ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

FACULITY OF INFORMATICS

DEPARTMENT OF INFORMATION SCIENCE

THE APPLICATION OF DATA MINING IN CRIME PREVENTION: THE CASE OF OROMIA POLICE COMMISSION

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

By

LEUL WOLDU ASEGEHGN

JULY 2003

ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

FACULITY OF INFORMATICS

DEPARTMENT OF INFORMATION SCIENCE

THE APPLICATION OF DATA MINING IN CRIME PREVENTION: THE CASE OF OROMIA POLICE COMMISSION

By

Leul Woldu

Name and Signiture of Members of the Examining Board

Chairman, Examining Board ______________________________________

W/t Ethiopia Tadesse, Advisor ______________________________________

Ato Dereje Teferi, Advisor ______________________________________

External Examiner ______________________________________

DEDICATION

Dedicated to my beloved father and mother.

ACKNOWLEDGEMENT

I would like to thank my advisors W/t Ethiopia Tadesse, W/o Rahel Bekele, and Ato Dereje Teferi for their constructive suggestions and overall guidance.

I would also thank all the community of the Department of Information Science for their support through out my stay in the department.

The staff of Research and Information Department of the Federal Police Commission deserves special thanks for their comments on the analysis part of this work.

I would also thank the Commissioner of the Oromia Police Commission and the staffs of the Computing Department for their good will for letting me use the data for this study and provide me with information about the crime prevention process in the respective region.

Table of Content

CHAPTER THREE

CRIME AND CRIME PREVENTION

CHAPTER FOUR

EXPERIMENTATION

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

LIST OF FIGURES

Page

LIST OF TABLES

Page

Abstract

Law enforcement agencies like that of police today are faced with large volume of data that must be processed and transformed into useful information and hence data mining can greatly improve crime analysis and aid in reducing and preventing crime.

The purpose of this study is to explore the applicability of data mining technique in the efforts of crime prevention with particular emphasis to the Oromia Police Commission and to build a model that could help to extract crime patterns. With this objective decision trees and neural network were employed to classify crime records on the basis of the values of attributes crime label (***CrimeLabel***) and crime scene (***SceneLabel***).

Results of the experiments have shown that decision tree has classified crime records at an accuracy rate of 94 percent when the attribute ***CrimeLabel*** is used as a basis for classification. Where as, in the same experiment, the accuracy rate of neural networks is 92.5 percent. On the other hand, in the case of classification of records on the values of the attribute ***SceneLabel*** decision tree has shown an accuracy rate of 85 percent while neural network revealed 80 percent.

In both experiments the output indicated that decision tree performed better. Besides, decision tree generated understandable rules that could be easily presented in human language and thus police officers can make use of these rules for designing crime prevention strategies. Thus, this experiment has proved that data mining is valuable to support the crime prevention process and particularly, decision trees seem more appropriate for the domain problem.

CHAPTER ONE

INTRODUCTION

1.1 Background

Human beings need to live and work in a place where they are safe. They want to ensure that there is a concerned body that protects their lives as well as their properties from potential hazards. In fact, one of the main functions of any government is to ensure that law and order for the security of its citizens are put in place (Wilson, 1963). In other words, far from formulation and enactment of laws for the prevention of crime, governments must establish agencies and organizations, which enforce these laws.

Accordingly, in many countries, if not all, there are organizations such as courts, prosecutions and police, which are responsible for the maintenance of law and order.

In Ethiopia, the police organization was established in 1942 under proclamation No. 6 as autonomous institution with the responsibility of preventing and investigating crime incidents. In 1966 the police institution was put under the then Ministry of Interior. Since its establishment, the police organization structure has been extended to the lower administration level, which is "woreda" and sometimes a "kebele" (Mesfin, 1999).

Due to the new constitution adopted in 1994, the Ethiopian government has been exercising federal political system and hence both the structure and authority of the police is changed accordingly. Based on the proclamation No. 1 article 50, regional governments are duly responsible to establish all the necessary administrative levels in their respective region. In light with this political sphere regional states established their own police institutions at a level of commission. As a result the Oromia Police Commission has become responsible to maintain law and order in its respective region together with other concerned agencies.

Along with the prevention and investigation of crime, police makes use of previous crime reports and data as an input for the formulation of crime prevention policies and strategic plans (Wilson, 1963). It is obvious from the out set that to make use of data and records, relevant data have to be kept and managed properly. For this reason the Oromia Police Commission have been collecting criminal records since its establishment and have maintained numerous criminal records consisting of fingerprints, names, photographs, and general descriptions of criminals.

Developments in the information and communication technologies have made it possible for organizations to collect, store and, manipulate massive amount of data (Corcoran and Ware, 2001). According to the Regional crime Analysis Program (2000), most law enforcement agencies, particularly police, today are faced with huge amount of data that must be processed and transformed into meaningful information. This indicates that police necessitate making use of these information and communication technologies to collect, store, process and transform large volume of raw data with the ultimate goal of extracting meaningful information.

However, in the case of Ethiopia, including Oromia Police Commission, until recent years there were no modern tools and techniques employed to facilitate the handling and processing of records (Mesfin, 1999). It is only in the last few years the commission has began to develop databases for departments that possess bulky data. The databases developed so far includes criminal database, traffic database, and personnel database.

The criminal database, which is used for this study, contains information about criminals and the crime committed by each criminal and it consists of about 50,000 records with more than 165 attributes. Criminal records from each "zone" are sent manually to the commission by filling the criminals profile form and then the computing section of the commission is responsible to enter all the records of the region to the database.

Crime records, here, refers to the data gathered and used by the police other than the administrative records of the police itself. According to the International City Managers Association (1961), the classes of crime records can be divided into three broad categories. These are:

- Cases or complaint records; which include information regarding complaints and reports received by the police from citizens and other agencies, and actions indicated by the police.
- Arrest records: This class contains all records about arrested offenders including their control, and disposition. The scope of arrest records covers every step from the person is arrested till s/he released.
- Personal Identification records; this major division of police records consists of records dealing with personal identification of criminals.

Among the above classes of criminal records maintained by the Oromia Police Commission, personal identification records are used for this study. The reason why these records are chosen is due to the fact that these are the only records stored electronically. Personal identification records include attributes of criminals such as name, age, sex, educational status, occupation, ethnic, address and so on.

In spite of the existence of databases with relatively large volume of data, the Oromia Police Commission has not yet exploited the information embedded in the data. The commissioner of the police commission noted that the current system of the commission is a manual system except the fact that the aforementioned databases are being used as repository. He added that crime prevention measures are being taken based on crime incidents although it would have been based on crime trends. This indicates that, currently, the crime prevention approach is based on the crime reports incoming to the police, which is a reactive approach although it would have been proactive. In fact, there are areas such as market places and offices where permanent patrolling is put in place; and in both cases the crime control effort does not depend on the study of the previous records.

There are several factors that contribute to this underutilization of the database systems. According to the head of the computing section of the Oromia Police Commission, one reason is lack of knowledge on what could be done using these databases or deficiency of appropriate tools that could make use of these databases and the other problem is due to lack of skilled manpower.

In areas where large volume of data is found, like that of police, information support technologies have become a mandatory to manage and process information for decision-making. Brown (2003) stated that law enforcement agencies are increasingly acquiring database management systems (DMBS), geographic information systems (GIS) and data mining to support their crime analytic capabilities. He added that crime analysts are using these systems to search through data, link records, and plot the results on maps.

Data mining is one such tool that has evolved to play a role as an instrument to discovery patterns buried in large databases. Data mining is the "exploration and analysis, by automatic or semiautomatic means, of large quantities of data in order to discover meaningful patterns and rules" (Berry and Linoff, 1997).

Therefore, police can deploy data mining tools to analyze and manage its bulky data so as to discover crime patterns. According to the Regional crime Analysis Program (2000), because of the complex nature of crime and enormous data there is no field in greater need of data mining technology than law enforcement.

Data mining tells important things unknown to the user or what is going to happen in the future. The central element to data mining is modeling. Modeling is simply the act of building a model in one situation where you know the answer and then applying it to another new situation (Thearling, 2003). There are a number of modeling techniques and the most commonly used data mining techniques are: Decision tree, neural networks, genetic algorithms, nearest neighbor method and rule induction (Ibid) (for details of these techniques refer to section 2.7 0f chapter 2).

1.2 Statement of the Problem

Crime is a complex social phenomenon and its cost is increasing due to a number of societal changes and the like, and hence, law enforcement organizations like that of police need to learn the factors that constitute higher crime trends (Wilson, 1963). To curb this social evil there is always a need for prudent crime prevention strategies and policies. Understanding and processing of criminal records is one method to learn about both crime and individuals who involve in misdeeds so that police can take crime prevention measures accordingly (Brown, 2003).

By processing and digesting criminal records police can extract crime patterns that could be invaluable in the process of crime prevention. According to the Megaputer Intelligence (2002) law enforcement agencies in this case police and other government organizations can benefit from crime pattern analysis by obtaining improved crime resolution rate, optimal and fair resource allocation based on dynamically changing crime patterns, swift and more up to date information, reduced officer training time and costs, and better crime prediction as well as prevention of offences. In fact, tools and techniques are prerequisite to process and interpret large volume of data so as to make use of them for decision-making purpose.

However, in the case of Oromia Police Commission there are no modern tools and techniques that can support in managing crime records properly and efficiently. As a result almost all the decision-making processes of the commission are not supported by tools and techniques that could extract patterns from previous crime records. Consequently, training programs, resource deployment, crime prevention and investigation strategies are being pursued on the basis of crime incidents rather than crime patterns and trends. Thus, one can observe the cost of those entire activities that do not rely on sound justifications.

However, recent technologies such as data mining can be immensely important to extract meaningful patterns that could be used as a sound justification in designing training programs, resource deployment, and formulation of crime prevention and investigation strategies (Regional crime Analysis Program, 2000). This indicates that by adopting data mining techniques the Oromia Police Commission could extract crime patterns, which could facilitate and support the process of crime control. Therefore, this study is launched to identify appropriate tools as well as to develop models that could extract crime patterns from the criminal database which supports the decision making process of crime prevention.

Particularly, the study attempted:

- To discover crime patterns in urban and rural areas and explores the nature of crime and characteristics of offenders in the respective areas.
- To extract the characteristics of offenders who involve in different levels of crime. It explores the characteristics of offenders who commit serious crimes that are costly to the society, as well as criminals that involve in medium and/or low crimes.

Therefore, the purpose of this work is to extract patterns of crimes as related to human factors or characteristics from the database of the Oromia criminal record database so that the performance of the commission could be optimized. The fact that the researcher has been working with the Federal Police Commission; it provides him the opportunity to know the problems the institution is currently facing.

The data mining techniques, which are used in developing the models, are neural networks and decision tree, which are among the most commonly, used techniques in the discipline (Thearling, 2003). A decision tree is a branched tree representation where each node shows an attribute and the branches denote the different values of the attribute (Han and Kamber, 2001). Decision tree is employed as a feature selection tool as well as a classifier.

Neural networks are designed to simulate the way a simple biological neurons system is believed to operate and have the capability to learn, memorize and create relationships among data (Bigus, 1996). Mutch (1999) describes neural networks as software that builds and implements a mathematical equation, which constitutes a model capable of predicting hidden or concealed information or knowledge.

The main reason for employing neural network and decision tree is that, these techniques are the most widely used data mining techniques and are robust to handle massive data and are not sensitive to missing and noisy data (Berry and Linoff, 1997). In addition to this, the researcher is familiar with these techniques and software packages that incorporate these techniques are relatively easily available.

A wide range of companies around the glob has deployed successful application of data mining (Thearling, 2003). Although the potential of data mining is with no doubt worth coming, as to the knowledge of the researcher so far, only five researches have been conducted at the Faculty of Informatics.

The first attempt was made by Gobena (1999) that was on the application of data mining technology and techniques in the Ethiopian Airlines and this work was extended by Henock (2002). Askale (2000) and Tesfaye (2002) also conducted other researches on the application of data mining in the financial industry specifically at the Dashen Bank and the Ethiopian Insurance company respectively. Moreover, Shegaw (2002) has also assessed the potential applicability of data mining technology in the Ethiopian context with particular reference to the health sector. Hence, this research is a continuation of the data mining researches carried out so far, however, with a different area of application, which is crime prevention.

1.3 Objective of the Study

1.3.1 General Objective

The general objective of the study is to examine the potential of data mining tools and techniques in developing models that could help in the effort of crime pattern analysis with the aim of supporting the crime prevention activities at the Oromia Police Commission.

1.3.2 Specific Objectives

To accomplish the above stated general objective, the following specific objectives are developed:

- To identify the nature of crime and the crime prevention process.
- To assess the potential of data mining techniques in crime prevention.
- To explore and choose among the various data mining software that support neural networks and decision tree technique to experiment with crime records.
- To build and train as well as test the performance of the model.
- To come up with a model that is capable to extract crime patterns.

- To interpret and analyze the results of the model that how strong is the model to extract crime patterns.
- To compare the decision tree and neural network data mining techniques and select the one which performs the best.
- Finally to forward recommendations based on the findings of the study.

1.4 Research Methodology

1.4.1 Exploration of the domain problem

In order to define the research problem properly, primary data was collected by interviewing concerned officers in the police commission as well as through observation (questions rose during the interview are presented in appendix C). Then based on the information obtained from these attempts, the overall crime prevention process of the Oromia Police Commission was described.

Relevant literatures on data mining techniques and crime were reviewed. The potential of data mining in general and particularly successful data mining applications in crime prevention were assessed.

1.4.2 Identification and Selection of Target Data Set

The Criminal records database of the Oromia police commission is the principal target data set to this study. These data are found and are collected from the Oromia Police Commission.

From this database a sample data of 5,500 records were drown. The sampling technique employed is random sampling; taking 1,100 records from every 10,000 records. Then, to partition the data to training and test dataset, cross-validation and percentage split methods were used. In the case of percentage split, 4,400 (80 percent) records were used for training and the remaining 1,100 (20 percent) were used for testing the performance of the model.

1.4.3 Data Preparation

After the data are collected, orders such as processing and cleansing are imposed in order to make the data more suitable for the particular data mining software, which are used in the study. This comprises attribute selection, defining target classes (attributes for classification), handling noisy data, accounting for missing data fields, coding text valued attributes and preparing the processed data in a file format acceptable to the weak software.

1.4.4 Building and Training Models

One of the critical tasks that have to be performed at this step is selection of software that supports the data mining techniques that are to be employed in the study. In conducting this research *weka* software was employed for reasons of accessibility and familiarity. W*eka* software package has different programs for different techniques and algorithms. However, in this study only two programs (sub packages) are employed. One is J48, which is a decision tree classifier, is used for decision tree construction. The other program is neural network, which is used to train and build neural network model.

Feature selection and model building have been made iteratively by modifying the values of the parameters of decision tree and neural networks in order to improve the performance of the model. In order to evaluate the accuracy of the model test data (records that were not used during the training process) were used.

1.5 Significance of the Study

This study attempted to extract crime patterns as related to criminal attributes and hence police officials can make use of these patterns in their day-to-day battle against crime.

Police officials in the crime prevention and investigation authority of the respective region can make use of the results of this study in order to make optimal deployment of resource in crime prevention. Moreover, the out put of the study shall be used for designing appropriate training programs and crime prevention and investigation strategies.

The out put of the study shall also be used as a benchmark for police officials as well as a source of methodological approach for studies dealing on the application of data mining on crime management as well as other similar areas.

1.6 Limitation of the Study

This research has taken the features of criminals, which are recorded by police officials. However, there are features that are important to learn about the behavior of criminals such as psychological make up of criminals that are not usually recorded by police officials. For this reason, this study is confined on the features of criminals, which are recorded and readily available in the criminal record database.

When a crime is committed there are individuals who are offenders of the crime, criminals and; victims of the crime. However, this study is limited to the study of criminal characteristics to extract crime trends. The case of victims is excluded from this study deliberately.

1.7 Thesis Organization

This thesis report consists of five 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 and third chapters are devoted to literature review of data mining technology as well as crime and crime prevention respectively.

Chapter four reports the experiment of the research. It comprises training, building and validation of the models. Results of the experiment are also analyzed and interpreted. The last chapter presents conclusions and recommendations.

CHAPTER TWO

DATA MINING TECHNOLOGY

2.0 Introduction

This chapter discusses the potential of data mining to discover knowledge from huge database. It also provides a brief historical development of the field. Besides, it presents a review of different data mining techniques, common tasks and basic steps.

2.1 What is data mining?

The convergence of computing and communication has resulted in a society that highly relies on information (Witten and Frank, 2000). Technological developments that aid to collect and store vast quantities of data have enabled organizations to capture and cumulate huge amount of data in their databases, within which, large amount of valuable information is buried (Fayyad, Piatesky-Shapiro and Smyth, 1996).

It is estimated that the amount of data stored in the world's database grows every twenty months at a rate of 100% (Witten and Frank, 2000). As the volume of data increases, the proportion of information in which people could understand decreases substantially. This reveals that the level of understanding of people about the data at hand could not keep pace with the rate of generation of data in various forms, which results with increasing information gap. Consequently, people begin to realize this bottleneck and to look into possible remedies.

