampling

Field to Balance:

- Driv_Sex
- Driv_Exp
- Driv_Exp_Cat
- Vehic_Age
- Vehic_Age_Cat
- Vehic_Type
- Road_Surf_Type
- Road_Cond
- Light_Cond
- Weather_Cond
- Acci_Type
- Acci_Cause
- ☒ Injury_Severity

Sampling Details:

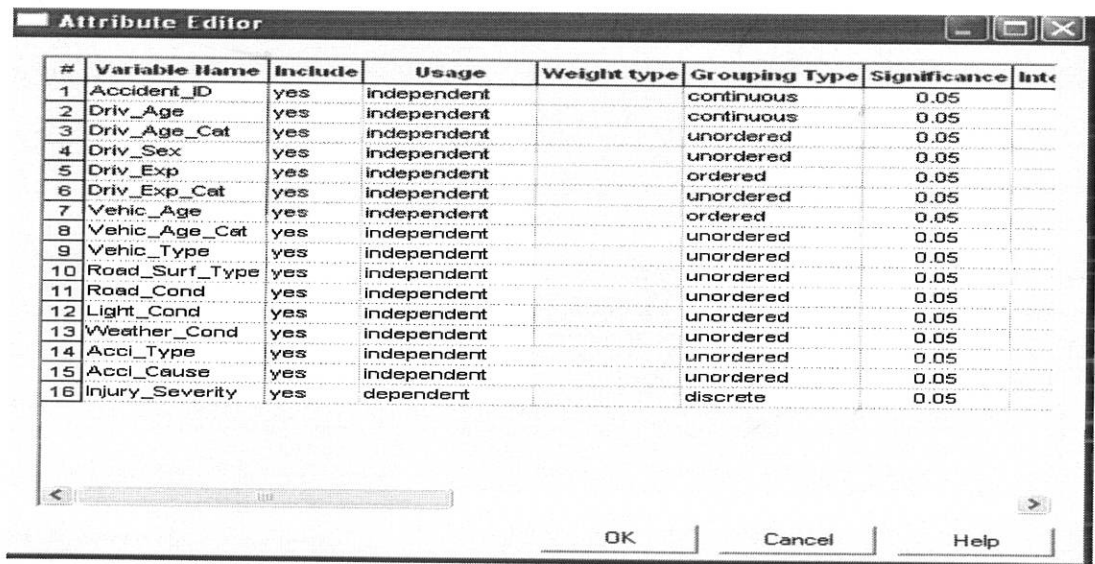
	Unique Value	Frequency	Split Percentage
1	F-Injury	341	25
2	P-damage	3206	25
3	Serious_Injury	402	25
4	Sli-Injury	709	25

Total Records: 4658
Unallocated Records: 0
Unallocated Portion: 0.0 %

OK Cancel Help

Figure 5.1: The knowledge studio data set partition dialog

editors allow manipulation, among other things, of the independent variables and the dependent variable used in the analysis, as well as grouping type, order display, usage, mapping, and missing value use.



The screenshot shows a window titled "Attribute Editor" with a table of 16 variables. The table has columns for #, Variable Name, Include, Usage, Weight type, Grouping Type, Significance, and Int. The variables are listed in the table below.

#	Variable Name	Include	Usage	Weight type	Grouping Type	Significance	Int
1	Accident_ID	yes	independent		continuous	0.05	
2	Driv_Age	yes	independent		continuous	0.05	
3	Driv_Age_Cat	yes	independent		unordered	0.05	
4	Driv_Sex	yes	independent		unordered	0.05	
5	Driv_Exp	yes	independent		ordered	0.05	
6	Driv_Exp_Cat	yes	independent		unordered	0.05	
7	Vehic_Age	yes	independent		ordered	0.05	
8	Vehic_Age_Cat	yes	independent		unordered	0.05	
9	Vehic_Type	yes	independent		unordered	0.05	
10	Road_Surf_Type	yes	independent		unordered	0.05	
11	Road_Cond	yes	independent		unordered	0.05	
12	Light_Cond	yes	independent		unordered	0.05	
13	Weather_Cond	yes	independent		unordered	0.05	
14	Acci_Type	yes	independent		unordered	0.05	
15	Acci_Cause	yes	independent		unordered	0.05	
16	Injury_Severity	yes	dependent		discrete	0.05	

Figure 5.3: Overview of the tree attributes editor

Decision Tree Model Building

Experiment One

The first experiment on the decision tree model building was based on the training set which is 75% (i.e.3493) of the data that has been prepared and imported to the KnowledgeSTUDIO software. And the variable "Accident-Severity" was set as dependent variable and all others were set as independent variables. The classification tree was built using all the default parameters suggested by the software (i.e. KnowledgeSTUDIO) using all the 13 variables excluding the three namely driver age, driver's experience, and vehicle age, variables from which derived attributes are drawn. Figure 5.4 shows the out put of the first decision tree.

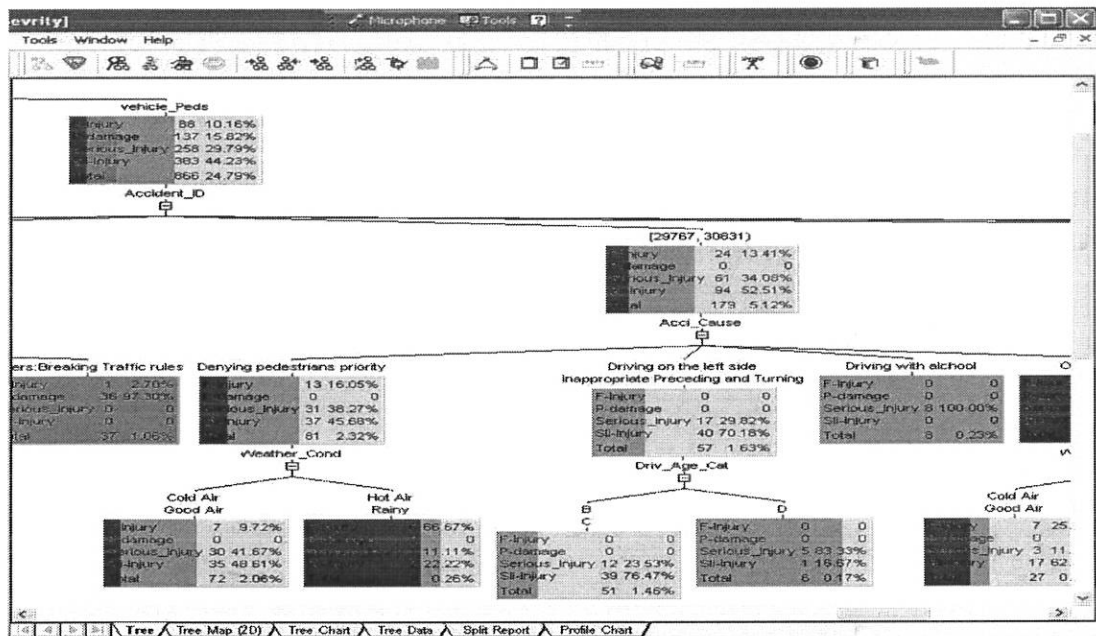


Figure 5.4: Output of the first Decision Tree

The first experiment shows that though each record in the dataset included 13 attributes with the inclusion of the dependent variable “Accident-Severity”, the decision tree constructed had only used 10 attributes. These 10 attributes namely accident type, accident cause, accident_Id, road surface type, weather condition, road condition, light condition, vehicle type, driver’s age category and driver sex are selected and ranked as an important attributes in such classification task by the software.

The other attributes namely Vehic_Age_Cat and Driv_Exp_Cat are considered by the software as statistically insignificant and are not ranked. This indicates that attributes that are not considered in the construction of the decision tree are not important to discriminate the records into the predefined classes (i.e. Fatal_Injury, Serious_Injury, Slight_Injury and P_Damage or No_Injury) provided to the model. But this may not be true always.

In addition to that, from the first experiment the validation statistics shown in figure 5.5 that is out of 1165 records of the validation dataset 1009 records were classified correctly, which is 86 % accuracy. And this accuracy does not necessarily show the adequacy of the model. So review of variables used should be considered in the subsequent experiments.

Statistics	
Total records	1165
Correctly predicted	1009
Percentage	86.61%

Figure: 5.5: prediction statistics of the first experiment

Thus it may be necessary some times to get confirmation from the domain experts that the attributes identified as important are acceptable in classifying potential accidents in the above mentioned classes. This is basically based on the suggestions made by StatSoft (2003) on the usefulness of combining the automatic methods for building trees with "educated guesses" and domain-specific expertise. Thus automatic methods and domain expert's expertise were employed in the subsequent tree building, refining and modifying the tree.

Experiment Two

Considering the relative importance of each variables for the classification of the accident severity, assessing the variables that are used for splitting the tree was the next important task done. This is because KnowledgeSTUDIO usually displays the most statistically significant variable or split at the root, and they might not be the most important variables to split the tree at the root.

