

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

BAYESIAN NETWORK FOR MODELING DETERMINANT FACTORS
INFLUENCING OFFENDERS TO COMMIT CRIME
(THE CASE OF ADDIS ABABA POLICE COMMISSION)

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT
FOR THE DEGREE OF MASTERS OF SCIENCE IN INFORMATION SCIENCE

BY

MOHAMMED ABRAR

JANUARY, 2009

ADDIS ABABA

DEDICATION

This research work is dedicated to my mother, Birke, and to my brothers, Hayatu and Abdulfetah, for their never ending support

ACKNOWLEDGEMENT

First of all, I would like to express my profound appreciation to Dr. Rahel Bekele for her precious advices and invaluable inputs that she has given me while conducting the research.

I forward my special thanks for the Addis Ababa Police Commission and the staffs for their good will for letting me use the data for this study and provide me with information about crime and criminals in the respective region. Especially, inspector Tesfaye Mengesha who has tried to the best of his effort in providing me valuable information regarding the crimes while conducting this research.

Finally, I forward my special thanks to my colleagues: Alemayehu, Amanuel and Behailu for their limitless cooperation and priceless discussions that we have made while conducting the research.

Mohammed Abrar

ABSTRACT

The identification of causes and phenomena associated with crime is one of the most popular goals in criminology, especially in view of its practical value and the belief that such identifications are useful when seeking to correct or control criminal behavior.

The utility of discovering causes must, however, be qualified. Understanding and processing of offenders' records is one method to learn about both crime and the individuals who involve in misdeeds so that police can take crime prevention measures accordingly.

Though data on criminals are continuously being gathered, they are not effectively being utilized for extracting patterns that can be used for effective management of crimes. This is mainly due to the inadequacy of the human brain to search for complex and multifactor dependencies in data and the lack of objectiveness in such analysis demanded a computerized approach.

Developments in the information and communication technologies have made it possible for organizations to collect, store and manipulate massive amount of data. One such development is Bayesian Network.

In this study, the main objective of the research is to develop a predictive model for factors that constitute higher crime trends in Addis Ababa which makes use of Bayesian Network modeling techniques. For this purpose, published literatures in related areas have been studied together with the review of different Bayesian Network modeling approaches. Different tools and techniques supporting such task were examined by taking into consideration their application to the problem domain. In addition, an experiment is conducted to explore the potential of Bayesian

network in modeling factors that constitute higher crime trend using personal identification record of criminals.

For the purpose of the experimentation 1572 criminal records were collected from the Addis Ababa Police Commission. The records were manually and automatically further preprocessed to make them compatible with software used. Important attributes that are considered relevant for the constructing predictive model for higher crime trends were selected.

After preprocessing the data, a learning classifier is used to learn from the training data and use this classifier to classify new data. A model is constructed for the best learned model from data. Based on the experimental data, a Bayesian performance prediction model was developed where **73.25 %** prediction accuracy was first observed. Further experiments and modification of the prediction model increased the level of prediction accuracy to **75.78 %**.

Finally, Three Phase Dependency Analysis in particular and Bayesian network in general is found applicable for modeling determinant factors for higher crime trends.

TABLE OF CONTENTS

DEDICATION	I
ACKNOWLEDGEMENT.....	II
ABSTRACT.....	III
TABLE OF CONTENTS	V
LIST OF TABLES.....	VIII
LIST OF FIGURES.....	IX
ACRONYM.....	X
CHAPTER ONE	
INTRODUCTION	1
1.1 Background.....	1
1.2 Statement of the Problem.....	2
1.3 Objective.....	7
1.4 Methodology.....	7
1.4.1 Attribute Selection.....	7
1.4.2 Data Collection/Creating Dataset	8
1.4.3 Data Preparation	8
1.4.4 Model Building/ Data Analysis	9
1.5 Scope and Limitation of the Study	9
1.6 Thesis Organization	10
CHAPTER TWO	
CRIME AND CRIMINALS.....	11
2.1 Introduction.....	11

2.2	Crime and Criminals.....	11
2.3	Importance of Crime Cause Identification.....	12
2.4	Trends of Crime in Ethiopia	15

CHAPTER THREE

LITERATURE REVIEW16

3.1	Introduction.....	16
3.2	Basics of Bayesian Network.....	16
3.2.1	Prior and conditional probability	16
3.2.2	Bayes’ Theorem.....	17
3.3	Bayesian Networks	18
3.3.1	D-separation.....	25
3.3.2	Learning in Bayesian Network	28
3.3.2.1	Parameter Learning.....	29
3.3.2.2	Structure Learning	30
3.3.2.2.1	Model building only from data (TPDA-I)	33
3.3.2.2.2	Structure Learning for Given Node Ordering (TPDA-II).....	34
3.4	Related Research Areas	35

CHAPTER FOUR

EXPERIMENTATION39

4.1	Introduction.....	39
4.2	Data Collection	39
4.3	Selection of Attributes	41
4.4	Data Preprocessing	43

4.5	Building The Bayesian Network Model	46
4.6	Testing The Network	47
4.7	Constructing the Bayesian Network Model.....	49
4.7.1	Constructing the Bayesian Network Model when node ordering is not given	49
4.7.2	Constructing the Bayesian Network Model for a given node ordering (TPDA-II)	51
4.8	Evaluation of the Model	57
 CHAPTER FIVE		
CONCLUSION AND RECOMMENDATION		58
5.1	Conclusion	58
5.2	Recommendation	60
REFERENCES.....		61
APPENDICES		66
DECLARATION		74

LIST OF TABLES

Table 4. 1: Attributes which exist in the personal identification record of the criminals but not used.....	41
Table 4. 2: Attributes that are selected for the purpose of the Research.....	45
Table 4. 3: The prediction performance for the best learned model from data	48
Table 4. 4: The Prediction accuracies for each of the Datasets used	49
Table 4. 5: Prediction accuracies for all partitioned datasets for TPDA-II.....	56
Table 4. 6: A costly misclassification for both classifiers.....	58

LIST OF FIGURES

Figure 3. 1:Conditional Independence Graph.....	21
Figure 3. 2: Conditional probability table using Bayesian Network in Java tool for nodes educational status, age and commitment of crime in group.	22
Figure 3. 3: Patterns for Paths through a Node.....	26
Figure 4.1:Processes for Building the Model	40
Figure 4. 2: Sample Training Dataset in Microsoft Access format after Preprocessing	46
Figure 4. 3: A report generated by BN PowerPredictor for both actual and predicted class labels with their assigned probabilities.	47
Figure 4. 4: A report that is generated when the learning process from the data is finished.....	50
Figure 4. 5:The best learnt Bayesian network model from data using TPDA-I	51
Figure 4. 6: A model after incorporating knowledge of the domain experts.....	53

ACRONYM

AIC Australian Institute of Criminology

BN Bayesian network

BNJ Bayesian Network in Java

CI Conditional Independence

CPT Conditional Probability Table

D-separation Direction Separation

DAG Direct Acyclic Graph

EM Expectation Maximization

TPDA Three Phase Dependency Analysis

WHO World Health Organization

CHAPTER ONE

INTRODUCTION

1.1 Background

Crime is a human experience as old as human race. It occurs at specific place, specific time and for specific reason. It can affect everyone and anyone at any time. Economies of countries and individuals world wide are affected by crime. Gladstone (1994) described the impacts of the crime on the economies of the country are incalculable. It places a massive burden on national economies, costing countries enormous amount of money each year in health care, law enforcement and lost productivity (WHO, 2007).

As of the report of Australian Institute of Criminology (2000), what is even worse is that the impacts of the crime do not only take into account primary or direct victims only. But this description also include people who were witnesses to the crime, family members, friends, neighbors and whole communities—secondary or indirect victims—who may suffer trauma as a result of crime (VCCAV 1994). It is difficult to provide an estimate of the extent to which crime affects individuals, families, friends and their relationships. In addition to this, they suggest that the impact of crime victimization can be long lasting and diverse. The consequences of crime can in some cases include physical injuries or death; many involve financial loss or property damage; and, less obvious but sometimes more devastating, psychological and emotional wounds.

The identification of causes and phenomena associated with crime is one of the most popular goals in criminology, especially in view of its practical value and the belief that such identifications are useful when seeking to correct or control criminal behavior. The utility of

discovering causes must, however, be qualified. Knowing what causes crime is not necessarily the same as reducing or addressing crime. Causation refers to factors or phenomena that have to be present (necessary or sufficient conditions) to precipitate crime. It is argued, however, that causal relationships are images created out of past experiences rather than observed reality. "No force can be directly observed coming from poverty, economic inequality, or deviant friends that drives a person to commit a crime. Conversely, no force can be observed coming from a well-developed superego, a stake in conformity, or a strong belief in the rules in society that restrains a person from committing a crime" (DiCristina, 1995).

Generally, groups of factors have been universally regarded as contributing to crime. The factors can be categorized as societal (income inequality, poverty, cultural norms, gender inequality) community based (poverty, unemployment, drug trade), relationship based (poor parenting, marital discord, peer pressure, parental conflict) and individual based (alcohol abuse, psychological disorder, crime history, victim of child maltreatment) (WHO, 2007).

However, more emphasis is still placed on the social context in which the person functions, taking limited note of the individual, characterized by his or her unique personality composition.

1.2 Statement of the Problem

In the period between 5000 BC and 1692 AD the high crime rate would have been blamed on demonic influences or during the period between 3500 BC and 1630 AD, on the zodiac or planetary influences (Letwin, 2002). Today, the causes of crime should be sought far beyond these factors. Although people have different perceptions of the reasons underpinning crime, few studies have attempted to explore the magnitude or the causes of crime (Glanz, 1996).

According to the UN-HABITAT (2007) report, the impacts of crime and violence are multidimensional. Apart from injury and death, victims of crime suffer long-lasting psychological trauma and subsequently live with the fear of crime. At the national level, crime and violence are impediments to foreign investment, contributing to capital flight, brain drain and hinder international tourism.

Fajnzyblber et al. (1999) stated that the concern with crime is well justified given its pernicious effects on economic activity and, more generally, on the quality of life of people who must cope with the reduced sense of personal and proprietary security. Despite the fact that violent crime is emerging as a priority in policy agendas worldwide, little is known regarding the economic, social, and institutional factors that make some countries have higher crime rates than others or make a country experience a change in its crime rate

Different stakeholders (including the domain experts) have different assumptions and estimations concerning the reasons behind the current causes for higher crime trends. The differences in assumptions might have come from considering the very diverse background of the criminals; therefore it is difficult to generalize the root causes unless a sound scientific procedure is adopted. As such, no strong solutions have been devised so far to address the problem.

According to Thakur (2003), the causes for the growing rate of crime include unemployment, economic backwardness, over population, illiteracy and inadequate equipment of police force.

Fleisher (1966) and Ehrlich (1973) examined the effect of unemployment rates, income levels, and income disparities on the incidence of crime. Though their findings on the effects of average income levels are contradictory, both authors find a significant crime-inducing impact of unemployment and income inequality.

Freeman (1992) found that youth in poverty are more likely to be arrested and go to jail than others. Tauchen and Witte (1994) found that in a sample of young men, going to work or school tends to reduce the probability of being involved in criminal activities. On the other hand, the effect of education on crime reduction is controversial in most studies. For example, Ehrlich (1975) found a positive relationship between the average number of school years completed by the adult population and property crimes.

According to the UN-HABITAT (2007) report, several factors influence the incidence of crime and violence. These include economic and political circumstances that produce opportunities and incentives for criminal behaviors and violent acts, as well as the situations that frame victimization. Their report also associated other factors with urban crime and violence include poverty, unemployment, income inequality, intergenerational transmission of violence as reflected in the continuous witnessing of parental abuse during childhood, the rapid pace of urbanization, poor urban planning design and management and growth in youthful population.

CS&CPC (2007) reported that crime is primarily the outcome of multiple adverse social, economic, cultural and family conditions. To prevent crime it is important to have an understanding of its roots. They also stated that these roots are complex and interrelated. They categorized crime as the consequence of economic factors, social environment, and family structures.

According to Don (2001), the debate about what causes crime arises because of failure to attend to the distinction between proximate and distal causes of crime. One proximate cause of involvement in crime is association with youthful offenders and the other distal causes crime is poor parenting.

To handle crime there is always a need for prudent crime prevention strategies and policies. To create effective prevention strategies and policies, law enforcement organizations like police needs to learn factors that constitute higher crime trend (Wilson 1963). Understanding and processing of offender records is one method to learn about both crime and the individuals who involve in misdeeds so that police can take crime prevention measures accordingly (Brown 2003). This could be a lot helpful in crime analysis to identify crime patterns and series, forecast future occurrences of crime, identify likely victims of crime, provide investigative leads, solve open cases, and provide supporting data for community policing programs and departmental planning efforts (Susan, 2004).

Though Data on criminals are continuously being gathered, due to lack of appropriate tool they are not effectively being utilized for extracting patterns that can be used for the effective management of crimes. Criminals' records analysis made by using traditional methods focuses on problems with only manageable number of variables and cases than may be encountered in real world. This is mainly because as the number of cases and variables increases, efficient analysis of the records becomes beyond the human capacity to discover new and unanticipated patterns and relationships that are hidden in conventional databases (Plate et al., 1997).

The problem of crime has harassed society down through the ages. The struggle between law observance and anti-social behavior is as old as man. To comprehend the question of crime and its dangerous influence upon mankind, it is necessary to understand man and the various factors which motivate his activity.

According to the Don (2001), criminal behavior cannot be explained by any one element. The make-up of man is so complicated, the changes of environment so intricate, that it is impossible

to focus the spotlight of understanding on any one factor as the cause of crime. The developments of nation, the economic, social and political factors which go hand in hand with its advancement, are valuable considerations in a study of crime causation. The value of research work into mental and physical characteristics of wrong-doers has also added much to the study of crime.

