

ADDIS ABABA UNIVERSITY
SCHOOL OF GRADUATE STUDIES
COLLEGE OF NATURAL SCIENCES
SCHOOL OF INFORMATION SCIENCE

PREDICTING ANTENATAL CARE FOLLOW UP OF WOMEN DURING PREGNANCY IN ETHIOPIA USING DATA MINING TECHNOLOGY

BY:

Teklehaimanot Assefa

OCTOBER 2015

ADDIS ABABA, ETHIOPIA

ADDIS ABABA UNIVERSITY
COLLEGE OF GRADUATE STUDIES
COLLEGE OF NATURAL SCIENCES
SCHOOL OF INFORMATION SCIENCE

PREDICTING ANTENATAL CARE FOLLOW UP OF WOMEN DURING PREGNANCY IN ETHIOPIA USING DATA MINING TECHNOLOGY

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

BY:

Teklehaimanot Assefa

ADVISOR:

Tibebe Beshah (PhD)

OCTOBER 2015

ADDIS ABABA UNIVERSITY
SCHOOL OF GRADUATE STUDIES
COLLEGE OF NATURAL SCIENCES
SCHOOL OF INFORMATION SCIENCE

PREDICTING ANTENATAL CARE FOLLOW UP OF WOMEN DURING PREGNANCY IN ETHIOPIA USING DATA MINING TECHNOLOGY

BY

Teklehaimanot Assefa

Name and Signature of Members of the Examining Board:

Name	Title	Signature	Date
1. Tibebe Beshah (PhD)	Advisor	_____	October, 2015
2. Million Meshesha (PhD)	Examiner	_____	October, 2015
3. Ato Getachew Gemanhe	Examiner	_____	October, 2015

Table of Contents

DECLARATION iv

DEDICATION v

ACKNOWLEDGEMENT vi

LIST OF TABLES vii

LIST OF FIGURES viii

LIST OF ABBREVIATIONS ix

ABSTRACT x

CHAPTER ONE 1

INTRODUCTION 1

1.1. Background of the study 1

1.2. Statement of the Problem 3

1.3. Objective of the Research 5

1.3.1. General Objective 5

1.3.2. Specific Objectives 5

1.4. Scope and limitation of the study 5

1.4. Methodology 6

1.4.1. Research Design 6

1.4.2. Understanding the Problem Domain 7

1.4.3. Understanding the Data 8

1.4.4. Data Preparation 8

1.4.5. Experimentations and Model Building 8

1.4.6. Evaluation of the Model 10

1.5. Significance of the Study 11

1.6. Thesis Organization 11

CHAPTER TWO 12

LITERATURE REVIEW 12

2.1. Overview of Data Mining technology 12

2.2. Data mining process models 15
2.3. Classification of data mining systems 19
2.3. Data Mining Tasks 19
2.3.1. Predictive modeling 19
2.3.2. Descriptive task 20
2.3.2.1. Clustering 20
2.4. Data mining Algorithms 21
2.4.1. Naïve Bayesian Classifier 23
2.4.2. Neural Network 24
2.4.3. Decision Tree 27
2.5. Classification Based on Association Rule Mining 33
2.5.1. Genetic Algorithms 33
2.5.2. AdaBoost 33
2.6. Application of Data Mining in Health Care 34
2.7. Related works 36
CHAPTER THREE 41
BUSINESS UNDERSTANDING, AND DATA PREPROCESSING 41
3.1. Business Understanding 41
3.1.1. Background business objective 41
3.1.2. Factors Affecting Antenatal Care Attendance 43
3.1.3. Business Success Criteria 45
3.1.4. Selecting the Target Dataset 45
3.2. Data Understanding 46
3.2.1. Data Source 46
3.2.2. Dataset exploration 46
3.2.3. Dataset description 47
3.3. Preprocessing 48
3.3.1. Data Transformation 50
3.3.2. Data format conversion 50
3.4. Performance investigation 51
3.4.1. Confusion matrix 52
3.4.2. Precision and recall 52

3.4.3. K-fold cross validation 54
CHAPTER FOUR 56
EXPERIMENTATION 56
4.1. Experimental Design 56
4.2. Feature Selection 57
4.3. Model Building Using J48 58
4.4. Model Building Using Naïve Bayes Classifier 63
4.5. Multilayer Perceptron 66
4.6. Discussion 68
4.6.1. Effect of attribute selection on j48, naïve Bayes, and neural network 68
4.6.2. Effect of pruning on J48 algorithm 69
4.6.3. Effect of test split and cross validation folds on J48, naïve Bayes and multi-layer perceptron 69
4.7. Evaluation and comparison of data mining algorithms 70
CHAPTER FIVE 75
5.1. Conclusion 75
5.2. Recommendations 77
References 78

DECLARATION

I, the undersigned, declared that this thesis work is my original work, has not been presented for a degree in this or any other universities, and all sources of materials used for this thesis work have been duly acknowledged.

Name : _____

Signature: _____

Date : _____

This thesis has been submitted for examination with my approval as a university advisor.

Name : _____

Signature: _____

Date : _____

DEDICATION

THIS THESIS IS DEDICATED TO MY FAMILY!

ACKNOWLEDGEMENT

My greatest gratitude goes to the Almighty God who blessed me with His abundant love and care and gave me the strength to accomplish this thesis and study program.

I would like to express my deepest gratitude to my advisor Dr. Tibebe Beshah for his commitment, constructive comments and suggestion throughout the thesis work. This study would not have been complete without his patience, invaluable advice and commitment.

I would like to thank my sister Selamawit Aseffa for her patience, care and encouragement while I was busy with my Thesis.

I would like to express my sincere gratitude to the following staffs at the School of Information Science for their unreserved support in many aspects whenever I required their professional assistance: Dr. Million Meshesha, and Tsegaye Berhanu, and to my friends Dr. Elias K. and Dr. Girma T.

Finally, I would like to thank my lovely family for their patience and encouragement.

LIST OF TABLES

Table 3.1: Attribute description…… 49
Table 3.2: Attribute with distinct values…….. 50
Table 3.3: Attribute with missing values…….. 50
Table 4.1: Comparison of unpruned J48 on attribute selection ……..….. 61
Table 4.2: Confusion matrix of unprunned J48…….……...…...……. 61
Table 4.3: Comparison of pruned J48 on attribute selection……..….. 63
Table 4.4: Output comparison on different folds and different test split percentage of pruned J48……...……. 64
Table 4.5: Output comparison of naïve Bayes classifier on all attributes and selected attribute……..……... 66
Table 4.6: Confusion matrices for Bayesian algorithm run with all attribute…….……. 66
Table 4.7:Output of naïve Bayes classifier on different folds and test split percentage…….. 67
Table 4.8: Multilayer perceptron output on different folds and test split percentage…….. 69
Table 4.9:Performance comparison of J48, naïve Bayes and multilayer perceptron……... 72

LIST OF FIGURES

Figure 2.1: KDD process…… 17
Figure 2.2: SEMMA steps…… 18
Figure 2.3: Hybrid model…… 19
Figure 2.4: How the artificial neural network works…… 26
Figure 2.5: Three layer neural network…… 28
Figure 2.6: Decision tree…… 31
Figure 3.1: Effect of SMOTE…… 51
Figure 3.3: Confusion matrix…… 54
Figure 3.4: Area under Roc Curve…… 56
Figure 4.1: Flow of experiments…… 59
Figure 4.2: Output of WEKA after unsupervised attribute filter…… 60
Figure 4.2: Output of unpruned J48 algorithm.....
Figure 4.3: Output of naïve Bayes Classifier…… 65
Figure 4.4: Output of neural network…… 68
Figure 4.5: Error rate of each fold for j48 algorithm…… 75
Figure 4.6: Error rate of each fold for naïve Bayesian algorithm…… 75

LIST OF ABBREVIATIONS

ADE	Adverse Drug Effect
ARFF	Attribute Relation File Format
CBA	Class Based Association
CLS	Concept Learning System
CRISP	Cross-Industry Standard Process
CSA	Central Statistics Agency
CSV	Comma Separated Variables
DMAIC	Define Measure Analyze Improve Control
EDHS	Ethiopia Demographic Health Survey
FN	False Negative
FP	False Positive
GIS	Geographic Information System
ID3	Iterative Dichotomiser 3
KDD	Knowledge Discovery in Databases
MDG	Millennium Development Goal
MOH	Minstry of Health
MMR	Maternal Mortality Ratio
ROC	Receiver Operating Characteristic
SMOTE	Synthetic Minority Oversampling Technique
STI	Sexually Transmitted Infections
TP	True Positive Rate
UNFPA	United Nations Population Fund
WEKA	Waikato Environment Knowledge Analysis
WHO	World Health Organization

ABSTRACT

The purpose of this research is to show applicability of data mining techniques in EDHS 2011 data set and predict factors affecting antenatal care by using data mining techniques. In this research J48 algorithm, naïve Bayes algorithm and neural network were used in different percentage split and cross validation folds test environment. WEKA toolkit is used to build the models and performance of the classification algorithms on the dataset was measured by precision, recall, error rate and ROC curves.

Despite low utilization of health care services, there is a considerable variation across different demographic and socio-economic variables. The explanation of this diversity may be complex if we handled it traditionally. Hence there is a need to apply data mining technology to extract hidden patterns and regularities.