People usually strive to innovate new technologies in order to unearth practical problems or to satisfy their desires or needs (Tesfaye, 2002). With the aim to fix the gap between the generation of data and rate of understanding of people about it, data mining can serve as a principal tool. According to Thearling (2003), data mining is the extraction of hidden patterns and useful trends from large database. It is a robust technology with substantial potential to help organizations concentrate on the most valuable information in their database. Witten and Frank (2000) have also noted that data mining is valuable to discover implicit, potentially useful information from huge data stored in databases via building computer programs that sift through databases automatically or semi-automatically, seeking meaningful patterns. The opportunity for the application of data mining has increased significantly as databases grew extremely and new machine with searching capabilities evolved.

According to Berry and Linoff (1997), data mining usually makes sense when there is large amount of data. For this reason most of the algorithms developed for data mining purpose requires large volume of data so as to build and train models that are responsible for different tasks of data mining such as classification, clustering, prediction, association and the like (details of these data mining tasks is provided in section 2.4 of this chapter). The need for bulky data can be explained by a couple of reasons. Primarily, in the case of small databases, it is feasible to capture appealing trends and relationships by introducing traditional tools such as spreadsheets and database query. The second reason is that most data mining tools and algorithms demand large amount of training data (data used for building a model) in order to generate unbiased models. The rationale is simple and straightforward, small training data results in unreliable generalizations based on chance patterns.

One of the most frequently raised questions among researchers and the intellectual community in the field of data mining is whether there is any difference between data mining and knowledge discovery (Tesfaye, 2002). According to Fayyad, et al (1996), Knowledge Discovery in Databases (KDD) is the "non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data." On the other hand, those authors have defined data mining as a single step in the process of Knowledge Discovery that involves the application of appropriate algorithm so as to discover meaningful trends from the database under investigation.

Nevertheless, many authors (Carbone, 1997; Piatetsky-Shapiro, 2000; Han and Kamber, 2001) believe that the term data mining has become more popular in industries, in media and the database researches as a synonym for knowledge discovery and hence the two terms are used interchangeably. This view is also applicable through out this study.

Some of the most frequently used terms in the process of knowledge discovery are instance (record or example) and attribute. Instance refers to the inputs or example to the data mining process that is going to be investigated based on its features, whereas, an attribute refers to the features or characteristics of an input, now to be more specific, an instance (Witten and Frank, 2000). Whenever the data set is presented in a tabular form; usually the instances are the rows of the table and the attributes are the columns.

2.2 Historical Overview of Data mining

In order to know the world and explain natural phenomenon, people have been gathering and analyzing data. Investigating this data has brought different theories, observations, and approaches that could help understand and know the natural world and its laws. Without the aid of machines, people had been analyzing data and looking for patterns (Berry and Linoff, 1997).

However, gradually, new technologies have begun to play a vital role to facilitate storage, analysis and processing of data. Specially, the advent of computer technology has revolutionized the way in which data are managed. These new methods of looking into data as well as the keen interest to learn from data have brought disciplines like that of data mining (Thearling, 2003).

Recently data mining has been drawn the attention of intellectuals, business article writers and software developers. Although data mining is the evolution of fields with a long history such as statistics, artificial intelligence and machine learning, it evolved to become widely known in the 1990s (ANGOSS, 2003).

Much of the tools and techniques of statistics are adopted in the study of data mining. However, although statistics is very useful technique, it is not capable to address all data mining problems (Berry and Linoff, 1997). For instance, some problems may demand learning from experience and statistical methods could not address such problems. Moreover, statistics usually employs sample data (part of the population data thought to be representative) to build statistical models and this method can miss much information about the population (Thearling, 2003).

The second longest family line of data mining is artificial intelligence. This field of study is developed on the basis of heuristics in contrast to statistics. It is an attempt to apply "human-thought-like" approach to statistical problems. The application of artificial intelligence has become pervasive when computers began to provide useful power at affordable prices (Ibid).

The other field of study that contributed a lot to data mining is machine learning, which is more properly described as the hybrid of statistics and artificial intelligence (ANGOSS, 2003).

Machine learning attempts to let computer programs learn about the data they study, such that programs make different decisions based on the qualities of the studied data, using statistics for fundamental concepts, and adding more advanced artificial intelligence heuristics and algorithms to achieve its goals (Carbone, 1997). This depicts that the application of machine learning in the study of large volume of data is a radical shift not only from statistics but also from artificial intelligence via merging both fields.

From the foregoing arguments it seems plausible to conclude that data mining, in many ways, is basically the adaptation of machine learning techniques to scientific and business problems. Data mining is the union of historical and recent developments in statistics, artificial intelligence, and machine learning. The tools and techniques borrowed from these fields of studies used together to extract previously unknown patterns buried in large database. Data mining is becoming popular in science and business areas where there is large amount of data.

2.3 Common Tasks of Data Mining

In practice, data mining can accomplish about six common tasks namely; classification, estimation, prediction, association, clustering, and description. However, many of the practical problems of scientific, economic and business interests can be mapped into one of these common tasks (Berry and Linoff, 1997).

A single data mining tool or technique is not equally applicable to all the above-mentioned tasks. Based on the nature of the problem under consideration and its proximity to the main divisions of data mining tasks, one need to choose the appropriate techniques among the numerous data mining techniques (Rudjer Boskovinc, 2001).

2.3.1 Classification

Classification is one of the most common data mining tasks, which is also pervasive in human life. Human beings usually classify or categorize in order to understand and communicate about the world. For any object or instance, classes are predefined according to the value of a specific field (Berry and Linoff, 1997). Classification comprises examining the features of unseen instances and assigning it into one of the predefined classes. In fact, in the case of data mining the objects that are going to be examined comes from a database. The task of classification incorporates updating each record by filling in a field with a class code of some kind (Witten and Frank, 2000). This task starts with instances often called a training set which consists of predefined classes and are used to train as well as build a data mining model so that the model can be applied to classify unseen objects.

2.3.2 Clustering

Clustering is another task of data mining in which groups of instances or objects that belong together are sought (Han and Kamber, 2001). The only difference between classification and clustering is in the case of clustering unlike classification there are no predefined classes. As there are no predefined classes and examples, in clustering records are grouped together on the basis of similarity among the instances (Beaza-yates and Ribeiro-Neto, 2000).

2.3.3 Association

Association learning is a data mining scheme in which any affinity grouping between features is sought, not just one that predict a particular class value (Witten and Frank, 2000). Association is suitable if the problem is to extract any structure from the data. The aim of association is to examine which instances are most likely to be grouped together. Market basket analysis is a typical practical application of association. Association rules can be developed in order to determine arrangements of items on store shelves in a given supermarket so that items often bought together will be found together (Berry and Linoff, 1997).

Association, unlike classification, can predict any field; not only the class and it can forecast more than one attribute's value at a time. For this reason we can find a number of association rules than classification rules.

2.3.4 Estimation, Prediction and Description

Estimation is usually applicable for instances which have continuous valued attributes such as income, height, or credit card balance. In the case of classification it deals with discrete valued outcomes such as yes or no. Estimation task comprises estimating the number of children in a family, a family's total household income, the life time value of customers, and the likelihood that someone will respond to a balance transfer solicitation (Berry and Linoff, 1997).

Prediction is another task of data mining that is almost similar to classification and estimation. In the case of prediction the records are classified according to some predicted future behavior or estimated future value. The accuracy of prediction is not known until classification is accomplished. Historical data is employed in order to build and train a model that describes the current observed behavior. The application of this model to actual inputs gives us a prediction of the future. Common prediction problems include predicting which customers will leave in the next three months, and forecasting which telephone subscribers will order a specific service (Berry and Linoff, 1997).

The last type of data mining task is description. Sometimes the data mining problem is simply to describe what is happening in a complicated database. This includes knowing the people the products or process that are applicable and which constitute the database (Berry and Linoff, 1997).

2.4 The Process of Data Mining

Many people think that data mining is simply employing software (Tesfaye**,** 2002). However, it is rather a process that involves a series of steps to transform the data prior to mining, evaluate and interpret the modeling results.

It is the process of discovering useful patterns in large amount of data that explain past events in such a way that the results of the model can be employed to predict about the future (Thearling, 2003). Like any other problem solving methods, the task of data mining should commence with a firm definition of the problem. Defining the data mining problem firmly enables to determine appropriate data mining process and modeling technique.

According to Berry and Linoff (1997), the most commonly employed steps in the process of data mining are; Identifying source of data, preparing data for analysis, building and training a computer model, and evaluating the computer model. This process is depicted in figure 2.1.

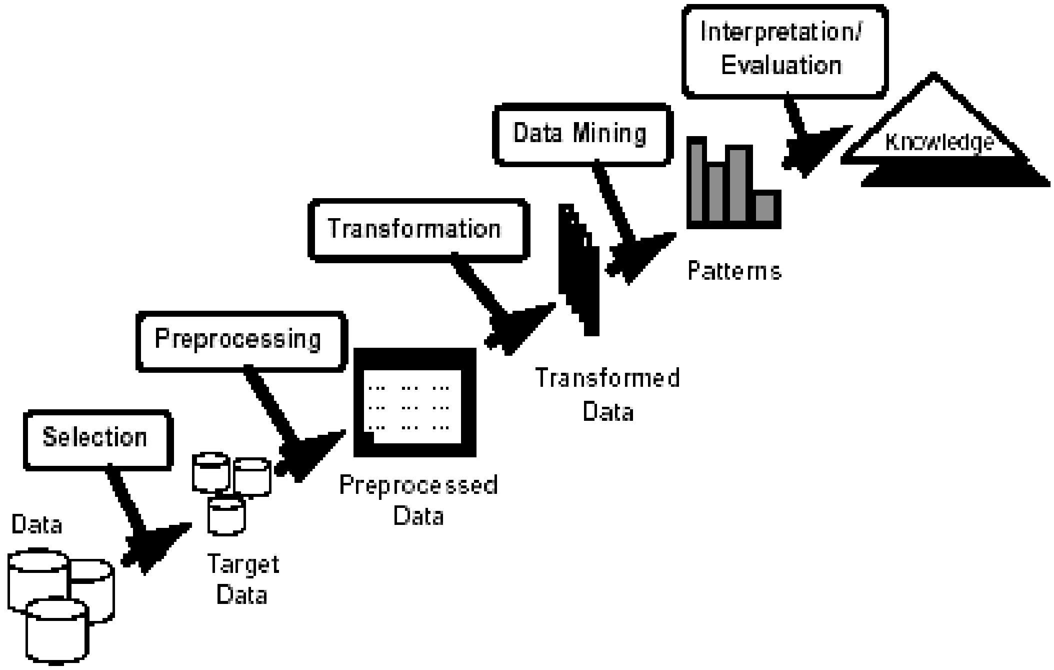

Figure 2.1 Graphical representations of the steps comprising the Knowledge Discovery Process (Fayyad, Piatetsky-shapiro and Smyth; 1996)

2.4.1 Identifying the Target Data Set (Selection)

Once the data mining problem is identified, the next step is to identify the target data set in which the solution of the problem can be found. To accomplish a successful knowledge discovery, reliable and consistent data is a prerequisite (Han and Kamber, 2001). Thus, identifying and selecting relevant and accessible data set is one of the most important steps in the process of knowledge discovery.

2.4.2 Preparing the Data for Analysis (Preprocessing and Transformation)

Real world databases usually contain incomplete, noisy and inconsistent data and such data may cause confusion for the knowledge discovery process. Hence, data cleaning is mandatory to improve the quality of data and so as to improve the accuracy and efficiency of the knowledge discovery process. Preprocessing is a routine task that usually consumes much of the efforts exerted in the entire data mining process (Han and Kamber, 2001).

To start with a data mining problem, it is important to bring all the data together into a set of instances with their corresponding attributes. Specially, in the case of relational databases (one type of data modeling technique in database management where records are placed in a tabular form), it is necessary to denormalize (merge related tables) the relations (tables) so that one can describe any relationship among attributes of a given instance (Witten and Frank, 2000). The process of integration and denormalization of relations will demand identifying primary keys (attributes that uniquely identify instances in a database) and understanding how attributes of instances are represented in different relations.

As different departments may use different methods of recording and representation, data mining applications that employ data from different computer systems needs a special care. Moreover, values of attributes that are textual, incomplete, and null value must be dealt at this step. In case where derived fields are required (like deriving age from date of birth, density from population and area) from existing ones, new fields that show the relationships that could yield useful results have to be incorporated (Berry and Linoff, 1997). This helps to capture any available information about instances as well as to simplify feeding data to the data mining software.

Data sets containing missing values, inaccurate values, duplicate data, noisy and inconsistent data have to be preprocessed before any data mining tool is applied. For instance, missing values are frequently indicated by out of entering values like negative one in a numeric field that are normally positive (Witten and Frank, 2000). Another approach to handle missing values is filling mean values, using the most probable value, filling manually, or ignoring the record containing missing values (Han and Kamber, 2001).

Simple tools like histogram can be used to see the graphical visualizations of data so that unusual values can be easily detected and preprocessed accordingly. Witten and Frank (2000) strongly recommend the need for consulting domain experts in the preprocessing of anomalies, missing values and the like.

Once the data is prepared it must be partitioned into training and test data set (Thearling, 2003). The training data set is part of the data that is used to build and train a model whereas the test data set is used to gauge the accuracy of the model in predicting unseen instances. The performance of the model is evaluated based on the outcome of the model for the test data.

2.4.3 Building and Testing the Model

The lion share of building and training a model is accomplished at this step. The details of building and training a model vary from technique to technique and hence there are no blue print procedures (Thearling, 2003). For this reason prior to building and training a model, a well-suited data mining technique has to be identified.

The problem of overfitting is another issue that deserves a due consideration at this step. According to Mitchell (1997) overfitting is an attempt to create overly complex data mining model that fits noise in the training data or unrepresentative features of the particular training data that decreases the generalization accuracy of the model over other unseen instances. In practice, most data mining models have a tendency to suffer from overfitting and the solution for this problem is to provide the model with the test dataset (Thearling, 2003). In fact, the model that yields a promising result for the training data set will at first come up with disappointing results on the test data. Thearling (2003) goes to say that the next step in the process is to refine the model to produce a second model that can work as well on the test set as it does on the training set. Of course, the detail process of tuning a model varies from technique to technique.

2.4.4 Evaluating the Model

In order to evaluate the performance of the model for new data, there is a need to examine the error rate on the data set that did not take part in the process of model formulation, test set (Thearling, 2003). Then a model with a high success rate or low error rate is considered as a good model. This holds true assuming that both the training and test set data are representatives of the underlying population (Witten and Frank, 2000). That is to say, data used to build and test the model are thought to generalize the data under investigation so that the newly incoming instances to be fed to the model will share some of the common features of the instances in the examples.

2.5 Data Mining and Data Warehousing

Data warehouse provides entire data about the organization as a whole. Fayyad, Piatetsky-shapiro and Smyth (1996) have defined data warehousing as "a popular business trend for collecting and cleaning transactional data to make them available for on-line analysis and decision support." The need for data warehousing emanates from the recognition of the fragmented data found at department or branch levels which does not give a full picture of an organization (Witten and Frank, 2000).

Data found at different branches and levels can have strategic importance if these databases are integrated. Hence, the trend these days is to create well-suited warehouses so that valuable data of an organization would be stored in integrated fashion, which is readily available for the knowledge discovery process. In fact, many of the steps involved in data warehousing will have to be undertaken to prepare the data for data mining process.

According to Fayyad, Piatetsky-shapiro and Smyth (1996) data warehousing lets data access on-line analysis; and one popular technique that facilitates this analysis is On-line Analytical Processing (OLAP). OLAP is a tool developed for multidimensional data analysis that provides an interactive data analysis.

OLAP is one of the most often confused techniques with data mining methods. However, data mining differs from OLAP methods in several different ways. According to Kim (2000) and Berry and Linoff (1997), the most significant distinction between data mining and OLAP lies on the approach, scope, and complexity of the techniques. OLAP is efficient in reporting on data whereas data mining focuses on extraction of patterns from a large database. In addition, OLAP applications often operate on summary data that have been aggregated in complex tables for fast and easy analysis. Data mining algorithm, on the other hand, runs on a database to extract trends.

2.6 The Application of data mining technology

Wide ranges of companies are employing data mining techniques as an important tool to improve the performance of their day-to-day activities (Javid, 1999). According to Gartner Group Advanced Technology Research, cited in (Thearling, 2003), data mining and artificial intelligence are at the top of five key technologies and "will clearly have a major impact across a wide range of industries within the next 3 to 5 years."

Data mining technology is applicable to any problem as long as there is huge data and well-defined understanding of the problem at hand (Thearling, 2003). One of the successful data mining applications to practical problems cited frequently is the Soybean Classification, which identifies the type of disease based on some diagnosis (Witten and Frank, 2000). The Labor Negotiation of the Canadian contract negotiations in 1987-88 is another example of the early success of data mining application (Witten and Frank, 2000).