Much criminological research involves trying to determine whether a particular factor increases the risk of involvement in crime when other possible risk factors are controlled (i.e. held constant). Of course, the discovery of a statistical association between some factor and crime never provides any guarantee that the factor in question causes crime, even when attempts have been made to control for other relevant factors. Identifying the causes of crime is never easy or certain (Don 2001).

Heckerman, (1996) explained that Bayesian network has the ability to reason under complex and uncertain situations. It is preferred than conventional methods based on exact inference because it produces models that are more accurate for prediction. They also encode the correlation between the input variables

This research focuses on developing a model to investigate how social, economical, and case related factors affect the extent of crime.

Based on the foregoing discussions, the following major questions guided the research work:

- What are the major factors that affect the criminals to commit the crime?
- How are these attributes dependent on each other?
- How can one build the Bayesian network based on these attributes?
- How accurate this would be?

1.3 Objective

The general objective of the research is to investigate the potential applicability of Bayesian belief network in developing a model that can support police officers, policy makers, and planners to identify the major determinants of crimes and interrelationship among the main factors in the effort of preventing and controlling crime.

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

- ▶ To identify the determinant factors that determines the seriousness of the crime.
- ▶ To model the interdependencies among these factors using the Bayesian network.
- ▶ To experiment with Bayesian network as uncertainty management technique.
- ▶ To see the applicability of the Bayesian network in pattern extraction on the personal identification record of the criminals
- ▶ Build and test models using the selected tools and technologies.

1.4 Methodology

This section provides a brief introduction to the approaches carried out in order to address the above research questions.

1.4.1 Attribute Selection

In order to reduce the complexity of the work being done in terms of identifying the most relevant crime label determinant factors, it is important to identify the common attributes to address the issue. Review of literatures, formal and informal interviews were conducted with

police officers to select important attributes that are considered relevant for the purpose of the research.

1.4.2 Data Collection

The major task that was performed at this stage was to identify and collect the relevant dataset for the purpose of the research work at hand. Since the specific steps and tasks at this stage determine the final result of the research work, much effort was made in creating the right dataset that is representative of the population by collecting data from all sorts of crime proportionally using random sampling method.

The data that that was used for the purpose of this research was collected from the Addis Ababa Police Commission. The personal identification records of offenders which contain offenders' education level, employment status, the kind of occupation they involve in, their age and other issues related to that specific crime is used for constructing the model. The number of records that are used for this research is 1572 records.

1.4.3 Data Preparation

This was a step where the collected data was arranged into a form that is suitable for the particular Bayesian-belief network tool. In other words it was just to prepare the data for analysis purpose so that it was compatible with requirements of the belief network software tool. At this stage pre-processing tasks like handling noisy data, unknown values, missing values and summarization of data was performed by taking into account the selected tool and techniques. Data summarization, for instance, included changing the values of some attributes to categorical labels. This has enabled to effectively apply Bayesian belief network.

1.4.4 Model Building

In order to build the intended model, the Bayesian network was used. Its different tools supporting such task were examined by taking into consideration their applicability to the problem domain. Among the publicly available modeling techniques, the BN PowerSoft Package was found to perform better than other softwares with respect to its efficiency and being able to incorporate domain knowledge. As a result it was used for both assessing the prediction performance of the classifiers as well as for building the model. For evaluating the prediction performance of the classifier the BN PowerPredictor was used. For constructing the model, the BN PowerConstructor was used.

After building the model the next task that was performed were evaluating or assessing and interpreting the results of the selected model. Experts in the area of crime were also consulted during the model building process. Based on the assessment and evaluation of such models the best model that had produced the best prediction accuracy was selected. And this has provided a good ground for result interpretation and recommendation.

1.5 Scope and Limitation of the Study

The scope of the research is limited to assessing the potential applicability of belief network technologies in supporting the crime prevention activities in Addis Ababa region. More specifically the research is delimited to modeling the determinant factors in committing crimes only for personal identification records of offenders which are more important than case records or arrest records. The major limitations while undertaking this research was the absence of full details about a given crime for some records. This had a constraint on the dimensionality and amount of data to be collected apart from creating workload on data preprocessing task.

In addition to this, due to sensitivity of the data the commission made a complete inspection so that the information being gathered does not dispose the personal identity of the criminals and lose their credibility. They were afraid that the data might be used for other purpose other than for academic purpose. This had limited the number of data that we collect, the kind of information they provide and the extent of time spent in collecting the required information.

The causes of any given crime attribute to the interaction of several psychological, social, neurophysiologic, and genetic variables. Unfortunately, many of these important attributes which are considered relevant by many scholars were not incorporated due to the unavailability of these attributes on the personal identification record of the criminals. Though the record keepings in Addis Ababa Police Commission are starting to be automated, many of the records are still manual which has created huge workload due to intensive preprocessing requirements so as to make data compatible with software package used.

1.6 Thesis Organization

This thesis is organized into five chapters. The first chapter briefly introduces the background of the area, the objectives, the methodology and the scope of the research. The second chapter deals with criminology and related literatures in the area of crime, causes of crime and other issues related in one way or another with the research area under investigation.

Chapter three introduces the basics of Bayesian network followed by a detailed description of learning in Bayesian networks. This chapter also included the applications of the Bayesian networks in the field of crime. Chapter four deals with experimental works related to development of predictive model for higher crime trend. Chapter five ends in giving concluding remarks as well as directions for future work.

CHAPTER TWO

CRIMINOLOGY

2.1 Introduction

The main aim of this research is to investigate the application of Bayesian network technique for modeling risk factors that influence offenders to commit crime using the personal identification record obtained from Addis Ababa Police Commission. To this end the chapter attempts to review crime, criminals and causes of crime as background for subsequent chapters discussed. Furthermore, the trend of crime in Ethiopia is also described.

2.2 Crime and Criminals

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

According to Andargachew (1988), a criminal is an individual 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 competent age in the light with 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 crime when it is prohibited by the criminals law. Criminal law, on the other hand, refers to a body of specific rules regarding human conduct, which have been explicitly stated by the 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 the sophisticated equipment in one place and then escape to another place (Thakur, 2003).

2.3 Importance of Crime Cause Identification

Just as there are many different types and kinds of violence, there are many different criminological explanations of violence. Siegel (2004) documents eight different explanations (personal traits, ineffective families, substance abuse, human instincts, regional values, cultural values, gangs and firearm availability). He further stated that it is customary to distinguish between instrumental (an attempt to improve one's financial or social position) and expressive (an attempt to vent rage, anger, or frustration) violence. It may be that humans are hard-wired to admire, mimic, or be violent, but biologists are continually telling us that there is no inborn tendency for violence, and anthropologists also tell us that history is full of stories of peace-loving, matrilineal tribes. However, psychological and sociological perspectives have the most to say about the causes of criminal violence as it exists today.

The term criminology derives from the Latin word 'crimen', translated as "offence", and is introduced by the anthropologist Paul Topinard in 1889 (Maquire, 1999). Criminologists use a variety of scientific methods to investigate criminal behavior and related topics. However, because crime penetrates almost all areas of social life, the scope of criminological studies is virtually unlimited. Therefore, as a discipline, criminology is defined as the scientific study of crime and (notions) of crime control (Nelken, 1994).

According to DiCristina (1995), the identification of causes and phenomena associated with crime is one of the most popular goals in criminology, especially in view of its practical value and the belief that such identifications are useful when seeking to correct or control criminal behavior. She further elaborated that the utility of discovering causes must, however, be qualified. Knowing what causes crime is not necessarily the same as reducing or addressing crime. Causation refers to factors or phenomena that have to be present (necessary or sufficient conditions) to precipitate crime. It is argued, however, that causal relationships are images created out of past experiences rather than observed reality. She further stated "No force can be directly observed coming from poverty, economic inequality, or deviant friends that drives a person to commit a crime. Conversely, no force can be observed coming from a well-developed superego, a stake in conformity, or a strong belief in the rules in society that restrains a person from committing a crime"

Svensson (2002) used the term "risk factors" when discussing factors that indicate an increased risk for criminal behavior and not causes of crime. DiCristina (1995) on the other hand, is of the opinion that it might be more appropriate to focus on probabilities of the association of certain phenomena, enabling to depict the cause and effect relationship among the phenomena.

According to CS&CPC (1996) statement individuals need to be responsible for their own actions. An understanding of root causes cannot and should not be seen as a way to absolve us from personal accountability. However, while individuals have an obligation to act responsibly and with respect for their fellow citizens, communities have a responsibility to address those conditions, which hinder healthy development and can become the breeding ground for crime.

In general the different role players need to consider the following aspects in order to come to a proper understanding of the crime situation and to develop strategies to address crime holistically.

To handle crime it demands the need for a multidimensional perspective and interactive approach. "Criminals" should not be regarded as being isolated from the rest of society, since criminal acts require complicity from others. This implies that all relevant factors should be taken into account, such as the offender's history and personality; the offender's mental state and physical condition at the time of the criminal event; the nature of the offence; the offender's reason for committing the crime; relation to and interaction with the victim; the harm suffered by the victim and society; as well as precipitating and situational factors contributing to the criminal event. The person will then be recognized as a unique human being by taking into account his/her biological and psychological make-up as well as the social context in which he/she functions. (Naude and Prinsloo, 1999)

In the process of creating an all inclusive solution the knowledge of the expertise of the criminologists can also be used. As important role-players, they are able to identify causes of crime by conducting offender-based research and offender or victim assessment, and can thus contribute to a holistic approach to the prevention of crime. (Naude and Prinsloo, 1999)

In the process of developing cost effective crime prevention programs and strategies, the government as well as the community must be involved in order to draft and impose an all-inclusive crime prevention policy. The approach must be multi-dimensional in order to include various crime prevention models and their programs. Preventive models and their programs must be based on the root causes of crime, factors enabling crime (crime opportunities) and specific

crimes. The uniqueness of crime patterns and causes in a specific area must be taken into consideration when the crime prevention plan is prepared. (DiCristina, 1995)

2.4 Trends of Crime in Ethiopia

In Ethiopia, the police organization was established in 1942 under the proclamation No. 6 as an autonomous institution with the responsibility of preventing and investigating crime incidents. In 1966 the police institution was put under the then Interior Minister. Since its establishment, the police organization structure extended to the lower administration level which includes “woreda” and sometime “kebele” level. (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 is changed accordingly. According to proclamation No.1 article 50, regional governments are duly responsible to establish all the necessary administrative levels in their respective region. As a result, Addis Ababa Police Commission has become responsible for maintaining 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).

In Ethiopia, crime statistics (of the Addis Ababa Police Commission) has shown that the rate of crime is increasing steadily. A sample survey conducted in the year 2000 by the research team of the Addis Ababa Police Commission has shown that in 2000 more than 51,000 crimes have been reported to the police (Addis Ababa Police Commission 2000).

CHAPTER THREE

LITERATURE REVIEW

3.1 Introduction

In this chapter different Bayesian network modeling techniques and methods are discussed in relation to graphical probabilistic model learning from data. Important issues related to learning classifiers from data and model building are also discussed in line with the Bayesian network.

3.2 Basics of Bayesian Network

3.2.1 Prior and conditional probability

The unconditional or prior probability associated with a proposition A is the degree of belief accorded to it in the absence of any other information. It is written as $P(A)$. Prior probabilities are used when there is no other information. For example, the probability that a typical crime being committed by a middle aged person, denoted by $P(\text{Age}=\text{Middle age})=0.5$, is the degree of belief accorded to a crime being committed by a middle aged person in the absence of any other detail. As soon as some new information is known, we must reason with the conditional probability of A given this new information. For example, the probability of the crime being committed by middle aged person given that the crime label rate is very serious, $P(\text{Age}=\text{Middle age} | \text{CrimelabRate} = \text{Very Serious})$.

Accordingly, the probability of some event A , given the occurrence of some other event B written $P(A|B)$ [read "the probability of A , given B "], where A and B are any propositions, is used to represent the conditional probability that A will occur given that B has already occurred. For instance, we might ask "what is the probability that a medium age people to involve in the serious crime?" This might be symbolized as $P(\text{Age} = \text{Middle age} | \text{CrimelabRate} = \text{serious})$.

Conditional probabilities can be defined in terms of unconditional probabilities. The defining equation is:

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \quad \text{Which holds whenever } P(B) > 0; \quad (3.1)$$

$$P(B/A) = \frac{P(B \cap A)}{P(A)} \quad \text{Which holds whenever } P(A) > 0; ; \quad (3.2)$$

From equation 3.1 and 3.2, we might derive the product rule

$$P(A \cap B) = P(A|B) * P(B) = P(B|A) * P(A) \quad (3.3)$$

Rearranging the product rule in equation 3.2 and 3.3 leads to the famous Bayes theorem:

3.2.2 Bayes' Theorem

Bayes' theorem is a theorem of conditional probability that allows estimates of probability to be revised continually on the basis of observations of occurrences of events.

Bayesian networks are network of relationships. And they are named "Bayes" after Reverend Thomas Bayes, 1702-1761, a British theologian and mathematician who wrote down a basic law of probability which is now called Bayes Rule:

$$\text{For any two events, A and B, } P(A/B) = \frac{P(B/A) * P(A)}{P(B)} \quad (3.4)$$

A generalized approach for two events.

$$P(A_i|B) = \frac{P(B|A_i) P(A_i)}{\sum_j P(B|A_j) P(A_j)}, \quad (3.5)$$

Theorems analogous to Bayes' theorem hold in problems with more than two variables. For example:

$$P(A|B \cap C) = \frac{P(A \cap B \cap C)}{P(B \cap C)} = \frac{P(C|A \cap B) P(A \cap B)}{P(B) P(C|B)} = \frac{P(A) P(B|A) P(C|A \cap B)}{P(B) P(C|B)}. \quad (3.6)$$

This can be derived in a few steps from Bayes' theorem and the definition of conditional probability.