The experiments have been conducted following the Cios et al. data mining process model. After understanding the Problem domain and data, data preprocessing were done to handle missing values, to handle outliers, select relevant attributes and to smooth unbalanced EDHS 2011 dataset. The model developed by using pruned J48 with selected attribute on 80 percentage split showed highest classification accuracy. Naïve Bayes with selected attribute showed lowest time complexity and multilayer perceptron showed the highest time complexity. Despite low utilization of health care services, there is considerable variation across different demographic and socio-economic variables. If residence is rural having education and at least once a week frequency of listening radio then yes to antenatal care. If residence is urban from Amhara region with no education as well as no frequency of watching television and listening to radio, then no to antenatal care. Finally recommendation is given to develop a knowledge base that helps health center to use the knowledge extracted using data mining.

CHAPTER ONE

INTRODUCTION

1.1. Background of the study

One of the major goals of the Ethiopian government in its five years achievements of the millennium is to minimize maternity and child death during birth. Researches done in the computing area tried to address this issue from different perspectives and angles. This research, tried to address the issue from a data mining point of view in order to identify the major factors that contribute to the aforementioned problems.

According to World Health Organization (WHO) [1] estimation, more than 500,000 mothers die each year because of pregnancy and related complications. Antenatal care is given to pregnant women to avoid this complication for the well-being of mother and child. Antenatal care is one of the most effective health interventions for preventing maternal morbidity and mortality particularly in places where the general health status of the women is poor [1].

According to WHO [8] only about 34 percent of births in Ethiopia are delivered by skilled attendants, and more than 10 percent of children die before their fifth birthday. While mortality rates have improved in recent years, there is still much to be done to save mother's and children's lives. Ethiopia's rates of maternal and newborn morbidity and mortality are among the highest in the world. Current estimates of maternal mortality stand at 871 deaths per 100,000 live births, or 25,000 maternal deaths per year Who Health Organization (WHO). But the government set to achieve the millennium development goal, and one of the criteria in the millennium goal is to reduce child mortality and to improve maternal health which will be achieved by proper antenatal care during pregnancy and early birth.

As mentioned in the Ethiopia society of population studies [2], finding the importance of maternal health care services in reducing maternal mortality and morbidity has received a significant recognition. Implementing and assuring utilization of effective maternity care for women in the developing world is not an easy task. In Ethiopia, like in other developing countries, most childbearing women are poor and live under harsh conditions.

Data mining is the analysis of observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner. The relationships and summaries derived through a data mining exercise are often referred to as models or patterns [3].

Data mining has been applied in different fields such as fraud detection in banking and insurance industry, customer behavior in e-commerce, and other fields. The popularity in the above fields initiate other fields for the implementation of data mining application. However, there are still misconceptions going on in the medical community about what data mining comprises. The generally accepted definition of data mining today is the set of procedures and techniques for discovering and describing patterns and trends in data [4]. Researchers believe that data mining plays major role in health sector for decision making, identification of better treatments and identification of at risk patient and society for intervention.

Data mining and knowledge discovery starts to appear in the early 1990s. There are algorithms that were capable of solving all the problem of searching for knowledge and tools which simplify the application of data mining tasks [1]. The current availability of advanced communication techniques has increased the availability of large collection digital data from the domain including health care, e-business and e-learning. But finding useful information from such a large data and extracting compact knowledge is a challenging issue [2]. Explosive growth and wider availability of data makes this time the data age. Powerful tools and multipurpose tools are badly needed to automatically uncover valuable information [7]. Data mining can be applied to any kind of data base as long as the data are meaning full for a target application. Data mining can be applied on database, data warehouse and transactional data. Data mining has several functionality. These include characterization and discrimination, associations, and correlations, classification and regression, clustering analysis, and outlier analysis [7].

Knowledge Management in healthcare can be regarded as the use of formal methodologies and techniques to facilitate the creation, identification, acquisition, development, preservation, dissemination and finally the utilization of the various facets of a healthcare enterprise's knowledge assets. The health care industry has evolved into an extended enterprise-an enterprise that is powered by sophisticated knowledge and information resources. Healthcare enterprises

can be regarded as ‘data rich’ as they generate massive amounts of data, such as electronic medical records, clinical trial data, hospitals records, administrative reports, benchmarking findings and so on. But, in the same breath it can be said that healthcare facilities are ‘knowledge poor’ because the healthcare data is rarely transformed into a strategic decision-support resource. For that matter, with the emergence of technologies such as knowledge management (KM) and Data Mining (DM) there now exist opportunities to facilitate the migration of raw empirical data to the kind of empirical knowledge that can provide a window on the internal dynamics of the healthcare enterprise [38].

1.2. Statement of the Problem

Antenatal care is the care given to woman during pregnancy. Antenatal period presents an important opportunity for identifying threats to the mother and unborn baby’s health, as well as for counseling on nutrition, birth preparedness, delivery care and family planning options after the birth [1].

According to the bench mark of WHO [8], antenatal care for the majority of normal pregnancies should consist of four visits. The first semester lasts for 12-16 weeks of pregnancy, the second semester stays for 24-28 weeks of pregnancy, the tri semester on 32 weeks of pregnancy, and the last semester on 36 weeks of pregnancy. This includes nutritional support, prevention and treatment of tuberculosis, treatment of sexually transmitted infections, prevention of HIV transmission from mother to child, and tetanus toxoid immunization.

Complications of pregnancy and childbirth are the leading causes of disability and death among women in the reproductive age in developing countries. The World Health Organization (WHO) estimates that about 529,000 women die worldwide every year in connection with pregnancy and childbirth. Nearly all (99 %) maternal, newborn, and child deaths occur in low and middle income countries [1]. In Ethiopia, the levels of maternal and infant mortality and morbidity are among the highest in the world [6].

Millennium Development Goal 4 aimed to reduce by two thirds, between 1990 and 2020, the mortality rate of children under five and also, Millennium Development Goal 5 planed that to

Reduce by three quarters between 1990 and 2020, the maternal mortality ratio; however, the target reduction of 75% between 1990 and 2020, which requires a 5.5 % annual decline rate, but this cannot be met in Sub-Saharan Africa (MDG report, 2009). To achieve the aforementioned development goal new technology supported antenatal care plays a crucial role as it works on mother and children.

Despite low utilization of health care services, there is considerable variation across different demographic and socio-economic variables. The explanation of this diversity may be complex if it is handled traditionally. Utilization of maternal health care services is affected by a multitude of factors (UNFPA). Traditional methods like statistics which are used in previous studies had shortcomings like conservative analysis strategies, when we compare with data mining which is more flexible about which methods are to be used in which order to mine data [10]. It is also difficult to handle variables other than numerical values in traditional method which is simple in data mining approach.

There is a wealth of knowledge to be gained from computerized health records [9]. Processing the demographic survey to exploit this wealth of knowledge will be difficult by traditional method compared to data mining. Data mining is best in doing on sparse, large and complex data to provide knowledge that facilitate the improvement of antenatal care in Ethiopia by applying sophisticated methods and techniques which are found in data mining approach.

The main factor that initiates to apply data mining in health sector is the highest maternal mortality rate and the lowest child healthcare status in sub-Saharan Africa, data overload and weakness of traditional statistical method to extract knowledge from huge amount of data. Huge amounts of data generated by healthcare transactions are too complex and voluminous to be processed and analyzed by traditional methods. Traditional method with sampling few thousand records was implemented in the early days because, decades ago when most computer systems or manual calculations could not handle an entire population data [10].

Data mining can improve decision-making by discovering patterns and trends in large amounts of complex data [5]. Predictive models are used to determine the future behavior of some entity.

These models are important for decision support because they provide information about processes and entities that are outside of the decision maker's control.

The aim of the study is to explore factors affecting antenatal care in Ethiopia with the help of data mining technology. To this end, enlisted below are the research questions that this study attempt to address:

- ✓ How one can represent the Ethiopian EDHS suitable for the data mining application in identifying the factors affecting ANC in Ethiopia?
- ✓ Which data mining algorithms implemented in health sector give the most accurate results in supporting antenatal care in EDHS 2011 dataset?
- ✓ What are the issues and challenges of using data mining technology in health sector especially in Ethiopia demographic health data for predicting the factor affects antenatal care?
- ✓ What are the factors affecting antenatal care in Ethiopia? Is there a regular pattern on the variables of the EDHS?)

1.3. Objective of the Research

1.3.1. General Objective

The general objective the research is identifying factors that predict antenatal care follow up of women during pregnancy in Ethiopia using data mining technology.

1.3.2. Specific Objectives

Specific objectives of this research are

- To understand the level of the problem from literature review
- To do the preprocessing task on the demographic data, Data Cleaning, Data Reduction, Data Transformation.
- To conduct an experiment and build model.
- To evaluate the accuracy of the model with different techniques.
- Report the analysis result and recommend for future work.

1.4. Scope and limitation of the study

The scope of this research is to investigate the application of data mining techniques for predicting the regularities of factor affecting antenatal care in Ethiopia on Ethiopia demographic health survey data set. In this research naïve Bayes, J48 and neural network were used. The proposed model was used to improve the health sector in different perspectives, such as to help for policy makers, health extension workers, and other parties involved in the health sector to make appropriate decision.

This research was aimed to include all-important information for solving the study problem. As described earlier this research only attempted to apply three DM techniques in predicting factors affecting antenatal care (J48, naïve Bayes, and neural network). The research was conducted on whether the women followed or not followed antennal care, as a two class problem. However, researches can also be conducted on the number of visits and the factors which affect each visit. Although this research was aimed to include large number of women throughout the country, the data used in this study was only 7751 women instance. Due to the number of women instance who have a recent birth five years preceding the survey was only 7513.

1.4. Methodology

It is necessary to have a methodology to guide a project, to improve understanding of the project as a phased approach, and to ensure that sufficient checks and phases not be overlooked. Research design and the process to be followed are the important steps to give emphasis in data mining researches. In this research, CRISP data mining process model was implemented because the model provides anyone from novices to data mining experts with a complete blueprint for conducting a data mining project [9].