There are also several successful systems developed and widely used to support the day-to-day activities of business organizations. Among these OpportunityExplorer is one that is used by A. C. Nielson (Carbone, 1997). IBM's intelligent Miner is another system which is being applied by retailers to know customers purchasing tendency and product popularity (Carbone, 1997). Government agencies and private companies are using data mining to enable financial management through analytical fraud detection (Garvin, 1998; Mangold, 1998). Recon, a successful data mining system of Lockheed Artificial Intelligence Center, could be cited as an example (Carbone, 1997).

According to research firm International Data Corp (IDC), cited in (Perez, 2001), worldwide revenue of analytic application solutions, which uses data mining for analysis, was expected to grow at a rate of 28% per annum, from US$2 billion in 1999 to more than US$6 billion in 2004. Particularly, IDC goes to say that Customer Relations Management (CRM) applications would be one of the fastest growing data mining application areas. IDC noted that in 1999 its revenue grew by 75% and is expected to be US$2.3 billion in 2004.

2.7 Overview of Data Mining Techniques

2.7.1 Widely Used Data Mining Techniques

According to Berry and Linoff (1997), learning and understanding of different data mining techniques is essential for the following reasons:

- In order to take the advantage of a specific technique, it is important to know the details of each technique.
- To determine the best applicable technique for the problem at hand.
- To know the advantages and disadvantages of a technique.

It is apparent that no one technique is applicable to all data mining problems. To determine the best technique suitable to the specific data mining problem, familiarity with the available techniques is necessary. According to Thearling (2003), the most commonly used data mining techniques are: Decision tree, neural networks, genetic algorithms, nearest neighbor method and rule induction.

Decision tree: is a technique in which records are presented in a tree like structure based on the values of their attributes. A depth discussion on decision tree is presented in section 2.7.2 of this chapter.

Artificial neural networks: are simulations of biological neural networks which are more suitable to model non-linear and complex relationships (Thearling, 2003). For details of this technique refer to section 2.7.3 of this chapter.

Genetic algorithm: is an optimization data mining technique which often adopts processes such as genetic combination and natural selection in order to find out set of parameters that best explain a predictive function (Thearling, 2003). Genetic algorithms employ "the selection, crossover, and mutation operators to evolve successive generation of solutions" said Berry and Linoff, (1997).

Nearest neighbor method: is a 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 (Thearling, 2003). On the basis of the number of records (k) it is often known as k-nearest neighbor. This technique is an instance-based learning in which a distance function is used to determine the class of a new instance. Although this technique is simple it often works very well (Witten and Frank, 2000).

Rule induction: is a technique that extracts useful sequence of 'if-then' rules from a database based on statistical significance (Thearling, 2003). The rule induction technique concentrates on the criterion and generation of rules with optimal accuracy.

2.7.2 How Decision Trees Work?

According to Han and Kamber (2001), a decision tree is "a flow-chart-like tree structure." Decision tree divide the records in the database into subsets based on the values of one or more fields. This process will be repeated for each subset recursively until all the instances at each node fall in to a single class (Berry and Linoff, 1997). The outcome of decision tree is a tree shaped structure that depicts a series of decision made at each step. Then these decisions are considered as rules for the task of classification (Thearling, 2003).

Decision tree are powerful and widely used data mining method for classification and prediction (Berry and Linoff, 1997). The strength and popularity of decision tree is due to the fact that in contrast to neural networks, it expresses the 'if-then' rules explicitly.

In applications where the accuracy of a classification or prediction of unknown instances is the only thing that matters, that is, in cases where how or why the model works is not important, both neural networks and decision tree can perform a good job (Witten and Frank, 2000). For instance, a bank firm could make use of data mining application to appraise credit application and it is more acceptable to both the bank official as well as the credit applicant to know that the application is rejected on the basis of a computer-generated rules than to declare that the decision has been made by a neural networks which provides no explanation for its action.

Decision tree methods are equally well suited at handling continuous and categorical variables. Categorical variables (take values in a pre-specified, finite set of possibilities), which pose problems for neural networks and statistical techniques, come ready-made with their own splitting criteria: one branch for each category. On the other hand, continuous variables (that take numbers either real or integer values) are equally easy to split by picking a number somewhere in their range of values.

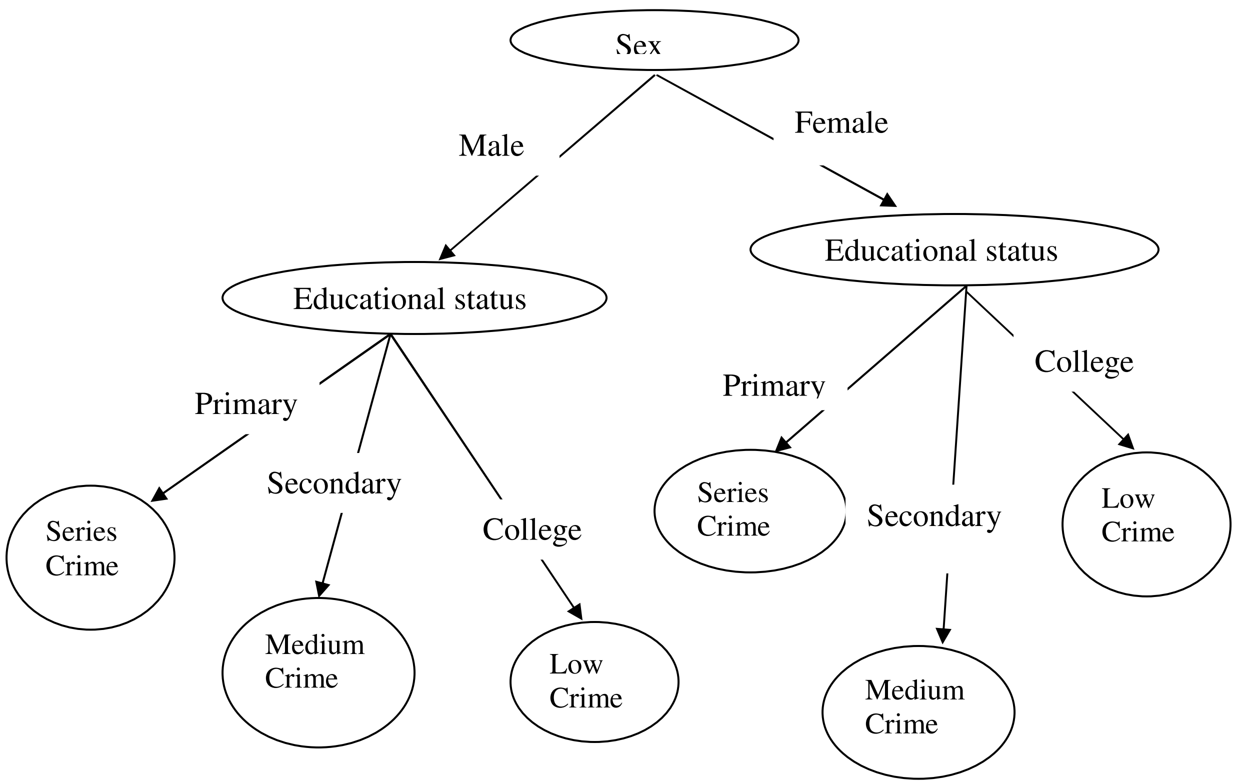

Figure 2.2: A hypothetical decision tree depicting how decision tree works and its general structure

An illustration of decision tree is presented above showing two attributes sex and educational status with two and three attribute values respectively as well as three classes; series, medium and low crimes.

In the above illustration the oval shaped figures show the attributes and the branching depict the different values of the attributes and the circles at the leaf nodes represent the classes. The node at the top most of the decision tree is known as a root node. Where as, the nodes at the bottom of the decision tree, which portray the classes, are called a leaf node. Nodes in between the root and the leaf nodes are called internal nodes.

Sticking to this hypothetical decision tree, an individual whose sex is male and his educational status is primary commits a series crime. Hence, the single path from the root node, that is, sex to each leaf node could express and generate a rule for the classification made as follows:

IF sex is male and educational status is primary

THEN commits series crime.

Although there are a number of decision tree algorithms developed so far, they all share the desirable features in presenting structural descriptions, i.e. understandable rules. The most commonly used algorithms for building decision tree are; CART, CHAID and C4.5 (Berry and Linoff, 1997). CART (Classification And Regression Tree) builds a binary tree by splitting the records at each node according to a function of a single input field. On the other hand, CHAID (Chi-squared Automatic Interaction Detection) is used for detecting statistical relationships between variables via building a decision tree (Ibid).

C4.5 is a 'divide-and-conquer' approach to decision tree induction that is the out growth of ID3 (Iterative Dichotomiser 3). It is receiving popularity and is readily available in several commercially available software packages. This study employs C4.5 and the rationale for choosing this algorithm is due to the fact that this algorithm incorporates additional qualities as stated in the subsequent sections. Moreover, the software package *weka,* which is readily available and familiar to the researcher, has this algorithm.

2.7.2.1 Attribute Selection Measure

Instances are evaluated and classified based on the values of their attributes. Of course, some of the attributes of an instance may be irrelevant to the process of classification and thus should be excluded. Moreover, there are attributes which best discriminates among the target classes (values of an attribute of an instance regarded as a class label) (Berry and Linoff, 1997). An attribute which best discriminates the instances belongs to the root node of the decision tree. Now the question is, how the attribute which best discriminates the instances could be selected? Different algorithms use slightly different methods to determine which attribute should be at the root node. For instance, CART employs diversity to measure the goodness of potential splitter and the best splitter has low diversity. Put differently, in using diversity the ultimate goal is to maximize:

diversity (before split) - [diversity (left child) + diversity (right child)]

In some commercially available CART packages like that of STarTree, three diversity measures are ready for the user to choose one (Ibid):

$$\min(p(c_1), p(c_2))\text{--}(2.1)$$

$$2p(c_1)p(c_2)\text{--}(2.2)$$

$$p(c_1)\log p(c_1) + p(c_2)\log p(c_2)\text{--------------------------------------}(2.3)$$

The result of these functions could be maximum or minimum when the probabilities of the classes are equal and when the set contains only a single class respectively.

C4.5 uses the last of the above measures (equation no.2.3) by extending to incorporate more than two classes and it is often known as information gain or entropy. According to (Bao, 2003), entropy specifies the minimum number of bits of information needed to encode the classification of an arbitrary member of instances. On the other hand, information gain measures the expected reduction in entropy. Entropy measures homogeneity of instances.

C4.5 uses the information gain measure to select the best attribute at each node of the tree and hence this method is a measure of the goodness of split. The attribute with the highest information gain or greatest entropy reduction is selected as the best attribute to discriminate instances in the resulting classification. In turn, this attribute, reduces the information required to make best classification and hence minimize the number of tests required to classify an instance.

Given a collection S, containing positive and negative examples of some target concept, the entropy of S relative to this Boolean classification is

$$\text{Entropy}(S) = -\ p_i \log_2 p_i - p_t \log_2 p_t \text{ - } (2.4)$$

Where p_i is the proportion of positive instances in S and p_t is the proportion of negative instances in S. In fact, in all calculations involving entropy conventionally 0log0 is defined to be 0.

Now let attribute A has V distinct values, $\{a_1, a_2, . . . , a_n\}$ and this attribute can classify S into V subsets $\{s_1, s_2 . . . s_v\}$, where S_j is consisting of instances in S that have values a_j of A. If A has chosen as the best test attribute for splitting, and then these subsets would belong to the partition grown from the node containing the set S. Suppose s_{ij} be the number of samples of class C_i in a subset by A, is computed as (Han and Kamber, 2001):

$$\text{Entropy (A)} = \sum_{j=1}^{V} \frac{s_{1j} + s_{2j} + \ldots s_{mj}}{S} I(s_{1j}, s_{2j}, \ldots s_{mj}) \text{ - - - - - - - - -(2.5)}$$

Where $(s_{1j} + s_{2j} + . . . s_{mj})/S$ acts as the weight of the j^{th} subset and is the number of samples in the subset divided by the total number of instances in S.

$$I(s_{1j}, s_{2j}, \ldots s_{mj}) = - \sum_{i=1}^{m} p_{ij} \log_2 (p_{ij}) \text{ - - - - - - - - - - - - - - - - - - - (2.6)}$$

Where $p_{ij} = s_{ij}/\mathrm{I}s_j\mathrm{I}$ and is the probability that a sample in s_j corresponds to class C_i.

The information gain that could be obtained from the splitting on A is given by:

$$\text{Gain (A)} = I(s_{1j}, s_{2j}, \ldots s_{mj}) - \text{Entropy(A)} \text{ - - - - - - - - -(2.7)}$$

However, in situations where attributes have a large number of possible values, a problem arises with the computation of information gain or entropy. Therefore, in reaction to this problem, C4.5 employs a modification of the measure called the gain ratio (Berry and Linoff, 1997).

Gain ratio = Gain/split info: $I(s_1, s_2, \ldots s_m)$ --------------------------(2.8)

The information values of a split, split info: $I(s_1, s_2, \ldots s_m)$, is the number of bits needed to determine which branch each instance is assigned to and the more branches there are, the greater this value is. Thus the gain ratio is calculated by dividing the original information gain defined in equation (2.8) by the information value of a split will be reduced for an attribute with multi values. Based on this measurement the best splitter is to choose the attribute that maximizes the gain ratio.

2.7.2.2 Practical Issues in Learning Decision Tree

2.7.2.2.1 Handling Missing Attributes Values

According to Bao (2003), there are two commonly used methods to handle missing values in constructing decision tree. One approach to handle the missing attribute value is to assign the value that is most common among training instances at each node. Alternatively, we might assign it the most frequent value among instances at the node that have a particular classification. Then the training instance using this estimated value could then be used directly by the existing decision tree learning algorithm.

The other, more complex procedure to handle missing value is to assign a probability to each of the possible values rather than simply assigning the most common value. These probabilities can be computed from the frequencies of the various values among the training instances at each node. C4.5 employs this approach in its attempt to handle missing attribute values (Bao, 2003).

2.7.2.2.2 Avoiding Over-Fitting Problems

Decision tree keep growing as long as new splits can be found that improves the ability of the tree to classify the instances of the training set into classes. Nevertheless, in some cases such process may end up with over-fitting problems which can lead to nonsense rules. Over-fitting is a significant practical difficulty for decision tree learning and many other learning methods. For instance, in one experiment of ID3 involving five different learning tasks with noisy, non-deterministic data, over-fitting was found to decrease the accuracy of learned decision tree by l0-25% on most problems (Bao, 2003). Therefore, there is a need to prune the tree in order to get more accurate rules that could be applicable to the general case.

According to Witten and Frank (2000), there are several approaches to avoiding over-fitting in decision tree learning and these can be grouped into two classes:

- ***Prepruning***: approaches that stops growing the tree earlier, before it reaches the point where it perfectly classifies the training data. It involves trying to decide during the tree building process when to stop developing subtree.
- ***Postpruning***: approaches that allow the tree to over-fit the data and then post prune the tree. That means a complete tree is first developed and then pruning afterward.

2.7.2.2.3 Generating Rules from Tree

One of the greatest strength of decision tree is its ability to generate rules (Bao, 2003). Although in some cases the decision tree may be large and complex, it is generally fairly easy to follow any one branch through the tree. Hence, the description for any particular classification or prediction is relatively straightforward. It is possible to find out a set of rules simply by traversing the decision tree and generating a rule for each leaf and making a combination of all the tests found on the path from the root to the leaf. This produces rules that are unambiguous in that it doesn't matter in what order they are executed.

2.7.3 Artificial Neural Networks (ANNs)

The study of neural networks commonly referred to as neural networks has been motivated from recognition of the fact that biological learning processes are entirely different from the way the traditional digital computers process information (Mitchell, 1997). The biological brain is made up of very complex and massive webs of inter-networked neurons. Although the brain is "very complex, nonlinear and parallel information processing," it is powerful enough to manage and control neurons in order to accomplish complex tasks such as pattern recognition, perception and motor control several times faster than the latest digital computer, which is prevalent today (Haykin, 1994).

Although the human brain contains inter-networked networks of about 10^{11} neurons and each neuron connected approximately to 10^{4} neurons, is quick enough to make complex decisions (Mitchell, 1997). According to Haykin (1994), intellectuals think that the information processing abilities of biological neural systems must emanate from highly parallel processes operating on representations that are dispersed across a number of neurons. Therefore, one motivation for the study of neural network is to simulate highly parallel information processing based on distributed representations.

Neural networks are considered to be black boxes because of their non-linear behaviors and are usually more complicated than other techniques (Berry and Linoff, 1997). They added that training a neural network is a further challenge requiring setting numerous parameters and the output of a neural network is not as easily understood by the user as the output seen by a decision tree tool. In spite of this, neural networks are proving their worth everyday in a wide variety of practical application. Besides, several algorithms have recently been developed for the extraction of rules from trained neural networks (Han and Kamber, 2001).

2.7.3.1 Structure of Neural Networks

The structure of neural network is very similar to the structure of the neurons in the human brain. All of the processing of a neural network is carried out by this set of neurons or units. Each neuron is a separate communication device, doing its own relatively simple job. A unit's function is simply to receive input from other units and, as a function of the inputs it receives, to compute an output value, which it sends to other units. The system is inherently parallel in that many units can carry out their computations at the same time.