Similarly,

$$P(A|B \cap C) = \frac{P(B|A \cap C) P(A|C)}{P(B|C)}, \quad (3.7)$$

For example, the

$$P(\text{Age} = \text{Medium} | \text{CrimelabRate} = \text{Serious}) = \frac{P(\text{CrimelabRate} = \text{Serious} | \text{Age} = \text{Medium}) * P(\text{Age} = \text{Medium})}{P(\text{CrimelabRate} = \text{Serious})}$$

Bayes' theorem is frequently used for reasoning about an uncertain hypothesis A given evidence B, and in that context P(A|B) is called the posterior (conditional) probability of A, and P(A) is called the prior (unconditional or marginal) probability of A. The table defining conditional probabilities for every possible combination of values that A and B can take is called a conditional probability table (CPT).

3.3 Bayesian Networks

Bayesian networks were introduced in the 1980's as formalism for representing and reasoning with models of problems involving uncertainty, adopting probability theory as a basic framework (Lucas, 1999). Over the last decade, the Bayesian network has become a popular representation

for encoding uncertain expert knowledge in expert systems (Heckerman, 1995). The field of Bayesian networks has grown enormously over the last few years, with theoretical and computational developments in many areas. Bayesian networks are also known as belief networks, causal probabilistic networks, causal nets, graphical probability networks, and probabilistic influence diagrams (Rahel, 2005).

A Bayesian network is used to model a domain containing uncertainty in some manner (Haulya and Erba, 2005). It is a graphical model for probabilistic relationships among a set of variables and is composed of directed acyclic graphs (DAGs) in which the nodes represent the random variables of interest, and the links represent informational or causal dependencies among the variables (Pearl, 1997). Here, each node contains the states of the random variable and it represents a conditional probability table. The CPT of a node contains probabilities of the node being in a specific state given the states of its parents (Jensen et al., 1996). Furthermore, edges reflect cause-effect relations within the domain. These effects are normally not completely deterministic. The strength of an effect is modeled as a probability.

Bayesian networks help us answer questions such as: What is the probability that a random variable will be in a given state if we have observed the values of some other random variables. They can also suggest what could be the best choice for acquiring new evidence since the route from the class label or any node to its successor or predecessor node is known. Conditional probabilities are important for building Bayesian networks. But Bayesian networks are also built to facilitate the calculation of conditional probabilities, namely the conditional probabilities for variables of interest given the data (also called evidence) at hand. (Cowell, 1999)

Pearl (1988) represented a Bayesian network by $BN = \langle N, A, \theta \rangle$, where $\langle N, A \rangle$ is a directed acyclic graph (DAG) each node $n \in N$ represents a domain variable and each arc $a \in A$ between nodes represents a probabilistic dependency between the associated nodes. Associated with each node $n_i \in N$ is a conditional probability table (CPT), collectively represented by $\theta = \{\theta_i\}$, which quantifies how much a node depends on its parents (Pearl, 1988).

Russel and Norvig (2003) stated that the structure of a Bayesian network is a graphical illustration of the interactions among the set of variables that it models. They explained the full specification as follows:

- (i) A set of random variables makes up the nodes of the network and which are also selected either by the domain experts or by reviewing different literatures relating to the problem domain;
- (ii) A set of directed links connects pairs of nodes. If there is a directed link from node A to node B, A is said to be a parent of B;
- (iii) Each set contains a finite set of mutually exclusive states;
- (iv) Each node A has a conditional probability table $P(A | \text{par}(A))$ that quantifies the effect of the parents on the node. If the variable A does not have any parent, then the table can be replaced by prior probabilities, i.e. $P(A)$;
- (v) The variables coupled with the directed edges construct a directed acyclic graph (DAG). DAG refers to a graph where all of the edges in the graph are directed and there are no cycles. The key advantage of not allowing cycles is that it makes possible very fast update algorithms, since there is no way for probabilistic influence to "cycle around" indefinitely

Bayesian network largely reduces the full joint probability table to probability table influenced by at most K others for some constant K (Russell and Norvig, 2003).

In a Bayesian network every node is associated with a table of conditional probabilities of the vertex given the state of its parents. We denote the conditional probability table using the notation $P(x_i|\text{par}(x_i))$, where lower case x_i denotes values of the corresponding random variable X_i and $\text{par}(x_i)$ denotes a state of the parents of X_i .

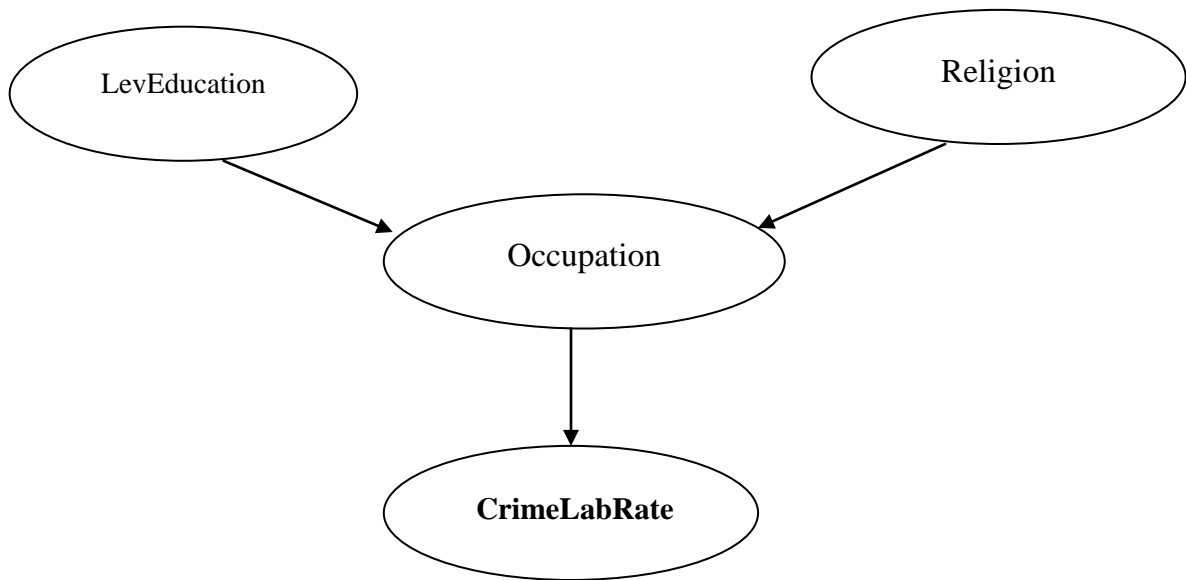


Figure 3. 1: Conditional Independence Graph

Figure 3.2 depicts a Bayesian network consisting of four discrete variables: **LevEducation**, **Religion**, **Occupation** and **CrimeLabRate**. The dependencies can be expressed in terms of conditional probability distribution for each variable in the Bayesian network. Each of the variables in the network has possible values they could take. These values are mutually exclusive. In the above Figure 3.1, the node ‘LevEducation’ has four states: ‘Illiterate’, ‘Elementary and Junior’, ‘Secondary’ and ‘College and Above’; ‘Religion’ has four states: ‘Orthodox’, ‘Muslim’, ‘Protestant’ and ‘Other’; ‘Occupation’ has four states: ‘Bussiness Man’, ‘Employed’, ‘Daily

worker' and 'Unemployed'; 'CrimeLabRate' has three states: 'Very Serious', 'Serious' and 'Medium'.

Using the probabilistic chain rule, the joint probability distribution can be written in the product form. For a node with no parents it needs only to compute the table for CPT containing its states. For example in figure 3.1, the nodes 'Religion' and 'LevEducation' contain CPT table containing only their own states. But for nodes which have parents they need to include their parent's state in their CPT table including theirs in the CPT table. For instance, the number of entries for the CPT of 'Occupation' is the products of its possible states multiplied by the number of entries of each of its parents. The numbers of entries for the 'Occupation' are 64 (4x4x4).

For each variable, we need to specify a table of conditional probability distributions, one for each configuration of states of its parents. $P(\text{Religion})$, $P(\text{LevEducation})$, $P(\text{Occupation} \mid \text{Religion}, \text{LevEducation})$ and $P(\text{CrimeLabRate} \mid \text{Occupation})$ are required to build the CPT tables for the Figure 3.1.

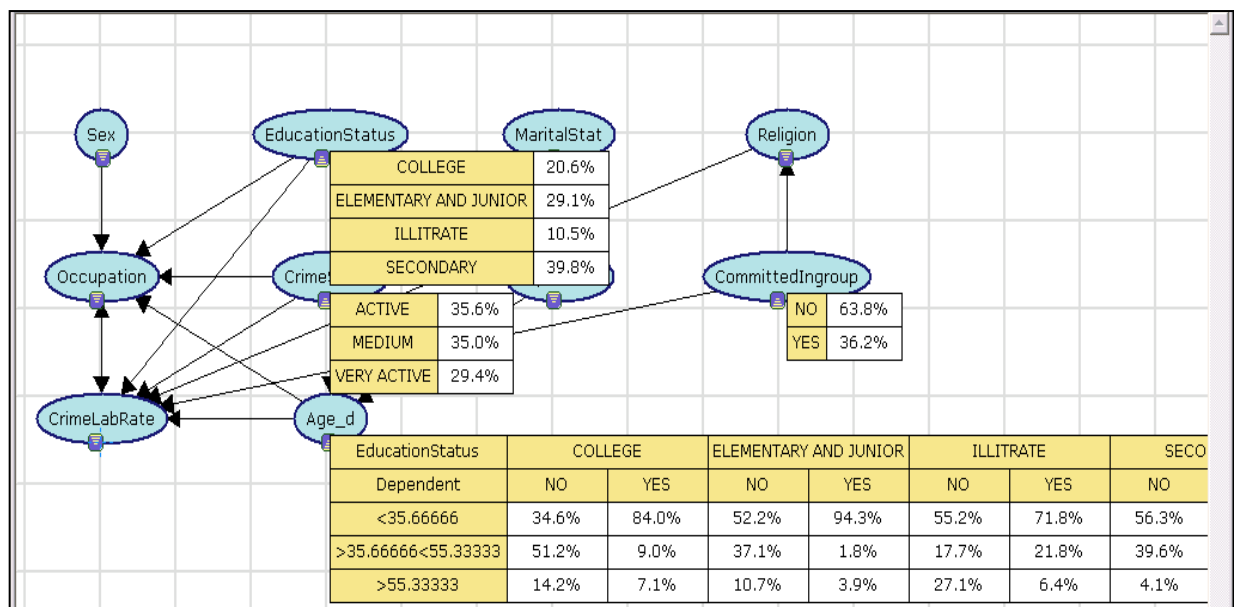


Figure 3. 3: Conditional probability table using Bayesian Network in Java tool for nodes educational status, age, crime scene and commitment of crime in group.

When we multiply the conditional for each variable, we get the joint probability which is defined by the product rule (chain rule) as follows:

$$\begin{aligned}
 P(a_1, a_2, \dots, a_n) &= P(a_1 | a_2, \dots, a_n) P(a_2, \dots, a_n) \\
 &= P(a_1 | a_2, \dots, a_n) P(a_2 | a_3, \dots, a_n) P(a_3, \dots, a_n) \\
 &= P(a_1 | a_2, \dots, a_n) P(a_2 | a_3, \dots, a_n) P(a_3 | a_4, \dots, a_n) P(a_4, \dots, a_n) \\
 &= P(a_1 | a_2, \dots, a_n) P(a_2 | a_3, \dots, a_n) \dots P(a_{n-1} | a_n) P(a_n) \\
 &= \prod_{i=1}^{i=n} P(a_i | \text{par}(a_i)) \dots \dots \dots (3.8)
 \end{aligned}$$

This property of joint probability distributions is called the general factorization property. Note that this product rule allows for any ordering of variables in the factorization.

For example,

$$P(\text{Religion}, \text{LevEducation}, \text{Occupation}) = P(\text{Religion}) * P(\text{LevEducation}) * P(\text{Occupation} | \text{Religion}, \text{LevEducation}) * P(\text{CrimeLabRate} | \text{Occupation})$$

$$P(\text{Religion}, \text{LevEducation}, \text{Occupation}) = P(\text{Religion} | \text{LevEducation}) P(\text{LevEducation}) * * P(\text{Occupation} | \text{Religion}, \text{LevEducation}) * P(\text{CrimeLabRate} | \text{Religion}, \text{LevEducation}, \text{Occupation})$$

The above two equations for the given attributes are equal using the concepts of independence and conditional independence. Dependence between two events occurs when the probability of an event depends on the knowledge of the other event.

In the above figure, figure 3.1, knowledge of the ‘Religion’ is not dependent on the value of ‘LevEducation’. As a result we denote their relationship as $P(\textit{Religion} \mid \textit{LevEducation}) = P(\textit{Religion})$, i.e., Religion is independent of the educational status of the criminal.