The study was done based on the EDHS data. The data includes all women in the range of reproductive (15-49) years old, this range is the WHO accepted range of fertility for women.

1.4.1. Research Design

There are different types of data mining models such as CRISP, SEMMA and DMICA. SEMMA (sample, explore, modify, model, and assessing) is one of most widely used models in

data mining. It is a five staged process model. The first step is sampling, the second stage is exploration of the data, the third stage is modification the data by creating, selecting and transforming the variable, the fourth step is modeling by allowing software to search automatically for combination of data and the last step is assessing the data by evaluating the usefulness.

CRISP is the other data mining process model with a six stage process. The first stage is business understanding focused on understanding the objective and define requirements, the second step is data understanding it starts with data collection, the third stage is data preparation, the fourth stage is modeling in this stage various modeling techniques are selected and applied, the fifth stage is evaluation of the model and the last stage is model deployment. But most of the time SEMMA and SAS enterprise miner are closely related. CRISP is comprehensive data mining methodology and process model that provides anyone from novices to data mining experts with a complete blueprint for conducting a data mining project, and the process model is also independent of both the industry sector and the technology used [9] [12].

The development of academic and industrial models led to hybrid models. One of the hybrid models is a six step KDP model developed by CIOS *et al*. The main differences against CRISP data mining process model is providing more general, research-oriented description of the steps, introducing a data mining step instead of the modeling step. For this research Cios *et al*. process model for Data Mining was used.

1.4.2. Understanding the Problem Domain

This step in Cios *et al*. process data mining model is the most important step in conducting data mining research. It focuses on understanding the problem domain and finally identifying the problem. In this research, discussions have been conducted with a domain expert, and related document such as; institutional delivery and skilled birth attendant utilization in Saharti Samre district by Yalem, utilization of antenatal care services among teenagers in Ethiopia by Tewodros, has been reviewed to have general overview of the business domain and the data to be analyzed.

Literatures were reviewed to gain an overall understanding of antenatal care, level of antenatal care utilization in Ethiopia, applications of data mining, importance and utilization of data

mining application in health care. Different literatures were also reviewed to investigate the approach of other researchers to find their gap and recommendations of related studies. Application of data mining in the health care and its effect in other countries were also reviewed from related literature.

1.4.3. Understanding the Data

To understand the general property of the data, sample data were taken from the dataset and discussed with domain expert to have general understanding of the data. In Cios *et al*. process model, there are guidelines to have the data understanding, collect the initial data, describe the data, explore the data, and verify data quality. In describing the data, examining issues such as the format of the data, the quantity of the data, the number of records and fields in each table, and the identities of the field were reviewed. In verifying the data, examining issues such as the format of the data, the quantity of the data, the number of records and fields in each table, and the identities of the field were conducted [9].

1.4.4. Data Preparation

In today's data-rich environment, decision makers draw conclusions from data repositories that may contain data quality problems [12]. One of the major problems in EDHS data is missing values. Since it can seriously affect the accuracy of conclusions drawn, researchers have described several approaches for dealing with missing values, such as fill the missing values with the mean value, fill the missing values with the mode value, and fill the missing values with the constant value. WEKA uses the first method [11].

Data preparation may include different iterative processes like data filtering and smoothing. The main task of Smoothing is to get rid of noises and outliers. To smooth a dataset there are several techniques to be incorporated [4]; moving average, median filtering, normalization or standardization, and fixing missing and empty values were the techniques used for data filtering. In this study WEKA used oversampling the minority class to balance the dataset.

1.4.5. Model Building

After the data were preprocessed and transformed, a model is built, by using the aforementioned data mining tool to come up with a representative and expressive data model. This model was generated with one of the data mining techniques, which was identified as most suitable. Identification of the data mining technology to be used for model generation was one of the major tasks that this research followed.

The modeling stage is the actual data analysis stage. Most data mining software includes online analytical processing and traditional statistical methods. Traditional statistical method cluster analysis, discriminate analysis and regression analysis. Non-traditional statistical analysis are neural networks, decision trees, link analysis and association analysis [1]. Generally, in this stage the Modeling Technique were selected. This task refers to choosing one or more specific modeling techniques, such as decision tree building with C4.5 or neural network generation with back propagation and generate Test Design after building a model. The model's quality and validity should be tested to run and determine the strength of the model.

WEKA is one such tool and platform that is freely available and is used to accomplish all aforementioned tasks. In this research, WEKA is used because nowadays, WEKA is recognized as innovative system in data mining [63].

Data mining tool selection is done after the definition of the problem. Selection of the appropriate data mining tool and techniques depend on the main task of the data mining process. In data mining there are tools for model building.

- WEKA tool kit is a widely used tool kit for data mining. WEKA contains large collection of learning algorithms written in JAVA. The author of this tool is University of Waikato in New Zealand. The programming language of WEKA is Java. Experimenter and Explorer are two mainly used interfaces during this experiment. One way of using WEKA is to apply a learning method to a dataset and analyze its output to learn more about the data on preprocessing stage.
- Tanagra is another open source data mining tool for academic and research purpose it is used for data exploration data analysis and statistical learning. Most of the time Tanagra is used by researcher to write data mining algorithm.
- KNIME is another open source data mining integration processing, analysis and exploration platform.

- Orange is also another open source data mining tool intended for both experienced user and researcher.
- Raid miner is open source data mining tool, use Java programming language and offers an integrating environment with visually appealing and user-friendly GUI.
- R miner was developed by worldwide development and use C, FORTRAN and R programming language, R's main problem is its language.

In this research WEKA, data mining tool is used to classify instance and identify pattern because WEKA tool kit achieved the best performance in percentage split test and cross validation test mode.

- WEKA is a set of state-of-the-art data mining algorithms and tools to in-depth analyses.
- It is used to train learned models to generate predictions and classification pattern on data set.
- WEKA is used to apply several different learners and compare their performance in order to choose one for prediction.
- WEKA allows analyzing the data sets saved in the *.ARFF, CSV, and c4.5* files which can be easily achieved by converting txt or .dta file. Knowledge explorer is also the other interface in WEKA which helps to build model [63].

1.4.6. Evaluation of the Model

After data mining algorithm build models, the model was tested by test data set with corresponding error rate, precision recall f-measure and area under ROC curves. Evaluation matrix measures true positive rate, true negative rate and accuracy of the model [10].

To evaluate the model, an assessment results in terms of business success criteria was summarized based on final statement about whether the project already meets the initial business objectives. The evaluation stage enabled the comparison of models and results from any data mining model by using a common benchmark [1]. Models that are developed in this research are evaluated based on squared error rate values, absolute error rate values, amount of seconds the algorithm takes to build the model, true positive rate, true negative rate, and experts' judgement.

Additionally, classification models that are developed in this research are evaluated using a test dataset based on their miss classification (false positive rate and false negative rate).

1.5. Significance of the Study

After the completion of this research:

- It will be helpful for decision makers in the health sector for making appropriate decision. When dataset is huge in quantity it helps to drive knowledge that will be useful for correct decision.
- It will be useful for policy planers to give special attention for low antenatal care level areas.
- For early detection and intervention maternal and child mortality.
- It helps for the achievement of the millennium development goal in reducing child mortality and improving maternal health.
- It will increase level and quality of antennal care for Safe pregnancy and delivery service

1.6. Thesis Organization

This thesis is organized into five chapters.

- ✓ The first chapter deals with introduction and overview of the study including background, statement of the problem, and objectives of the research.
- ✓ The second chapter is devoted to conceptual literature review of data mining technology as well as review of related articles of antenatal care respectively.
- ✓ Chapter three explains the research methodologies, algorithm used as well as the WEKA software used in this study.
- ✓ Chapter four presents the experimentation phase of the study. It comprises training, building and validation of the models. Results of the experiment are also analyzed and interpreted.
- ✓ The last chapter is devoted for the final conclusions and recommendations based on the research findings.

CHAPTER TWO

LITERATURE REVIEW

In this chapter, different literatures were reviewed to have an overview of data mining and its applications, techniques that are helpful to find pattern in prediction, challenges in applying data mining in health sector and related work done on application of data mining in health sector.

2.1. Overview of Data Mining technology

Data is raw. It simply exists and has no significance beyond its existence, information is data that has been given meaning, and knowledge is the appropriate collection of information. Cheap computers and storage technologies have dramatically increased the availability of digital data in every sector. Like other sectors data captured in the health sector has increased dramatically, and thus huge amount data are the basic input that administrators and decision makers use to build theories and models or pattern. But the analysis capability to build pattern and model from this available data increased very slowly when we compared with the availability of the data. The development of methods and techniques for making sense of the data is the important issue here. The basic problem addressed by the KDD process is one of mapping low level (raw and bulky data) into compact and readable forms, and data mining is at the core of this process [42].

Data mining is a relatively new concept that emerged in the middle of 1990's as a new approach to data analysis and knowledge discovery. But it starts to appear in health sector late 2000 even the application was seen before in financial institutes, credit and saving fraud detections, in space science, customer relation management in marketing and many other fields. But in other literature Florence Nightingale invented polar-area diagrams in 1855 to show that many army deaths could be traced to unsanitary clinical practices and were therefore preventable. She used the diagrams to convince policy-makers to implement reforms that eventually reduced the number of deaths Nightingale were able to personally collect, sift through and analyze the mortality data during their times because the volume of information was manageable. Today, the

size of the population, the amount of electronic data gathered, along with globalization and the speed of disease outbreaks make it almost impossible to accomplish what the pioneers did [20].

Data mining is defined as the process of discovering meaningful new correlations, patterns and trends by filtering large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques [23]. The discovered knowledge can be used by the healthcare administrators to improve the quality of service.