In artificial neural networks, neurons are grouped in layers, often classified as input, hidden, and output layers (Frohlich, 1999). Inputs layer is a processing element that receives the input to the neural network and hidden layers are processing elements between a neural network's input layer and its output layer. On the other hand, output layer is the processing element that produces neural network's output.

There could be a number of input, hidden and output neurons in each corresponding layer. For example, in the following neural network (figure 2.2), there are three inputs, three hidden and three output neurons. In fact, the network consists of one input, hidden and output layer.

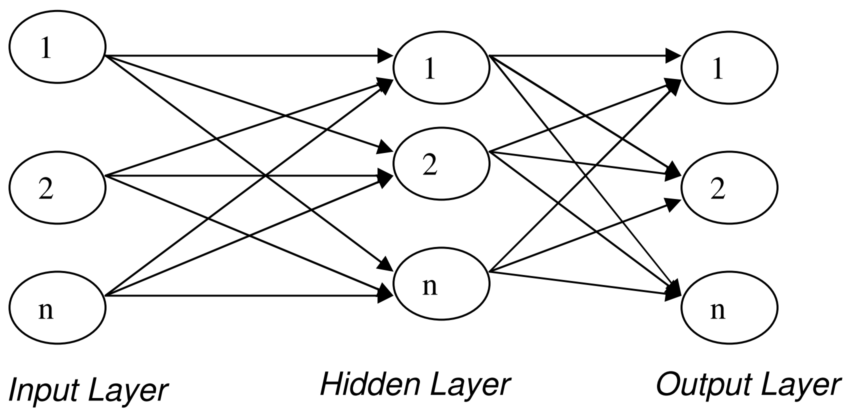

Figure 2.3: A simple neural network

2.7.3.2 Classification of Neural Networks

Depending on the pattern of connectivity, two types of networks can be distinguished: feed forward and recurrent networks.

- ***Feed forward networks:*** are employed in situations when we can bring all of the information to bear on a problem at once, and we can present it to the neural network (Bigus, 1996). In this type of network, the data flows through the network in one direction, and the answer is based solely on the current set of inputs. In other words, they have no feedback connection i.e. they have no connections through weights extending from the outputs of a layer to the inputs of the same or previous layer. Feed forward networks have no memory; their output is solely determined by the current inputs and the values of the weight.

- ♦ ***Recurrent networks:*** are networks with feedback connections. According to Haykin (1994), in some configurations, recurrent networks re-circulate previous outputs back to inputs; hence, their output is determined both by their current input and their previous outputs.

Neural networks could also be discussed in terms of their learning mechanisms. The learning rule is the very heart of a neural network. It determines how the weights are adjusted as the neural network gains experience. Many networks use some variation of the Delta rule (to search the hypothesis space of possible weight vectors to find the weights that best fit the training examples) for training. One type of the rule most widely used is back propagation.

Training algorithms can be classified as supervised and unsupervised training (Ibid).

- ***Supervised training:*** is the process of learning in which the network learns from examples, i.e. a pair of input vector with target vector containing the desired output. The network is provided with instances whose class are known and let learn from these instances. The learning algorithm takes the difference between the correct output and the prediction of the neural networks so that the prediction next time would be closer to the correct answer (Mitchell, 1997). The neural network has to be given examples many times in order to learn and make correct prediction. The back propagation algorithm employs the supervised learning paradigm.

- ***Unsupervised training:*** is the process of learning in which the training set consists solely of input vectors. In this approach of training there are no target outputs and it is impossible to determine what the result of the learning process will look like (Frohlich, 1999). The network is simply provided with a number of inputs and the network organizes itself in such a way as to come up with its own classification of inputs.

The back propagation algorithm is the algorithm considered in this study since it is the most widely used algorithm in the study of neural networks. Often, back propagation algorithm comprises the following three steps (Mitchell, 1997).

- The input pattern is presented to the network whereby the input pattern is propagated through the network until they reach the output units. This forward pass would produce the actual or predicted output.
- Then the desired output would be given as part of the training set, so that the actual output can be subtracted from the desired output in order to give the error signal.
- In the last step the errors are passed back through the neural network by computing the contribution of each hidden processing unit and deriving the corresponding adjustment needed to produce the correct output. The connection weights are then adjusted and the neural network is said to have learned from an experience.

The above steps are repeatedly carried out for all examples in the data until the termination condition of the iteration is satisfied.

CHAPTER THREE

CRIME AND CRIME PREVENTION

3.1 Introduction

In this chapter the researcher presents the definition of crime, criminals and situations where an act could be a crime. Possible causes of crimes cited in the literature and crime prevention attempts by the law enforcement organizations are also reviewed. In addition, the trend of crime in Ethiopia and practical examples of the data mining technology in the area of crime prevention are also described.

3.2 Crime and Criminals

Thakur (2003) has defined crime as an act or omission of an act, which is punishable by law. However, an act that is considered as a crime in one place and time may not be true in another place or time.

According to Andargachew (1988), a criminal is an individual person who has violated the legally forbidden act. In fact, there are some factors that have to be taken into account to convict whether a person should be considered as a criminal or not. Among these, an individual should be of competent age in light with the law of the land; and there must be a well-predefined punishment for the particular act committed.

Sutherland and Cressey, cited in (Andargachew, 1988), stated that an act would be considered as a crime when it is prohibited by the criminal law. Criminal law, on the other hand, refers to a body of specific rules regarding human conduct, which have been explicitly stated by political authority.

Crime has increasingly become as complex as human nature. Modern technological advancement and tremendous progress in communication have facilitated criminals of every corner of the world to commit a crime using sophisticated equipment in one place and then escape to another place (Thakur, 2003). These days the glob is facing the proliferation of problems such as illicit drug trafficking, smuggling, hijacking, kidnapping, and terrorism.

Crime has adversely affected the societies of both civilized as well as developing countries by declining the quality of life, endangering human right and fundamental freedom and posing a serious challenge to the community. Although the level and intensity of the problem might vary from nation to nation no country has remained unaffected.

3.3 Trend of Crime in Ethiopia

In Ethiopia, crime statistics (of the Federal Police Commission) has shown that the rate of crime is increasing steadily. A sample survey conducted in the year 1996 by a research team of the Federal Police has shown that in 1986 about 51,869 crimes were reported to the police (Federal Police Commission, 1996). Taking the total population of the country during this period, this figure indicated that one crime was committed among 800 people during this year. Whereas the research report has shown that in the year 1994 about 96,995 different crimes were reported. This reveals that during this period one crime is committed among 568 people. As compared to the year 1986 the total number of crimes committed in the 1994 have shown a significant increment.

The national crime statistics report compiled by the Federal Police Commission in 2003 for the year 2002 has shown that about 219,539 crimes were reported to police through out the country and out of this 51 percent were committed in urban area while the remaining 49 percent were occurred in rural areas.

Moreover, the national crime statistics report has indicated that the Oromia regional state accounts 22.49% of the total crimes reported to the police in the entire country in the year 2002. This figure indicates that the Oromia region is one of the crime prone regions in the country.

3.4 Crime Prevention

The causes for the growing rate of crime include unemployment, economic backwardness, over population, illiteracy and inadequate equipment of the police force (Thakur, 2003). The form, seriousness and size of the crime, may rely on the form of a society and thus its nature changes with the growth and development of the social system (Ibid*)*. In every generation it has its own most critical, new and special problems of crime, although the crime problem is as old as man himself. In addition to this, the techniques employed to commit crime are new in the sense that they make use of modern knowledge and technique. The rise in crime both national and transnational is generally thought as the result of interplay between socio-economic changes.

The circumstances surrounding the individual offender such as his personality, physical characteristics intelligence, family background, environmental surrounding such as peer groups, neighbors etc have been subject of the study of crime (Andargachew, 1988). Hence, understanding the attributes of criminals will be helpful to design and implement prudent crime prevention strategies.

Crime, so to say, is one of the most critical social evils a society can face. Hence maintaining law and order is one of the principal functions of any government (Wilson, 1963). For this reason, governments usually establish organizations such as courts, prosecutions and police, which are responsible for the maintenance of law and order in their respective country.

These agencies and other related organizations are responsible to curb the rate and occurrence of crimes. To do so, crime prevention agencies need to issue and implement crime prevention strategies. Theoretically, it is argued that crime prevention is better than cure for the following reasons (Thakur, 2003):

- Prevention safeguards the life and property of the society whom the police are in duty to protect.
- It avoids a good deal of trouble to the victim both physical and mental.
- Crime prevention rules out litigation, which follows in the process of detecting a crime.
- Prevention also saves the police from the trouble of recording crime at all odd hours of the day and night and of taking immediate action for the investigation.

Thakur (2003) suggested that intent and opportunity are two major factors that lead to the occurrence of a crime. An individual cannot commit a crime unless and otherwise s/he gets an opportunity even though s/he is intended to commit a crime. Therefore, the best strategy for crime prevention is to provide a system that denies any opportunity for a criminal to commit a crime.

However, these days Law enforcement and investigating agencies have recognized the tremendous value in extracting hidden knowledge embedded in their data warehouses which could be valuable in the process of combating crimes (Megaputer Intelligence, 2002). The police departments want to reveal frequent crime patterns from historical reports to help them investigate new cases.

According to Megaputer Intelligence (2002), the analysis of crime patterns and trends is very important for police officers and analysts can learn from historical crime patterns and enhance crime resolution rate. It also helps to prevent future incidents by putting in place preventive mechanisms based on observed patterns. Another possible advantage is, it can reduce the training time for officers assigned to a new location and having no prior knowledge of site-specific patterns to assist them in investigations. In light with the crime patterns extracted from previous records, police can deploy scarce resources to the right place at the right time.

3.5 Data mining and crime prevention

Most law enforcement agencies today are faced with large volume of data that must be processed and transformed into useful information (Brown, 2003). Data mining can greatly improve crime analysis and aid in reducing and preventing crime. Brown (2003) stated "no field is in greater need of data mining technology than law enforcement."

One potential area of application is spatial data mining tools which provides law enforcement agencies with significant capabilities to learn crime trends on where, how and why crimes are committed (Veenendaal and Houweling, 2003). Brown (2003) developed a spatial data mining tool known as the Regional Crime Analysis Program (ReCAP), which is designed to aid local police forces (e.g. University of Virginia (UVA), City of Charlottesville, and Albemarle County) in the analysis and prevention of crime. This system provides crime analysts with the capability to sift on data to catch criminals. It provides spatial, temporal, and attribute matching techniques for pattern extraction.

One of the early applications of data mining in the law enforcement was by the FBI of the federal government of the United State for the investigation of the Oklahoma City bombing case. The FBI adopted data mining techniques to scrutinize large volume of data gathered from various sources to track down criminals (Berry and Linoff, 1997). Similarly, Berry and Linoff (1997) reported that the Treasury Department of the United States adopted data mining technology to extract suspicious money laundering or fraud patterns. This is particularly aiming at detecting criminals involving in money laundry or fraud.

The West Midlands, U.K., police department employed data mining in its daily battle against crime (SPSS, 2003). When any crime is committed, physical descriptions of the criminals as well as their modus operandi (MO) (information about a criminal and a crime recorded by the police when a crime is committed) is recorded. It uses two Kohonen networks to cluster similar physical descriptions and MOs. Then clusters are combined to look at if groups of similar physical descriptions are alike with groups of similar MOs. If a good match is found, and criminals are known for one or more of the offenses, then it is possible to conclude that the new crime is committed by the same individual.

In fact, the West Midlands further investigated the clusters using statistical tools and techniques to validate the significance of its conclusion. Moreover, the system is also extended to investigate the behavior of repeat offenders, with the goal of identifying crimes that seem to fit their behavioral pattern (SPSS, 2003).

In describing the invaluable role of data mining in crime control Adderley, an Inspector in the West Midlands, U.K. police department, said that "if I can build a model using data in which the offender is known, then we can apply that model to the database of unsolved crimes. Once we determine that such models work, this could give us an invaluable aid in linking known criminals to specific crimes quickly" (Cited in SPSS, 2003).

The aforementioned attempts of data mining applications in the area of crime prevention reveal the rich potential of the field in the efforts against crime. However, much of the attempts have not taken into account the attributes of criminals or offenders to classify crime records and extract rules for crime patterns. Thus, this work is unique in the sense that it employs the attributes of offenders so as to come up with certain rules that could support the process of crime prevention.

CHAPTER FOUR

EXPERIMENTATION

4.1 Introduction

This chapter deals with the identification of source of data, data cleaning and preprocessing of the data employed in this study. Efforts exerted and techniques used in the process of data preparation and model building are also described. Besides, it presents results of the experiment and its interpretations.

4.2 Data Collection

The data employed in this research was collected from the Oromia Police Commission. A full backup of the database of the criminal database of the Oromia Police Commission was taken and all the necessary operations of data selection and data preparation were carried out.

Originally, information about the offender and the corresponding crime is recorded when the individual is caught by police. In fact, all policemen are provided with a centralized form or record format that should be filled when a crime is committed and/or an offender is caught.

The criminal database of the Oromia Police Commission contains more than 50,000 records; however, still there are a number of records stored manually awaiting to be entered to the automated system. Although theoretically large volume of data is more important to train data mining models, due to time constraint to preprocess the data, a sample data containing 5,500 records are taken for this study. The sampling method employed here is random sampling i.e. taking 1,100 records from every 10,000 records randomly.

4.3 Data Preprocessing

The criminal record database system of the Oromia Police Commission, which is employed for the purpose of this study, suffers from a number of limitations. These include missing values, outliers and encoding inconsistency in various attribute values. In fact, as stated by Witten and Frank (2000), one of the critical problems in building data mining models is limitations in the data itself. Thus, an optimal model could be constructed once a comprehensive, clean and automated data is well prepared.

As mentioned in section 2.4 of chapter 2, knowledge discovery comprises some defined steps and data preprocessing is one of these steps. The purpose of this step is to cleanse the data and to transform it into a form that is suitable to the subsequent steps. The data preprocessing includes a number of tasks and the common tasks are presented and discussed as follows.

4.3.1 Attribute selection

Theoretically, decision tree could determine relevant attributes for classification automatically using the concept of information gain or entropy with out manual efforts. However, in real world databases, it is common to find a number of irrelevant attributes (attributes that could not be helpful for some task of classification or other purpose) that could be easily known their irrelevance before adopting any complicated technique. Thus, it is important to exclude those attributes that are not important for analysis in order to simplify the task of decision tree.

The criminal record database contains more than165 attributes and to decide on the relevant attributes for this study, a discussion was made with crime analysis experts at the Federal Police

Commission. The discussion resulted in 160 attributes shown in Appendix A (list of attributes, descriptions and their possible values).

Thus, records consisting of these attributes were preprocessed and prepared as stated in the subsequent sections before the data were provided to the decision tree algorithm. Then the decision tree algorithm calculates the information gain to select valid attributes and makes classification based on the selected attributes.

As there is no such mechanism to select attributes in the case of neural networks decision tree are used to select attributes for neural networks (Berry and Linoff, 1997). For this reason attributes selected by the decision tree are supplied to the neural networks.

4.3.2 Data Preparation for Analysis

One of the most important tasks in data mining is preparing the data in a way that is suitable for the specific data mining tool or software package. Usually, real world databases contain incomplete, noisy and inconsistent data and such unclean data may cause confusion for the data mining process (Han and Kamber, 2001). Thus, data cleaning has become a must in order to improve the quality of data so as to improve the accuracy and efficiency of the data mining techniques.

The criminal record database of the Oromia Police Commission is a Microsoft Access Database, which comprises information about offenders, crimes, victims and results (court verdicts). However, this information was kept in different tables and hence to prepare the data for analysis it was important to denormalize or merge some of the tables. For this reason information about offenders, which is located in the offender table and information about crime incidents found in crime table were merged in order to create aggregate information about offenders and the corresponding crimes.

The next stage was handling inconsistent and missing values. The inconsistency was mainly due to typing errors and some of them were originally unrecorded. Theoretically, there are several ways suggested in the data mining literature to handle missing values such as calculating the average of continuous valued attributes and filling this mean value to missing attribute values (Han and Kamber, 2001). Nevertheless, in this study records with too many inconsistent and missing values were excluded from the study and were replaced by other records. On the other hand, for some of the records an attempt was made to refer the manual documents and to enter the corresponding values. The rationale for choosing this approach is due to the fact that there is no deficiency of data and hence the performance of the model could be improved by supplying records that does not contain too much missing values.

4.4 Defining the target attributes

Some data mining techniques need predefined classes in order to train and build classification models. In such cases, the data preparation task is incomplete until the target attributes are well defined. In other words, the training set should be pre-classified so that the data mining algorithms know what the user is looking for.

This study comprises two segments of experiment i.e. classification of records on the basis of target classes namely: crime scene (***SceneLabel***) and crime label (***CrimeLabel***). ***SceneLabel*** refers to the place where the crime is committed i.e. whether it is urban or rural. On the other hand, ***CrimeLabel*** refers to the label or seriousness of crime i.e. whether it is serious, medium or low. Each record in the data set was classified into one of the attribute values of each of the two attributes.