Accordingly, based on experts opinion the second variable, i.e. Accident_Id, used to split the tree at the node next to the root was not considered as being the most important variable to consider the level of injury severity, hence the next most statistically significant variable was taken. By using the 'Next split' command that displays the next most statistically significant split, the researcher was able to obtain a tree, which according to the domain experts made use of a good variable for splitting the records at the root. The next figure shows the second decision tree built after the above measures.

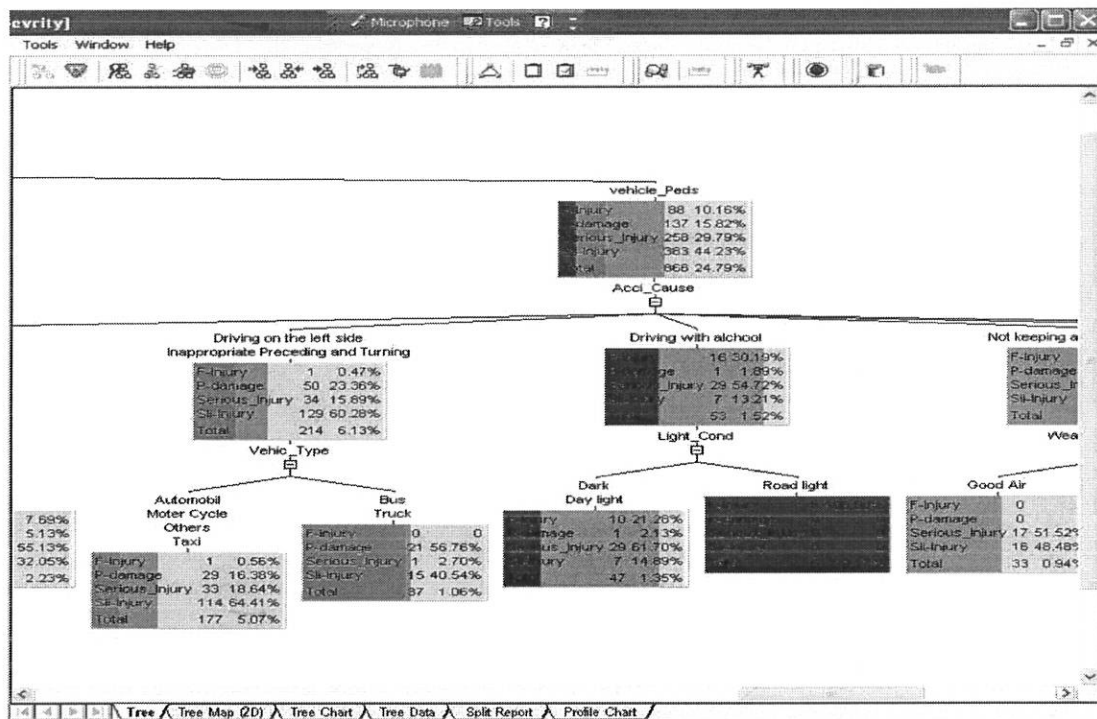


Figure 5.6: Output of the second Decision Tree

In this experiment, the number of variables used was reduced to nine by excluding accident_id which is not significant in this specific research. Thus the tree was splitted first by the variable accident type, having 5(five) distinct values and out of which 'vehicle_ped' and 'turn over' contribute much for fatality and serious injuries. The next most important variable used for split was 'accidentous cause'. And as it revealed from the decision tree nodes value, 'denying pedestrians priority' and 'over speeding' were the most significant cause for such fatality and serious injuries.

Assessment and validation of the result was the next task after building the tree. An evaluation of the results was made on both the training and testing set. And the validation shows good accuracy in general though it is less than the first accuracy which includes attributes that does not contribute much to the classification task. The result of the validation of the decision tree built given below in form of confusion matrix.

Confusion Matrix - Injury_Severity					
	Predicted				
		F-Injury	P-damage	Serious_Injury	Sli-Injury
	F-Injury	37	31	1	22
	P-damage	7	744	2	35
	Serious_Injury	9	11	8	67
	Sli-Injury	12	37	1	141

Statistics	
Total records	1165
Correctly predicted	930
Percentage	79.83%

Figure 5.7: Validation results from the second decision tree

The confusion matrix depicts that out of the total records provided to the software tool for testing (i.e. 1165), 37, 8, 141 and 744 records were classified correctly in the class of fatal injury, property damage, serious injury and slight injury respectively. On the other hand, 54 records were incorrectly classified as serious injury(1), slight injury(22) or property damage(31) while actually they were supposed to be in the fatal injury class and 44 records were classified incorrectly as fatal injury, serious injury, and slight injury while actually they were in the property damage class. This result reveals that from the total records (i.e. 1165), 930 were classified correctly while the remaining 235 records were classified incorrectly. Hence this indicates that records whose class is property damage were classified with a minimum error as compared with the records in the class fatal and serious injury.

Although the performance of this training scheme was not bad in terms of accuracy careful consideration of the variables need more attention to get better model that can provide sound rule with better accuracy. Hence, the researcher made use of the above selected variables for further tests by building several decision trees through varying the number and combination of the variables. Some of the results obtained from the experiments conducted are presented in table 5.1

Se.No	Number of Variables utilized	Accuracy
1	10 (ten)	87.47%
2	9 (nine)	79.83 %
3	8 (eight)	84.56%
4	7(seven)	87.47%
5	6 (Six)	85.37%

Table 5.1: Results of the Tests Conducted

From additional experiments conducted, figure 5.8 shows the 'best' decision tree obtained by the use of 7 (seven) variables out of 13 variables that were considered for model building. As indicated above with the confirmation of the domain expert and their statistical significance in splitting a tree, the seven attributes, namely accident type, accident cause, vehicle type, driv_age_cat, road condition, road surface type, and light condition are considered to be most important in this specific classification task.

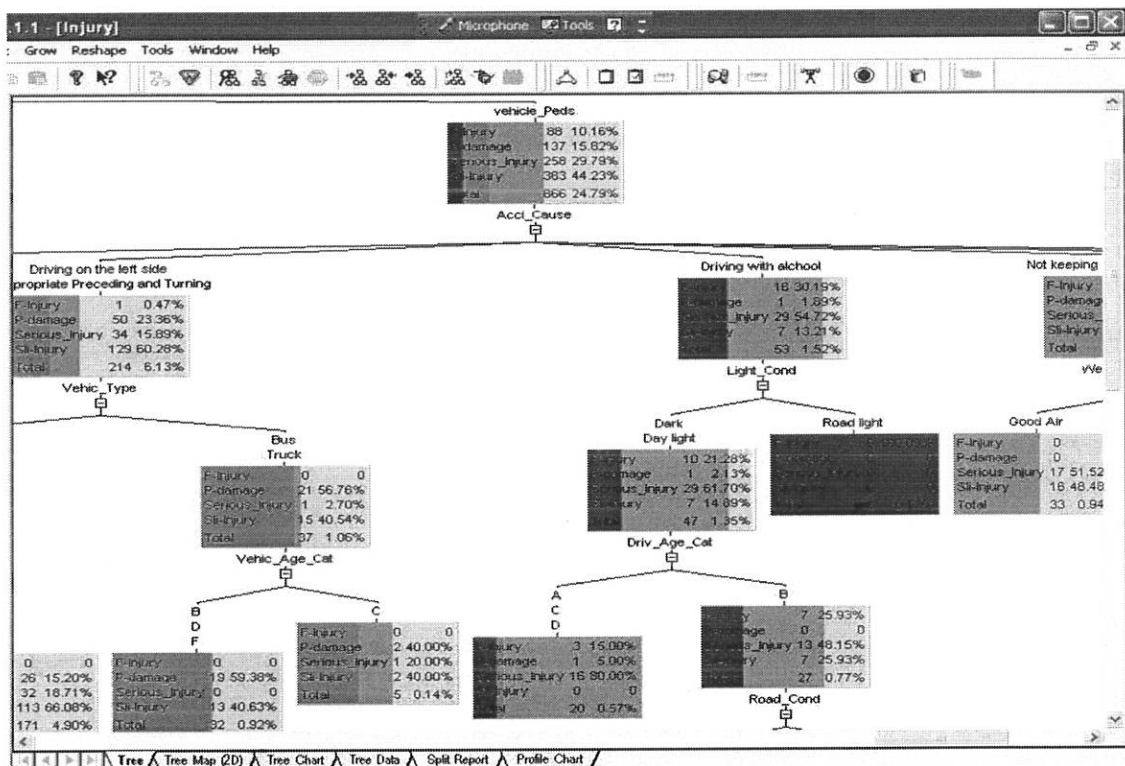


Figure 5.8: 'Best' decision tree with 7 variables

From the above decision tree shown in figure 5.8 out of 3493 training set records, it was classified by accident type into five categories of turn over, vehicle_fixobj, Vehicle_Vehicle, Others, and Vehicle_Ped having 258(7.39%), 690(19.5%), 1,359(38.91%), 320(16.24%) and 866(24%) respectively. With respect to fatality and serious injuries the two most important type of accidents are vehicle_ped and turn_over. To illustrate this, out of 866 vehicle_ped accidents, F_injury, p_damage, Serious_injury, and Sli_injury takes 88(10.16%), 137(15.82%), 258(29.79) and 383(44%) respectively.