In a similar manner events are conditionally independent, if for every pair (a, c), $P(A = a \mid B, C = c)$ remains constant as B varies. We, therefore, say that A is conditionally independent of B given C. We can therefore drop B from the conditional probability $P(A \mid B, C)$ altogether and rewrite the representation as:

$$P(A \mid B, C) = P(A \mid C) \tag{3.9}$$

Using this equation we get $P(\textit{CrimeLabRate} \mid \textit{LevEducation}, \textit{Religion}, \textit{Occupation}) = P(\textit{CrimeLabRate} \mid \textit{Occupation})$, i.e., *CrimeLabRate* is conditionally independent of *Religion* and *LevEducation* given *Occupation*. Independence and conditional independence can be illustrated from the Bayesian network. Considering predecessors (LevEducation, Religion) and their successor at the bottom (CrimeLabRate, Occupation). Given their parents, each node or variable is conditional independent of its predecessor if there does not exist a directed arc from the predecessors to successors. Since there does not exist cycles in the network such relationship exists among variables in the network. Figure 3.1 can be used to illustrate these relationships among variables. For example, since there does not exist an arc from *Religion* to *LevEducation*, *Religion* and *LevEducation* are independent of each other. Similarly the lack of an arc from the *Religion* to *CrimeLabRate* and from *LevEducation* to *CrimeLabRate* can be an evidence for that the rate of crime label is conditionally independent of the religion and the educational status of the criminal given their successor occupation of the criminal.

As we have already discussed above, the main advantage of Bayesian networks is the ability to define the conditional independencies first, before specifying numerically the actual conditional probability distributions. A general conditional independence property of Bayesian networks is that any variable X in the network is conditionally independent of its non-descendants $\mathbf{ND}(X)$ given its parents $\mathbf{par}(X)$ (Pearl, 1988). That is, if a variable's parents become known, then any information about nodes that are not on a directed path from X will be irrelevant. This is the so-called directed Markov property of Bayesian networks.

3.3.1 D-separation

Bayesian networks encode the dependencies and independencies between variables. Under the causal Markov assumption, each variable in a Bayesian network is independent of its ancestors given the values of its parents. With the causal Markov assumption, we can check some conditional independence in Bayesian networks. For the general conditional independence in a Bayesian network, Pearl (1979) proposed a concept – d-separation (direction separation) for the purpose. D-separation is a graphical property of Bayesian networks and has the following implication: If two sets of nodes X and Y are d-separated in Bayesian networks by a third set Z (excluding X and Y), the corresponding variable sets X and Y are independent given the variables in Z . The definition of d-separation is as follows: two sets of nodes X and Y are d-separated in Bayesian networks by a third set Z (excluding X and Y) if and only if every path between X and Y is “blocked”, where the term “blocked” means that there is an intermediate variable V (distinct from X and Y) such that:

- The connection through V is “tail-to-tail” or “tail-to-head” and V is instantiated

- Or, the connection through V is “head-to-head” and neither V nor any of V’s descendants have received evidence (blocked).

The graph patterns of “tail-to-tail”, “tail-to-head” and “head-to-head” are shown in the figure below.

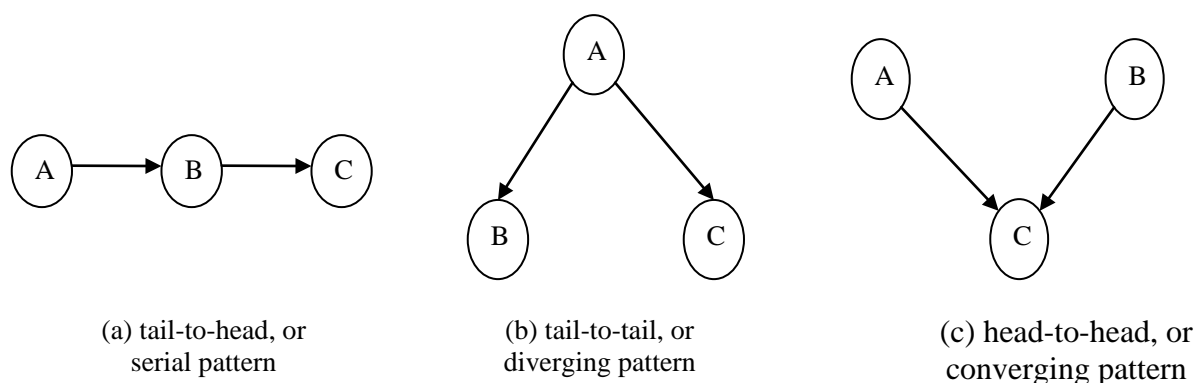


Figure 3. 4: Patterns for Paths through a Node

In **serial (tail-to-head connections)**, figure 3.3, A has control over B which then has control over C. In same way, in figure 3.1, *LevEducation* has causal effect on *Occupation* which in turn causes *CrimeLabRate*. Apparently, the evidence on the *LevEducation* will affect the certainty of the variable *Occupation* that in turn affects the certainty of variable *CrimeLabRate*. Analogously, the evidence on the *CrimeLabRate* will affect the certainty of the *LevEducation* through the variable *Occupation*. On the contrary, if the state of *Occupation* is given, then the link is blocked, and variables *LevEducation* and *CrimeLabRate* become independent. Influence can or can not pass from *LevEducation* to *CrimeLabRate* and vice versa unless *Occupation* is instantiated. In other words, a path between nodes *LevEducation* and *CrimeLabRate* is closed, given some evidence

Occupation, if *LevEducation* and *CrimeLabRate* are conditionally independent given *Occupation*. *LevEducation* and *CrimeLabRate* are then said to be D-separated (Direction separated).

In a **diverging (tail-to-tail) connection**, the influence can pass between all the children of the variable *A* unless the state of the variable *A* is given. If the state of the variable *A* is known, then the variables *B* and *C* become independent from each other. Therefore, influence may run between *A*'s children unless *A* is instantiated.

In **converging (head-to-head) connection**, if there is nothing known about the variable *C* other than what may be deduced from the knowledge of its parents *A* and *B*, then the parents are said to be independent. The independence implies that evidence on one of the parents has no effect on the certainty of the others. For example, in figures 3.1, the evidence on either *LevEducation* or *Religion* has no effect on the certainty of *Occupation* and *CrimeLabRate*. If there is any other kind of evidence influencing the variable *LevEducation*, then the parents become dependent. Therefore, evidence may only be transmitted through a converging connection if either the *LevEducation* or *Religion* or one of its descendants has received evidence. The evidence can be direct evidence on the variable *CrimeLabRate*, or it can be evidence from one of its children.

The three connections explained above wrap all the forms in which evidence may be transmitted through a variable. It is observed that one can decide for any pair of variables in a causal network whether or not they are dependent once knowing the evidence entered into the network.

Before concluding about the d-separation, the following Bayesian network semantics are worth mentioning. By applying Bayes' theorem, the direction of the arcs can be reversed as long as a directed cycle is not induced. While changing the arc, directionality may change the d-separation properties of the network, the overall joint probability distribution will be invariant.

Therefore, technically, networks differing only in arc directionality can be considered equivalent. However, semantics are conventionally used to make particular configurations of arc directions unique. While not entailed by the underlying theories, the addition of semantics is convenient. The most common interpretation of an arc is: if A is a parent of B, then A is said to exert a causal influence on B, or precede B temporally, and not the other way around (Jensen, 1996).

3.3.2 Learning in Bayesian Network

A Bayesian network may be hand-constructed by a domain expert, that is, the domain expert draws the dependencies between the nodes. The conditional probabilities can then be assessed by the expert, learned from data, or obtained using a combination of both techniques (Neapolitan, 2004). However, eliciting Bayesian networks from experts can be a laborious and difficult procedure in the case of large networks. For the purpose of this research a Bayesian network model and the probabilities will be learnt from data. Different researchers have developed methods that could learn the conditional probability distributions from data (parameter learning) as well as the DAG (structure learning).

Though data on criminals are continuously being gathered, due to lack of appropriate tool they are not effectively being utilized for extracting patterns that can be used for effective management of crimes (Wilson, 1963). Such a data collection usually contains highly valuable information about the relationships between the variables discerned, be it implicitly. If a comprehensive data set is available, a Bayesian network can be learnt from the data, that is, it can be developed without explicit access to knowledge of human experts.

Learning a Bayesian network from data involves the tasks of structure learning, that is, identifying the graphical structure of the network, and parameter learning, that is, estimating the

conditional probability distributions to be associated with the network's digraph. In many learning algorithms, the two tasks are performed simultaneously and, as a consequence, are not easily distinguished (Heckerman, 1996).

3.3.2.1 Parameter Learning

Whenever we have a known network structure and complete data, the joint probability distribution (parameter) can be learned from data. Learning Bayesian networks with known network structure and complete data is the most studied case in the literature, since the network structure is already defined and the algorithm needs to estimate only the parameters (Spiegelhalter et al., 1993). Parameter learning is achieved simply by calculating the conditional probability table (CPT) entries using estimation techniques such as Maximum Likelihood Estimation and Bayesian estimation. An approach to parameter learning with complete data is described in Heckerman (1999) and Krause(1999).

Methods for handling the missing values are needed to either construct the structure of the Network or approximate the parameters. Different methods to approximate the missed values exist for learning the parameters when the random sample is incomplete (i.e., some variables in some cases are not observed).

One class of approximations is based on Monte-Carlo or sampling methods. These approximations can be extremely accurate, provided one is willing to wait long enough for the computations to converge. Monte-Carlo methods yield accurate results, but they are often intractable (for example, when the sample size is large). Another approximation that is more efficient than Monte-Carlo methods and often accurate for relatively large samples is the Gaussian approximation (e.g., Kass et al., 1988; Kass and Raftery, 1995).

3.3.2.2 Structure Learning

According to Rahel (2005), the process of learning Bayesian networks from data takes four different forms, in terms of whether the structure of the network is known and whether the data is complete. These are:

- Unknown network structure and complete data
- Known network structure and complete data
- Unknown network structure and incomplete data
- Known network structure and incomplete data.

Learning with complete data indicates that the training data contains no missing values, while, learning with incomplete data indicates that some piece of information in the data are not known (missing). One point to remember here is that each case has its own learning algorithms.

There are two general approaches to graphical probabilistic model learning from data, the *search & scoring* methods and the *dependency analysis (Constraint Based)* methods.

Search & scoring methods, these algorithms view the learning problem as to search for a structure that can fit the data best. They start with a graph without any edges, and then use some search method to add an edge to the graph. After that, they use some scoring method to see if the new structure is better than the old one. If it is, they keep the newly added edge and try to add another one. This process continues until no new structure is better than the previous one. Different scoring criteria have been applied in these algorithms to evaluate a structure, such as Bayesian scoring method (Cooper and Herskovits, 1992; Heckerman *et. al.*, 1994; Ramoni and Sebastiani, 1996), entropy based method (Herskovits, 1991), minimum description length method (MDL) (Suzuki, 1996; Lam and Bacchus, 1994), and minimum message length method (Wallace

et. al., 1996). Most of these algorithms apply heuristic search methods. To reduce the search space, many of these algorithms require node ordering.

In the second approach, the learning problem is viewed differently. Since a structure encodes many dependencies of the underlying model, the algorithms of this approach try to discover the dependencies from the data, and then use these dependencies to infer the structure. These algorithms are referred as *CI* (Conditional Independence) based algorithms or **constraint-based algorithms** (Spirtes and Glymour 1996; Cheng *et al.* 1997). The dependency relationships are measured by using some kind of conditional independence test. The BN structure encodes a group of conditional independence relationships among the nodes, according to the concept of *d-separation* (Pearl, 1988). Using some statistical tests (such as Chi-squared test and mutual information test), it can find the conditional independence relationships among the attributes and use these relationships as constraints to construct a BN.

Different algorithms have been developed to handle both approaches. According to Cheng *et al* (1998), although some of these algorithms can give good results on some benchmark data sets, there are still several problems related to:

- *Node ordering requirement.* Many of the previous algorithms assume that node ordering is available. Unfortunately, in many times this is not the case.
- *Lack of efficiency.* Some recent algorithms do not need node ordering, but they are generally not very efficient. All practicable dependency analysis based algorithms require exponential numbers of Condition Independence (CI) tests.
- *Lack of publicly available learning tools.* Although there are many algorithms for this task, only a few Bayesian network learning systems are publicly available.

One such algorithm that is capable of fulfilling all these requirements is the Three Phase Dependence Analysis (TPDA) algorithm.

The most significant advantage of these algorithms is that unlike all other practicable dependency analysis based algorithms; these algorithms can avoid exponential complexity on conditional independence (CI) tests.

Rahel (2005) underlined that in a preliminary investigation made for the purpose of selecting an appropriate learning algorithm from those which are publicly available, **TPDA** algorithm is found to perform better and hence its selection for her research work. Accordingly, the algorithm is adopted in this research, too.

TPDA can be used to learn Bayesian networks in the general case where node ordering is not provided and it can also be used to learn Bayesian networks in a special case where node ordering is given. These two algorithms are named TPDA-I and TPDA-II respectively (Rahel, 2005). Both *TPDA-I* and *TPDA-II* algorithms have been incorporated into the *Bayesian Network PowerConstructor* system.

For the purpose of the modeling the experiment, a Belief Network PowerSoft software package is used. Cheng (2001) stated that this software package includes BN PowerConstructor, BN PowerPredictor and a Data Preprocessor. Besides its efficiency and scalability, this system has the following features.

- **Accessibility.** The system supports most of the popular desktop database and spreadsheet formats, including Ms-Access, dBase, Foxpro, Paradox, Excel and text file formats. It also supports remote database servers like Oracle, SQL-server through ODBC.

- Reusability. It can be easily integrated into other Bayesian network, data mining or knowledge base systems for Windows 95/98/NT/2000.
- Supporting domain knowledge. Complete ordering, partial ordering and causes and effects can be used to constrain the search space and therefore speed up the construction process.
- Automatic feature subset selection and model selection in PowerPredictor by using a wrapper approach.
- Supporting misclassification cost function definition in PowerPredictor software uses the publicly available Three Phase Dependence Analysis (TPDA) learning algorithm.