Data mining is also defined as the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner. Observational means historical not experimental in other way; objectives of the data mining exercise play no role in the data collection strategy, data mining uses data that have already been collected for some purpose. This is one way in which data mining differs from much of statistics [3]. For administrative decision making digested model and information from large complex data, help the expert for easy and precise decision.

Many authors use data mining and knowledge discovery interchangeably, but data mining is one step in knowledge discovery process, data mining adapt different techniques from machine learning, pattern recognition, statistics, and artificial intelligence [13].

Data mining is an interdisciplinary exercise. Statistics, database technology, machine learning, pattern recognition, artificial intelligence, and visualization, all play a role in data mining process. It is difficult to define sharp boundaries between these disciplines, it also is difficult to define sharp boundaries between each of them and data mining [61], statistics in particular has much in common with KDD, discovering knowledge from data is also statistical endeavor but one can find pattern that appear statistically important but not this issue properly raised in KDD. KDD can also be viewed as encompassing a broader view of modeling than statistics.

Data mining can be applied in different data repositories. In principle data mining should be applied to any data base. Relational data base one type data repository which we can apply data mining. Relational database is collection of interrelated data which is known as database.

Relational database is collection of table, each of which assigned a unique name. Data warehouse is another repository where data mining can be applied. A data warehouse is a repository of information collected from multiple sources, stored under a unified schema, and that usually resides at a single site. Data warehouses are constructed via a process of data cleaning, data integration, data transformation, data loading, and periodic data refreshing. We can also apply data mining in transactional data base which is a record consists of a file where each record represent a transaction. There are also different type of database in which we can apply data mining, object– Relational database, temporal database, spatial database and many other repositories [13][11].

Data mining has been applied in many real life applications, in science it has been applied in astronomy to perform image analysis, classification and cataloging of sky objects. In marketing by analyzing customer database identify different customer group and forecast future trend of customer need. In investment companies use stock market and other database to decide when and where to use their capital effectively. In fraud detection also Data mining applications attempt to detect fraud and abuse, often establish norms and then identify unusual or abnormal patterns of claims by physicians, laboratories, clinics, or others financial institutions [13].

In related fields evolving from database is data warehousing, its main task is collecting and cleaning transactional data to make the data available online. This collection and cleaning is also the first step in KDD. OLAP is a tool that is used for analysis of data warehouse online for simplifying supporting interactive data analysis. But KDD has a step beyond the two steps in OLAP [13].

Different types of data mining tools are available in the marketplace, each with their own strengths and weaknesses. Information experts in the organization need to be aware of the different kinds of data mining tools available and recommend the purchase of a tool that matches the organization's current objective. The majority of those tools are available in both windows and UNIX version.

2.2. Data mining process models

Like many other fields data mining need a standard framework to identify business problems and set data mining task. According to Rüdiger data mining is a complex process which needs a proper combination of various data mining tools, techniques, and human expert. The process model will help experts to have a common reference ground and will increase the understanding of complex data mining issue. For new data analyst, the process model provides guidance, helps to structure the project, and gives advice for each task of the process. Even experienced analysts can benefit from check lists for each task to make sure that nothing important has been forgotten (used as a check list). There are different industry standards for data mining process. Cross industry standard process for data mining is one of the standards used in benchmarking data mining projects [12].

The CRISP-DM reference model for data mining provides an overview of the life cycle of a data mining project. It contains the phases of a project, their respective tasks, and their outputs. The first phase in this data mining process model is understanding problem domain which give an overview of project objective and requirement and what the data mining project will bring, in this step a data mining problem will be defined, the second phase is understanding the data, initial data observation in order to get familiar with the data, the third phase is data preparation to have a final data set, the fourth phase is modeling where various modeling techniques are selected and applied, the fifth step is evaluation. Before proceeding to final deployment of the model, it is important to more thoroughly evaluate the model [25].

Some process models include additional phase, according to Fayyad *et al*. [13] some process model add the seventh step which is interpreting mined patterns, possibly returning to any of steps 1 through 6 for further iteration. This step can also involve visualization of the extracted patterns and models or visualization of the data given the extracted models [13].

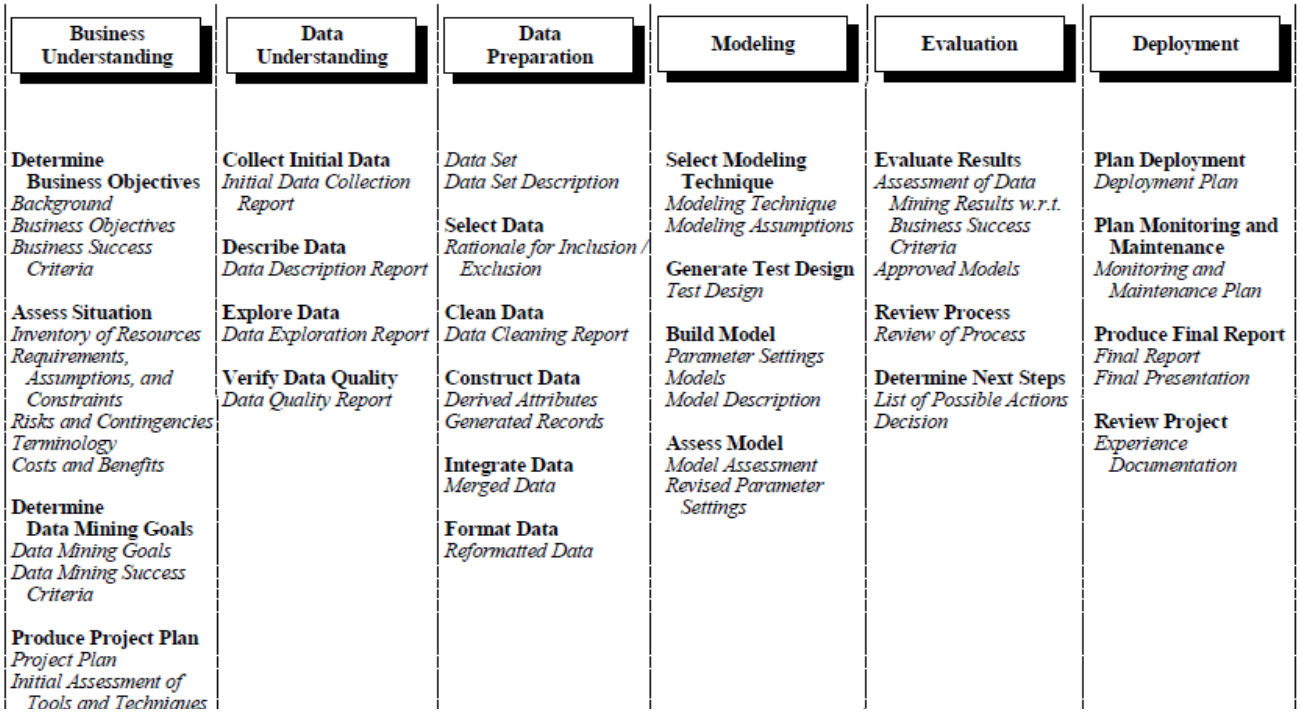


Figure: 2.1 CRISP data mining process

There are also other data mining process model DMAIC (define – measure - analyze – improve – control) is the one which emphasize on eliminating defect and waste, quality controlling problem of all kind in manufacturing and service delivery, management and other business activity. The other framework is SEMMA (simple – explore – modify – model – assess) refers to a methodology that clarifies data mining process. The SEMMA method divides data mining into five stages: (1) sample—to draw a statistically representative sample of data; (2) explore—to apply exploratory and statistical and visualization techniques; (3) modify (or manipulate)—to select and transform the most significant predictive variables; (4) model—to model the variables to predict outcomes; and (5) assess—to confirm a model's accuracy [11].

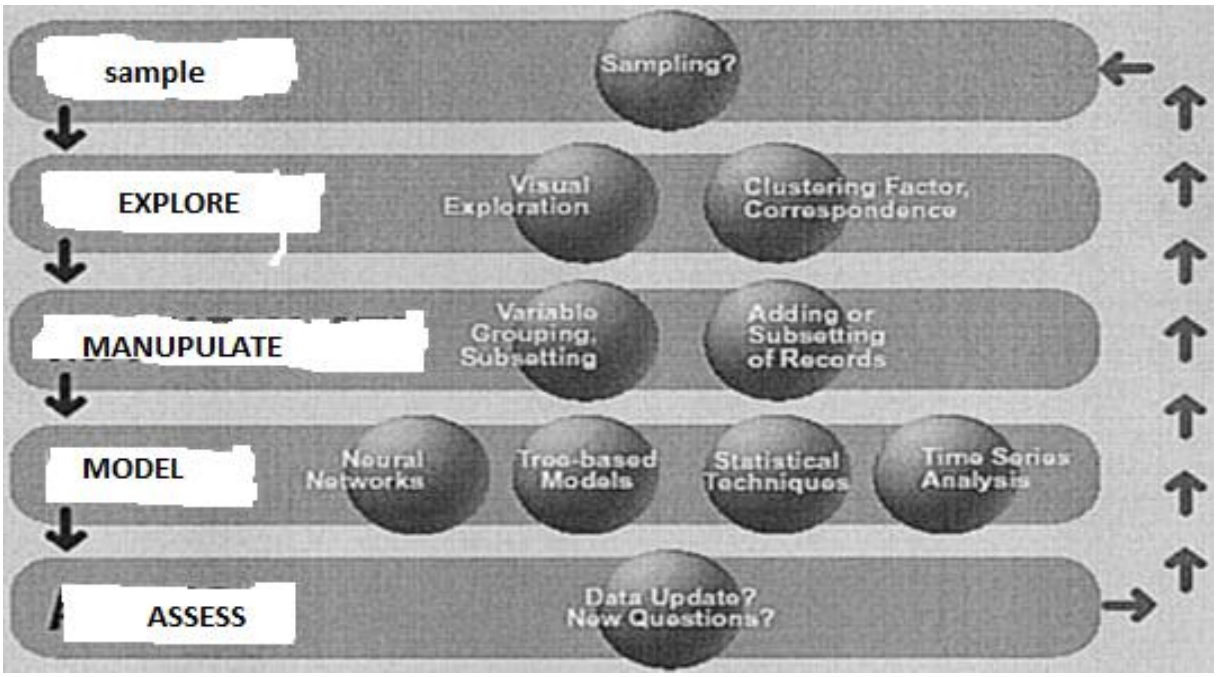


Figure: 2.2 SEMMA steps

The development of academic and industrial models led to hybrid models. One of the hybrid models is a six step KDP model developed by CIOS et al. The main differences against CRISP data mining process model is providing more general, research-oriented description of the steps, introducing a data mining step instead of the modeling step.