Similarly out of 258 turn over accidents, F_injury, p_damage, Serious_injury, and Sli_injury shares 61(23.64%), 143(55.43%), 29(11.24%) and 25(9.69%) respectively.

From this we can understand that these two types of accidents namely vehicle_ped and Turn over are contributing a lot to fatalities and serious injuries and this in turn should be considered by the traffic office in formulating policies.

Next to the accident type, Acci_cause was the next important attribute in splitting the tree. Again taking in to consideration the two important accident types, denying pedestrian priority, driving with alcohol and over speeding are determinant cause of turn over and vehicle-ped accident types.

To illustrate this out of 391 vehicle_ped accidents caused by denying pedestrians priority, 208(53%) results in either fatal injury or serious injury. And out of 53 accidents caused by driving with alcohol, 45 (85%) results in either fatality or serious injury. The detail rules generated from this tree are attached as an annex.

The decision tree constructed in this manner was validated on the training set and as it showed a good result on the dataset that it used to train, it was validated on the testing set and the output of confusion matrix (figure 5.9) shows that, out of the total 1165, 1019 (87 %) were classified correct and (13 %) 146 were classified incorrectly.

Confusion Matrix - Injury_Severity					
Actual		Predicted			
		F-Injury	P-damage	Serious_Injury	Sli-Injury
	F-Injury	32	31	12	16
	P-damage	7	776	5	0
	Serious_Injury	2	7	39	47
	Sli-Injury	3	7	9	172

Statistics	
Total records	1165
Correctly predicted	1019
Percentage	87.47%

Figure 5.9: Validation results from the 'best' decision tree

The confusion matrix shown in figure 5.9 shows that out of the total records provided to the program, 32, 776, 39, and 172 records were classified correctly in the class fatal, property damage, serious injury and slight injury respectively. This shows that the accuracy rate for the

prediction of property damage and slight injury records is better than the others. Hence records whose class is either property damage or slight injury were classified with a minimum error as compared with the records in the fatal and serious injury. Although several attempts were made, it was not possible to obtain a good tree with a sound rule and an accuracy of more than 87%. Thus, the decision tree presented in figure 5.9 which is the one selected as a good one for the present research, classified 87% of the records correctly and 13 % of records incorrectly.

Selection of a good model from subsequent experiments conducted was made by taking into consideration the soundness of rules generated as well as the number of records which are misclassified, as this type of error has to be minimized in a road safety area. Accordingly the above decision tree was found to be the 'best' possible model to be selected from among a number of experiments conducted.

Generating Rules from Decision Tree

After successive experiments in building the best decision tree model, the next step is to generate, rules by tracing through the branches up to leafs. A rule is a correlation found between the main variable (dependent) and the others (independent). Thus rules were generated using KnowledgeSTUDIO (i.e. the software used) from the decision tree constructed. The tool also allows the rules to be presented in many types. Some of the types supported are SQL, generic, Java and English. For the purpose of the present research the rules generated are in the generic type as they are found to be easily understandable. Some of the rules extracted from the decision tree are selected and presented as follows while the other rules are attached as an annex.

Generic Rules

RULE #1

(Whole Tree)

```
Injury_Severity = F-Injury 0.0781115879828
Injury_Severity = P-damage 0.676394849785
Injury_Severity = Serious_Injury 0.0815450643777
Injury_Severity = Sli-Injury 0.163948497854
```

RULE #2

if

```
Acci_Type = Turn Over
Acci_Cause = Denying pedestrians priority, Driving with alchool,
Inappropriate Preceding and Turning or Over Speeding
then
```

```
Injury_Severity = F-Injury 0.535714285714
```

Injury_Severity = P-damage 0.321428571429
Injury_Severity = Serious_Injury 0.0357142857143
Injury_Severity = Sli-Injury 0.107142857143

RULE #3

if
 Acci_Type = vehicle_Peds
 Acci_Cause = Denying pedestrians priority or Not keeping appropriate
distance
then
 Injury_Severity = F-Injury 0.155405405405
 Injury_Severity = P-damage 0.00675675675676
 Injury_Severity = Serious_Injury 0.371621621622
 Injury_Severity = Sli-Injury 0.466216216216

RULE #4

if
 Acci_Type = vehicle_Peds
 Acci_Cause = Driving with alchool
then
 Injury_Severity = F-Injury 0.333333333333
 Injury_Severity = P-damage 0.0833333333333
 Injury_Severity = Serious_Injury 0.375
 Injury_Severity = Sli-Injury 0.208333333333

RULE #5

if
 Acci_Type = vehicle_Peds
 Acci_Cause = Over Speeding
then
 Injury_Severity = F-Injury 0.137254901961
 Injury_Severity = P-damage 0.21568627451
 Injury_Severity = Serious_Injury 0.137254901961
 Injury_Severity = Sli-Injury 0.509803921569

RULE #6

if
 Acci_Type = vehicle_Peds
 Acci_Cause = Denying pedestrians priority or Not keeping appropriate
distance
 Road_Cond = Dry
then
 Injury_Severity = F-Injury 0.112781954887
 Injury_Severity = P-damage 0.00751879699248
 Injury_Severity = Serious_Injury 0.383458646617
 Injury_Severity = Sli-Injury 0.496240601504

RULE #7

if
 Acci_Type = vehicle_Peds
 Acci_Cause = Denying pedestrians priority or Not keeping appropriate
distance
 Road_Cond = slippery
then
 Injury_Severity = F-Injury 0.533333333333
 Injury_Severity = P-damage 0
 Injury_Severity = Serious_Injury 0.266666666667
 Injury_Severity = Sli-Injury 0.2

From the general rule, it is easy to see that, there is a 7.8 % chance that Injury_Severity will be F-Injury, a 67.63 % chance that Injury_Severity will be P-damage, a 8.15% chance that Injury_Severity will be Serious_Injury and a 16.39% chance that Injury_Severity will be Sli-Injury. And this reveals that about 33% of the accidents, results in injury of different level half of which is either fatal or serious injury. In addition to this accident types involving pedestrians and single vehicle turn over mostly results in either fatal or serious injuries. Denying pedestrian priority and over speeding were also the top most determinant factors in injury severity.

It is also apparent from rule four (4) that if accident cause is driving with alcohol and accident type is vehicle_peds there is high probability of an accident resulting in fatalities or injuries. Over speeding is also another important factor especially if associated with vehicle_peds type of accidents.

In general the rules presented above indicate the possible conditions in which an accident will result in either of the injury severity classes. Moreover, the rules generated have indicated that attributes such as 'accident cause', 'accident type', 'driver age', 'road surface type', 'road condition', 'vehicle type', and 'light condition' are found to be important variables for classification of accident severity. They are playing significant role in all experiments being placed at the higher level of the tree which indicates their statistical significance than other variables like 'sex', 'weather condition' and 'accident_id'. And this will help the traffic office and road authority in focusing their attention on determinant causes of accidents, like denying pedestrian priority and over speeding, resulting in fatal and serious injuries while revising road traffic rules and regulations and allocating resources in their effort of improving road safety.

Building a Predictive Model

After identifying important variables and their relation in determining the severity level of a given accident building and testing predictive model was the next step in this research. A predictive model refers to a structure and process for predicting the values of specified variables in a dataset. The model was based on the variables identified by the decision tree found to be relevant for the classification of injury severity. After assessing the accuracy of the model, recommendation was made so as to make more experiments and make use of it to classify future data tuples or objects for which the class label is not known.

In building a decision tree based predictive model, KnowledgeSEEKER algorithm, Adjusted P-value Bonferroni Adjustment measure and reduced error pruning method were used. Since the tree algorithms selected will determine how splits are calculated, considering their feature KnowledgeSEEKER was selected for its flexibility, powerfulness and ability in handling large amount of variables with either a continuous or discrete dependent variable over HeatSEEKER algorithm

As to the measure type the adjusted measure specifically Adjusted P-value Bonferroni Adjustment measure is selected for its ability in reducing the possibility of producing chance results and puts all variables on a common statistical footing. The reduced error method starts with a complete tree and run the test data through it, noting the number occurring in each value of the dependent variable at each node. For each non-leaf node, count the number of errors if the sub-tree is kept and the number if it becomes a leaf through pruning. The pruned node will often make fewer errors on the test data than the sub-tree makes. The difference between the numbers of errors (if positive) is a measure of the gain from pruning the sub-tree. From all the nodes, the pruning method chooses the one with the largest difference as the sub-tree to prune.