The whole purpose of defining these classes was to investigate which section of the society (individuals having some values for the given attributes) is being involved in crime acts in both urban and rural areas. A crime could be committed in urban areas or in rural areas and there is a need for police officers to know the nature of crimes that are occurring in the respective areas. In addition, it is also equally important to know the type of people who involve in the crime committed in the corresponding areas and crimes. Thus, to assist the police commission in deploying appropriate resource allocation and crime prevention methods, the study also attempted to identify which crime is committed.

The first experiment deals with classification of records on the basis of where a crime is committed i.e. urban or rural. These classes are designated by the attribute name of ***SceneLabel***. In order to train and build a model that classify records into urban and rural, the ***SceneLabel*** attribute of the dataset was considered as a target attribute containing classes of urban and rural. In the criminal record database for the attribute ***SceneLabel*** only numbers 1 or 2 are used representing urban and rural respectively. The numeric values of the ***SceneLabel*** attribute are transformed into their corresponding nominal values since a target class is supposed to be nominal in the weka's implementation of C4.5 decision tree algorithm.

With this objective records whose crime scene (urban or rural) is known were provided to the weka software (i.e. classifier.j48.J48). As mentioned above 5,500 records were provided to the weka and out of this, 3,252 (59.2%)records were found to be in the class ‘urban’, while the remaining 2,242(40.8) records were in the class ‘rural’. Table 4.1 below shows the distribution of records in each class.

Table 4.1: Distribution of records based on crime scene.

Class (***SceneLabel***)	Number of records in each class	Share of each class in percent
Urban	3252	59.2%
Rural	2248	40.8%
Total number of records	5,500	100%

The other segment of this study is classification of crime records on the basis of seriousness of the crime. There is a trend in the police institution to classify crimes in to three broad categories i.e. serious, medium and low crimes (Federal Police Commission, 1996). This classification of crimes is based on the cost of a crime to the society at large. In other words, crimes associated with too much cost to the society are classified into serious crimes. Nevertheless, in the criminal database there are no such well-defined classes. Rather the database contains 54 different crime types and it is coded (an integer value in the range of 1 to 54 is assigned for each crime type) to make the task of data entry more ease. Thus, there was a need to group all the crime types in to three categories based on the seriousness of the crime and this was done by consulting the crime analysis experts at the Federal Police Commission.

Therefore, based on their recommendation the most serious crimes such as murder, crimes against public institutions and the like are labeled as serious crimes. On the other hand, crimes like that of rape, theft, fraud and crimes against individual persons are classified as medium crimes. Lastly, crimes such as smuggling trade, drug use and transmission and the like are grouped in the low crime class. Then on the basis of the above justification the values of the attribute ***CrimeLabel*** are transformed into one of these classes (i.e. serious, medium, and low).

Table 4.2: Distribution of records based on crime label.

Classes (***CrimeLabel***)	Number of records in each class	Share of each class in percent
Serious	2562	46.6%
Medium	1431	26%
Low	1507	27.4%
Total number of records	5,500	100%

Table 4.2 depicts the distribution of records on the target class of ***CrimeLabel***, which is consisting of three classes, i.e. serious, medium and low class. So out of the 5,500 records 2,562 (46.6%) records are in the class 'serious', 1431(26%) records in the class 'medium' and the remaining 1507(27.4%) records are in the class 'low'.

4.5 Data preparation and organization

Like any other software, Weka needs data to be prepared in some formats and file types. The data sets provided to this software were prepared in a format that is acceptable for weka software. Weka accepts records whose attribute values are separated by commas and saved in an ARFF (Attribute-Relation File Format) file format (a file name with an extension of ARFF i.e. FileName.arff).

In order to prepare such file format the records from the Microsoft Access database were exported to Microsoft Excel and saved as a Comma Delimited (CMV) file. Then the next step was to open this file using the WordPad program. Weka file starts with the dataset's name followed by list of attributes. In fact, the dataset's name should be preceded by the symbol '@' and the word 'relation' (for example; @relation crime where crime is the name of the dataset) and each attribute name also starts with the same symbol and the word 'attribute' and following the name of the attribute including its possible values. If the variable or attribute is nominal a list of possible values contained in a brace is required otherwise generic words like 'real' to mean continuous are used.

By default the last attribute in the list of the attributes name in the dataset designates the target class. However, it is also possible to choose any attribute name as a target class no matter its position in the list while running the program.

For instance, the following are valid ways of data presentation for weka:

@relation crime

@attribute age real

@attribute ZoneID {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15}

@attribute SceneID {rural, urban}

-

-

-

@data

Once the list of attributes is completed, the word '@data' is used to indicate the beginning of the data and finally each record was prepared by listing the values of each attribute separated by comma and missing values are represented by a "?" (question mark). Then it was saved with a proper file name and extension.

Weka has the facility to extract a random sample and then test the accuracy of the classifier on disjoint collection of cases. It provides three options to partition the dataset in to training and test data. These are:

- Preparing distinct files for training dataset and test dataset
- Cross validation with possibility of setting variety number of folds (the default was 10 fold)
- Percentage split

The last two options have been used for this research. In the first case the default value of all parameters were considered and hence a ten fold cross validation. In cross validation the data is split into some number of partitions of the data, in this case 10 approximately equal proportions, and each in turn was used for testing while the remainder was used for training. This process repeats 10 times and at the end, every instance has been used exactly once for testing. Finally the average result of the 10 fold cross validation is considered (Witten and Frank, 2000).

In the second option, which was a percentage split, out of the total data prepared 80% (4,400) records were employed for model building and the remaining 20% (1,100) records were used for validation set or testing.

4.6 Decision Trees

4.6.1 Decision Tree Model Building Using Weka Software

The decision tree software employed for the purpose of this research was the weka software package, which contains several classifiers, clusters and association algorithms. As stated in section 2.7.2 of chapter two the C4.5 decision tree algorithm was used in this study and weka's implementation of this decision tree algorithm is called J48 (Witten and Frank, 2000). In fact, J48 implements a later and slightly improved version of which is known as C4.5 Revision 8, which was the last public version of this family of algorithms before C5.0 (new version of C4.5); a commercial implementation was released (Ibid).

Weka software is organized in hierarchy, which comprises packages and sub-packages. Thus the J48 algorithm is under the j48 package that is a sub-package of classifiers, which is part of the over all weka package. While the sub-package name is j48, the program to be executed within it is called J48.

The data preprocessed in the preceding steps are well suited to the weka software to train a classifier model. Thus, to start building the model the weka file that contains the dataset was opened and the J48 program was run. In fact, there are some parameters that can be set based on ones interest. In this research two distinct procedures were adopted. In the first case the model was constructed using the default values of the program. On the other hand, some of the parameters were modified when found necessary. On this premises the following experiments were conducted.

Experiment 1: Classification of Records Using the *SceneLabel* Target Class

In order to classify the records based on their values for the attribute ***SceneLabel***, the model was trained by using the default values of the program. A ten fold cross validation was employed to partition the dataset into training set and test set; including the default values of the parameters.

Figure 4.1 depicts while the J48 decision tree program was on training. As shown in the test options the method employed for partition data set into training and test was based on the cross-validation which was set by default. In addition to this, there are other parameters under the more options button and in this case all are set to their default values.

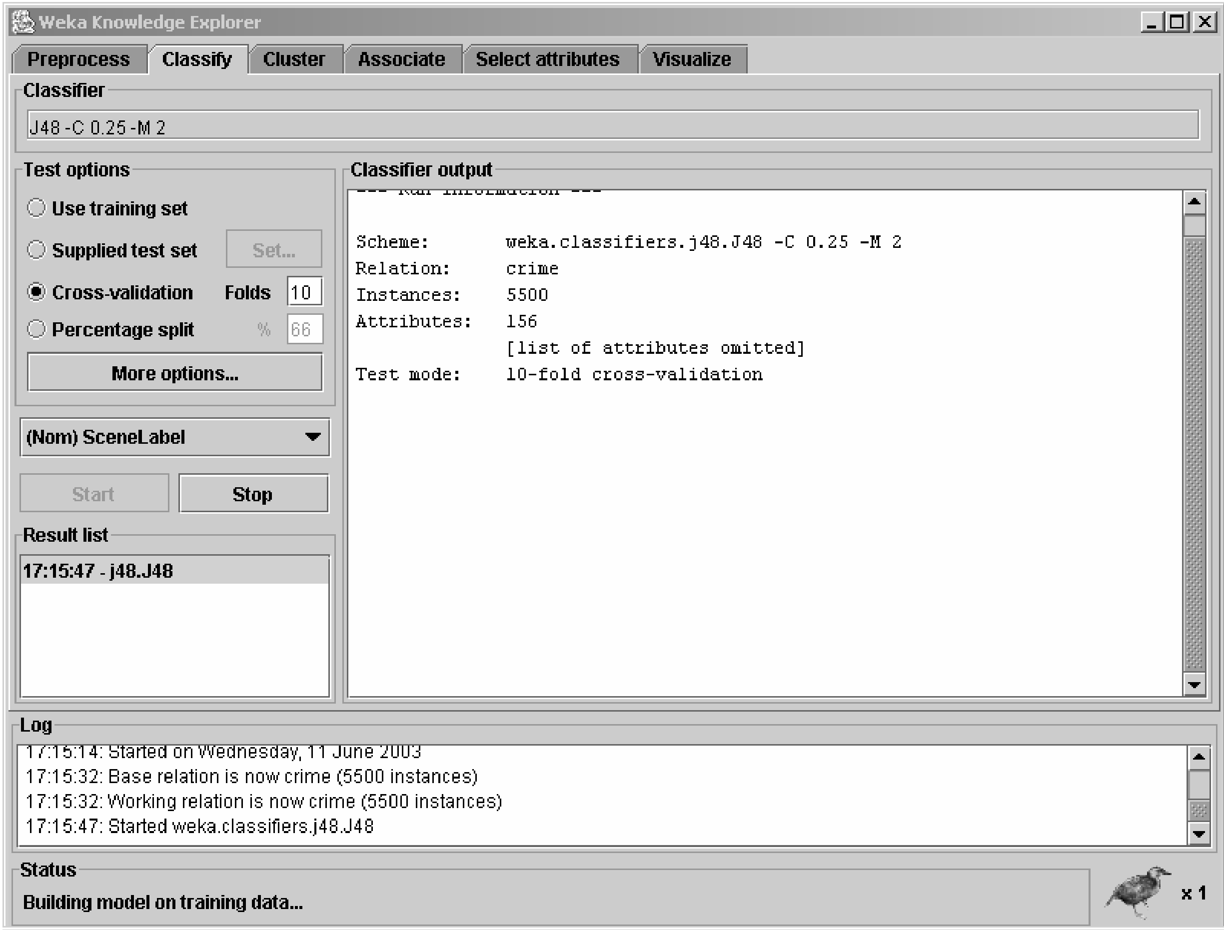

Figure 4.1: J48 decision tree program on training

Although each record in the dataset contains 158 (excluding ***incident id*** and the target class ***SceneLabel***) attributes the decision tree algorithm has selected only 22 attributes (the list of these attributes are presented in appendix A.1) to make the classification task. This indicates that attributes that are not considered in the construction of the decision tree are not important to discriminate the records in to the predefined classes (urban and rural) provided to the model. The output of the training is presented in the confusion matrix (Table 4.3) below.

Table 4.3: output from the J48 decision tree learner

Actual	**Predicted**		**Total**	**Score**
	Rural	Urban		
Rural	2,026	222	2248	91.13%
Urban	104	3,148	3,252	96.80%
Total	2130	3370	5,500	94.07%

The confusion matrix depicts that out of the total records provided to the program, 2,026 (91%) and 3148 (97%) records were classified correctly in the class of rural and urban respectively. On the other hand, 222 (7%) records were incorrectly classified as urban while actually they were supposed to be in the rural class and 104 (3%) records were classified incorrectly as rural while actually they are in the urban class. This portrays that from the total records 5,174 (94%) records were classified correctly while the remaining 326 (6 %) records were classified incorrectly. Hence, this indicated that records whose class is urban were classified with a minimum error as compared with the records in the class rural.

J48 pruned tree

```
MAOEMPE <= 0
|  MOEDCU <= 0
|  |  FOEDS <= 0
|  |  |  MAOEOTH <= 0
|  |  |  |  MAOEMG <= 0
|  |  |  |  |  MAOHAU <= 1
|  |  |  |  |  |  MOEDJ <= 0
|  |  |  |  |  |  |  MAOEPW <= 0
|  |  |  |  |  |  |  |  TOM <= 0
|  |  |  |  |  |  |  |  |  FOEDI <= 0: rural (291.0/16.0)
|  |  |  |  |  |  |  |  |  FOEDI > 0
|  |  |  |  |  |  |  |  |  |  FOMNM <= 0: urban (195.0/10.0)
|  |  |  |  |  |  |  |  |  |  FOMNM > 0: rural (15.0/4.0)
|  |  |  |  |  |  |  |  TOM > 0
|  |  |  |  |  |  |  |  |  MAOEMUN <= 0: urban (2986.0/179.0)
|  |  |  |  |  |  |  |  |  MAOEMUN > 0
|  |  |  |  |  |  |  |  |  |  MOEDI <= 1
|  |  |  |  |  |  |  |  |  |  |  TOFF <= 0: rural (56.0/13.0)
|  |  |  |  |  |  |  |  |  |  |  TOFF > 0: urban (5.0/1.0)
|  |  |  |  |  |  |  |  |  |  MOEDI > 1: urban (8.0)
|  |  |  |  |  |  |  MAOEPW > 0
|  |  |  |  |  |  |  |  MOEDI <= 0: rural (209.0/5.0)
|  |  |  |  |  |  |  |  MOEDI > 0
|  |  |  |  |  |  |  |  |  MAOHAU <= 0: urban (89.0/3.0)
|  |  |  |  |  |  |  |  |  MAOHAU > 0: rural (7.0/2.0)
|  |  |  |  |  |  MOEDJ > 0
|  |  |  |  |  |  |  MAOEMF <= 1
|  |  |  |  |  |  |  |  MAOHAU <= 0
|  |  |  |  |  |  |  |  |  MAOEMF <= 0
|  |  |  |  |  |  |  |  |  |  MOA915r <= 0: rural (52.0/10.0)
|  |  |  |  |  |  |  |  |  |  MOA915r > 0: urban (4.0)
|  |  |  |  |  |  |  |  |  MAOEMF > 0
|  |  |  |  |  |  |  |  |  |  MORC <= 1: urban (21.0/2.0)
|  |  |  |  |  |  |  |  |  |  MORC > 1: rural (2.0)
|  |  |  |  |  |  |  |  MAOHAU > 0: rural (169.0/5.0)
|  |  |  |  |  |  |  MAOEMF > 1
|  |  |  |  |  |  |  |  TOFF <= 0: urban (28.0/3.0)
|  |  |  |  |  |  |  |  TOFF > 0: rural (3.0/1.0)
|  |  |  |  |  MAOHAU > 1
|  |  |  |  |  |  MOA1930 <= 0: rural (164.0/3.0)
|  |  |  |  |  |  MOA1930 > 0
|  |  |  |  |  |  |  MAOEMUN <= 1
|  |  |  |  |  |  |  |  MAOHN <= 1: urban (31.0/5.0)
|  |  |  |  |  |  |  |  MAOHN > 1: rural (2.0)
|  |  |  |  |  |  |  MAOEMUN > 1: rural (164.0)
|  |  |  |  MAOEMG > 0: rural (119.0/9.0)
|  |  |  MAOEOTH > 0
|  |  |  |  MOEDP <= 0: rural (84.0)
|  |  |  |  MOEDP > 0
|  |  |  |  |  MAOHOTHE <= 0: urban (5.0/1.0)
|  |  |  |  |  MAOHOTHE > 0: rural (2.0)
|  |  FOEDS > 0: rural (91.0/1.0)
|  MOEDCU > 0: rural (186.0/4.0)
MAOEMPE > 0: rural (512.0/11.0)
```

Figure 4.2: A decision tree generated by the default values of the program.

Figure 4.4 shows the decision tree constructed when the different parameters of the program are set to their default values. The numeric value in the logical expression indicates the number of offenders (the minimum value is always 0) who were involved in each crime incident having the specified attribute. This is due to the fact that a crime could be committed by a group of individuals or a single person. Thus, if a crime is committed by group of individuals then those individuals might have different values for each attribute. At the end of the tree (leaf node) the numbers contained in a bracket refers to the number of records classified correctly and wrongly in each class.

Although this training scheme has shown a good performance in terms of accuracy, since the decision tree is somehow bushy and complex, it has difficulty to generate understandable rules. The size of the tree and the number of leaves produced from this training is 55 and 28 respectively. As a result one needs to traverse through all the nodes of the tree in order to come out with valid rule sets.

Therefore, to make ease the process of generating rule sets or to make it more understandable, the researcher attempted to modify the default values of the parameters so as to minimize the size of the tree and number of leaves. With this objective, the parameter minNumObj (minimum number of instances in a leaf) was set to 20, which were 2 in its default value and this figure was set after a number of trial were made. This means the process of classifying records proceeds until the number of records at each leaf reached 20.