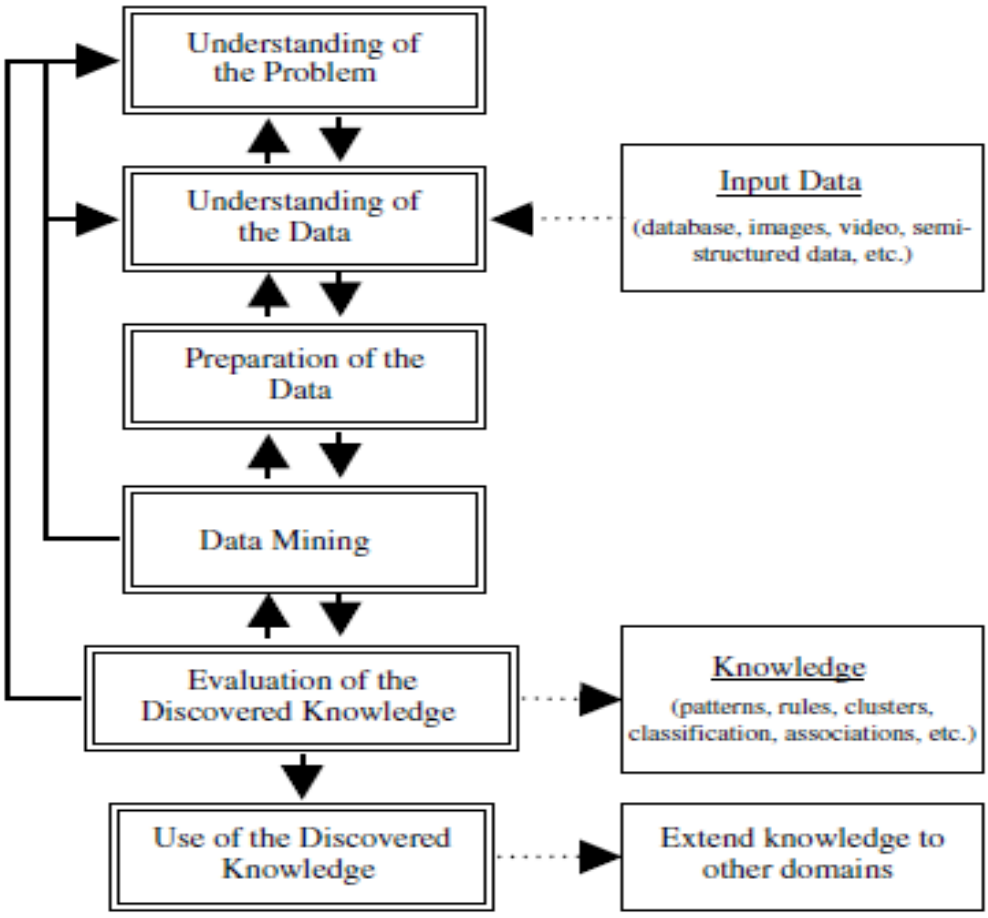


Figure 2.3 Hybrid model

The first step is the understanding of the problem domain, which involves working closely with domain expert to define problem and determine the project goal. The second step is understanding of the data, this step includes collecting sample data and then data are checked for completeness, redundancy, missing values, and plausibility of attribute value. The third step is preparation of the data. This step concerned on deciding which data will be used as input for DM methods in the subsequent step. The fourth step is Data mining. Here the data miner uses various DM methods to derive knowledge from preprocessed data. The sixth step is evaluation of the discovered knowledge. This includes understanding the result, and checking whether the discovered knowledge is novel and interesting. The final step is use of the discovered knowledge, this final step consists of planning where and how to use the discovered knowledge [65].

2.3. Classification of data mining systems

Data mining systems are categorized according to the data type of the data source mined. In an organization a huge amount of data are available, we need to classify the data according to its type (audio/video, text format) and mining also applied on spatial data, multimedia data, and time series data [72][42].

Another classification of data mining system is based on the kind of knowledge discovered. This classification is based on the kind of knowledge discovered or data mining functionalities, such as characterization, discrimination, association, classification, clustering [42]. Clustering is segmenting a diverse group into a number of similar subgroup or clusters. Classification is finding a model (function) that describes and distinguishes class behavior. Characterization and discrimination generalize, summarize and contrast data characteristics [71].

Data mining techniques used can also classify data mining system. Data analysis approaches used such as machine learning, neural networks, genetic algorithms, statistics, visualization, database oriented or data warehouse-oriented [42].

2.3. Data Mining Tasks

The task of data mining is very diverse and distinct because there are many patterns in a large database. Different kinds of method and techniques are needed to find different kinds of pattern. Based on the pattern we are looking for data mining tasks are categorized in summarization, classification, clustering, link analysis and trend analysis [18].

2.3.1. Predictive modeling

The task of data mining can be modeled as predictive or descriptive model. Predictive model is a supervised learning technique that involves analysis of the dependency of some attribute values upon the values of other attributes in the same item, and the development of a model then makes prediction about unknown data values by using the known values, for instance Classification, Regression, Time series analysis, and Prediction [70] [72].

2.3.1.1. Classification

Classification is the derivation of a function or model which determines class of an object based on its attribute. Classification function or model is constructed by analyzing the relation between the attribute and the class of the object in training set. The classification function or model can be used to classify future objects and develop a better understanding of the class of the object in the database [18].

2.3.1.2. Trend Analysis

Trend analysis is discovering interesting pattern in the evolution of the object. A model or a function is constructed to simulate the behavior of the object which can be used to predict the future behavior, for instance we can predict the profit of a company for this year based on the past trend [18].

2.3.2. Descriptive task

Descriptive method can be defined to discover interesting regularities in the data, to uncover patterns and find interesting clusters in the bulk of data. The goal of a descriptive data mining model is to discover patterns in the data and to understand the relationships between attributes represented by the data Ex. Clustering, Summarization, Association rule, link analysis [70] [72].

2.3.2.1. Clustering

Clustering is the identification of class, also called clusters. The objects are clustered so that the intra class similarity or similarity between objects in the same cluster are maximized and inter class similarity or similarity between objects in different class is minimized. Similarity between objects is measured with different mechanism. For example we can cluster objects into several groups based on sex, age, education background and other attribute [18].

2.3.2.2. Summarization

Summarization is the abstraction or generalization of data or avoiding unnecessary detail. A set of task relevant data is summarized and abstracted, resulting simple and general overview of the data and usually with aggregation of information. For example instead pointing out what each

semester of antenatal care constitutes, we can simply point out the number of antenatal care attend. Summarization can go up to different level of abstraction level and can be viewed from different angles [18].

2.3.2.3. Association

Association is the discovery of related object which are coming together frequently or with same terms like yngjiang object togetherness**.** Association rule revealed the association relationship among object. For example in the case of EDHS data education background associated with good knowledge about antenatal care. The association rules can be useful for making commodity management, marketing and advertising [18].

2.3.2.4. Link Analysis

Using link analysis one can examine the associations between objects. Link classification provides category of an object, not just based on its features, but also on connections in which it takes part, and features of objects connected with certain path. Example of link analysis in medicine is task of predicting disease type based on people's characteristics or predicting age of people based on disease they are infected with and based on age of people they have been in contact with. Link analysis can be used in order to understand where patients go to receive the healthcare treatment and to identify the components or resources in service that must be addressed [19].

2.4. Data mining Algorithms

Before applying data mining techniques in any data set, understanding what kind of algorithm are found and how to apply this algorithm is important. In data mining algorithms are classified in to two major categories according to their task, descriptive (unsupervised learning) and predictive (supervised learning) [11].

Descriptive data mining is explanatory in nature. It is unsupervised way of learning, group the data set by maximizing inter group similarity and minimizing intra group similarity. Descriptive data mining help to understand pattern existing on a data set which is not with labels like

supervised learning and containing complex and nonlinear interaction. The most common clustering techniques are K-means and its variant, hierarchical clustering and density based clustering [4] [73]. In descriptive data mining task hierarchical clustering algorithm successively merge the most similar group until the termination condition holds. In partional clustering user requires specifying the number of clusters (K). Partitional algorithm successively relocates the object by measuring the similarity between the object and the centroid, until the number of cluster equal K. for both methods the ultimate goal of the techniques is optimization of inter class similarity and minimization of intra class similarity, which measured based on distance between objects [7].