The description of predictive models build employing the above logic, procedure, algorithm and measure is presented as follows. They were built just following the predictive model wizard from 'insert menu' of KnowledgeSTUDIO.

As it is clearly indicated in KnowledgeSTUDIOS manual, to build a decision tree based predictive model it is simply to specify the tree to be used. The tree itself is a training phase (Angoss Software Corporation, 2001). And it is also possible to use validation data set of the decision tree for predictive models. This is because predictive modeling based on decision tree is basically an extension of selecting the best decision tree. And the best tree is selected after several experiments by changing the number and combination of attributes. But again by adjusting number of variables and percentage of partitions and different predictive models were built and the best one with good accuracy was selected.

The first experiment was done using all attributes selected for decision tree experiments, and using the test data set partitioned for validating the best decision tree. The pruning method used was critical value. The accuracy of the resulting model is presented as follows in figure 5.10

```

Dependent Variable: Injury_Severity
Independent Variables:
  Driv_Age_Cat, Driv_Exp_Cat, Vehic_Type, Road_Surf_Type, Road_Cond,
  Light_Cond, Weather_Cond, Acci_Type, Acci_Cause
Hidden branches: Disabled
Training Results:
  Variance / Entropy Explained: 0.300190
  Records correctly predicted: 1855
  Percentage correctly predicted: 75.87 %
Algorithm: KnowledgeSEEKER
Pruning: Enabled
  Classification: critical-value
  Dataset used to prune: Auto partitioning

```

Figure 5.10: Accuracy result of the first experiment on predictive model

The next experiment was done by selecting only two attributes that are statistically significant and located at the top of the decision tree namely accident type, accident cause. And selecting the 60:40 partitioning ratio. That means 60% for training and 40 % percent for testing. The accuracy of the resulting model is presented as follows in figure 5.11.

```

Dependent Variable: Injury_Severity
Independent Variables:
  Acci_Type, Acci_Cause
Hidden branches: Disabled
Training Results:
  Variance / Entropy Explained: 0.290117
  Records correctly predicted: 2116
  Percentage correctly predicted: 75.73 %
Algorithm: KnowledgeSEEKER
Pruning: Enabled
  Classification: reduced-error
  Dataset used to prune: Auto partitioning
  Nodes existed prior to prune: 24
  Nodes remain: 6
  Nodes were pruned: 18

```

Figure 5.11: Accuracy result of the second experiment on predictive model

The third experiment is used seven attributes that are used to build the best decision tree. This attributes are top most statistically useful attributes. And the pruning method used is reduced-error pruning method. Here follows the resulting predictive model accuracy.

Dependent Variable: Injury_Severity

Independent Variables:
 Driv_Sex, Vehic_Type, Road_Surf_Type, Road_Cond, Light_Cond,
 Acci_Type, Acci_Cause

Hidden branches: Disabled

Training Results:
 Variance / Entropy Explained: 0.352823
 Records correctly predicted: 2535
 Percentage correctly predicted: 77.76 %

Algorithm: KnowledgeSEEKER

Pruning: Enabled
 Classification: reduced-error
 Dataset used to prune: Auto partitioning
 Nodes existed prior to prune: 59
 Nodes remain: 11
 Nodes were pruned: 48

Figure 5.13: Accuracy result of the third experiment on predictive model

Thus, from the model one can learn that if accident type is involving pedestrian, denying pedestrians priority, driving with alcohol and over speeding contribute much than other causes for an accident to result in fatality or serious injury. The predictive accuracy of the model built using the independent variables identified and the dependent variable 'injury severity' is 77.76 %.

From the previous results it is easy to understand that attributes selected as determinant in building the best decision tree with reduced-error pruning method produces better accuracy. Thus it is selected as the best predictive model from the experiments conducted. The summary of the experiment's accuracy comparison is presented in table 5.2

<i>Exp.no</i>	<i>Attributes used</i>	<i>Accuracy</i>
1	Driv_Exp_Cat, Driv_sex, Driv_Age_Cat, Acci_type, Acci_cause, Road_Cond, Road_Surf_Type, Vehic_type, Light_Cond	75.87%
2	Acci_type, Acci_cause	75.73%
3	Acci_type, Acci_cause, Road_Cond, Road_Surf_Type, Driv_sex, Vehic_type, Light_Cond	77.76%

From the above summary it is easy to learn that the third experiment is better in accuracy employing seven attributes, namely Acci_type, Acci_cause, Vehic_type, Road_Cond, Driv_age_cat, Road_Surf_Type, and Light_Cond. And identifying these determinant factors will help the Addis Ababa Traffic Office in policy making. Specifically it will be used to gear the focus and allocation of resources to wards improving road safety and decreasing fatalities and severity of injuries resulted from an accident.

5.4 Evaluation

Next to building and validating models, assessing the degree to which the model meets the business objectives was considered. It is through consultation of domain experts. Moreover, the whole process has been reviewed and the next steps were determined. The model-building phase was divided into two. An iterative decision tree model building using KnowledgeSEEKER algorithm which are used to generate rules was the first phase. The variables used for the decision trees constructed were those variables, which described an accident. Each of the decision tree constructed were used to generate rules that were analyzed in order to establish their business worth as meaningful rule sets that discovered information to help in the road safety programs.

The analysis, which was closely undertaken with domain experts, revealed that the variables that were used were good attributes to classify injury severity into the predefined classes (i.e. fatal, serious, slight, and property damage). Each classification decision tree constructed under different experiments conducted were validated based on a dataset set aside for that purpose from the original dataset that was collected. That means after partitioning the data set in to training and validation, the validation data set was used to validate each tree through 'validate' command from tools menu of KnowledgeSTUDIO. This validation results were obtained in the form of confusion matrix.

But decision of selecting the best decision tree was based on the soundness of the rules generated as well as the number of misclassified records of different levels of injury severity. The decision tree depicted in figure 5.8 was the 'best' from the others in that it provided rule sets that were sounder to the problem domain. That means it provides most determinant factors in classifying accidents in their results. But the model is by no means the ultimate model; therefore it is the researcher's belief that more experimentation could yield more meaningful trees that yield more sound. The validation test conducted on the results of the

selected decision tree revealed that the over all accuracy of classifying accidents is 87.47%. Once the model that generated the soundest rule to classify accidents in to a predefined class was selected, the next activity was to test the predictive accuracy of the model with the use of the independent variables identified as important and 'injury severity' as the dependent variable so as to be able to use the model to classify future data tuples or objects for which the class label is not known.

Training results showed that the most important variables for the problem were accident type, cause of an accident, driv_age_cat, road condition, vehicle type, and light condition, road surface type. The model built in the present research was intended to support an effort in improving road safety at Addis Ababa Traffic office. Specifically it will used by the office in policy formulations. Therefore as previous researches reviled, the researcher also believes that the use of decision tree method is better as it can be easily explainable.

Decision makers at the traffic office can easily detect and explain the level of injury severity of a given accident with decision trees because all the reasons for the different levels severity can be traced back on the decision tree. For example attributes like accident type and accident cause are most important in determining the severity of an accident. That means these decision trees can provide the possible combinations of determinant factors in classifying injuries at a different level of injury severity. But further experiments need to be conducted to assess if the use of a combination of the techniques could yield better results. However, since no analysis technique can replace experience and knowledge of the domain expert consideration of such expertise should get due attention.

As it is apparent from the above discussion application of data mining technology in road traffic accident analysis or in modeling traffic data records can help to understand driver's behavior, roadway conditions, and weather conditions that were causally connected with the different injury severity. This can in turn help decision makers to formulate better traffic safety control policies. Conclusions and recommendations drown from the research findings are presented in the next chapter.

Chapter Six

Conclusion and Recommendation

6.1 Conclusion

Till now the capability to gather, store and manipulate data, accelerated by information technology, exceeds the ability to analyze and understand a huge volume of data. But the new generation of computerized methods is helping the effort of examining and analyzing very large datasets automatically and efficiently, thereby extracting useful information and knowledge, which are valuable for decision making.

The objective of this research undertaking was to explore the possible application of data mining technology at Addis Ababa Traffic Office, for developing a classification model. Such a classification model could support the traffic officers at Addis Ababa Traffic Office in making decisions in traffic control activities. Specifically it helps decision makers to understand driver's behavior, road and weather conditions and other related issues causing accidents resulting in fatalities or serious injuries so as to formulate better traffic safety control policies. Such a model developed also could help in minimizing accident severities at the region, and thereby improve the quality of life.

Data mining technology basically is an iterative processes consisting of four basic steps; data collection, data preparation, model building, and evaluation. Accordingly different problems created instance where there was a need to go back and forth between the different steps.