TPDA algorithms can be used to construct the network when the data is complete and incomplete. But for the purpose of this research, learning the Bayesian network structure when we have a complete dataset is considered. Network structure learning for complete data can be accomplished using two approaches. One approach is using only the available dataset . This is to mean experts are not consulted in the process of constructing the model. In another approach, experts' opinions are integrated in the construction process in addition to the learning the model from the data.

Model building without experts involvement (TPDA-I) and with addition of their domain knowledge (TPDA-II) for complete data are discussed as follows

3.3.2.2.1 Model building only from data (TPDA-I)

According to Rahel (2005), Bayesian learning methods with unknown network structure and complete data, involves **three** phases. First, manual selection of the model variables and their

possible values are performed. Second, automatic determination of the structure of the graph based on a dataset is performed. This can be implemented with the approaches either **Constrained-based** or **Score-based**. Third, automatic calculation of the conditional probability distribution (parameter) is performed.

TPDA learning algorithm with unknown structure (**without a given node ordering**) takes a database table as input and constructs a Bayesian network structure as output. Since node ordering is not given as input, this algorithm has to deal with two major problems (Rahel, 2005):

- I. How to determine if two nodes are conditionally independent, and
- II. How to orient the edges in a learned graph.

As described by Cheng et al. (1997), the algorithm has four phases: *drafting* (connecting arcs according to the mutual information of each node pair), *thickening* (adding the missed arcs), *thinning* (removing unwanted arcs) and *orienting edges* (adjusting the directions of the arcs).

3.3.2.2 Structure Learning for Given Node Ordering (TPDA-II)

This algorithm takes as input both a table of database entries and a node ordering and constructs a Bayesian network structure as output.

The **first three phases** of this algorithm are the same as the TPDA algorithm described in the previous section. However, the last phase (orienting edges) described above, is not implemented in this algorithm, since the direction of the arcs are decided by the node ordering provided. The main features involved in these three phases, according to the discussions of Rahel (2005), are the following:

- I. When direct cause and effect relations are available, it uses them as a basis for generating a draft in phase I.
- II. In phase II, the algorithm will try to add an arc only if it agrees with the domain knowledge.
- III. In phase III, the algorithm will not try to remove an arc if it is already specified by experts.

3.4 Related Research Work

In recent years applicability of the Bayesian network is increasing due to capacity to reason for complex and uncertain situation. Some of the researches that were conducted in areas related to crime and those predictions employed using Bayesian network are reviewed below.

Leul (2004) used the data mining techniques for extracting patterns that exist in personal identification record of criminals. In his research, neural network and decision tree were used for extraction of the patterns. He divided the dataset into training and testing datasets. After training the classifier using both neural network and decision tree, the decision tree provided better result as compared to neural network over the test data. And he obtained satisfactory result.

Kelli (2005) used the Bayesian network modeling technique to model the criminal behavior linking the action of an offender on the scene of the crime to his or her psychological profile. A technique has been developed to reduce the search space of possible BN structures by modifying the greedy search K2 learning algorithm to include a-priori conditional independence relations among nodes. Once the BN model is constructed, an inference algorithm is used to predict the offender profile from the behaviors observed on the crime scene. The overall predictive accuracy of the model obtained by the modified K2 algorithm was 79%, showing a 15% improvement with respect to a model obtained from the same data by the original K2 algorithm. In fact, the predictive accuracy is found to increase with the confidence level provided by the BN.

Huygens (2004) used the Bayesian network for reasoning in legal system. It compiled the evidences obtained and tried to reason based on the evidences for exact cause of the accident so that the juries make correct prediction by taking car accident as a sample case.

Kwan et al. (2007) presented a reasoning model based on the probability distribution using a Bayesian Network. By setting out probability distributions over hypotheses for computer forensics analyses, they quantified the evidential strengths of such hypotheses for enhancing the reliability and traceability on the analytical results of computer forensics examinations. To study the validity of the proposed model, a real court case about BT technology has been fitted to the calculations. In order to detach the subjective views, a survey is carried out to collect the expertise of 31 experienced law enforcement agencies. Their responses were aggregated to generate some more objective assignments to the prior probabilities to be used. The outcome demonstrates a high propagated probability of 92.7%, which is in accordance with the actual court verdict of guilty. That presents computer forensics a real scientific science with quantifiable analyses.

Norman and Martin (2000) introduced what they believe, a previously unreported fallacy, which they refer to as the jury observation fallacy. They stated that there is a basic misunderstanding about the belief in probability of guilt when a prior similar conviction by a defendant is revealed after the jury returns a not guilty verdict. Specifically, it is widely believed that the information about the prior conviction might suggest to external observers that the jury verdict is wrong (the belief is that probability of guilt increases). In fact, using very reasonable (and indeed conservative) assumptions they tried to show, using Bayesian reasoning, that such a response is irrational in many situations. To explain the Bayesian argument without exposing readers to any of the mathematical details they have used Bayesian Networks (BNs) and a tool (Hugin) to

execute them. Hence, a secondary objective of this paper is to show that there is a way of making all of the implications of Bayesian reasoning clear to lay people, without them having to understand any of the underlying mathematics. The implications of this in the legal profession are profound. Courts could eventually accept Bayesian arguments just as they accept forensic evidence without having to resort to explanations from first principles. Additionally, the results presented suggest that there may be reason for disquiet about the use of previous convictions as a basis for selecting suspects as is common Police practice.

Violations of system security policy by authorized computer users present a major threat to information security. Kathryn et al. (1996) presented an innovative use of human behavior models for detecting and responding to intruders. A promising approach to detection and response is to model behavior of normal users and threats, and apply sophisticated inference methods to detect patterns of behavior that deviate from normal behavior in ways suggesting a possible security threat. This paper presents an approach, based on multi-entity Bayesian networks, to modeling user queries and detecting situations in which users in sensitive positions may be accessing documents outside their assigned areas of responsibility. Such unusual access patterns might be characteristic of users attempting illegal activities such as disclosure of classified information. They present a scalable proof of concept behavior model, provide an experimental demonstration of its ability to detect unusual access patterns in simulated situations, and describe future plans to increase the realism and fidelity of the model.

In addition, Rahel (2005) experimented on the application of the Bayesian network for computer-assisted learner group formation based on personality traits. This study addressed two main issues in relation to forming effective heterogeneous learner groups to improve student performance. One was the task of developing a performance prediction model without administering exams and

the other was the development of a software tool to form effective heterogeneous groups. It took mathematics as the subject of the experimental study. The level of prediction accuracy was 78.4%.

In summing up, Bayesian networks have been found to have the following advantages (Rahel, 2003):

- They handle incomplete data sets without difficulty because they discover dependencies among all variables;
- One can also learn about causal relationships between variables using Bayesian networks and the strength of the causal relationships with probabilities
- Considering the Bayesian statistical techniques, Bayesian networks facilitate the combination of domain knowledge and data. Prior or domain knowledge is crucially important if one performs a real-world analysis; in particular, when data is inadequate or expensive. The encoding of causal prior knowledge is straightforward because Bayesian networks have causal semantics;
- Independencies can be dealt with explicitly. They can be articulated by an expert, displayed graphically, and reasoned about, yet they remain robust to numerical expressions.
- Bayesian Network structure represents the inter-relationships among the attributes. Humans can easily understand the network structures and experts can modify them to obtain a better predictive model.

CHAPTER FOUR

EXPERIMENTATION

4.1 Introduction

This chapter discusses the application and implementation of Bayesian network for modeling determinant factors influencing offenders to commit crime. The experimentation is conducted on records which were collected from the Addis Ababa Police Commission. In the following subsections the main concepts to be discussed are: data collection, preprocessing, training and testing the classifier, model building using the TPDA-I and TPDA-II and finally evaluation of the classifiers that are learned from TPDA-I and TPDA-II using the BN PowerSoft software.

4.2 Data Collection

To develop a Bayesian network model from the data, we need to have the data itself in the first place. The data for this experiment was collected from the Addis Ababa Police Commission. The data that is collected from the commission was from the personal identification record of the criminals. Addis Ababa Police Commission was selected for three reasons. Firstly, it is central repository for documents that are collected from all sub-cities of the city. Secondly, the documents are in well organized manner that enables proper collection of the data. Thirdly, they deal with crimes that have huge impact on the victims, society and on the country at large. This is mainly due to the fact that crimes that have lower crime label rate are handled by the police stations residing at the sub-cities of the city.

The general architecture that is followed in constructing the Bayesian network model is depicted on the next page.

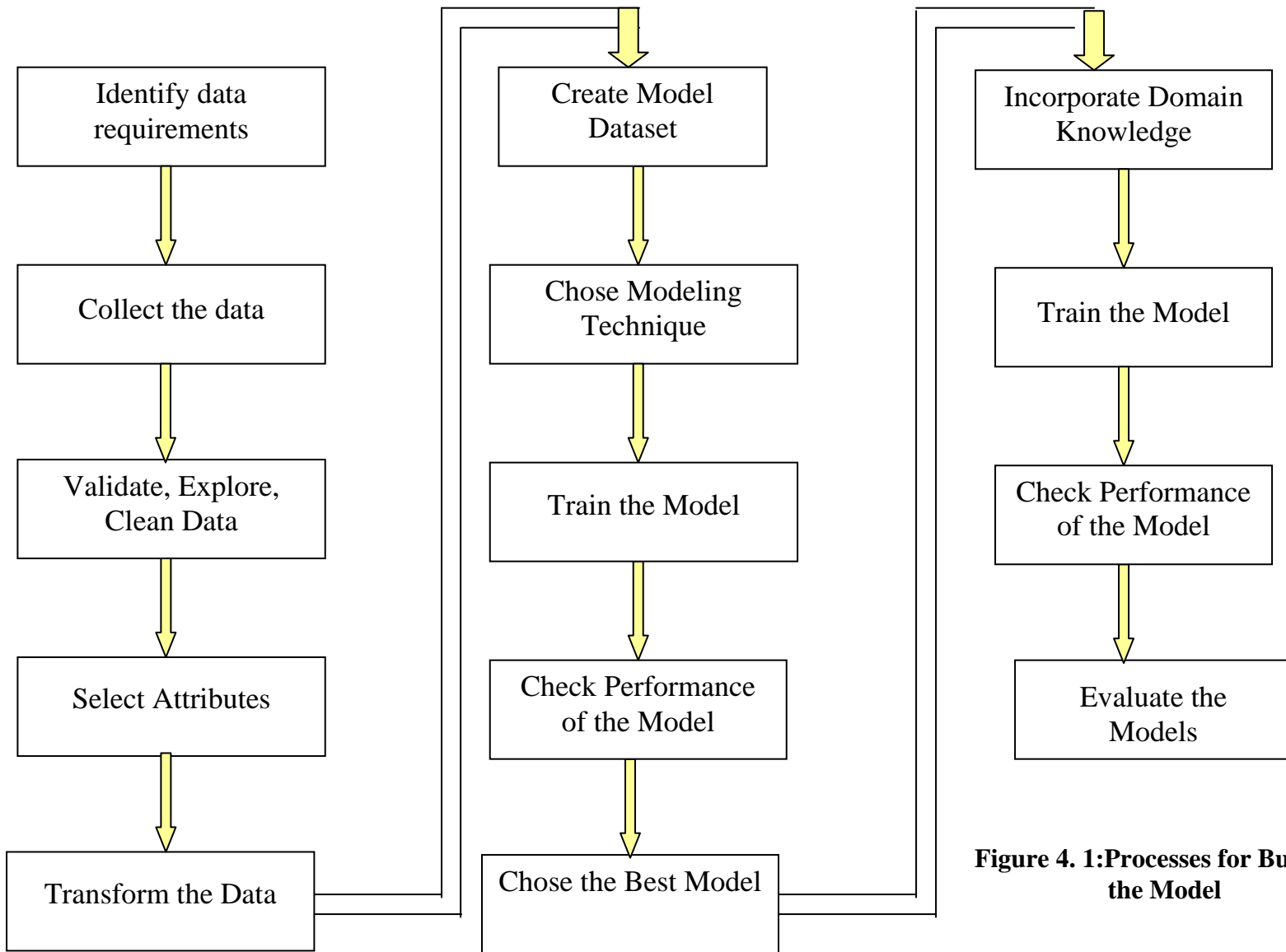


Figure 4. 1:Processes for Building the Model

The personal identification record of criminal contains much information regarding the crime. It has included the reports that are obtained from the criminals, witnesses, inspectors and also other information that are relevant for the analysis of crimes. Undoubtedly, all these information contribute their share in the analysis of crimes. But only information regarding criminals are selected due to the emphasis of the research on identifying risk factors that influence criminals than witnesses or inspectors.

All of the collected records regarding the criminals were first in manual format. The process of changing the manual data into an electronic format was time taking, tedious and challenging. Moreover before working with the tool, further preprocessing was conducted for handling outliers, missing values and other preprocessing tasks.

4.3 Selection of Attributes

Identification of important attributes that are relevant for the correct prediction of factors that constitute higher crime trend and for the identification of the interdependencies among these attributes requires careful selection. The attributes that are selected for the purpose of this research with the discussion of the domain experts, review of literatures and their easiness for collection from criminal record are the following attributes:

- **Gender:** Referred to the sex (masculinity or femininity) of the convicted criminal.
- **Age:** Referred to the age of the convicted criminals.
- **Educational Status :** Referred to the highest level of education the criminal had attained
- **Marital Status:** Referred to the status whether the criminal has been married or not

- **Religion:** Referred to the belief and worship of the convicted criminal.
- **Occupation:** Referred the means the criminal is using to gain money.
- **Crime Scene:** Referred to the location where the crime is committed.
- **Committed in group:** Referred to a crime being committed by more than one criminal.
- **Dependence:** Referred to relying on the family for living.
- **Crime Label Rate:** Referred to the extent or rate of the impact of a crime

Attribute Name	Its Description
Accuser Name	Name of the accuser who have convicted the criminal
Criminal Name	The name(s) of the criminal(s) who committed the crime
Crime Time	The time at which the crime was committed
Ethic	The Ethic background of the criminal
Crime Date	The date the crime was committed
Report Date	The date the crime was reported to the police
Case Number	A unique number assigned for each crime
Name of the Inspector	The name of the inspector who is handling the case
Date Sent	The date the case was sent to the court
Court Name	The court that is assigned for handling the case
Sentence Date	The date at which the case was given a sentence
Sentence type	The judgment by the court for the case

Table 4. 2: Attributes which exist in the personal identification record of the criminals but not used for the research purpose.