In fact, this has its own disadvantage as some of the instances are going to be classified incorrectly. In other words, records in a given leaf could be in different class and there could be attributes, which could farther split the records in the same node into disjoint classes.

```
J48 pruned tree
------------------

MAOEMPE <= 0
|   MAOEMUN <= 4
|   |   FOEDS <= 0
|   |   |   MAOEOTH <= 0
|   |   |   |   MAOEMG <= 0
|   |   |   |   |   MAOHAU <= 1
|   |   |   |   |   |   MOEDJ <= 0
|   |   |   |   |   |   |   MAOEPW <= 0
|   |   |   |   |   |   |   |   TOM <= 0
|   |   |   |   |   |   |   |   |   FOEDI <= 0: rural (291.0/16.0)
|   |   |   |   |   |   |   |   |   FOEDI > 0: urban (210.0/21.0)
|   |   |   |   |   |   |   |   TOM > 0
|   |   |   |   |   |   |   |   |   MAOEMUN <= 0: urban (2989.0/181.0)
|   |   |   |   |   |   |   |   |   MAOEMUN > 0: rural (70.0/25.0)
|   |   |   |   |   |   |   MAOEPW > 0
|   |   |   |   |   |   |   |   MOEDI <= 0: rural (291.0/5.0)
|   |   |   |   |   |   |   |   MOEDI > 0: urban (96.0/8.0)
|   |   |   |   |   |   MOEDJ > 0
|   |   |   |   |   |   |   MAOHAU <= 0
|   |   |   |   |   |   |   |   MAOEMF <= 0: rural (56.0/14.0)
|   |   |   |   |   |   |   |   MAOEMF > 0: urban (53.0/9.0)
|   |   |   |   |   |   |   MAOHAU > 0: rural (170.0/6.0)
|   |   |   |   |   MAOHAU > 1
|   |   |   |   |   |   MOA1930 <= 0: rural (164.0/3.0)
|   |   |   |   |   |   MOA1930 > 0: urban (36.0/10.0)
|   |   |   |   MAOEMG > 0: rural (215.0/11.0)
|   |   |   MAOEOTH > 0: rural (93.0/4.0)
|   |   FOEDS > 0: rural (92.0/1.0)
|   MAOEMUN > 4: rural (162.0/1.0)
MAOEMPE > 0: rural (512.0/11.0)
```

Figure 4.3: A decision tree generated from the J48 decision tree learner after some of the parameters are modified.

Another change introduced is the method of partitioning a dataset. In this learning scheme a percentage split is used to partition the dataset into training and testing data and this parameter was set to 80, which is to mean 80% for training and 20% for testing. The purpose of using this parameter was to assess the performance of the learning scheme by increasing the proportion of testing dataset. The result of this learning scheme is summarized and presented in table 4.5. The output of this experiment portrays that due to the adjustment of some of the parameters, the size of the tree has reduced to 31 and the number of leaves has become 16.

Figure 4.5 depicts the decision tree generated after some of the parameters are modified. This decision tree has proven that as the value of the parameter minNumObj (minimum number of instances in a leaf) is increased, then the bushiness of the tree started to reduce; however at the expense of declining in accuracy. In other words, the numbers of misclassified instances have increased to 8 percent, which was 6 percent in the previous one.

Table 4.4 Output from the J48 decision tree learner after some of the parameters are modified.

Actual	**Predicted**		**Total**	**Score (Accuracy rate)**
	Rural	Urban		
Rural	393	40	433	90.76%
Urban	44	623	667	93.40%
Total	437	663	1,100	92.36%

Out of the 1,100 test data provided to the program 1,016 (92%) records were classified correctly and the remaining 84(8%) records were classified incorrectly. The results of this training scheme seem more ease to generate less complex rules, however, it revealed low accuracy rate as compared to the previous one.

Therefore, although the decision tree developed in the first learning scheme is relatively complex, it seems more accurate to classify records. In addition, the attributes employed in this classification are also more meaningful for decision-making processes. For instance, attributes like education level, age, sex, occupation, and habit of offenders were found important attributes for the task of classification.

Experiment 2: Classification of Records Using the CrimeLabel Target Class

This experiment uses the attribute ***CrimeLabel*** to classify records. The attribute consists of three classes; serious, medium and low. Every record contains 158 (excluding ***incident id*** and the target class ***CrimeLabel***) attributes and were supplied to the decision tree; however, the decision tree algorithm has selected only 20 attributes to make the classification task.

Initially, the experiment was carried out using the default values provided by the program and by modifying some of the parameters. Using the default values of the parameters the learning scheme has been conducted and resulted in a decision tree containing 402 nodes and 283 leaves. The accuracy of this learning scheme was 86 percent which indicates that out of the total number of records supplied 4,734 (86%) records were classified correctly while the remaining 766 (14%) records were classified incorrectly. The summary of the output is presented in table 4.5. Moreover, the results of the experiment has shown that about 92 percent of the records in the class of serious crime were correctly classified while about 80 and 85 percent of the records in the medium and low crime classes respectively were classified correctly.

Table 4.5: Output from the J48 decision tree learner using the default values of the parameters.

Actual	**Predicted**			**Total**	**Score**
	Serious	Medium	Low		**(Accuracy rate)**
Serious	2325	134	103	2562	91.75%
Medium	218	1137	76	1431	79.46%
Low	167	68	1272	1507	84.41%
Total	2710	1339	1451	5,500	86.07.36%

The decision tree generated in this learning scheme assured that as the size of the tree keeps increasing, it becomes difficult to analyze, interpret and generate rule sets. For this reason, the decision tree discovered in this training scheme has become complex and unease to interpret. To address such difficulties an attempt was made to modify some of the parameters. The parameter minNumObj (minimum number of instances in a leaf) was set to 25, which were 2 in its default value and this figure was set after trying different values. This means the process of classifying of records proceeds as long as the number of records at each leaf node is reached 25.

The percentage split option was used to partition the dataset into training and testing data and this parameter was set to 80, 80% for training and 20% for testing. The result of this learning scheme is summarized and presented in table 4.6. This result portrays that due to the adjustment of some of the parameters, the size of tree reduced to 74 and the number of leaves has become 49.

Table 4.6: Output from the J48 decision tree learner by adjusting some values of the parameters.

Actual	**Predicted**			**Total**	**Score (Accuracy rate)**
	Serious	Medium	Low		
Serious	476	24	20	520	91.54%
Medium	43	233	17	293	79.52%
Low	48	14	225	287	78.40%
Total	567	271	262	1,100	84.90%

The output of this experiment depicted that out of the total test data (1,100 records) provided 85 percent records were classified correctly. The results of the experiment has also shown that about 92 percent of the records in the class of serious crime were correctly classified while about 80 and 78 percent of the records in the medium and low crime classes respectively were classified correctly. As compared to the training scheme carried out with the default values for the different parameters, the level of complexity of the rule that could be generated from this decision tree could be easily understandable.

An attempted has been also made to conduct the experiment by excluding some of the attributes to see if the accuracy of the learning scheme could be improved. For instance, by dropping the attributes zone and woreda, the accuracy of learning scheme was found to be 85 percent, which is similar to the above.

Although several attempts were made, it was not possible to improve the performance of the classification scheme above 85%. Thus, this indicates that about 15% records are wrongly classified and this may account to the error made by the crime analysis experts in classifying the data employed in this study.

4.6.2 Generating Rules from Decision Tree

From the decision tree developed in the aforementioned experiments, it is possible to find out a set of rules simply by traversing the decision tree and generating a rule for each leaf and making a combination of all the tests found on the path from the root to the leaf node (Bao, 2003). This produces rules that are unambiguous in that it doesn't matter in what order they are executed. The following are some of the rules extracted from the decision tree depicted in figure 4.4.

1. *IF MAOEMPE<=0 and MOEDCU<=0 and FOEDS<=0 and MAOEOTH<=0 and MAOEMG<=0 and MAOHAU<=1 and MOEDJ<=0 and MAOEPW<=0 and TOM>=0and MAOEMUN<=0 and TOFF>=0*

   *THEN **Urban** (2986.0/179.0)*

2. *IF MAOEMPE<=0 and MOEDCU<=0 and FOEDS<=0 and MAOEOTH<=0 and MAOEMG<=0 and MAOHAU>=1 and MAO1930>=0 and MAOEMUN>1*

   *THEN **Rural** (164.0/0*

3. *IF MAOEMPE<=0 and MOEDCU<=0 and FOEDS<=0 and MAOEOTH<=0 and MAOEMG<=0 and MAOHAU<=1 and MOEDJ<=0 and MAOEPW>=0 and MOEDI<=0*

   *THEN **Rural** (209.0/5.0)*

4. *IF MAOEMPE<=0 and MOEDCU<=0 and FOEDS<=0 and MAOEOTH<=0 and MAOEMG<=0 and MAOHAU>=1 and MAO1930>=0 and MAOEMUN<=1 and MAOHN<=1*

   *THEN **Rural** (35.0/5.0)*

5. *IF MAOEMPE<=0 and MOEDCU<=0 and FOEDS<=0 and MAOEOTH<=0 and MAOEMG<=0 and MAOHAU>=1 and MAO1930>=0 and MAOEMUN<=1 and MAOHN<=1*

   *THEN **Rural** (169.0/5.0)*

6. *IF MAOEMPE<=0 and MOEDCU<=0 and FOEDS<=0 and MAOEOTH<=0 and MAOEMG<=0 and MAOHAU<=1 and MOEDJ<=0 and MAOEPW>=1 and MOEDI>=1 and MAOHAN<=0*

   *THEN **Urban** (89.0/3.0)*

7. *IF MAOEMPE<=0 and MAOEMUN<=4 and FOEDS<=0 and MAOEOTH<=0 and MAOEMG<=0 and MAOHAU>1 and MOA1930 <= 0:*

   *THEN **Rural** (164.0/3.0)*

The rules presented above indicate the possible conditions in which a crime record could be classified in each class. The numeric value in the logical expression indicates the number of offenders involved in each crime incident having the specified attribute. For instance, in rule 7 the value of MAOEMUN (Male Adult Offender employment Unemployed) is less than or equal to 4 and this indicates that as the number of offenders having this characteristics is less or equal to this figure; in addition to the other stated attribute values are supposed to commit a crime in rural areas.

Moreover, the numbers contained in a bracket refers to the number of records classified correctly and wrongly in each class. For example, in rule 1 the numbers in the bracket are to mean that this rule has classified 210 records correctly and 21 records wrongly.

These rules have indicated that attributes such as age, sex, occupation, education level, unemployment, and habit are the basis for classification.

For instance, in the case of rule 1:

If there are both male and female offenders and

If none of the male offenders are college and junior school education level and

If there are no females with education level of secondary school, and

If the male offenders are not employed in government, private, and other institutions and not self employed and

If one or none of the male offenders has an alcohol habit,

Then the crime is committed in urban area.

This rule classifies 2986 records correctly and 179 records wrongly. The rule indicated that crimes committed in urban areas are associated with sex, educational level, occupation, and alcohol habit of offenders. According to this rule, resource deployment and other crime prevention activities should be based on the analysis of these attributes.

Rule 2 also depicted that male offenders in the age group of 19 and 30 whose educational level is not college/university and who have alcohol habits and who are unemployed and female offenders who are not with education level of secondary school tend to commit crime in rural areas.

Having such interpretation of rules could help the police to design crime prevention policies and provide counseling services for target groups. For instance, police can design counseling, advising and awareness creating courses of action for such groups cited in rule number 2 since individuals in this age group could be found easily in schools and other educational institutions.

Moreover, from the decision tree constructed using the target class of crime label (***CrimeLabel***) rules can be generated which could be helpful to justify why a crime record is classified into one of the predefined classes. The following rule can be presented as an example from the decision tree displayed in appendix B (Classification on the target class ***SceneLabel***):

1. IF MAOEMUN<=4 and MAOEMG<=1 and MAOHOTHE<=1 and FAOEPW<=0

and TOM>0 and MAOHAU>1 and MOA1930<=0:

*THEN **Low** (164.0/3.0)*

2. IF MAOEMUN<=4 and MAOEMG<=1 and MAOHOTHE<=1 and FAOEPW<=0

and TOM<=0 and SceneLabel = rural and Woreda = 4:

*THEN **Low** (162.0/1.0)*

3. IF MAOEMUN<=4 and MAOEMG<=1 and MAOHOTHE<=1 and FAOEPW<=0 and TOM>0 and MAOHAU<=1 and MAOEMPE>0 MAOHS>0:

THEN Serious (166.0/1.0)

4. IF MAOEMUN<=4 and MAOEMG<=1 and MAOEOTH<=1 and FAOEPW<=0 and TOM>0 and MAOHAU<=1 and MAOEMPE>0 MAOHS<=0:

THEN ***Low*** *(343.0/16.0)*

5. IF MAOEMUN<=4 and MAOEMG<=1 and MAOHOTHE<=1 and FAOEPW<=0 and TOM>0 and MAOHAU<=1 and MAOEMPE>0 MAOHS>0:

THEN ***Serious*** *(166.0/1.0)*

6. IF MAOEMUN<=4 and MAOEMG<=1 and MAOHOTHE<=1 and FAOEPW<=0 and TOM>0 and MAOHAU<=1 and MAOEMPE>0 and MAOEOTH>0:

THEN ***Medium*** *(90.0/5.0)*

7. IF MAOEMUN <= 4 and MAOEMG <= 1 and MAOHOTHE <= 1 and FAOEPW <= 0 and TOM > 0 and MAOHAU <= 1 and MAOEMPE <= 0 and MAOEOTH <=0 and MOMNM<0 and MOEDS <= 0 and MAOEPW <= 0 and FORC <= 0:

THEN ***serious*** *(1916.0/296.0)*

The above rules indicate how a given record could be classified based on some attribute values. Hence, having these rules, instances were classified into the predefined classes. In fact, in classifying crime records into the predefined classes, attributes such as occupation, habit, unemployment, sex, age, address, religion, marital status, educational level of offenders, and the area where the crime is occurred have been the basis for classification.

An interesting finding identified from the above rules, particularly from rule 4 and 5 is the association of level of crime and habit of smoking. As it is observed in rule 5 if the number of male offenders who are unemployed is less or equal to four and the number of male offenders working in government, private and other institution and have alcohol habit is less or equal to one and have smoking habit then tend to commit serious crimes. While rule 4 indicated that male offender having these attributes however nonsmokers involve in low crimes.

Hence, this indicates that the cost of crime associated with smokers is much higher and police should take this into account while providing counseling and awareness rising course of action to discourage such bad habits and thus curb the cost of crime.

4.7 Neural Network Model Building Using Weka Software

The second data mining technique employed in this study is neural networks. To build the neural network model, Weka software package is used and it employs back propagation algorithm in developing a model.

In order to train a model there is usually a need to prepare the dataset in a form, which is suitable to the particular data mining technique and software. As explained in sections 4.3, 4.4 and 4.5 an attempt has been made to clean and preprocess the data for decision tree. Much of the tasks mentioned in these sections are also applicable for neural networks with a little variation. In this section some of the additional data preparation needs of neural networks are described.

Neural networks technique can only process datasets when the values of attributes (variables) are numeric. Only the target class could be nominal valued variable. Moreover, the values of each attribute must be between 0 and 1 or between -1 and 1. Therefore, to make the dataset suitable for the neural network, values of all attributes were changed into numbers between the acceptable ranges (in this case between 0 and 1). This process is usually known as normalization (Witten and Frank, 2000).

In order to normalize all attribute values to lie between 0 and 1 the following formula was applied:

$$a_i = v_i\text{-min } v_i/(\text{max } v_i\text{-min } v_i)$$

Where v_I is the actual value of attribute $_I$, and the maximum and minimum are taken over all instances in the dataset.

Once the attributes are normalized the next step was to prepare the dataset in an ARFF file format similar to the procedure stated in section 4.5 of this chapter. Then the network was trained by presenting a set of instances repeatedly through all the training set addressing each instance in turn and making necessary corrections. When the entire list of records has been presented, Weka starts over at the beginning of the list and the training process was repeated until the termination condition of the iteration process is reached.

Using the neural network program (sub package) of the weka software, two experiments were conducted i.e. classification of records on target classes of ***SceneLabel*** and ***CrimeLabel***.

In order to make comparison on the performance of decision tree and neural networks the same dataset provided to the decision tree is used for the neural network. The attributes selected by the decision tree developed earlier were provided to the neural network and hence training and testing of the model is based on these attributes. In conducting this experiment the parameters of the neural networks program were set to their default values.

The results of the experiment are presented in table 4.7 and this indicates that out of the total test set (1100 records) provided to the program 1017 (92.5%) instances were correctly classified while the remaining 83 (7.5%) instances were classified incorrectly. Looking to records in each class, 95.7 percent of the records in the class of rural were classified correctly while 87.5 percent of the records in the class of urban were classified correctly.

Table 4.7: Output of neural networks for the classification task on target class SceneLabel.

Actual	**Predicted**		**Total**	**Score (Accuracy rate)**
	Rural	Urban		
Rural	638	29	667	95.65%
Urban	54	379	433	87.53%
Total	692	403	1,100	92.46%

To improve the performance of the model some of the parameters were revised however the outcome of these attempts resulted in poor accuracy rate.