Supervised learning estimates future value based on the pattern on the current data set. Unlike descriptive data mining techniques classification is supervised way of model building (i.e. the classifier is told to which class the record belongs), supervised data mining use class labels to train and test the model then use the model for unseen data set. Classification and prediction are two forms of data analysis that can be used to extract models describing important data classes or to predict future data trends. Classification predicts categorical (discrete, unordered) labels, prediction models continuous valued functions. For example, we can build a classification model to categorize antenatal care as either properly taken or not properly taken, or a prediction model to predict the expenditures in dollars of potential customers on computer equipment given their income and occupation. Classification is a two-step process, the first step is learning or training step, in this step a classifier is built by learning trends from the data set, classification algorithms identifies rules and build classification rules. The second step is testing. In this step the accuracy of classification rules built in the first step will be evaluated by test data set. The accuracy of a classifier on a given test set is the percentage of test set tuples that are correctly classified by the classifier. Accuracy of a predictor refers to how well a given predictor can guess the value of the predicted attribute for new or previously unseen data. In prediction the model constructed predicts a continuous-valued function, or ordered value, as opposed to a categorical label [7].

The other measure of the model strength include the following, Speed refers to the computational costs involved in generating and using the given classifier or predictor. Robustness is the ability of the classifier or predictor to make correct predictions given noisy data or data with missing

values. Scalability refers to the ability to construct the classifier or predictor efficiently given large amounts of data. Interpretability: This refers to the level of understanding and insight that is provided by the classifier or predictor. Interpretability is subjective and therefore more difficult to assess [3].

2.4.1. Naïve Bayesian Classifier

Given a set of objects, each of which belongs to a known class, and each of which has a known vector of variables, our aim is to construct a rule which will allow us to assign future objects to a class, given only the vectors of variables describing the future objects [16]. Naïve or simple Bayesian classifier is probabilistic statistical classifier, which assumes all attribute of a record in the data set are independent, naïve Bayesian made this assumption to reduce computational complexity. Independency in the data base means any change of an attribute of a record will not affect other attributes of the same record which is not real most of the time. But the assumption made the algorithm easy and ready to handle a data set with many attribute like EDHS. The classifier use small training data set to develop accurate classification rules in the learning step. It requires only the calculation of the frequencies of attributes and attributes outcome pair(s) in the training data set. The major drawback is the assumptions all attribute are conditionally independent. For example assuming educational background and knowledge about antenatal care is independent in EDHS data set unrealistic [11].

In health care data mining is become increasingly popular. There are a lot of factors which initiate application of data mining in health sector, the huge amount of data generated by the health sector are too large and complex to compute. The other motivational factor for applying data mining in health sector is the realization of data mining can generate information and knowledge that is very useful to experts [46].

Naïve bays technique is widely apply to data mining because its simplicity and robustness. Bayesian network is a conditional probability which is the likelihood of something happen given some evidence observation E where a dependence relationship exist between C and E this probability is denoted by p(C/E)[44].

$$p(C\backslash E) = \frac{p(E\backslash c).p(c)}{p(E)}$$

P(C) = prior probability of hypothesis c

P (E) = prior probability of training D

P (E/C) = probability of E given C [4]

$$P(A/B) = \frac{P(A \text{ and } B)}{P(B)}$$

The conditional relationship allows an investigation to gain probability information about either c or E [43]. Given a set of objects each of which has a known vector of variables, the objective is to construct rules which will allow us to assign future object to a class. Highly correlated variables have been eliminated, which means relation between any variables assumed to be independent. Naive Bayes handles missing values naturally as missing at random [16]. The algorithm replaces sparse numerical data with zero. If we chose to manage our data preparation binning method will be selected [44].

2.4.2. Neural Network

Neural network was developed in early 20th century. Neural network was used widely because of the introduction of other algorithms like decision tree vector space machine was not realized. This helped the algorithm to be popular and also applied in health sector as it was similarly applied in other sectors. Neural network works analogously like human brain. It consists of computational nodes which work as human brain, mostly the input and the output nodes appear in neural network, the input layer is used to take input value and pass it to the next layer for processing. Multilayer perceptron with back propagation which is also available in WEKA is widely used in neural network. Neural network properly handle noise data but its learning process is time taking and computationally complex and expensive. Because it requires many parameters including the optimum number of hidden layer nodes that are empirically determined, and its classification performance is very sensitive to the parameters selected. The other shortcomings is it lacks explanatory power or model transparency, which make the model difficult for medical domain expert and decision maker to understand the knowledge found [11].

There is a potentially useful knowledge stored in health data and possible extracted to useful pattern. But achieving this task is impossible without the help data mining. Data mining allows organizations and institutions to get more out of existing data at minimal extra cost [18]. Perceptron in neural network can be trained by adjusting the width of input with supervised learning to have a pattern that will be useful in classification or prediction. Neural network can map pattern to the associated output pattern, can learn by example, and possess the capability to generalize training set of the input value already classified in predefined class. Choosing number of nodes for each layer will depend on problem Neural Network is trying to solve, types of data neural network is dealing with, quality of data and some other parameters. Number of input and output nodes depends on training set in hand [49]. Advantages of neural network includes high tolerance to noisy data, it can be used with having little knowledge of relationship between attribute, well suited for continuous value input and output, and well work on pathology and laboratory data [60].

There are different kinds of learning in neural network. Supervised learning is learning method in neural network which every input pattern that is used to train the network is associated with an output pattern. Unsupervised learning is the other learning method without predefined input output pattern. The network must discover features, regularities, correlation, or category of the input data automatically.

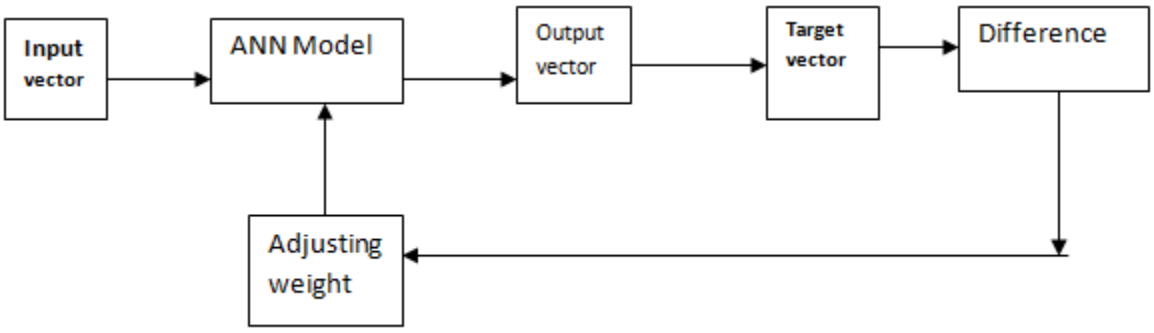


Figure: 2.4 How the artificial neural network works

2.4.2.1. Feed forward neural network

In feed forward network information flows forward only from input layer through hidden layer to the final output layer, no feedback loop. Feed forward neural network consist of three layer, an input layer, hidden layer and output layer. In each layer there are perceptrons simulating the human neuron. Perceptron receive input from either the outside world or the previous layer. The simplified process of training feed forward network is input data is presented to the network and propagate through the network until it reaches to the output layer. After propagation the predefined output is subtracted from the actual output and error value for the network is calculated. In most case back propagation trains the network, for adjusting the weight, it starts with subtracting the weight between the output layer perceptrones and the last hidden layer perseptrones. After the back propagation finished adjusting the weight the forward process start again, this process continues until the error between the predefined and the actual output [48].

2.4.2.2. Back propagation algorithm

It is a method of training neural network how to perform a given task. Artificial neural network are organized in layers inputs are travel forward and error backward. Back propagation algorithm use supervised learning; we provide the algorithm with example of input and desired output. After training difference between actual and expected result is calculated and finds the error. The idea behind back propagation is minimizing this error [48]. Back propagation algorithm learns by iteratively processing a training data set. This algorithm propagates the input first, and then the net input and output of each unit in the hidden and output layers are computed. The net input to a unit in the hidden or output layers is computed as a linear combination of its inputs. The outputs of the units connected to it in the previous layer. Each connection has a weight and using this weight the algorithm activates the process [60].