4.4 Data Preprocessing

This is a step where the collected data was arranged in a form that was suitable for further processing of the data using the BN PowerPredictor or BN PowerConstructor. The process of converting the raw records to be suitable for constructing the model using the Bayesian network involved manual preprocessing and automatic preprocessing. The manual preprocessing involved removal of noisy and irrelevant data from the collection that may have occurred in mistakes in data entry, omission of values and inconsistencies. This has to be performed in view of the fact that the BN PowerSoft works only with complete data and those data records which have incomplete data for some attributes are disregarded from the collected data.

For automatic preprocessing of data, the Belief Network PowerSoft has its own data preprocessor tool called Data PreProcessor. This tool is used for the preprocessing of the training data which is then used for testing the prediction accuracy and for the construction of the model. This tool has three functions: converting data from other desktop database formats to Microsoft JET/Access (*.MDB) format (as required by BN PowerPredictor or BN PowerConstructor), for detecting and discretizing data fields that contain continuous data and for dividing the training data into internal training set and internal test set (as required by BN PowerPredictor).

For instance, the attribute Age was first a continuous value beginning from 16 up to 75 years of age. This was difficult for BN PowerSoft either for learning Bayesian network classifiers or constructing the Bayesian network model. Hence the attribute Age is discretized, using the equal width discretization approach into three categorical labels of $\text{Age} < 35.666$, $35.666 < \text{Age} < 55.3$ and $\text{Age} > 55.333$ representing the young, middle-aged and the old respectively.

The equal width discretization approach divides the range of values between the highest and smallest values into convenient classes, which must be mutually exclusive and are usually equal. Its values are automatically selected by the Data Preprocessor, the user only provides the number of classes to be performed.

Yet another attribute, *Crimescene*, the location where the crime is committed can go up to ten sub-cities, 'woredas' and even to 'kebele' levels. With the consultation of the domain experts, this attribute is also made to have three categorical values: "*Very Active*", "*Active*" and "*Medium*". These categorical labels indicate the business activities of the location as well as the population size of the sub city.

The attribute "*LevEducation*" describes the highest level of education the criminal has attained. This attribute has so many values beginning from illiteracy up to the masters level. This is also changed to categorical values: "*Illiterate*", "*Elementary and junior*", "*Highschool*" and "*College and above*" category so as to enable proper interpretation of the outcome of the Bayesian software.

The class label, "*CrimeLabRate*", is used for the prediction as well as network construction purpose. The "*CrimeLabRate*" represent the extent to which the crime has affected the victim, society and the country. It had so many values ranging from being a false witness up to killing an individual. This also required proper adjustments to work with the software package and for making proper interpretation of the network. With discussion of the inspectors these values also changed to "Medium", "Serious" and "Very Serious" signifying their impact on the society and the number of years imprisonment for committing these crimes.

After making proper rearrangement for these attributes and others, the attributes that are selected and their categorical labels are shown in the table below.

Attributes	Short forms	Possible values
Gender:	Sex	Male , Female
Age :	Age	Young, Middle-aged, Old
Educational level	LevEducation	Illiterate, Elementary and Junior, Secondary, College and above
Committed In group	CommittedIngroup	Yes, No
Marital Status	Marritalstat	Married, Single, Divorced
Religion	Religion	Orthodox, Muslim, Protestant, Other
Occupation	Occupation	Unemployed, Business Man, Daily Worker, Employed
Crime Scene	CrimeScene	Very Active, Active, Medium.
Dependence	Dependence	Yes, No
Crime Label Rate	CrimeLabRate	Medium, Serious, Very Serious

Table 4. 2: Attributes that are selected for the purpose of the Research

After preprocessing, the data is used for testing the prediction accuracy of the classifier and for constructing Bayesian network model for the personal identification record of the criminals. The following is a sample of the training dataset that was used for both testing the classifier accuracy and constructing the model.

Sex	Age	EducationStatus	MaritalStat	Religion	Occupation	CrimeScene	Dependent	Committedingroup	CrimeLabRate	Age_d
M	18	College and Above	Single	orthodox	Unemployed	Very Active	yes	Yes	Serious	<35.66666
M	38	College and Above	Married	Orthodox	Business Man	Medium	Yes	Yes	medium	>35.66666<55.33333
M	27	Illiterate	Married	Muslim	Daily Worker	Medium	no	Yes	very serious	<35.66666
M	34	Elementary and Junior	Married	orthodox	Business man	Medium	no	No	very serious	<35.66666
M	60	Secondary	Married	Orthodox	Business man	Active	no	No	medium	>55.33333
M	19	Elementary and Junior	Single	orthodox	Daily Worker	Very Active	yes	No	Serious	<35.66666
M	28	Secondary	Single	Muslim	Business man	Very Active	yes	No	Serious	<35.66666
M	38	Secondary	Married	Orthodox	Business man	Very Active	no	No	Medium	>35.66666<55.33333
M	48	Illiterate	Married	orthodox	Business man	Active	no	No	medium	>35.66666<55.33333
F	25	Secondary	Married	Orthodox	Unemployed	Active	No	No	serious	<35.66666
M	46	College and Above	Single	orthodox	Unemployed	Active	yes	No	medium	>35.66666<55.33333
M	28	Secondary	Single	protestant	Unemployed	Active	yes	No	medium	<35.66666
M	26	Elementary and Junior	Single	Muslim	Daily Worker	Active	no	No	medium	<35.66666
M	21	Secondary	Single	Orthodox	Business Man	Very Active	Yes	Yes	Very Serious	<35.66666
M	46	College and Above	Married	orthodox	Unemployed	Active	no	No	medium	>35.66666<55.33333

Figure 4. 2: Sample Training Dataset in Microsoft Access format after Preprocessing

4.5 Building The Bayesian Network Model

To work with BN PowerSoft package and its learning algorithm, the dataset is fed to the PowerPredictor which partitions the data internally to training and testing sets. For testing the accuracy and to give each data a chance to be part of the test set, the dataset was divided to ten partitions (10-fold cross validation). The reason 10-fold cross validation was selected is because it is highly recommended whenever we have a smaller or medium datasets and it is also used as a default for many packages (Cheng et. al, 1997). Using the 10-fold cross validation, each partition was used exactly once for testing purpose and the remaining parts were used as training sets. These processes were repeated ten times so that each partition was used for testing purpose only once. Then the average of these ten experiments is reported as the prediction accuracy for experimentation. When 1572 records were fed to the network, the BN PowerSoft divided the data internally into training and test sets. The purpose of partitioning the dataset is to learn from the patterns in the training set and check the accuracy and deviation of the test set from the already

trained dataset for the actual class label. Models were constructed using both the dataset (with out expert opinion) and combination of expert knowledge and the dataset for improving the predication accuracy of the classifier.

4.6 Testing The Network

The BN PowerPredictor is first trained with the training dataset from which it learns the classifiers. To test the prediction accuracy of the learned classifier, a test set is fed to the PowerPredictor. The BN PowerPredictor confers with the model based on the evidence values it has obtained from the training set. The PowerPredictor evaluated the probability values assigned for the three categories of seriousness of the crime . It then took the crime label rate category having the maximum probability value and saved it together with the corresponding data record or class label. The following figure shows the previously assigned (actual) class label as well as predicted class label that is obtained after the BN PowerPredictor learned from the data and used it for classification purpose accordingly.

SEX,"AGE","EDUCATIONSTATUS","MARITALSTAT","RELIGION","OCCUPATION","CRIMESCENE","DEPENDENT","COMMITTEDINGROUP"	"CRIMELABRATE","BNPP_OUTPUT","Prob. MEDIUM","Prob. SERIOUS","Prob. VERY SERIOUS"
M,"19","Elementary","single","orthodox","daily worker","Very Active","yes","no","serious","VERY SERIOUS"	,6.763377E-02,2770713,655295
M,"29","college","single","Muslim","Business man","Very Active","yes","no","serious","SERIOUS"	,1980757,7677172,3.420696E-02
M,"37","secondary","married","orthodox","Business man","Active","no","no","serious","SERIOUS"	,3772359,3819908,2407733
M,"22","secondary","single","orthodox","daily worker","Medium","no","no","very serious","VERY SERIOUS"	,4.378574E-02,3991289,5570853
F,"45","secondary","married","orthodox","employed","Very Active","no","no","very serious","VERY SERIOUS"	,4.015682E-02,3749565,5848866
M,"17","secondary","single","orthodox","unemployed","Medium","yes","no","very serious","SERIOUS"	,1071383,7316253,1612365
M,"21","secondary","single","orthodox","Business man","Very Active","yes","no","very serious","VERY SERIOUS"	,1815595,3989608,4194797
M,"34","secondary","single","orthodox","unemployed","Very Active","Yes","no","serious","SERIOUS"	,2195053,6953514,8.514334E-02
M,"19","secondary","single","orthodox","daily worker","Active","Yes","no","serious","VERY SERIOUS"	,3.006691E-02,3459325,6240006
M,"27","Junior","single","orthodox","unemployed","Very Active","Yes","Yes","serious","SERIOUS"	,1.481281E-03,9858858,1.263294E-02
M,"26","secondary","single","orthodox","unemployed","Medium","Yes","no","medium","SERIOUS"	,1071383,7316253,1612365
M,"38","Elementary","Married","orthodox","Business man","Active","no","no","medium","MEDIUM"	,7321984,1953915,7.241009E-02
M,"38","secondary","single","orthodox","Business man","Medium","no","Yes","serious","MEDIUM"	,4995747,4504585,4.996686E-02
M,"31","illitrate","Married","orthodox","Business man","Active","no","no","medium","VERY SERIOUS"	,3001299,2980821,401788
M,"31","illitrate","Married","orthodox","Business man","Medium","no","no","medium","MEDIUM"	,5902559,2008242,2089198
M,"18","Junior","single","muslim","daily worker","Active","Yes","Yes","serious","SERIOUS"	,4.593956E-03,9851496,0102565

Figure 4. 3: A report generated by BN PowerPredictor for both actual and predicted class labels with their assigned probablities.

Upon completion of the prediction, it produced the confusion matrix comparing predicted rate of crime label with that of actual rate of the crime label. Then the average of all the prediction performances evaluated for all ten partitions of the dataset is reported as a result for the prediction performance of the classifier.

For this experiment, the best prediction accuracy of the classifier that is learned from the data among the datasets that are used for 10-fold cross validation is shown in table below.

		Predicted			Accuracy
		Medium	Serious	Very Serious	
Actual	Medium	471	47	59	81.63
	Serious	61	331	53	74.38
	Very Serious	59	80	283	67.06

Table 4. 3: The prediction performance for the best learned model from data

From the table 4.3, it can be observed that a misclassification that is highly costly is the classifying those criminals who commit very serious crimes into medium crime committers. For instance, in table 4.3 such misclassification accounts for only 4.1%.

In similar manner, for each of the remaining nine experiments the actual class labels values and the predicted values for the class labels are depicted in the appendix. The percentage prediction performance of the classifier for each dataset and the average of these prediction accuracies are depicted in the following table.

	Datasets										
	1	2	3	4	5	6	7	8	9	10	Average
Accuracy	72.71	74.17	73.41	73.82	71.88	75.14	73.96	70.89	74.86	71.61	73.25

Table 4. 4: The Prediction accuracies for each of the Datasets used

From the table 4.4, the average prediction performance of the classifier without experts involvement in the network construction is 73.25% at 95 % confidence interval. Though the result of this experiment is convincing, the accuracies was further enhanced with incorporation of the knowledge of the domain experts.

4.7 Constructing the Bayesian Network Model

4.7.1 Constructing the Bayesian Network Model when node ordering is not given (TPDA-I)

TPDA learning algorithm with unknown network structure (**when node ordering is not given**) takes a database table as input and constructs a Bayesian network structure as output. Since node ordering is not given as input, this algorithm has to determine if nodes are independent or not and orient the edges in the learned graph (Rahel, 2005).

Bayesian network PowerConstructor is part of the BN PowerSoft software package. It is used for building the model either from data only or the combination of the experts' domain knowledge with the data. Among the several publicly available learning algorithms, the Three Phase

Dependency Analysis (TPDA) algorithm that is embedded in Bayesian network PowerConstructor is used in developing the model.

To construct the Bayesian network model, a data that is preprocessed and partitioned is fed to the BN PowerConstructor. Using the Three Phase Dependency Analysis that is embedded in it, the BN PowerConstructor constructs the model of cause and effect relationships among the attributes that are learned from the data. The following is a report that is generated for the dependencies among the attributes and which is also used internally by the constructor for building the graphical network structure of the cause and effect relationship among the attributes.

```
Sex -> Occupation  Sex -> CrimeScene  Sex -> CrimeLabRate
EducationStatus -> Occupation  EducationStatus -> Age_d
MaritalStat -> Dependent  Religion <- CommittedIngroup
Religion -> CrimeLabRate  Occupation <- Dependent
Occupation -> CrimeLabRate  Occupation <- Age_d
CrimeScene -> CrimeLabRate  Dependent -> Age_d
CommittedIngroup -> CrimeLabRate  CrimeLabRate <- Age_d
```

Figure 4. 4: A report that is generated when the learning process from the data is finished.