Addis Ababa traffic control division has a function of reducing accidents as one of its major activities. In doing so periodic registration of accidents and a small scale analysis has been done so far. But the severity of the case, that is the magnitude of accidents, needs some more advanced techniques in order to reach at measurable and actionable recommendations to help decision makers in strategic planning and decision-making. Moreover, the identification of major determinants of road traffic accident and severity nature and fatality of the injury will help save lives and support the effort of National Police Commission in reducing crime and improving the quality of life at large.

Since this research was intended specifically to identify the possible application of decision tree in the transport sector, other techniques like neural network should also be considered. In order to support traffic control system of Addis Ababa City, several models were built by employing decision tree approaches, and extracting rules. The best performing decision tree classifier was chosen taking into account the soundness of the rules it generated and also the number of false negatives it reduced, then its predictive accuracy was evaluated and analyzed.

A dataset having 4658 records of vehicle accidents was used to build and test decision tree classification models. The determinant variables for injury severity level of an accident that were identified by the present research were: 'accident cause', 'accident type', 'road condition', 'vehicle type', 'light condition', 'road surface type' and 'driver age'. The classification accuracy of the decision tree was tested, and it showed an accuracy of 87.47%.

➤ Although as tried to mention above the classification accuracy of decision tree should be compared with other techniques, decision tree was found to be more appropriate to the problem area since it provided with an explanation as to the decision it supplied for the actions that it recommended. In addition to this further experiments should be made to test if the combination of the models could result in a better classification performance.

To conclude, results from the study have shown that the problem of injury severity due to vehicle accident could be supported by the use of data mining, in particular with the use of decision trees technique.

6.2 Recommendations

Through this research work an attempt has been made to find out the potential applicability of data mining technology to support the traffic control activity at Addis Ababa City based on historical data accumulated on accidents. Thus, based on the findings of this research work, the researcher would like to make the following recommendations particularly in relation to the possible application of data mining technology in supporting traffic accident severity analysis at Addis Ababa City Traffic Office.

Even though, this research was conducted as an academic exercise its results are found to be promising to be applied in addressing practical problems of injury severity analysis resulted

from an accident. Hence, based on the findings of this study, the following recommendations are forwarded:

- ◆ An efficient data-mining task basically requires an availability of electronic data. This is because it is computer based solution for a given business problem. Having records available in electronic format facilitate the process of updating, reduce the time required to key in the records and avoid errors in keying in data. To this end, although there is a good start, Addis Ababa Traffic Office should take measures to store all its records with all the necessary attributes in an electronic format and to make all decisions based on collected records.
- ◆ Data mining techniques could contribute a lot in identifying potential accidents that will result in fatal and serious injuries. Thus, it could be more important to use data mining technique as a tool for the decision making process especially in revising the road traffic rules and regulations and creating awareness for drivers and the citizens. In other words, Addis Ababa Traffic office could optimize its effort in improving road safety by employing data mining technology.
- ◆ In this research, an attempt has been made to assess the applicability of data mining technology to support the identification of determinant factors for an accident and injury severity level resulting from an accident by using some set of variables that were considered important by experts. However it is open for investigation to consider the effect of other variables towards the issue under treatment so as to build models with better performance and accuracy than the models built in this research work.
- ◆ Further experimentation on a larger dataset should also be made in order to derive more appropriate rules, which are good classifiers of potential accidents resulting as a fatal, serious, and slight or non-injury risk so as to be able to predict the outcome of records of accidents.
- ◆ The present experiment provides a way to only classify accidents resulting as fatal, serious, slight, or non-injury. Further experiments should be able to find out other interesting pattern of an accident.
- ◆ Although in this study encouraging results were obtained, a small number of records was used for training and testing classifiers. Hence, it is appropriate to

conduct the experiments with large training and testing datasets as well as making a number of trials to come up with more accurate and better performing classifiers.

- ◆ Further tests by extending this research or viewing it from different angle should also be made on the same dataset to see if other techniques like neural network or a combination of one or more techniques will result in a better classification performance.

References

- Abebe Asrat (2004). *Basic Points Requiring Close Attention on Road Safety In Ethiopia*.
Unpublished document. Addis Ababa.
- Addis Ababa Transport Authority (2004). *Addis Transport*. Biannual Magazine.1 (1). Addis Ababa.
- Angoss Software Corporation. (2001). *Knowledge Studio Data Mining Software User Guide*.
Available URL:<http://www.angoss.com>
- Berry, M. J. and Linoff, G. (2000). *Mastering Data Mining: The Art and Science of Customer Relationship Management*. New York: John Wiley & Sons, Inc.
- Brand, E. and Gerritsen, R.(1998). *Classification and Regression*. Available at URL:
<http://www.dbmsmag.com>.
- Berson, A. , Smith, S. and Thearling K. (1999). *An overview of data mining techniques*.
Available at URL: <http://businessintelligence.ittoolbox.com/document.asp>
- Chong, M., Abraham A., and Paprzycki M. (2002). *Traffic Accident Analysis Using Decision Trees and Neural Networks*. Available at URL:
<http://arxiv.org/ftp/cs/papers/0405/0405050.pdf>
- Chong, M., Abraham A., and Paprzycki M. (2004). *Traffic Accident Analysis Using Machine learning Paradigms*. Available at URL:
<http://falklands.globat.com/~softcomputing.net/informatica1.pdf>
- Chung, H.M.,Gray P. Mannino, M. (1998). *Introduction to data mining and knowledge discovery*.
- City Government of Addis Ababa, Transport Authority. Facts about Addis Ababa City Transport. Available at URL: <http://www.telecom.net.et/~aata/>
- Conerly, B.G, Busby K. and Mansfield E. (2000). *Data Mining and Visualization of the Alabama Accident Database*. University Transportation Center for Alabama.
Tuscaloosa, Alabama.
- Coppock, David S. (2003). *Data Mining and Modeling. Application of Data mining Techniques*. DM Review. com
- CRISP-DM. (2000).CRISP-DM 1.0: *Step-by-step data mining guide*. Available URL:
<http://www.crisp-dm.org>.
- Ethiopian Roads Authority. (2002). *Report on annual rural traffic movement in Ethiopia*.

- Fayyad, Usma, Piatetsky-shapiro, G. and Smyth, P. (1996). *From Data Mining to knowledge Discovery in Databases*. Available URL:
<http://citeseer.nj.nec.com/fayyad96from.html>
- Gerritsen, R. and Brand, E. (1999). *Assessing Loan Risks: A Data Mining Case Study*. Available at URL: <Http://www.dbmsmag.com>
- Growe, Glenn A. (1999). *Comparing Algorithm and clustering data: Components of the data mining process*. Masters thesis, Grand Valley State University, Department of Computer science and information systems. Michigan.
- Hallberg, Örjan. (2004). *Analysis and prediction of traffic accidents*. Available at URL:
<http://home7.swipnet.se/~w-78067/www.hir.nu>
- Han, Jiawei and Kamber, Micheline. (2001). *Data Mining: concepts and Techniques*. San Fransisco; Morgan kufman Publishers.
- Hand, D.J., Mannila, H. and Smyth, P. (2005). *Principles of Data Mining*. Available at URL: <http://mitpress.mit.edu/books/chapters/026208290Xchap1.pdf>
- Henok Wobishet (2002). *The Application of Data Mining to Support Customer Relationship Management at Ethiopian Airlines*. Unpublished Master's Thesis. Addis Ababa University. Addis Ababa.
- Hong, S. and Weiss, S. (2003). *Advances in Predictive Models for Data Mining*. IBM T. J. Watson Research Center. Available at URL:
http://www.research.ibm.com/dar/papers/pdf/weiss_prli_2000.pdf
- Klerfors, D. (1998). *Artificial Neural Network*. Available at URL:
<http://www.burks.brighton.ac.uk/burks/foldoc/42/7.htm>
- Leul Woldu. (2003). *The Application of data mining in crime prevention: The case of Oromia Police Commission*. Master's Thesis. Addis Ababa University. Addis Ababa.
- Levin, Nissan and Jacob Zahavi, (1999). *Data Mining*. Available at URL:
<http://www.kdnvggets.com>
- Lloyd-williams, cichael (1997). "Discovering the hidden secrets in your data- the data mining approach to information" *Information Research*, 3(2). Available at URL:
<http://informationr.net/ir/3-2/paper36.html>
- Mangold, John (1998). *Turning straw into Gold: How Data mining Transforms Data from an expense into an Asset*. DM Review Magazine.
- Mao, R. (2001). *Adaptive-Fp: An Efficient and Effective Method for Multi-Level Multi-Dimensional Frequent Pattern Mining*. Unpublished Master's Thesis. Simon