The second classification task was based on the values of attribute ***CrimeLabel***. This experiment was conducted by setting the parameters of the program to their default values. The output of this experiment is presented in table 4.9. As presented in this table out of the total datasets supplied to the neural network program 4394 (80%) records are correctly classified while the remaining 1106 (20%) records were wrongly classified. Moreover, the results of the experiment have shown that about 91 percent of the records in the class of serious crime were correctly classified while about 64 percent of the records in the medium crime class are classified correctly.

Table 4.8: Output of the experiment on neural networks for the classification task on target class CrimeLabel.

Actual	**Predicted**			**Total**	**Score (Accuracy rate)**
	Serious	Medium	Low		
Serious	2328	165	69	2562	90.87%
Medium	481	912	38	1431	63.73%
Low	212	141	1154	1507	76.58%
Total	567	271	262	15,500	79.89%

The classification of records using the attribute ***CrimeLabel*** resulted in significant number of records, which are incorrectly classified. Thus, in order to improve the performance of the model an attempt was made to modify some of the parameters. With this objective the researcher modified the number of hidden units to 6 and the learning rate (the rate at which the network weights were adjusted) and momentum (that tends to keep the change in the same direction from one iteration to the next) to 0.4 and 0.5 which were 0.3 and 0.2 in the default value respectively. However, this trial has diminished the performance of the model and the accuracy rate obtained from this experiment was 73 percent. Moreover, an attempt was also made to set the number of hidden units at different level and other parameters such as learning rate and momentum are also modified. Nevertheless, it was not possible to improve the performance of the model in spit of all these efforts.

Another problem faced during the experiment of neural network was that there was no any structural description that explicitly described how and why the classification is made. In fact, for this reason neural networks are usually known as “black boxes” (Witten and Frank, 2000). Nevertheless, recently attempts are being made to incorporate tools to the neural network that extracts rules or justifications that reveals why and how instances are classified into some classes (Han and Kamber, 2001).

4.8 Comparison of Decision Tree and Neural Networks

One of the purposes of this study was to compare the decision tree and neural networks data mining techniques and to select the one, which performs the best. Accordingly, each experiment carried out in this research has employed both decision tree and neural networks. In all experiments the same datasets were used. The output of these experiments indicated that the classification task of records using the crime scene (***SceneLabel***), both decision tree and neural network have performed well. Decision tree has shown an accuracy rate of 94 percent while neural networks classified 92.5% records correctly. Whereas in the classification task of records using the target class of ***CrimeLabel***, the accuracy of decision tree was 85 percent while, that of neural networks was 80 percent.

These results revealed that in the first experiment (i.e. classification using the target class (***SceneLabel***)) both techniques have performed well even though the accuracy of decision tree exceeds slightly. Moreover, in terms of ease and simplicity to the user the decision tree is more self-explanatory. It generates rules that can be presented in simple human language.

In the case of classification using the attribute ***CrimeLabel*** the accuracy of decision tree was about 85 percent whereas the accuracy of neural networks has declined to 80 percent. Still decision tree performed better. Therefore, it is plausible to conclude that the decision tree data mining technique is more appropriate to this particular case than the neural network.

To evaluate the importance of the results obtained from the classification task and the rules generated from the decision tree, the results were provided to the crime analysis experts at the Federal Police Commission. They have suggested that the results of the experiment are appealing to initiate crime prevention methods. According to them, the accuracy rate and combination of rules obtained in both experiments is promising, although it could have been more important if the accuracy rate in the classification of records using the attribute ***CrimeLabel*** could be improved.

Hence, the Oromia Police Commission could employ the results of this experiment as an input for the decision making process on resource deployment, designing training programs and crime prevention and investigation strategies polices. For instance, sites identified, as crime prone deserves much resource and well trained policemen. Individuals with some defined attributes that are thought to be criminals could be consulted and advised accordingly. Moreover, when offenders are under custody a counseling service could be provided based on their crime involvement. In addition to this, detectors could use this model to scrutinize suspected individuals before a depth investigation is commenced.

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

The purpose of this study was to explore the applicability of data mining techniques in the process of crime prevention for the Oromia Police Commission. In doing so, 5,500 sample data were taken randomly from the criminal record database of the Oromia Police Commission to build and test the model.

The results of the experiments carried out in this research using decision tree and neural networks have revealed that the technique of data mining is applicable in the process of crime prevention. Taking the crime scene (***SceneLabel***) attribute, the decision tree data mining technique has classified crime records at accuracy rate of 94 percent. Moreover, the decision tree has classified crime records into some crime levels (serious, medium and low crime) at accuracy rate of 85 percent.

On the other hand, the results of experiment carried out using neural networks has reported 92.5 percent accuracy rate on the target class of crime scene (***SceneLabel***) and 80 percent accuracy rate on the target class of crime level (***CrimeLabel***).

Although both techniques have shown promising results, the decision tree data mining technique was found more appropriate to the crime prevention problem as the accuracy rate was relatively higher in both experiments. Moreover, decision tree seems applicable due to the fact that in contrast to neural networks, it expresses the rules explicitly. These rules can be expressed in human language so that anyone can easily understand how and why a classification of instances is made.

The tree generated from the experiments on decision tree have shown that attributes of offenders such as sex, occupation, educational level, and age could be employed to classify whether the crime is committed in rural or urban areas. On the other hand, in the task of classification of records into the three crime levels, (i.e. serious, medium and low) attributes such as sex, age, occupation, religion, woreda, education level and marital status has become a basis for classification.

To sum up, the results of this experiment could be employed as an input for the decision making process on resource deployment, designing training programs, and crime prevention and investigation methods accordingly. To mention some, sites identified, as crime prone may be worthy of much resource and well trained policemen. Moreover, when offenders are under custody a counseling service could be provided based on their crime involvement. In addition to this, detectors could use this model to scrutinize suspected individuals before a depth investigation is commenced.

5.2 Recommendations

Basically, this research was conducted for an academic purpose. However, that the results of this study are found to be promising to be applied to address practical problems of crime prevention. The results of this study have shown that the data mining technology particularly the decision tree technique is well applicable in the efforts of crime management. Hence, based on the findings of this study, the following recommendations can be forwarded.

- Data mining techniques could contribute a lot in identifying areas that are crime prone and the characteristics of offenders who involve in different levels of crimes. Thus, it could be more important to use the data mining technique as a tool for the decision making process. In other words, the Oromia Police Commission could optimize its crime prevention efforts by employing data mining technology.

- Although both decision tree and neural networks reported promising results and hence could be applied in the area of crime prevention, decision tree tends to perform better. Hence, it would be more optimal for the Oromia Police Commission to employ the model developed with this technique.

- Although features such as physical descriptions of offenders, geographic information of crimes and the like are important in the analysis of crime patterns, these were not incorporated in the crime records database of the Oromia Police Commission. Hence, in order to exploit the potential advantage of data mining, it would be more important to incorporate such additional features. For instance, if one knows the specific site where a crime is committed such as nearby school, hotels, market etc, then, a due attention may be given to such sites. Moreover, physical descriptions of an offender could help to search and match suspected person for a specific crime by comparing with the information contained in a database.

- The way in which crime incidents are designated as serious, medium or low crime needs a revision. Some crimes may be grouped erroneously in a crime group that does not belong to it. For instance, crimes that are threats of these days of the international community such as terrorism are designated as medium crime that lacks a sound justification. This may be one reason why the model has reported relatively a low accuracy rate in this particular experiment.

- Although in this study encouraging results were obtained, a sample data was used for training and testing classifiers due to time constraint. Hence, it is appropriate to conduct the experiments with large training and testing datasets as well as making a number of trials to come out with more accurate and better performing classifiers.

References

Andargachew, Tesfaye. (1988). *The crime problem and its correction*: Volume I Addis Ababa.

ANGOSS (2003). A brief history of data mining. http://www.data-mining-software.com/data_mining_history.htm

Askale, Worku. (2001). *Possible application of data mining technology in supporting loan disbursement activity at Dashen Bank S.C*. Masters Thesis Addis Ababa University, Addis Ababa.

Berry, M. and Linoff, G. (1997). *Data mining techniques: For marketing, sales and customer support*. New York. John Wiley and Sons, Inc.

Baeza-Yates, R. and Ribeiro-Neto, B. (1999). *Modern Information Retrieval*. Harlow, England: Addison Wesley, 1999.

Bigus, J.P. (1996). *Data mining with neural networks in solving business problems: From application to Development to decision support*. New York: McGraw-Hall.

Brown, D. (2003). The Regional Crime Analysis Program (RECAP): A Framework for Mining Data to Catch Criminals. http://vijis.sys.virginia.edu/publication/RECAP.pdf

Bao, H (2003). Knowledge Discovery And Data Mining Techniques And Practice.

http://www.netnam.vn/unescocourse/knowlegde/3-1.htm

Butler, A. (1992). *Police Management*. Second edition.

England: Dartmouth Publishing Company limited.

Carbone, P. (1997). Data Mining or “Knowledge Discovery In Databases”: An Overview. http://www.mitre.org/pubs/data_mgt/Papers/DMHdbk.pdf.

Corcoran, J. and Ware, J. (2001). Predicting the Geo-Temporal Variations of Crime and Disorder. http://www.ojp.usdoj.gov/nij/maps/Conferences/02conf/Ware.doc.

Deogun, J.s. (1998). Feature selection and effective classifiers. Journal of American society for information science.Vol.49 (5).

DRE.(1995). Federal Negarit Gazeta of the democratic republic of Ethiopia. No-1.

Frohlich, J. (1999). Neural Net Overview. http://rfhs8012.fh-regenburg.de/~saj39122/jfroehl/diplom/e-text.html

ØÁ‰L ±lþS ÷¸>N. (1996). bwNjL#bwÈT _ÍtŸnTÂ bmNgD T‰ðK XNQSÝs¤ §Y ytkÂwn _ÂT# xÄþS xbÆ”””

ØÁ‰L ±lþS ÷¸>N. (2003). bwNjL#bwÈT _ÍtŸnTÂ bmNgD T‰ðK XNQSÝs¤ §Y

Fayyad, U, Piatesky-Shapiro, G and Smyth, P. (1996). Knowledge Discovery and Data Mining: Towards a Unifying Framework. ftp://ftp.research.microsoft.com/pub/dtg/fayyad/kdd96/fayyad-intro.ps.

Haykin, S. (1994). Neural Networks: A comprehensive Foundation. USA. Prentice-Hall, inc.

Han, J. and Kamber, M (2001). *Data Mining: Concepts and Techniques*. San Francisco: Morgan Kaufmann publishers.

Henock, Woubshet. (2002). Application of data mining techniques to support customer Relationship management at Ethiopian airlines. Masters Thesis Addis Ababa University, Addis Ababa.

Gobena, Mikael. (2000). Flight Revenue Information Support System for Ethiopian Airlines. Masters Thesis Addis Ababa University, Addis Ababa.

Garvin, M. (1998). Data Mining and the Web: What They Can Do Together. http://www.dmreview.com/master.cfm?NavID=198&EdID=420

International City Managers association. (1961) *Municipal Police Administration*, USA, Cushing Malloy, Inc and Ann Arbor, Michigan.

Javid, S. (1999). Data Mining in the Next Millennium. http://www.dmreview.com/master.cfm?NavID=198&EdID

Kim, H. (2000). Data Mining for Business Solution. http://www.ciadvertising.org/student_account/fall_00/adv391k/hyojin /termpaper/comparison.html

LIS - Rudjer Boskovic Institute. (2002). Introduction to Data Mining. http://dms.irb.hr/project/proj_main.php

Mitchell, M. (1997). *Machine Learning*. New York: The McGraw-Hill Companies, Inc.

Masuda, Y. (1995). *The Information Technology Revolution: Parameters of the Postindustrial, society*. Basil Blackwee Ltd.

Mutch, J. (1999). Technology: Unlocking the neural network.

http://rmisweb.com/rmisartc/riskins/0199b.htm.

Mesfin, G/Michael (1999). *A computer based criminal records management system case study of the Addis Ababa police*. Masters Thesis Addis Ababa University (SISA), Addis Ababa.

Megaputer Intelligence, Inc. (2002). Crime Pattern Analysis: Megaputer Case Study.

http://www.elon.edu/facstaff/mconklin/cis230/cases/crime_pattern_case.pdf

Mangold, J. (1998). Turning Straw into Gold: How Data Mining Transforms Data from an Expense into an Asset.

http://www.dmreview.com/master.cfm?NavID=198&EdID=826

Perez, B. (2001). Data mining technology use grows.

http://robotics.stanford.edu/users/roonnyk/kohavilnSCMP.pdf

Regional Crime Analysis Program. (2000). The Design and Implementation of Regional crime analysis program: Small department edition (ReCAP SDE)

http://www.sys.virginia.edu/capstone/past/cap2000/11-01ReCAP.doc

Rudjer Boskovinc institute (2001). Overview of technology and its place in Business.

http://www.vancouver.wsu.edu/fac/roseg/MIS375/fall2002projects/Datamining.doc.

Raghavan, V.v; Deogun, J.s. and Sever, H. (1998). Knowledge discovery and data mining: Introduction. Journal of American society for information science.Vol.49(5).

Shegaw Anagaw. (2002). *Application of data mining technology to predict child mortality patterns: The case of (Butajira Rural Health Project) BRHP.* Masters Thesis Addis Ababa University, Addis Ababa.

SPSS Inc. (2003). West Midlands Police Department. http://www.spss.com/success/pdf/WMPAPP0299.PDF.

Thearling ,K. (2003). An introduction to data mining. http://www3.shore.net/~kht/text/dmwhite.htm.

Thakur, C. (2003). Crime Control. http://ncthakur.itgo.com/chand3c.htm

Tesfaye, Hintsay. (2002). *Predictive Modeling Using Data Mining Techniques In Support to Insurance Risk Assessment*. Masters Thesis Addis Ababa University, Addis Ababa.

Veenendaal and Houweling (2003). Gut feelings, crime data and GIS. http://robotics.stanford.edu/users/ronnyk/kohaviInSCMP.pdf

Wilson , O.W. (1963). *Police Administration*. USA, McGraw Hill Company.

Witten, I. and Frank, E. (2000). *Data mining: Practical Machine Learning Tools and Techniques with Java Implementations.* San Francisco: Morgan Kaufmann publishers.

Appendices A

List of attributes, their possible values and descriptions

No.	Field Name (Attribute)	Set values (Data Type)	Description
1	IncidentId	Number	Computer Generated key No
2	TOFM	Number	Total Offender Found Male
3	TOFF	Number	Total Offenders Found Female
4	TOM	Number	Total Offenders Male
5	TOF	Number	Total Offenders Female
6	MOA915	Number	Male offenders Age 9-15
7	FOA915	Number	Female offenders Age 9-15
8	MOA1618	Number	Male offenders Age 16-18
9	FOA1618	Number	Female offenders Age 16-18
10	MOA1930	Number	Male offenders Age19-30
11	FOA1930	Number	Female offenders Age 19-30
12	MOA3150	Number	Male offenders Age 31-50
13	FOA3150	Number	Female offenders Age 31-50
14	MOA51	Number	Male offenders Age 51 Plus
15	FOA51	Number	Female offenders Age 51 plus
16	MOAUNK	Number	Female offenders Age Unknown
17	FOAUNK	Number	Male offenders Age Unknown
18	MOEDI	Number	Male Offender Education Illiterate
19	FOEDI	Number	Female Offender Education Illiterate
20	MOEDP	Number	Male Offender Education Primary
21	FOEDP	Number	Female Offender Education Primary
22	MOEDJ	Number	Male Offender Education Junior
23	FOEDJ	Number	Female Offender Education Junior
24	MOEDS	Number	Male Offender Education Secondary

25	FOEDS	Number	Female Offender Education Secondary
26	MOEDCU	Number	Male Offender Education College/University
27	FOEDCU	Number	Female Offender Education College/University
28	MOEDUNK	Number	Male Offender Education Unknown
29	FOEDUNK	Number	Female Offender Education Unknown
30	MOMM	Number	Male Offender Marital Married
31	FOMM	Number	Female Offender Marital Married
32	MOMNM	Number	Male Offender Marital Not Married
34	FOMNM	Number	Female Offender Marital Not Married
35	MOMD	Number	Male Offender Marital Divorced
36	FOMD	Number	Female Offender Marital Divorced
37	MOMS	Number	Male Offender Marital Separated
38	FOMS	Number	Female Offender Marital Separated
39	MOMUNK	Number	Male Offender Marital Unknown
40	FOMUNK	Number	Female Offender Marital Unknown
41	MONE	Number	Male Offender Nationality Ethiopian
42	FONE	Number	Female Offender Nationality Ethiopian
43	MONOTH	Number	Male Offender Nationality Others
44	FONOTH	Number	Female Offender Nationality Others
45	MONUNK	Number	Male Offender Nationality Unknown
46	FONUNK	Number	Female Offender Nationality Unknown
47	MORC	Number	Male Offender Religion Christian
48	FORC	Number	Female Offender Religion Christian
49	MORI	Number	Male Offender Religion Islam
50	FORI	Number	Female Offender Religion Islam
51	MORP	Number	Male Offender Religion pagan
52	FORP	Number	Female Offender Religion pagan
53	MOROTH	Number	Male Offender Religion Others
54	FOROTH	Number	Female Offender Religion Others