Figure 4.4 shows the cause and effect relationships among the different attributes of the criminal. For instance, between seriousness of a crime, *CrimeLabRate*, and the occupation of the criminal, *Occupation* is depicted as *Occupation->CrimeLabRat*. This means the effect, seriousness of crime, is dependent on the cause, the kind occupation the criminal involved in. Similarly, the relationship between the level of education and employment is depicted as *EducationStatus -> Occupation*. This is to mean that the kind of occupation criminals are involved in depends on their educational background. From the figure 4.4, the seriousness of the crime, *CrimeLabRate*, depends on sex, age and occupation of the criminal. Using the above cause and effect relationship

report, the following network structure is learned from the dataset with out the interference of the domain experts

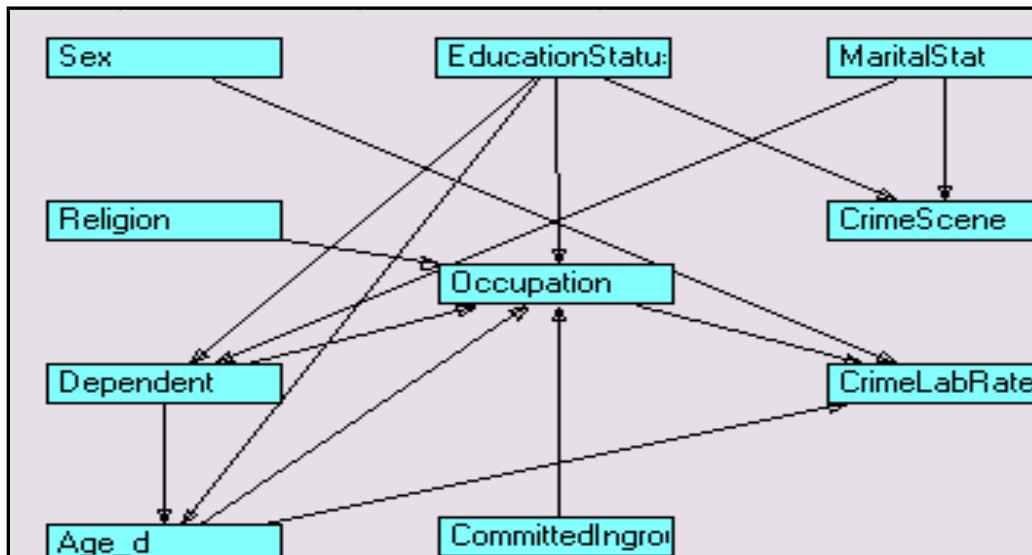


Figure 4. 5: The best learnt Bayesian network model from data using TPDA-I

Figure 4.5 shows that seriousness of the crime depends on the age, occupation and sex of the criminals. These dependencies among attributes of the offenders continue up to the root nodes.

Experts were consulted to integrate their domain specific knowledge in order to improve the prediction accuracy as well as the applicability of the network. Doing so has required another approach that constructs the dependencies among the attributes when the node ordering is given.

4.7.2 Constructing the Bayesian Network Model for a given node ordering (TPDA-II)

One of the advantages of the Bayesian network is it enables to incorporate the knowledge of the domain experts to improve the accuracy of the prediction and applicability of the network. The reason for doing so is that links may have been established between attributes which are

independent in the general population; links may exist where direction should have been the opposite from what appears in the network; links may exist where attributes are not directly related; an expert may also expect a variable to have several more parents than actually appearing on the network. This means automatic learning methods alone may not be sufficient. An option considered under such circumstances, is to reinforce the learning using the knowledge of the human/domain expert (Rahel, 2005).

Model building using the TPDA-I algorithm is without a given node ordering (i.e., human experts knowledge are not incorporated in the construction of the model). Such construction has already been discussed in section 4.7.1. Building a model when experts are consulted for the ordering of the nodes and cause and effect relationships among attributes led to TPDA-II. TPDA-II is the construction of the Bayesian network model following the same processes that have been performed for the model construction using the TPDA-I except that the order of the nodes, cause and effect relationship, forbidden links are all initially provided by the experts in the domain.

The experts are let to reason about the existence or not of: a direct cause and effect relationship between the attributes, indirect relationship, dependence or independence given certain conditions. To incorporate their knowledge, BN PowerSoft constructor has different features that enable domain knowledge assimilation. Among which, one of them is order of the nodes and cause and effect relationships are provided by the experts. In complete ordering the expert creates the complete order of dependencies among attributes by moving up and down the attributes. The attributes at the top does not depend on any of the attributes below it. Using this mechanism, they integrate their belief in the constructing the belief network. For example, they suggested that the sex of the criminal does not depend on the any of the attributes in any way. So it is placed at top of all attributes signifying that it does not depend on any of the attributes below it.

Further experts' knowledge incorporation is possible using the causes and effects identifications between all of the attributes. For example, the kind occupation a criminal involved depends on the educational status of the criminal. Moreover the links that were inadvertently built using the TPDA-I were also removed from the network.

After making such rearrangements, the BN PowerConstructor once again was fed the best learned dataset that is used for constructing the TPDA-I. The PowerConstructor produced the following figure.

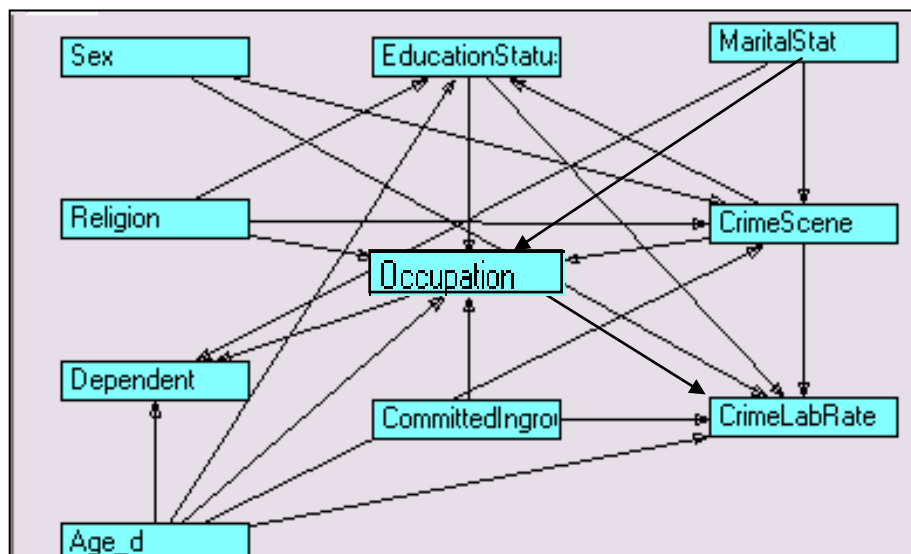


Figure 4. 6: A model after incorporating knowledge of the domain experts

Some of the major changes made in the modified network as compared to the original learned network are the following:

- In the original network, only three attributes were believed to have a direct impact on the seriousness of the crime. But three more attributes, which in the original network believed to have either an indirect effect or no effect at all, were changed to have a direct impact on

the extent and rate of crime label rate. These are committing crime in group, crime scene and educational status of the criminal.

- In the original learnt network, a crime committed alone or in-group has no impact on rate of crime label. But according to the experts it has an impact on the label rate of the crime committed
- Marital status of the criminal, *Maritalstat*, has no either direct or indirect impact on the crime rate on the original network. But it is believed to have an indirect influence on the extent of the crime rate through the occupation

With the provision of the original model, experts were let to reason the cause and effect relationship among attributes. The reasons for changes made are due to the consensus among many of the experts on the following issues:

- ☆ The higher the education level of the criminals the less would be their involvement in crime. Since crime limits a person's ability to appreciate the consequences of their actions. The less their education level the more likely to offend, more likely to offend frequently, more likely to commit more serious offences and more likely to persist in crime
- ☆ Crimes committed in group or alone determine the crime label rate. The successful commission of many kinds of crime requires a certain measure of knowledge and skill. Delinquents are a valuable source of information about various techniques and opportunities for committing crime.
- ☆ Certain places have higher crime rates. First, returns on crime are likely to be higher due to the greater concentration of wealthier victims, more opportunities to commit various types of

crime, and a more developed second-hand market for the disposal of stolen items. Second, the probability of arresting criminals is less.

- ☆ Married people have the responsibility to finance their family. Having proper finance also depends on the kind of occupation the individuals are engaged in. In real practice, married people with high wages than the others are less likely to involve in serious crime.

In addition to these, all have common consensus on the following points

- ☆ Three different but not mutually inconsistent explanations for the effect of income inequality on crime have been put forward. On one account, income inequality motivates individuals to offend because it creates a sense of relative deprivation amongst those who are poor. According to a second, inequality causes crime in an area because it brings those motivated to offend in close spatial contact with attractive targets for crime. According to a third, the effect of inequality on crime stems from the fact that high levels of inequality result in poverty becoming concentrated in certain areas. Since children from poor households are at higher risk of involvement in crime, the spatial concentration of poverty brings actual and potential offenders into more frequent contact with each other. This further increases the rate of involvement in crime

To their experience, they suggested that education, occupation, crime scene, sex, age and crimes committed in groups have direct impact on the rate and extent of crime.

- Employment is a fundamental issue related to crime and violence rates among young people. They suggested that unemployed youths are disproportionately more likely to be perpetrators, as well as victims of crime and violence.

- Poverty, unemployment and income inequality have all consistently been found to render areas crime-prone.

Model construction using TPDA-II followed the same procedures like that of TPDA-I except domain knowledge incorporation to the network construction. For TPDA-II, the same dataset and number attributes are used in constructing the Bayesian network model. Accordingly, the new learning process is conducted with changes made by the experts.

Following the aforementioned steps, the prediction results were changed significantly for all partitions. The confusion matrix for each of these cases is discussed as follows: The following table summarizes the prediction accuracies and the riskiest predictions for each case obtained from the 10-fold cross validation:

	Datasets										
	1	2	3	4	5	6	7	8	9	10	Average
Accuracy	75.21	76.39	76.59	77.77	75.14	77.04	76.52	73.52	76.11	73.55	75.78

Table 4.5: Prediction accuracies for all partitioned datasets for TPDA-II

Comparing the average prediction performances of the classifiers obtained using 10-fold cross validation when the order of the node is determined and when the order of the nodes is not determined. From the table 4.4 and table 4.5, it can be seen that the average performance prediction has increased from **73.25 % to 75.78 %**. To this end, the average predictive accuracy of the model is improved at the elicitation of domain expert knowledge.

	Datasets										
	1	2	3	4	5	6	7	8	9	10	Average
TPDA-I	4.501	4.086	4.501	4.432	4.363	4.086	4.501	6.025	4.778	4.363	4.564
TPDA-II	4.086	3.670	4.363	3.947	4.086	3.878	4.086	5.471	4.363	3.947	4.190

Table 4. 6: A costly misclassification for both classifiers

From table 4.6, it is observed that the average costly misclassification has decreased from 4.564% to 4.190 %. Though TPDA-II has comparatively produced better result as compared to the TPDA-I with regard to costly misclassifications, both have produced very good result. This is to say, the number of criminals that are wrongly predicted in the medium class label rate which should have actually been predicted in very serious class label rate is very small in both classifiers.

4.8 Evaluation of the Model

From TPDA-I and TPDA-II, an average predictive accuracies of **73.25 % and 75.78 % are** achieved respectively. Moreover, the risk of classifying medium level crime committers as very high label has reduced from 4.564% to 4.190 %. From the results obtained, it can be seen that the Bayesian Network is a powerful predictor even in the absence of a domain expert. With a proper intervention of a domain expert, the Bayesian Network can perform even better. On the other hand, predicting the crime label rates require a more accurate prediction mechanism. In fact this has a lot to do not only with the model but also the data itself. As far as a comprehensive and properly documented data with proper preprocessing conducted on it, Bayesian Network's performance is reliable.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATION

5.1 Conclusion

Bayesian network can be used to model any situation where uncertainties and complexities highly exist (Heckerman, 1996). Learning the structure of the Bayesian network model that represents a domain can reveal insights into its underlying causal structure, criminal modeling being one case. Moreover, it can also be used for prediction of quantities that are difficult, expensive, or unethical to measure (Norsy, 2007)

In this research, Bayesian Networks technology had been used to model patterns that exist in personal identification record of criminals. The objective of this research was to investigate the potential applicability of Bayesian Network technology in developing a model for determinant factors influencing offenders to commit crime.

A total of 1572 records were collected. These records were further preprocessed so as to make the data compatible with software used for modelling.

The belief network modelling software that was employed for the purpose of this experiment was the Belief Network PowerSoft. A Three-Phase-Dependency-Analysis algorithm that is embedded in BN PowerConstructor was used for developing the model. The BN PowerPredictor was also used for predicting the performance of the classifier.

Model building and testing were performed twice: when the order of the nodes was not given (learning the structure of the model merely from data) and when the order of the nodes was

provided by the domain experts. The former approach is named as TPDA-I and the latter one as TPDA-II.

For both TPDA-I and TPDA-II, to avoid inconsistencies during the selection of the test and training datasets, the experiment was carried out by splitting the data using the 10-fold cross-validation

The average of all the 10 test datasets' results was taken as a predictive performance of the model. Accordingly, the predictive accuracy of the model when the order of the nodes was not provided (i.e. TPDA-I) was **73.25 %** at 95% confidence level.