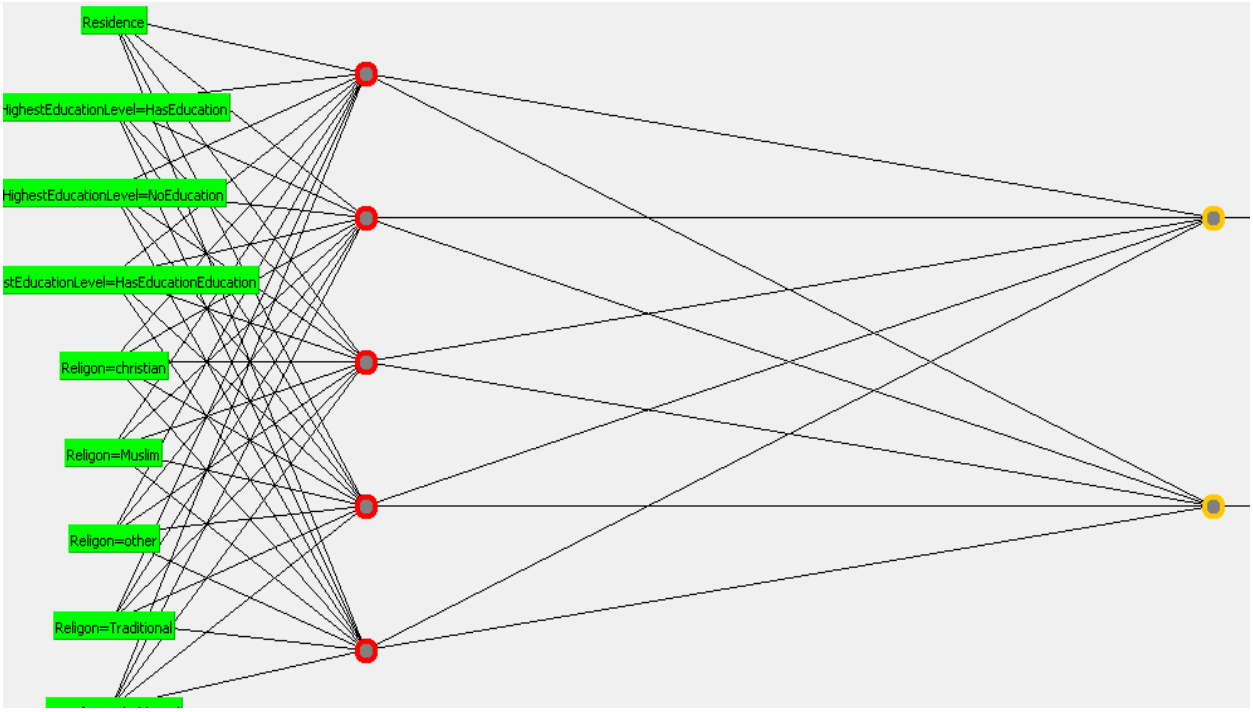


Figure 2.5. : Three layer neural network

The above figure shows the output of multilayer preceptron. Neural network has an input layer (green) which has perception equal to the attribute of the data set, output layer (yellow) which has two perceptron which is equal to the number of class and hidden layer (red) which is the average of input perceptron and output perceptron.

2.4.3. Decision Tree

Ross Quilan introduced a decision tree algorithm (iterative DICHOTOMiSER(ID3)) in 1979. C4.5 is a successor of ID3. Decision tree work like divide and conquer manner. Selection of root attribute is key process in decision tree. The root attribute best split the given record. This method seeks an attribute whose sorting result is closest to the pure partition by the class. For example in EDHS data set the education background of woman perfectly separate the data set for being taking antenatal care, education background will be the root node, and continuously breaking the data set hierarchically.

The major advantage of decision tree is unlike neural network the algorithm output is simple and allows user to easily understand the output or the model, which help domain expert to make appropriate decision by the output of the algorithm. The major drawback of decision tree is its

difficulty to apply on a data set with many attributes like EDHS. But tree pruning approach is applied to such decision tree. These approaches resolve the problem of over fitting [11].

C4.5 is a descendent of CLS and ID3, like CLS and ID3 generate classifier expressed as decision tree, but it can also constructed classifier in more comprehensible rule set form. Complex decision trees can be difficult to understand, for instance because information about one class is usually distributed throughout the tree. C4.5 introduced an alternative formalism consisting of a list of rules of the form "if educated and educated husband and live in Addis and husband help in house. Then class properly take antenatal care", where rules for each class are grouped together. The principal disadvantage of C4.5's rule sets is the amount of CPU time and memory that they require. C4.5 first grows an initial tree using the divide-and-conquer algorithm. Attributes can be either numeric or nominal in C4.5 [16].

However data mining and its application in medicine and public health is relatively young field of study data mining have been applied in health sector rapidly [18]. Data mining application in health care is crucial because health sector is rich with information and use of information technology allows atomization of process for extraction of data that help to get intensive knowledge and regularities [17].

Decision tree is a classifier which iteratively defines partional of instance space, it consists of nodes that form a rooted tree [42]. Decision tree is one of the approaches in inductive learning. In creating decision tree uses divide and conquer method [45]. A node with outgoing edge is called an internal node and all other node are testing node. Each leaf is assigned to one class representing the most appropriate target. Numeric attribute in decision tree can be geometrically interpreted as collection of hyper plane [42]. There are three phases in developing rule induction creating an initial large rule from the training set, second pruning the tree to remove less important node in decision making, and third processing the pruned tree to improve its understandability [45]. Rules from the tree represent knowledge in the form of IF-THEN rules, one rule is created for each path from the root to the leaf, and each attribute value pairs along a path form a conjunction [54]. Most of the time decision tree is applied to inductive learning for the reason of, decision tree is a good generalization for unseen instances, it has low computational complexity that is proportional to the number of observed training instances, and the resulting decision tree provides a representation of the concept that is readable to human [52].

2.4.3.1. Overfitting and pruning in decision tree

When algorithm build decision tree because of noise and outlier the tree building algorithm might not solve the problem on hand or output biased result. Tree pruning method address this problem of over fitting and worse time and space complexity. Pruned trees tend to be smaller, less complex and have less computational complexity. There are two common approach of tree pruning. Prep running which is done by halting tree construction early by deciding not split further. Post pruning which removes sub tree from a fully grown tree. Post pruning has higher computational complexity than pre pruning because pruning done after the whole tree were build [61]. There are five common type of post pruning. Error complexity pruning is a pruning method which is done in two stages, whole tree building and after that finding a sub tree which misclassify the attribute. The error-complexity method takes account of both the number of errors and the complexity (size) of the tree. Critical value pruning is the other post pruning method which relies on estimating the importance a nod from calculation done in tree creation step. During tree building goodness of split measure determine the attribute at a node. The value of the measure reflects how well the chosen attribute splits the data between classes at that node. The third pruning method is minimal-error pruning which develop a single tree which should give a minimum error rate when classifying independent set of data. The fourth pruning method is reduced error pruning which produce a serious of pruning trees by using the test data directly. The fifth type of tree pruning is pessimistic error pruning which aims to avoid the necessity of a separate test data. The misclassification rates produced by a tree on its training data are overly optimistic and, if used for pruning, produce overly large trees [45].

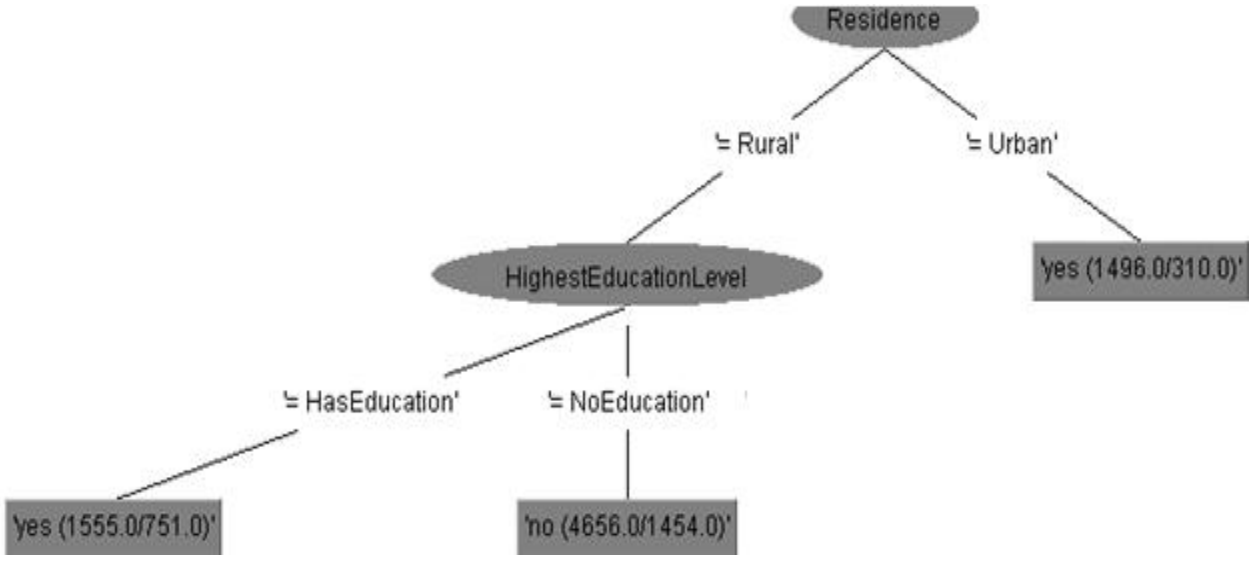


Figure 2.6.: Decision tree

The figure describes decision tree that shows whether or not the woman take antenatal care during pregnancy. Decision tree incorporate nominal attributer. Given this decision tree the analyst can predict the attribute that affect the customer whether to take antenatal care or not [42].

Tree complexity is explicitly controlled by the stopping criteria and pruning method. Tree complexity is measured by the total number of node, total number of leaves, tree depth and number of total attribute used. And each path from tree to one of its leave can be transformed into rule [42].

2.4.3.2. ID3 algorithm

Id3 is a simple decision tree algorithm developed by Ross Qulaln. The main idea behind this decision tree algorithm is to build a decision tree of a given set by using top dawn greedy search to check each attribute of every tree node. Shannon Entropy has been used in ID3 algorithm to calculate the Information Gain contained by data, which helps to make Decision Tree [50]. In ID3 information theory is used to determine the most informative attribute which will be used the root for the node. The measure of the information content of a message is the inverse of the probability of the receiving message [51].

$$\text{Information (m)} = \frac{1}{probablity\,(m)}$$

$$\text{Information (m)} = -\log_2(probablity(m))$$

Where m is a message, and log in base 2 selected because the number of bit required encoding the message is 2 bits [51].

Information content of a message or attribute should be related to the degree of uncertainty. Message with a high probability of arrival are not as informative as message with low probability arrival. Different message have different probability of arrival, the more uncertain a receiver about the source message the more informative the message will be [51].

$$\text{Entropy} = -\sum pi \log_2 pi$$

If the entropy of the attribute is 0, it is a homogeneous node and there is no need to classify further. If the entropy of the attribute is 1, it is a heterogeneous node and there is a need to classify further [51].