55	MORUNK	Number	Male Offender Religion Unknown
56	FORUNK	Number	Female Offender Religion Unknown
57	MAOEMG	Number	Male Adult Offender Employment Government
58	FAOEMG	Number	Female Adult Offender Employment Government
59	MAOEMPE	Number	Male Adult Offender Employment Private Employee
60	FAOEMPE	Number	Female Adult Offender Employment Private Employee
61	MAOEMF	Number	Male Adult Offender Employment Farming
62	FAOEMF	Number	Female Adult Offender Employment Farming
63	MAOEMUN	Number	Male Adult Offender Employment Unemployed
64	FAOEMUN	Number	Female Adult Offender Employment Unemployed
65	MAOEPW	Number	Male Adult Offender Employment Private Worker
66	FAOEPW	Number	Female Adult Offender Employment Private Worker
67	MAOENGO	Number	Male Adult Offender Employment NGO
68	FAOENGO	Number	Female Adult Offender Employment NGO
69	MAOESTR	Number	Male Adult Offender Employment Street
70	FAOESTR	Number	Female Adult Offender Employment Street
71	MAOESTUD	Number	Male Adult Offender Employment Student
72	FAOESTUD	Number	Female Adult Offender Employment Student
73	MAOEOTH	Number	Male Adult Offender Employment Others

74	FAOEOTH	Number	Female Adult Offender Employment Others
75	MAOHUNK	Number	Male Adult Offender Employment Unknown
76	FAOHUNK	Number	Female Adult Offender Employment Unknown
77	MJOEMG	Number	Male Juvenile Offender Employment Government
78	FJOEMG	Number	Female Juvenile Offender Employment Government
79	MJOEMPE	Number	Male Juvenile Offender Employment Private Employee
80	FJOEMPE	Number	Female Juvenile Offender Employment Private Employee
81	MJOEMF	Number	Male Juvenile Offender Employment Farming
82	FJOEMF	Number	Female Juvenile Offender Employment Farming
83	MJOEMUN	Number	Male Juvenile Offender Employment Unemployed
84	FJOEMUN	Number	Female Juvenile Offender Employment Unemployed
85	MJOEPW	Number	Male Juvenile Offender Employment Private Worker
86	FJOEPW	Number	Female Juvenile Offender Employment Private Worker
87	MJOENGO	Number	Male Juvenile Offender Employment NGO
88	FJOENGO	Number	Female Juvenile Offender Employment NGO
89	MJOESTRE	Number	Male Juvenile Offender Employment Street
90	FJOESTRE	Number	Female Juvenile Offender Employment

			Street
91	MJOESTUD	Number	Male Juvenile Offender Employment Student
92	FJOESTUD	Number	Female Juvenile Offender Employment Student
93	MJOOTHERS	Number	Male Juvenile Offender Employment Others
94	FJOOTHERS	Number	Female Juvenile Offender Employment Others
95	MJOEUNK	Number	Male Juvenile Offender Employment Unknown
96	FJOEUNK	Number	Female Juvenile Offender Employment Student
97	MAOHDU	Number	Male Adult Offender Habit Drug Use
98	FAOHDU	Number	Female Adult Offender Habit Drug Use
99	MAOHAU	Number	Male Adult Offender Habit Alcohol Use
100	FAOHAU	Number	Female Adult Offender Habit Alcohol Use
101	MAOHS	Number	Male Adult Offender Habit Smoker
102	FAOHS	Number	Female Adult Offender Habit Smoker
103	MAOHGS	Number	Male Adult Offender Habit Glue Sniffing
104	FAOHGS	Number	Female Adult Offender Habit Glue Sniffing
105	MAOHGAM	Number	Male Adult Offender Habit Gambling
106	FAOHGAM	Number	Female Adult Offender Habit Gambling
107	MAOHN	Number	Male Adult Offender Habit None
108	FAOHN	Number	Female Adult Offender Habit None
109	MAOHOTHE	Number	Male Adult Offender Habit Others
110	FAOHOTHE	Number	Female Adult Offender Habit Others
111	MAOEUNK	Number	Male Adult Offender Habit Unknown
112	FAOEUNK	Number	Female Adult Offender Habit Unknown
113	MJOHDU	Number	Male Juvenile Offender Habit Drug Use
114	FJOHDU	Number	Female Juvenile Offender Habit Drug Use

115	MJOHAU	Number	Male Juvenile Offender Habit Alcohol Use
116	FJOHAU	Number	Female Juvenile Offender Habit Alcohol Use
117	MJOHS	Number	Male Juvenile Offender Habit Smoker
118	FJOHS	Number	Female Juvenile Offender Habit Smoker
119	MJOHGS	Number	Male Juvenile Offender Habit Glue Sniffing
120	FJOHGS	Number	Female Juvenile Offender Habit Glue Sniffing
121	MJOHGAM	Number	Male Juvenile Offender Habit Gambling
122	FJOHGAM	Number	Female Juvenile Offender Habit Gambling
123	MJOHN	Number	Male Juvenile Offender Habit None
124	FJOHN	Number	Female Juvenile Offender Habit None
125	MJOHOTHE	Number	Male Juvenile Offender Habit Others
126	FJOHOTHE	Number	Female Juvenile Offender Habit Others
127	MJOHUNK	Number	Male Juvenile Offender Habit Unknown
128	FJOHUNK	Number	Female Juvenile Offender Habit Unknown
129	MJOLPAR	Number	Male Juvenile Offender Living with Parents
130	FJOLPAR	Number	Female Juvenile Offender Living with Parents
131	MJOLMOT	Number	Male Juvenile Offender Living with Mother
132	FJOLMOT	Number	Female Juvenile Offender Living with Mother
133	MJOLFAT	Number	Male Juvenile Offender Living with Father
134	FJOLFAT	Number	Female Juvenile Offender Living with Father
135	MJOLREL	Number	Male Juvenile Offender Living with Relatives

136	FJOLREL	Number	Female Juvenile Offender Living with Relatives
137	MJOLOTH	Number	Male Juvenile Offender Living with Others
138	FJOLOTH	Number	Female Juvenile Offender Living with Others
139	MJOLSTR	Number	Male Juvenile Offender Living with Street
140	FJOLSTR	Number	Female Juvenile Offender Living with Street
141	MJOLOTHERS	Number	Male Juvenile Offender Living Others
142	FJOLOTHERS	Number	Female Juvenile Offender Living Others
143	MJOLUNK	Number	Male Juvenile Offender Living with Unknown
144	FJOLUNK	Number	Female Juvenile Offender Living with Unknown
145	PEOM	Number	Male Ethnic Oromo
146	PEOFF	Number	Female Ethnic Oromo
147	PEAM	Number	Male Ethnic Amara
148	PEAF	Number	Female Ethnic Amara
149	PETM	Number	Male Ethnic Tigray
150	PETF	Number	Female Ethnic Tigray
141	PESM	Number	Male Ethnic Southern
152	PESF	Number	Female Ethnic Southern
153	PEOTHM	Number	Male Ethnic Others
154	PEOTHF	Number	Female Ethnic Others
155	PEUNKM	Number	Male Ethnic Unknown
156	PEUNKF	Number	Female Ethnic Unknown
157	ZoneID	Nominal {1,2,3- - - 15}	Zone Id of Crime
158	WoredaID	Nominal	Woreda Id of Crime

		{1,2,3,- - - 25}	
159	***CrimeLabel***	Nominal {serious, medium, low}	Type of Crime
160	SceneLabel	Nominal {rural, urban}	Type of Crime Scene

Appendix A.1

List of attributes, their possible values and descriptions (Attributes selected by the decision tree for the experiment on the target class SceneLabel/

No.	Field Name (Attribute)	Set values (Data Type)	Description
1	TOM	Number	Total Offenders Male
2	TOFF	Number	Total Offenders Found Female
3	MOA915	Number	Male offenders Age 9-15
4	MOA1930	Number	Male offenders Age19-30
5	MOEDI	Number	Male Offender Education Illiterate
6	FOEDI	Number	Female Offender Education Illiterate
7	MOEDP	Number	Male Offender Education Primary
8	MOEDJ	Number	Male Offender Education Junior
9	FOEDS	Number	Female Offender Education Secondary
10	MOEDCU	Number	Male Offender Education College/University
11	FOMNM	Number	Female Offender Marital Not Married
12	MORC	Number	Male Offender Religion Christian
13	MAOEMG	Number	Male Adult Offender Employment Government
14	MAOEMPE	Number	Male Adult Offender Employment Private Employee
15	MAOEMUN	Number	Male Adult Offender Employment Unemployed
16	MAOEPW	Number	Male Adult Offender Employment Private Worker
17	MAOEOTH	Number	Male Adult Offender Employment Others
18	MAOEMF	Number	Male Adult Offender Employment Farming
19	MAOHAU	Number	Male Adult Offender Habit Alcohol Use
20	MAOHN	Number	Male Adult Offender Habit None

21	MAOHOTHE	Number	Male Adult Offender Habit Others
22	SceneLabel	Nominal {rural, urban}	Type of Crime Scene

Appendix A.2

List of attributes, their possible values and descriptions (Attributes selected by the decision tree for the experiment on the target class of CrimeLabel)

No.	Field Name (Attribute)	Set values (Data Type)	Description
1	TOM	Number	Total Offenders Male
2	MOA3150	Number	Male offenders Age 31-50
3	MOA1930	Number	Male offenders Age19-30
4	MOEDP	Number	Male Offender Education Primary
5	MOEDJ	Number	Male Offender Education Junior
6	MOEDS	Number	Male Offender Education Secondary
7	MOMNM	Number	Male Offender Marital Not Married
8	FORC	Number	Female Offender Religion Christian
9	MORI	Number	Male Offender Religion Islam
10	MAOEMG	Number	Male Adult Offender Employment Government
11	MAOEMPE	Number	Male Adult Offender Employment Private Employee
12	MAOEMUN	Number	Male Adult Offender Employment Unemployed
13	MAOEPW	Number	Male Adult Offender Employment Private Worker
14	FAOEPW	Number	Female Adult Offender Employment Private Worker
15	MAOEOTH	Number	Male Adult Offender Employment Others
16	MAOHS	Number	Male Adult Offender Habit Smoker
17	MAOHAU	Number	Male Adult Offender Habit Alcohol Use
18	WoredaID	Nominal {1,2,3,- - -	Woreda Id of Crime

		25}	
19	***CrimeLabel***	Nominal {serious, medium, low}	Type of Crime
20	SceneLabel	Nominal {rural, urban}	Type of Crime Scene

Appendix B

A decision tree generated from the J48 decision tree learner after some of the parameters are modified for the target class CrimeLabel.

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

J48 pruned tree

```
MAOEMUN <= 4
|   MAOEMG <= 1
|   |   MAOHOTHE <= 1
|   |   |   FAOEPW <= 0
|   |   |   |   TOM <= 0
|   |   |   |   |   SceneLabel = rural
|   |   |   |   |   |   WoredaID = 1: serious (10.0/4.0)
|   |   |   |   |   |   WoredaID = 2: serious (7.0/2.0)
|   |   |   |   |   |   WoredaID = 3: low (6.0/3.0)
|   |   |   |   |   |   WoredaID = 4: low (162.0/1.0)
|   |   |   |   |   |   WoredaID = 5: low (82.0/1.0)
|   |   |   |   |   |   WoredaID = 6: low (3.0/1.0)
|   |   |   |   |   |   WoredaID = 7: low (82.0/2.0)
|   |   |   |   |   |   WoredaID = 8: serious (2.0/1.0)
|   |   |   |   |   |   WoredaID = 9: medium (2.0)
|   |   |   |   |   |   WoredaID = 10: serious (2.0/1.0)
|   |   |   |   |   |   WoredaID = 11: low (1.0)
|   |   |   |   |   |   WoredaID = 12: serious (2.0/1.0)
|   |   |   |   |   |   WoredaID = 13: low (1.0)
|   |   |   |   |   |   WoredaID = 14: serious (2.0/1.0)
|   |   |   |   |   |   WoredaID = 15: serious (2.0)
|   |   |   |   |   |   WoredaID = 16: low (1.0)
|   |   |   |   |   |   WoredaID = 17: serious (1.0)
|   |   |   |   |   |   WoredaID = 18: low (0.0)
|   |   |   |   |   |   WoredaID = 19: low (0.0)
|   |   |   |   |   |   WoredaID = 20: low (1.0)
|   |   |   |   |   |   WoredaID = 21: medium (1.0)
|   |   |   |   |   |   WoredaID = 22: low (0.0)
|   |   |   |   |   |   WoredaID = 23: low (0.0)
|   |   |   |   |   |   WoredaID = 24: low (0.0)
|   |   |   |   |   |   WoredaID = 25: low (0.0)
|   |   |   |   |   SceneLabel = urban: serious (40.0/23.0)
|   |   |   |   TOM > 0
|   |   |   |   |   MAOHAU <= 1
|   |   |   |   |   |   MAOEMPE <= 0
|   |   |   |   |   |   |   MAOEOTH <= 0
|   |   |   |   |   |   |   |   MOMNM <= 0
|   |   |   |   |   |   |   |   |   MOEDS <= 0
|   |   |   |   |   |   |   |   |   |   MAOEPW <= 0
|   |   |   |   |   |   |   |   |   |   |   FORC <= 0: serious (1916.0/296.0)
|   |   |   |   |   |   |   |   |   |   |   FORC > 0: low (25.0/13.0)
|   |   |   |   |   |   |   |   |   |   MAOEPW > 0: low (36.0/21.0)
|   |   |   |   |   |   |   |   |   MOEDS > 0
|   |   |   |   |   |   |   |   |   |   MOA3150 <= 0: low (103.0/18.0)
|   |   |   |   |   |   |   |   |   |   MOA3150 > 0: serious (186.0/15.0)
```

```
|  |  |  |  |  |  |  |  |  MOMNM > 0
|  |  |  |  |  |  |  |  |  |  MAOHAU <= 0
|  |  |  |  |  |  |  |  |  |  |  MAOHOTHE <= 0
|  |  |  |  |  |  |  |  |  |  |  |  MOEDP <= 0
|  |  |  |  |  |  |  |  |  |  |  |  |  MOEDJ <= 0
|  |  |  |  |  |  |  |  |  |  |  |  |  | MOA1930<=0:serious (40.0/21.0)
|  |  |  |  |  |  |  |  ||  |   |  |MOA1930> 0: medium (626.0/40.0)
|  |  |  |  |  |  |  |  |  |  |  |   |  MOEDJ > 0: serious (40.0/24.0)
|  |  |  |  |  |  |  |  |  |  |  |  MOEDP > 0
|  |  |  |  |  |  |  |  |  |  |  |  |  MORI <= 0
|  |  |  |  |  |  |  |  || | | |SceneLabel=rural:medium (34.0/21.0)
|  |  |  |  |  |  |  || | | | |SceneLabel=urban:serious (153.0/30.0)
|  |  |  |  |  |  |  |  |  |  |  |  |  MORI > 0: medium (30.0/12.0)
|  |  |  |  |  |  |  |  |  |  |  MAOHOTHE > 0: serious (102.0/15.0)
|  |  |  |  |  |  |  |  |  |  MAOHAU > 0
|  |  |  |  |  |  |  |  |  |  |  MOA1930 <= 0: serious (170.0/6.0)
|  |  |  |  |  |  |  |  |  |  |  MOA1930 > 0: medium (27.0/16.0)
|  |  |  |  |  |  |  |  MAOEOTH > 0: medium (90.0/5.0)
|  |  |  |  |  |  |  MAOEMPE > 0
|  |  |  |  |  |  |  |  MAOHS <= 0: low (343.0/16.0)
|  |  |  |  |  |  |  |  MAOHS > 0: serious (166.0/1.0)
|  |  |  |  |  |  MAOHAU > 1
|  |  |  |  |  |  |  MOA1930 <= 0: low (164.0/3.0)
|  |  |  |  |  |  |  MOA1930 > 0: serious (40.0/20.0)
|  |  |  |  FAOEPW > 0: medium (190.0/24.0)
|  |  MAOHOTHE > 1: medium (273.0/22.0)
|  MAOEMG > 1: low (173.0/5.0)
MAOEMUN > 4: low (163.0/2.0)
```

Appendix C

Unstructured questions for interview

1. How is crime prevention activities pursued in your region?
2. Do you have any automated systems to support your day-to-day activities?
3. If your answer for question number 2 is yes, what are they?
4. What are some of the current functions of these automated systems if any?
5. How do you formulate crime prevention and investigation policies including training and resource deployment (i.e. what are the basis for these)?

Declaration

The thesis is my original, has not been presented for a degree in any other university and that all sources of materials used for the thesis are duly acknowledged.

Leul Woldu Asegehgn

July 2003

The thesis has been submitted for examination with our approval as university advisors.

Ethiopia Tadesse (W/t)

Dereje Teferi (Ato)