When the orders of the nodes were provided by the experts, a better performance was obtained. The average predictive accuracy obtained for TPDA-II increased to **75.78 %** at 95% confidence level. This is mainly due to the intervention of the domain experts in process of model building.

Though a better result was obtained using TPDA-II (i.e. when the knowledge of the expert incorporated) to TPDA-I (learning only from data), both TPDA-I and TPDA-II have produced a convincing result in modelling the factors for higher crime trends.

Experimental result show that the Bayesian network in general and Bayesian Network PowerSoft software package in particular found to be applicable in identifying the patterns for the higher crime trends for the personal identification record of the criminals. Even though the BN PowerSoft is good predictor in the absence of the domain experts, its accuracy can further be enhanced with a proper intervention of a domain expert.

Despite the fact that a comprehensive and properly documented dataset were not used, the results obtained using the smaller dataset size and numbers of attributes are encouraging for further research work to be conducted in the area of crime and criminals using the Bayesian network.

The Police can make use of the result of this research work to identify issues or areas where to give emphasis to so that the rate and extent of crime be reduced. Furthermore, the result of this research work would enable to identify the kind of support and guidance that could be given to the criminals so that they could change to a productive society. Moreover the research would enable to device a mechanism to handle crime from its root level since we can track from root cause to class label using the cause and effect relationship among the attributes using the model built.

5.2 Recommendation

In this research, the researcher has tried to present the potential applicability of the Bayesian network technology in modeling the impact of economical and environmental factors that has influenced criminals in committing the crimes in Addis Ababa region. In spite of many limitation, this research has revealed that Bayesian network is applicable for building a model that is representative of patterns that are shared by many criminals .Based on the experimental result found and the experience acquired the following recommendations are forwarded

- This research mainly focused on the environmental and social factors that are related to the crime. But for the research to have extensive implication different views from different areas need to be taken in. For instance, the psychological/personality disorder, substance abuse (Addiction), history of previous crime and family condition of the criminals need to be included for better accuracy.
- Different algorithms that are either score based or constrain based or the combination of both can be tested for better accuracy of prediction. Since each approach has its own

advantages, a combined approach could improve the predictive accuracy of the learned model. Hence it is recommended.

- The size of the dataset has an impact on the performance of learning classifier. The more dataset size is provided to the learning tool, the better to spot the relationship among the attributes. This research is conducted on the smaller dataset. The applicability and prediction performance of the predicted model can further be enhanced by using larger dataset size.
- In this research the model was constructed only using the complete data set. But learning is also possible for incomplete dataset using different approximation algorithms which handle missing values.
- BN can be used for reasoning in legal system. Compiling the evidences obtained and then reasoning based on the evidences for exact cause of the accident so that the jury make correct prediction

REFERENCES

1. Antoni, Y. (2003). *Does crime affect unemployment? The role of social networks*. ANNUAL ECONOMIC STATISTICS STIQUE. – No 71-72.
2. Badenhorst, J. J. (1992). *Paradigms, philosophy, theory, methodology and geographic thought*. University of Venda.
3. Berry, G. and Carter, M. (1992). *Assessing crime prevention initiatives: The first steps*, **Crime Prevention Unit**. London.
4. Cheng, D.A. and Liu, W. (1997). *An algorithm for Bayesian belief network construction from data*. Proceedings of AI & STAT'97, Ft. Lauderdale, Florida.
5. Cheng, D.A. and Liu, W. (1997). *Learning belief networks from data: an information theory based approach*. Proceedings of the Sixth ACM International Conference on Information and Knowledge Management.
6. Chow, C.K. and Liu, C.N. (1968). *Approximating discrete probability distributions with dependence trees*. IEEE Transactions on Information Theory, 14:462-467,
7. Cooper G.F, Herskovitz E. (1992). *A Bayesian method for the induction of probabilistic networks from data*. Machine Learning
8. D. Heckerman, D. Geiger, and D. Chickering, (1995). *Learning Bayesian networks: The combination of knowledge and statistical data*. *Machine Learning*, 20(3):197-243, Kluwer.
9. D. Heckerman(1995). *A tutorial on learning with bayesian networks*. Technical report, Microsoft Research.
10. D.M. Chickering, D. Heckerman, and C. Meek. (2004). *Large-sample learning of bayesian networks is np-hard*. *Journal of Machine Learning Research*, 5:1287.

11. D.M. Chickering. (2002). *Optimal structure identification with greedy search*. Journal of Machine Learning Research, 3:507.
12. Daryle Niedermayer, et al. (1998).: *An Introduction to Bayesian Networks and their Contemporary Applications*:
13. DiCristina, B. (1995). *Method in Criminology*. Harrow & Heston, New York
14. Dimitris, M. (2003). *Learning Bayesian Network Model Structure from Dat*. Carnegie Mellon University, Pittsburgh.
15. Fajnzylber P. Lederman D and Loayza N (2002). , *A simple reduced-form model of criminal behaviour*. World Bank Research.
16. Fourie, M. (1999). "*Addiction, crime, death*", Servamus, Special community edition, p 33.
17. Friedman NIR, Linial M, Nachman I, Pe'er D. (2000). *Using Bayesian network to analyze expression data*.
18. Glanz, L. (1996). *The causes of crime in South Africa*. Pretoria.. UNISA,
19. Heckerman, D. (1995). *A tutorial on learning Bayesian networks*. Technical Report MSR-TR-95-06, Microsoft Research.
20. Heckerman, D. (1996). *Bayesian networks for data mining*, Data Mining and Knowledge Discovery
21. Heckerman, D. (1999). *A Tutorial on learning with Bayesian networks*. In (Ed.), Learning in graphical models, ed. M. Jordon, 301-354. Cambridge, Massachusetts: MIT Press.
22. Huygens, P.E.M. 2004. *Use of Bayesian Belief Networks in legal reasoning*. 17th BILETA Annual Conference, Free University, Amsterdam.
23. J. Pearl and T. S. Verma. (1991). *A theory of inferred causation*. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference*. Morgan Kaufmann, San Mateo, CA.

24. James, B. J. and Kimberly, P. (1998). *Hate Crimes: Criminal Law and Identity Politics*. Oxford and New York: Oxford University Press.
25. Jensen, F. V. (1996). *An introduction to Bayesian network*. London. U.K: University College London Press.
26. Jensen, F. V. ,Lauritzen, S. L., & Olesen, K. G. (1990). *Bayesian updating in causal probabilistic networks by local computations*. Computational Statistics Quarterly
27. Kelli, A.C. (2005). *Bayesian Network Modeling of Offender Behavior Profiling*
28. Kerner, H J, Weitekamp E G M, Stelly, W and Thomas J (1997). *Patterns of criminality and alcohol abuse: Results of the Tuebingen Criminal Behaviour Development Study*, Criminal Behaviour and Mental Health.
29. Leul, W. (2003). *Application of data mining in crime prevention : A Case of Oromia Police commission* Unpublished Master's Thesis at the Addis Ababa University, Addis Ababa
30. Maquire, B, & Radosh, F. (1999), *Introduction to Criminology*, West/Wadsworth, Belmont,
31. Mesfin, G. (1999): *A Computer Based Offender Records management System Case study of the Addis Ababa Police*: Unpublished Master's Thesis at the Addis Ababa University, Addis Ababa.
32. Microsoft Research (1998).Tutorial Learning Bayesian Networks from Data. AAAI series
33. Naude, C M B, Prinsloo, J H and Martins, J H (1999), *Crimes against the South African business sector*, Unpublished report, Unisa, Pretoria.
34. Naude, C M B, Prinsloo, J H and Snyman H F. (1999). *The International Crime Victim Survey in Windhoek Namibia*. Unpublished research report, Institute for Criminological Sciences (Unisa), Pretoria.
35. Nelken, D. (1994). *The Futures of Criminology*, Sage, London

36. Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*, Morgan Kaufmann.
37. Pelsier, A and C De Kock, "Violence in South Africa: A note on some trends in the 1990s", *Acta Criminologica*, Vol 13, No 1, 2000, p 86.
38. Prinsloo, J H and C M B Naude(2000). *The African Renaissance--A criminological perspective*. *Acta Criminologica*
39. Prinsloo, J., (199)hais also a Research Fellow at the School of criminal Justice, Grand Valley State University, Michigan, USA
40. Rahel, B. (2005): *Computer-Assisted Learner Group Formation Based on Personality Traits. Computer*. Dissertationsschrift zur Erlangung des Grades eines Doktors der Naturwissenschaften am Fachbereich Informatik der Universität Hamburg
41. Rocha-Silva, L and I Stahmer. (1997). "*Dangerous liaison. Alcohol, drugs and crime*", *Crime and conflict*, Vol 9 (Winter), p 1.
42. Russell, S. and Norvig, P. (1995). *Artificial Intelligence: A Modern Approach*. Prentice-Hall
43. Susan, D.K. (2004). *Developing an Information System for a Police Service in Developing Country: The Case of Zambia Police service*. Zambia
44. Svensson, R. (2002). *Strategic offences in the criminal career context*. *The British Journal of Criminology*
45. Norsys Software Corp.(2007). *Tutorial on Bayesian Networks with Netica:*

APPENDICES

Appendix A

The Predicted Accuracies for Each of the Test Datasets Using TPDA-I

(i) Partition 1

		Predicted			Accuracy
		Medium	Serious	Very Serious	
Actual	Medium	464	49	75	78.91
	Serious	73	319	56	71.21
	Very Serious	65	76	267	65.44

(ii) Partition 2

		Predicted			Accuracy
		Medium	Serious	Very Serious	
Actual	Medium	463	48	72	79.42
	Serious	67	334	49	74.22
	Very Serious	59	78	274	66.67

(iii) partition 3

		Predicted			Accuracy
		Medium	Serious	Very Serious	
Actual	Medium	462	51	77	78.31
	Serious	61	329	49	74.94
	Very Serious	65	81	269	64.82

(iv) partition 4

		Predicted			Accuracy
		Medium	Serious	Very Serious	
Actual	Medium	465	51	75	78.68
	Serious	61	332	51	74.77
	Very Serious	64	76	269	65.77

(v) partition 5

		Predicted			Accuracy
		Medium	Serious	Very Serious	
Actual	Medium	453	54	82	76.91
	Serious	71	324	53	72.32
	Very Serious	63	83	261	64.13

(vi) Partition 6

		Predicted			Accuracy
		Medium	Serious	Very Serious	
Actual	Medium	471	47	59	81.63
	Serious	61	331	53	74.38
	Very Serious	59	80	283	67.06

(vii) Partition 7

		Predicted			Accuracy
		Medium	Serious	Very Serious	
Actual	Medium	469	51	77	78.56
	Serious	61	327	49	74.83
	Very Serious	65	73	272	66.34

(viii) Partition 8

		Predicted			Accuracy
		Medium	Serious	Very Serious	
Actual	Medium	443	57	71	77.58
	Serious	73	317	53	71.56
	Very Serious	87	80	265	61.34

(vii) Partition 9

		Predicted			Accuracy
		Medium	Serious	Very Serious	
Actual	Medium	473	53	73	78.96
	Serious	56	337	37	78.37
	Very Serious	69	75	271	65.30

(vii) Partition 10

		Predicted			Accuracy
		Medium	Serious	Very Serious	
Actual	Medium	451	55	87	76.05
	Serious	69	322	53	72.52
	Very Serious	63	83	261	64.13

APPENDIX B

THE PREDICTED ACCURACIES FOR EACH OF THE TEST DATASETS USING TPDA-II

(i) Partition 1

		Predicted			Accuracy
		Medium	Serious	Very Serious	
Actual	Medium	476	43	69	80.95
	Serious	67	329	50	73.77
	Very Serious	59	70	281	68.54

(ii) Partition 2

		Predicted			Accuracy
		Medium	Serious	Very Serious	
Actual	Medium	469	45	63	81.28
	Serious	61	347	47	76.26
	Very Serious	53	72	287	69.66

(iii) Partition 3

		Predicted			Accuracy
		Medium	Serious	Very Serious	

Actual	Medium	479	41	67	81.60
	Serious	55	339	43	77.57
	Very Serious	63	69	288	68.57

(iv) Partition 4

		Predicted			
		Medium	Serious	Very Serious	Accuracy
Actual	Medium	485	42	64	82.06
	Serious	51	349	46	78.25
	Very Serious	57	61	289	71.01

(iv) Partition 5

		Predicted			
		Medium	Serious	Very Serious	Accuracy
Actual	Medium	471	45	71	80.24
	Serious	62	337	48	75.39
	Very Serious	59	74	277	67.56

(v)

Partition 6

		Predicted			Accuracy
		Medium	Serious	Very Serious	
Actual	Medium	477	44	53	83.10
	Serious	58	348	50	76.32
	Very Serious	56	71	289	69.47

(vi) Partition 7

		Predicted			Accuracy
		Medium	Serious	Very Serious	
Actual	Medium	487	45	69	81.03
	Serious	55	339	46	77.05
	Very Serious	59	65	279	69.23

(vii) Partition 8

		Predicted			Accuracy
		Medium	Serious	Very Serious	
Actual	Medium	455	51	65	79.68
	Serious	70	340	47	74.40
	Very Serious	79	73	274	64.32

(viii) Partition 9

		Predicted			Accuracy
		Medium	Serious	Very Serious	
Actual	Medium	479	47	67	80.78
	Serious	51	346	34	80.28
	Very Serious	63	72	285	67.86

(ix) Partition 10

		Predicted			Accuracy
		Medium	Serious	Very Serious	
Actual	Medium	467	46	74	79.56
	Serious	60	338	47	75.96
	Very Serious	57	74	281	68.20

DECLARATION

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

Mohammed Abrar

January 2009

The thesis has been submitted for the examination my approval as a university advisor.

Dr. Rahel Bekele

January 2009