- Fraser University. Available at URL:
<http://www.sal.cs.uiuc.edu/~hanj/pubs/theses.html>
- Mekitew Mola.(2000). *Traffic Information System: The case of Addis Ababa Traffic Police Department*. Master's Thesis. Addis Ababa University. Addis Ababa.
- Moss, Larisa T. (2003). *Defining Data Mining*. Available at URL:
<http://www.businessintelligence.com/ex/asp/code.b4/xc/article.html>
- Moxon, B. (1996). *Defining Data Mining*. California. Miller Freeman, Inc.
- Plate, Tony et al. (1997). *A comparison between neural networks and other statistical techniques for modeling the relationship between tobacco and alcohol and cancer*. Available at URL: <http://citseer.nj.nec.com/plate96comparison.html>
- Road Transport Authority (2005). *Zena Road Transport Authority*. Biannual Magazine, No. 1
- SafeCarGuide (2004). *International Injury & Fatality Statistics*. Available at URL:
<http://www.safecarguide.com/exp/statistics/statistics.htm>
- Shegaw Anagaw.(2002). *Application of data mining technology to predict child mortality patterns: The case of Butajira Rural Health Project (BRHP)*. Master's Thesis. Addis Ababa University, Addis Ababa.
- Sima, J. (1998). *Introduction to Neural Networks*. Technical Report, Academy of Science of the Czech Republic. Available at URL:
http://neuron.tuke.sk/math.chtf.stuba.sk/pub/vlado/NN_books_texts/sima_introduction.pdf
- Siebes A.(2004). *Data Mining Projects*. Available at URL:
<http://homepages.cwi.nl/~arn/college/college10/webcoll10.ps.gz>
- StatSoft (2003). *Neural Networks*. Available at URL:
<http://www.StatSoftinc.com/textbook/steneunet.html>
- Stergion, C. and Siganos, M. (1996). *Neural Networks*. Available at URL:
<http://www.doc.ic.ac.uk>
- Taddele, D. and Larson, C.P (1991). *The occurrence and driver characteristics associated with motor vehicle injuries in Addis Ababa, Ethiopia*. Journal of Tropical Medicine and Hygiene,94, 395-400. ♣
- Tesfaye Hintsay (2002). *Predictive Modeling Using Data Mining Techniques in Support of Insurance Risk Assessment*. Unpublished Master's Thesis. Addis Ababa University. Addis Ababa.
- Thearling, K. (2003). *An introduction to data mining*.
<http://www.thearling.com/text/dmwhite/dmwhite.htm>

Tracy, D. (2000). *Data Mining for Golden Opportunities*. Sandhills Publishing Company.

Two Crows Corporation (1999). *Introduction to Data Mining and KnowledgeDiscovery*.

Available at URL: <http://www.twocrows.com/intro-dm.pdf>

WHO (2004). *World report on road traffic injury prevention*. Switzerland. Geneva.

Witten, I.H and Frank, E. (2000). *Data Mining: practical machine learning tools and techniques with java implementations*. Morgan Kaufmann publishers. San-Francisco.

Wondwossen Mulugeta (1999). *Correlates of car traffic accident: the case of Addis Ababa in 1990 E.C.* Unpublished paper. Addis Ababa University, Addis Ababa. ✓

Appendices

Annex 1: Rules Generated

Generic Rules

RULE #1

(Whole Tree)

Injury_Severity = F-Injury 0.0781115879828
Injury_Severity = P-damage 0.676394849785
Injury_Severity = Serious_Injury 0.0815450643777
Injury_Severity = Sli-Injury 0.163948497854

RULE #2

if

Acci_Type = Turn Over

then

Injury_Severity = F-Injury 0.3
Injury_Severity = P-damage 0.5
Injury_Severity = Serious_Injury 0.0714285714286
Injury_Severity = Sli-Injury 0.128571428571

RULE #3

if

Acci_Type = Vehicle-FixObj

then

Injury_Severity = F-Injury 0.093896713615
Injury_Severity = P-damage 0.859154929577
Injury_Severity = Serious_Injury 0
Injury_Severity = Sli-Injury 0.0469483568075

RULE #4

if

Acci_Type = Vehicle-Vehicle

then

Injury_Severity = F-Injury 0.0152838427948
Injury_Severity = P-damage 0.901746724891
Injury_Severity = Serious_Injury 0.0218340611354
Injury_Severity = Sli-Injury 0.061135371179

RULE #5

if

Acci_Type = others

then

Injury_Severity = F-Injury 0.0480769230769
Injury_Severity = P-damage 0.951923076923
Injury_Severity = Serious_Injury 0
Injury_Severity = Sli-Injury 0

RULE #6

if

Acci_Type = vehicle_Peds

then

Injury_Severity = F-Injury 0.11875
Injury_Severity = P-damage 0.18125
Injury_Severity = Serious_Injury 0.25
Injury_Severity = Sli-Injury 0.45

RULE #7

```

if
    Acci_Type = Turn Over
    Acci_Cause = Denying pedestrians priority, Driving with alchool,
    Inappropriate Preceding and Turning or Over Speeding
then
    Injury_Severity = F-Injury 0.535714285714
    Injury_Severity = P-damage 0.321428571429
    Injury_Severity = Serious_Injury 0.0357142857143
    Injury_Severity = Sli-Injury 0.107142857143

RULE #8
if
    Acci_Type = Turn Over
    Acci_Cause = Driving on the left side or Others:Breaking Traffic
rules
then
    Injury_Severity = F-Injury 0.1875
    Injury_Severity = P-damage 0.8125
    Injury_Severity = Serious_Injury 0
    Injury_Severity = Sli-Injury 0

RULE #9
if
    Acci_Type = Turn Over
    Acci_Cause = Not keeping appropraite distance
then
    Injury_Severity = F-Injury 0
    Injury_Severity = P-damage 0
    Injury_Severity = Serious_Injury 0.4
    Injury_Severity = Sli-Injury 0.6

RULE #10
if
    Acci_Type = Turn Over
    Acci_Cause = Denying pedestrians priority, Driving with alchool,
    Inappropriate Preceding and Turning or Over Speeding
    Vehic_Type = Automobil, Others or Taxi
then
    Injury_Severity = F-Injury 0.65
    Injury_Severity = P-damage 0.2
    Injury_Severity = Serious_Injury 0
    Injury_Severity = Sli-Injury 0.15

RULE #11
if
    Acci_Type = Turn Over
    Acci_Cause = Denying pedestrians priority, Driving with alchool,
    Inappropriate Preceding and Turning or Over Speeding
    Vehic_Type = Bus or Truck
then
    Injury_Severity = F-Injury 0.25
    Injury_Severity = P-damage 0.625
    Injury_Severity = Serious_Injury 0.125
    Injury_Severity = Sli-Injury 0

RULE #12
if
    Acci_Type = Turn Over
    Acci_Cause = Not keeping appropraite distance
    Vehic_Age_Cat = B or D
then

```

```

    Injury_Severity = F-Injury  0
    Injury_Severity = P-damage  0
    Injury_Severity = Serious_Injury  0.8
    Injury_Severity = Sli-Injury  0.2

RULE #13
if
    Acci_Type = Turn Over
    Acci_Cause = Not keeping appropriate distance
    Vehic_Age_Cat = C or F
then
    Injury_Severity = F-Injury  0
    Injury_Severity = P-damage  0
    Injury_Severity = Serious_Injury  0
    Injury_Severity = Sli-Injury  1

RULE #14
if
    Acci_Type = Vehicle-FixObj
    Acci_Cause = Denying pedestrians priority or Driving with alcohol
then
    Injury_Severity = F-Injury  0.322580645161
    Injury_Severity = P-damage  0.483870967742
    Injury_Severity = Serious_Injury  0
    Injury_Severity = Sli-Injury  0.193548387097

RULE #15
if
    Acci_Type = Vehicle-FixObj
    Acci_Cause = Driving on the left side, Inappropriate Preceding and
Turning, Not keeping appropriate distance, Others:Breaking Traffic rules or
Over Speeding
then
    Injury_Severity = F-Injury  0.0549450549451
    Injury_Severity = P-damage  0.923076923077
    Injury_Severity = Serious_Injury  0
    Injury_Severity = Sli-Injury  0.021978021978

RULE #16
if
    Acci_Type = Vehicle-FixObj
    Acci_Cause = Denying pedestrians priority or Driving with alcohol
    Driv_Exp_Cat = B, C or F
then
    Injury_Severity = F-Injury  0.05555555555556
    Injury_Severity = P-damage  0.6666666666667
    Injury_Severity = Serious_Injury  0
    Injury_Severity = Sli-Injury  0.2777777777778

RULE #17
if
    Acci_Type = Vehicle-FixObj
    Acci_Cause = Denying pedestrians priority or Driving with alcohol
    Driv_Exp_Cat = D
then
    Injury_Severity = F-Injury  0.692307692308
    Injury_Severity = P-damage  0.230769230769
    Injury_Severity = Serious_Injury  0
    Injury_Severity = Sli-Injury  0.0769230769231

RULE #18

```

```

if
    Acci_Type = Vehicle-FixObj
    Acci_Cause = Driving on the left side, Inappropriate Preceding and
Turning, Not keeping appropriate distance, Others:Breaking Traffic rules or
Over Speeding
    Light_Cond = Dark or Day light
then
    Injury_Severity = F-Injury 0.0564971751412
    Injury_Severity = P-damage 0.926553672316
    Injury_Severity = Serious_Injury 0
    Injury_Severity = Sli-Injury 0.0169491525424

RULE #19
if
    Acci_Type = Vehicle-FixObj
    Acci_Cause = Driving on the left side, Inappropriate Preceding and
Turning, Not keeping appropriate distance, Others:Breaking Traffic rules or
Over Speeding
    Light_Cond = Road light
then
    Injury_Severity = F-Injury 0
    Injury_Severity = P-damage 0.8
    Injury_Severity = Serious_Injury 0
    Injury_Severity = Sli-Injury 0.2

RULE #20
if
    Acci_Type = Vehicle-Vehicle
    Acci_Cause = Denying pedestrians priority
then
    Injury_Severity = F-Injury 0
    Injury_Severity = P-damage 0.754716981132
    Injury_Severity = Serious_Injury 0.0377358490566
    Injury_Severity = Sli-Injury 0.207547169811

RULE #21
if
    Acci_Type = Vehicle-Vehicle
    Acci_Cause = Driving on the left side, Inappropriate Preceding and
Turning, Not keeping appropriate distance, Others:Breaking Traffic rules or
Over Speeding
then
    Injury_Severity = F-Injury 0.0102827763496
    Injury_Severity = P-damage 0.928020565553
    Injury_Severity = Serious_Injury 0.0179948586118
    Injury_Severity = Sli-Injury 0.0437017994859

RULE #22
if
    Acci_Type = Vehicle-Vehicle
    Acci_Cause = Driving with alcohol
then
    Injury_Severity = F-Injury 0.1875
    Injury_Severity = P-damage 0.75
    Injury_Severity = Serious_Injury 0.0625
    Injury_Severity = Sli-Injury 0

RULE #23
if
    Acci_Type = Vehicle-Vehicle

```



```

    Acci_Cause = Driving on the left side, Inappropriate Preceding and
    Turning, Not keeping appropriate distance, Others:Breaking Traffic rules or
    Over Speeding
    Vehic_Age_Cat = B, C or D
  then
    Injury_Severity = F-Injury 0.0106761565836
    Injury_Severity = P-damage 0.903914590747
    Injury_Severity = Serious_Injury 0.0249110320285
    Injury_Severity = Sli-Injury 0.0604982206406

RULE #24
if
  Acci_Type = Vehicle-Vehicle
  Acci_Cause = Driving on the left side, Inappropriate Preceding and
  Turning, Not keeping appropriate distance, Others:Breaking Traffic rules or
  Over Speeding
  Vehic_Age_Cat = F
then
  Injury_Severity = F-Injury 0.00925925925926
  Injury_Severity = P-damage 0.990740740741
  Injury_Severity = Serious_Injury 0
  Injury_Severity = Sli-Injury 0

RULE #25
if
  Acci_Type = others
  Acci_Cause = Denying pedestrians priority, Driving on the left side,
  Driving with alcohol, Inappropriate Preceding and Turning, Not keeping
  appropriate distance or Others:Breaking Traffic rules
then
  Injury_Severity = F-Injury 0.0238095238095
  Injury_Severity = P-damage 0.97619047619
  Injury_Severity = Serious_Injury 0
  Injury_Severity = Sli-Injury 0

RULE #26
if
  Acci_Type = others
  Acci_Cause = Over Speeding
then
  Injury_Severity = F-Injury 0.15
  Injury_Severity = P-damage 0.85
  Injury_Severity = Serious_Injury 0
  Injury_Severity = Sli-Injury 0

RULE #27
if
  Acci_Type = others
  Acci_Cause = Denying pedestrians priority, Driving on the left side,
  Driving with alcohol, Inappropriate Preceding and Turning, Not keeping
  appropriate distance or Others:Breaking Traffic rules
  Vehic_Type = Automobil, Bus, Moter Cycle, Others or Truck
then
  Injury_Severity = F-Injury 0
  Injury_Severity = P-damage 1
  Injury_Severity = Serious_Injury 0
  Injury_Severity = Sli-Injury 0

RULE #28
if
  Acci_Type = others

```

```

    Acci_Cause = Denying pedestrians priority, Driving on the left side,
    Driving with alchool, Inappropriate Preceding and Turning, Not keeping
    appraite distance or Others:Breaking Traffic rules
    Vehic_Type = Taxi
then
    Injury_Severity = F-Injury 0.0869565217391
    Injury_Severity = P-damage 0.913043478261
    Injury_Severity = Serious_Injury 0
    Injury_Severity = Sli-Injury 0

RULE #29
if
    Acci_Type = vehicle_Peds
    Acci_Cause = Denying pedestrians priority or Not keeping appraite
distance
then
    Injury_Severity = F-Injury 0.155405405405
    Injury_Severity = P-damage 0.00675675675676
    Injury_Severity = Serious_Injury 0.371621621622
    Injury_Severity = Sli-Injury 0.466216216216

RULE #30
if
    Acci_Type = vehicle_Peds
    Acci_Cause = Driving on the left side or Inappropriate Preceding and
Turning
then
    Injury_Severity = F-Injury 0
    Injury_Severity = P-damage 0.302631578947
    Injury_Severity = Serious_Injury 0.118421052632
    Injury_Severity = Sli-Injury 0.578947368421

RULE #31
if
    Acci_Type = vehicle_Peds
    Acci_Cause = Driving with alchool
then
    Injury_Severity = F-Injury 0.333333333333
    Injury_Severity = P-damage 0.0833333333333
    Injury_Severity = Serious_Injury 0.375
    Injury_Severity = Sli-Injury 0.208333333333

RULE #32
if
    Acci_Type = vehicle_Peds
    Acci_Cause = Others:Breaking Traffic rules
then
    Injury_Severity = F-Injury 0
    Injury_Severity = P-damage 1
    Injury_Severity = Serious_Injury 0
    Injury_Severity = Sli-Injury 0

RULE #33
if
    Acci_Type = vehicle_Peds
    Acci_Cause = Over Speeding
then
    Injury_Severity = F-Injury 0.137254901961
    Injury_Severity = P-damage 0.21568627451
    Injury_Severity = Serious_Injury 0.137254901961
    Injury_Severity = Sli-Injury 0.509803921569

```

```

RULE #34
if
    Acci_Type = vehicle_Peds
    Acci_Cause = Denying pedestrians priority or Not keeping appropriate
distance
    Road_Cond = Dry
then
    Injury_Severity = F-Injury 0.112781954887
    Injury_Severity = P-damage 0.00751879699248
    Injury_Severity = Serious_Injury 0.383458646617
    Injury_Severity = Sli-Injury 0.496240601504

RULE #35
if
    Acci_Type = vehicle_Peds
    Acci_Cause = Denying pedestrians priority or Not keeping appropriate
distance
    Road_Cond = slippery
then
    Injury_Severity = F-Injury 0.533333333333
    Injury_Severity = P-damage 0
    Injury_Severity = Serious_Injury 0.266666666667
    Injury_Severity = Sli-Injury 0.2

RULE #36
if
    Acci_Type = vehicle_Peds
    Acci_Cause = Driving on the left side or Inappropriate Preceding and
Turning
    Driv_Exp_Cat = B or C
then
    Injury_Severity = F-Injury 0
    Injury_Severity = P-damage 0.1875
    Injury_Severity = Serious_Injury 0.0625
    Injury_Severity = Sli-Injury 0.75

RULE #37
if
    Acci_Type = vehicle_Peds
    Acci_Cause = Driving on the left side or Inappropriate Preceding and
Turning
    Driv_Exp_Cat = D or F
then
    Injury_Severity = F-Injury 0
    Injury_Severity = P-damage 0.386363636364
    Injury_Severity = Serious_Injury 0.159090909091
    Injury_Severity = Sli-Injury 0.454545454545

RULE #38
if
    Acci_Type = vehicle_Peds
    Acci_Cause = Driving with alcohol
    Driv_Exp_Cat = B or F
then
    Injury_Severity = F-Injury 0.142857142857
    Injury_Severity = P-damage 0.285714285714
    Injury_Severity = Serious_Injury 0.571428571429
    Injury_Severity = Sli-Injury 0

RULE #39

```

```

if
    Acci_Type = vehicle_Peds
    Acci_Cause = Driving with alchool
    Driv_Exp_Cat = C or D
then
    Injury_Severity = F-Injury 0.411764705882
    Injury_Severity = P-damage 0
    Injury_Severity = Serious_Injury 0.294117647059
    Injury_Severity = Sli-Injury 0.294117647059

RULE #40
if
    Acci_Type = vehicle_Peds
    Acci_Cause = Over Speeding
    Vehic_Age_Cat = B or D
then
    Injury_Severity = F-Injury 0.175
    Injury_Severity = P-damage 0.275
    Injury_Severity = Serious_Injury 0.15
    Injury_Severity = Sli-Injury 0.4

RULE #41
if
    Acci_Type = vehicle_Peds
    Acci_Cause = Over Speeding
    Vehic_Age_Cat = C or F
then
    Injury_Severity = F-Injury 0
    Injury_Severity = P-damage 0
    Injury_Severity = Serious_Injury 0.0909090909091
    Injury_Severity = Sli-Injury 0.909090909091

```

Annex 2: Accident data recording form



08/54

U. S. DEPT. OF AGRICULTURE
BUREAU OF PLANT INDUSTRY
WASHINGTON, D. C.

19
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531

12. 1. 2. 5. 7 27. 11 መ. 18 187

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											

35 አዲስ የፖሊስ ሰዎች

[illegible]

28 11/16/00 ...
11/16/00 11/16/00 11/16/00

100
1000

1000	+	100
------	---	-----

1000 + 100 = 1100

1100	+	100
------	---	-----

1100 + 100 = 1200

25
03-127- 88.00 7.87- 07311.0 0.40m3

0	000
---	-----

30 እርሳን

品名	数量	単位	金額	備考
1	1	個	100	
2	1	個	100	
3	1	個	100	
4	1	個	100	
5	1	個	100	
6	1	個	100	
7	1	個	100	
8	1	個	100	
9	1	個	100	
10	1	個	100	
11	1	個	100	
12	1	個	100	
13	1	個	100	
14	1	個	100	
15	1	個	100	
16	1	個	100	
17	1	個	100	
18	1	個	100	
19	1	個	100	
20	1	個	100	
21	1	個	100	
22	1	個	100	
23	1	個	100	
24	1	個	100	
25	1	個	100	
26	1	個	100	
27	1	個	100	
28	1	個	100	
29	1	個	100	
30	1	個	100	
31	1	個	100	
32	1	個	100	
33	1	個	100	
34	1	個	100	
35	1	個	100	
36	1	個	100	
37	1	個	100	
38	1	個	100	
39	1	個	100	
40	1	個	100	
41	1	個	100	
42	1	個	100	
43	1	個	100	
44	1	個	100	
45	1	個	100	
46	1	個	100	
47	1	個	100	
48	1	個	100	
49	1	個	100	
50	1	個	100	
51	1	個	100	
52	1	個	100	
53	1	個	100	
54	1	個	100	
55	1	個	100	
56	1	個	100	
57	1	個	100	
58	1	個	100	
59	1	個	100	
60	1	個	100	
61	1	個	100	
62	1	個	100	
63	1	個	100	
64	1	個	100	
65	1	個	100	
66	1	個	100	
67	1	個	100	
68	1	個	100	
69	1	個	100	
70	1	個	100	
71	1	個	100	
72	1	個	100	
73	1	個	100	
74	1	個	100	
75	1	個	100	
76	1	個	100	
77	1	個	100	
78	1	個	100	
79	1	個	100	
80	1	個	100	
81	1	個	100	
82	1	個	100	
83	1	個	100	
84	1	個	100	
85	1	個	100	
86	1	個	100	
87	1	個	100	
88	1	個	100	
89	1	個	100	
90	1	個	100	
91	1	個	100	
92	1	個	100	
93	1	個	100	
94	1	個	100	
95	1	個	100	
96	1	個	100	
97	1	個	100	
98	1	個	100	
99	1	個	100	
100	1	個	100	

105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 543
 544
 545
 546
 547
 548
 549
 550
 551
 552
 553
 554
 555
 556
 557
 558
 559
 560
 561
 562
 563
 564
 565
 566
 567
 568
 569
 570
 571
 572
 573
 574
 575
 576
 577
 578
 579
 580
 581
 582
 583
 584
 585
 586
 587
 588
 589
 590
 591
 592
 593
 594
 595
 596
 597
 598
 599
 600
 601
 602
 603
 604
 605
 606
 607
 608
 609
 610
 611
 612
 613
 614
 615
 616

1000

9 እዚን የፈጸሙ ተሽከርካሪዎች ዓይነት

[illegible]

10 እዚጋ የፈጸሙ ተሽከርካሪዎች ንብረትነት

[illegible]

11 የተሸከርካሪው ጉድለት

ተ.ረ. ቁ.	የተጠቃሚው ፊርማ	አደጋው ያለበትበት ጊዜ			የጥፋት ቁ.	ጠቅላላ
		የጥፋት	የሰዓት	የቀን		
1	የጥፋት ፊርማ					
2	የጥፋት ፊርማ					
3	የሰዓት ጥፋት					
4	የጥፋት ፊርማ					
5	የሰዓት ጥፋት					
6	የጥፋት ፊርማ					
7	የሰዓት ጥፋት					

12 አዲስ የተጻፈ መዝሙር መገንጠል

ተረፍ ቁጥር	የግንባታው ስም	አገልግሎት ያስጠቀሱት ጉዳት				ጥቅም
		የጉዳት	ከሰድስት ወራት በላይ	የአስተዳደር ጉዳት	ተግባራዊ ጉዳት	
1	ከግብርና የግንባታ ጉዳት					
2	አገልግሎት የሰጠ የጉዳት					
3	የግብርና ጉዳት					
4	የጉዳት ጉዳት					

13 እደጋው የተከፈለውበት አካባቢ

15 ARZUT HAN		ARZUT HAN			
DATE	NAME	ARZUT HAN			
		DATE	NAME	DATE	NAME
1	ARZUT HAN				
2	ARZUT HAN				
3	ARZUT HAN				
4	ARZUT HAN				
5	ARZUT HAN				
6	ARZUT HAN				
7	ARZUT HAN				
8	ARZUT HAN				
9	ARZUT HAN				
10	ARZUT HAN				
11	ARZUT HAN				

15 တောတဲက အနာမိဘ

Date		Time		Location		Weather		Remarks	
Day	Month	Hour	Minute	Place	Altitude	Temp	Wind	Clouds	Notes
1	1	10	15	AT	PA				
2	1	10	15	AT	PA				
3	1	10	15	AT	PA				
4	1	10	15	AT	PA				
5	1	10	15	AT	PA				
6	1	10	15	AT	PA				
7	1	10	15	AT	PA				
8	1	10	15	AT	PA				
9	1	10	15	AT	PA				

፲፭ የመኒግ. ንግሥት ዓይነት

2020-2021		2021-2022				2022-2023	
2020-2021	2021-2022	2022-2023	2024-2025	2026-2027	2028-2029	2030-2031	2032-2033
1	2020-2021						
2	2021-2022						
3	2022-2023						
4	2024-2025						
5	2026-2027						
6	2028-2029						
7	2030-2031						
8	2032-2033						

19 44077 0-25

Date	Policy No.	Amount of Payment			Total
		Jan	Feb	Mar	
1	1000000				
2	1000000				
3	1000000				
4	1000000				

14 800727 11428

10 Yaw Yaw Aha		Aha Yaw Aha Yaw			299C
299C	299C	299C	299C	299C	
1	299C	299C	299C	299C	299C
2	299C	299C	299C	299C	299C
3	299C	299C	299C	299C	299C
4	299C	299C	299C	299C	299C
5	299C	299C	299C	299C	299C

16 ဒုတိယကမ္ဘာပွဲတွင် အနိုင်ရရှိခဲ့သော နိုင်ငံများ

No	Pictogram	ARCO Pictogram				Ref
		Ref	WLS Pictogram	WLS Pictogram	WLS Pictogram	
1						
2						
3						
4						
5						
6						
7						
8						

18 १००७१२ ५३५

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

DATE	PROJECT NAME	ARROW PROJECTS T-47				REMARKS
		TYPE	STATUS	COMPLETION	PERCENTAGE	
5	ARROW PROJECTS T-47					
6	ARROW PROJECTS T-47					
7	ARROW PROJECTS T-47					

Declaration

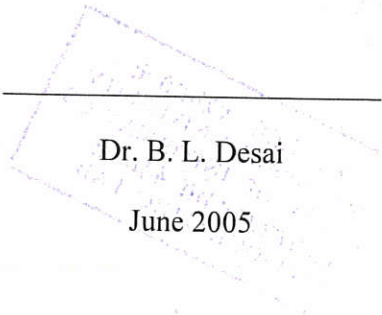
I, the undersigned, declare that this thesis is my original work and has not been presented as a partial degree requirement for a degree in any other university and that all sources of materials used for the thesis have been duly acknowledged.



Tibebe Beshah Tesema

June 2005

The thesis has been submitted for examination with my approval as university advisor



Dr. B. L. Desai

June 2005