To minimize a decision tree depth, it is important to select the best attribute for splitting the tree node. In formation gain is one of the method applied in depth reduction during decision tree building [52].

$$\text{Information gain } (S_1, S_2, S_3 \ldots Sr) = \text{ENTROPY(S)} - \sum_{j=1}^{r} \frac{sj}{s} ENTROPY(Sj)$$

2.4.3.3. J48 Algorithm

J48 is the implementation of ID3 algorithm developed by WEKA project team. With additional future of tree pruning, missing data handling, continuous attribute value range and derivation of rule. J48 represent a leaf for instance belong to the same class. Both the discrete and continuous attributes are handled by J48 algorithm. J48 algorithm also handles the missing values in the training data. Pruning is performed for decreasing classification errors which are being produced by specialization in the training set. Pruning is performed for the generalization of the tree [15].

Another article describes J48 as an open source implementation of C4.5 algorithm in WEKA data mining tool. C4.5 algorithm generates a classification decision tree for the training data set by recursively partion the data [55].

2.4.3.4. Support Vector Machine

Vladmier Vapkinkn and co-workers at AT&T Bell laboratory introduced the support vector machine (SVM) in 1992 based on the statistical learning theory and design to solve two classification problem, for example attend antenatal care and not attend antenatal care. SVM works by first representing data set in high dimensional feature space then search for the (linear) optimal separating hyper plane where the margins between two different object (or record) is maximal.

The major advantage of SVM is its classification accuracy, but SVM is not always the best classification algorithm for every data set. The major drawback of SVM is basically designed to resolve two – class classification problem. To address this problem the method reduces a multi class problem into multiple binary problems. The other drawback is compared with other classification algorithm the training step of SVM is extremely slow and requires extensively computational resource. According to Mohammed J. Zaki [15] the maximum error margin models constructed by these machines and the quadratic programming solution process of these machines are not readily understood to the non-specialists of the technology and for the domain expert [15].

The mapping function can be either classification function or regression function. For regression a hyper lanes are constructed to separate the class in the training set. Goal of SVM is to achieve maximum separation (margin) between the two (or more) classes. SVM training stage find geometric interpretation, after training crate the same type of decision hyper space. Involves complex and time demanding calculations [69].

2.4.3.5. Regression Analysis

Regression is a learning function that maps a data item to real-valued prediction variable. For a set of items, regression is the analysis of the relationship of dependency between the values of the attribute. A model in form of a function (y = b + wx), response variable y and single

predictive variable x is automatically produced and this function can predict attribute value for new item. For example regression is applied in health sector to estimate the probability of a patient survival given the result of diagnostic test [17] [13].

2.5. Classification Based on Association Rule Mining

Classification Based on Association (CBA) was developed and proposed by Bing Liu and his colleagues in 1998. Classification based on association use different approach to data mining algorithm association. CBA outperform the widely used classification algorithm C4.5. This performance difference occurs because CBA generates (like association mining) and utilizes all class-related rules for prediction; while association mining is designed to extract all existing rules, including class-related rules from a dataset. C4.5 cannot generate all class-related (If-Then) rules. The algorithm first find classification association rules (CARs) based on a prior algorithm, build classifier based on CARS by using training set and the apply classifier to predict which class a new item belongs [11] [64].

2.5.1. Genetic Algorithms

Genetic algorithms are based on the principle of genetic modification, mutation and natural selection. The genetic algorithm creates a number of random solutions to the problem. All these solutions may not be good, a group of solutions can be skipped entirely, and it can come down to the overlapping solutions. Poor solutions are discarded, and the good ones retained. Good solutions are then being hybridized, and then the whole process is repeated [19]. Implementation of genetic algorithms is by applying operators (reproduction, crossover, selection) with the feature of mutation to enhance generation of potentially better combinations on discretized data; Randomly select parents, Reproduce through crossover, Select survivors for the next generation through fitness function [69].

2.5.2. AdaBoost

Ensemble learning deals with methods which employ multiple learners to solve a problem. The AdaBoost algorithm proposed by Yoav Freund and Robert Schapire is one of the most important ensemble methods [6]. The rationale behind the ensemble approach is that multiple classifiers (or

classification models) working together can yield better classification accuracy than the use of a single classifier. As a simple example, if Classifiers A, B, and C predict that a hard-to-classify women (women1) has taken antenatal care and Classifiers D and E predict that women 1 doesn't take antenatal care, then, by using a voting strategy, women 1 is determined taken antenatal care. Better than ensemble approach AdaBoost use weighted majority voting. The rationale behind this idea is that classifiers showing good classification results during the iterative training process have higher voting weights than others in final classification decisions [11].

2.6. Application of Data Mining in Health Care

In health care data mining is becoming increasingly popular for several reasons: the extremely large amount of data or data over load in the sector; the need for health organization to make decision based on the analysis of clinical and financial data; and the power to generate information that is fundamentally useful to all parties involved in the health industry.

Traditional decision making in health care is based on the ground information, lesson learned in the past, Health care data, resource management data, and transformed data. Health care organizations must have the ability to analyze data, and data mining techniques can be of tremendous use in health sector. Data mining can be used for data modeling for health care application, executive information system for health care, forecasting treatment cost and demand of resource, Public health informatics, and health insurance. Without data mining it is difficult to realize the full potential of data collected within health care organization as data under analysis is massive, high dimensional, distribute and uncertain. For successful health care organization it is important to empower the management and staff with data ware housing based on critical thinking and knowledge management tool for strategic decision making. Data ware housing can be supported by decision support tools such as data mart, OLAP and data mining tool [1] [22].

The demand and use of data mining in health care is tremendous. With the merged application of data mining via WEKA with J48 and GIS, health centers can determine methods that would lead to policy suggestion to health institution for proper utilization health care resources [7].

Non-invasive diagnosis and decision support is application of data mining in health industry. Every patient cannot afford for expensive, persistent and aching diagnostic and laboratory methods. For instance to discover cervical cancer using biopsy in women is a difficult task. K-means gathering algorithm was utilized by Thangavel *et al*. [7]. To detect the cervical cancer in women and he originated that the results of gathering data is more accurate than those of medical results [7].

Some medicines and vaccines which were declared to be safe for human sometimes injure the users, this kind of adverse drug effect (ADE) can be modeled with the use of data mining application. Data mining is used to investigate the drugs side effects in their data by US Food and Drug Administration. Around 67% of ADEs was discovered [7].

Data mining applications can greatly benefit the health care industry but with certain limitations. Health care data mining can be limited by the accessibility of data, secondly data problems which included missing data, corrupted data, inconsistent data, or nonstandard data such as piece of recorded in different format in different data source may arise, and sufficient exhaustive of data will certainly yield patterns of some kind that are a product random fluctuations [5][22].

Data mining in healthcare can be limited in data access, since the raw inputs for data mining frequently exist in different settings and systems, like administrations, clinics, laboratories etc. Therefore, data must be collected and integrated before data mining can take place. Building of data warehouse before data mining begins can be a very expensive and time consuming process. Healthcare organizations that develop data mining must use big investment resources, especially time, effort and money [19].

Data mining techniques are used to ‘mine’ healthcare data repositories to inductively derive decision-quality healthcare knowledge. Such data-derived knowledge can enable healthcare managers and policy-makers to infer inherent, yet invaluable, operative principles, values, know-how and strategies pertinent towards the improvement of the operational efficacy of the healthcare facilities. The operational efficacy of a healthcare enterprise can be significantly increased by (a) procuring diverse facets of empirical knowledge from the seemingly placid healthcare data repositories, and (b) by operationalizing the procured empirical knowledge to

derive a suite of packaged, value-added strategic healthcare decision support services that aim to impact strategic decision-making, planning and management of the healthcare facilities [38].

Data mining can be applied in health sector like it has been applied in many other fields. Data mining support the sector in treatment effectiveness by identifying pattern found in the treatment record. Health experts can simply identify which course of treatment is effective and accordingly apply the effective way of treatment. Data mining is also applied in health care management for decision making and strategic planning, data mining support decision makers and strategic planners in health sector to identify which disease is chronic and present danger for the community, which state is highly affected and need quick intervention, and appropriate knowledge for proper management. Data mining also applied in customer relationship management in health sector, better customer management will be achieved by fulfilling customer expectation. Data mining of patient data can help set reasonable expectations about waiting times, reveal possible ways to improve service, and provide knowledge about what patients want from their healthcare providers [17].

2.7. Related works

Yalem [66] conducted a research on determinants of antenatal Care, institutional delivery and skilled birth attendant utilization in Saharti Samre district, Tigray, Ethiopia. A total of 1113 women participated in the quantitative survey. The methodology of the study has combined both quantitative and qualitative methods. The quantitative study was a community-based cross-sectional survey using a structured questionnaire. Cluster sampling technique was utilized to select the study subjects who were mothers who gave birth once in the past five years prior to the survey period. For the qualitative study focus group discussions (FGDs) was chosen as the tool for data collection. The objective of the study was to explore the trends and magnitudes of antenatal, institutional delivery and skilled birth assistance utilization and to identify determinants for selection of this service in Saharti Samre district, Tigray, Ethiopia. His study finally found that women who received antenatal care for their recent births were 54%. Only 4% of mothers gave birth for their recent child in the health facility and only 6% mothers were assisted by skilled birth attendants [66].

Bereket [67] conducted another research on mothers' utilization of antenatal care and their satisfaction with delivery services in selected public health facilities of Wolaita Zone, Southern

Ethiopia. The methodology used was cross-sectional survey with exit interview conducted in Wolaita Zone. A total of 363 postpartum mothers were enrolled in the study. Among 363 postpartum mothers from thirteen public health facilities (one hospital and 12 health Centers) participated in the exit survey 317(87.3%) were in age group 20-34 and 354 (97.5%) were either married or living together with partner. Finally the research concluded that ANC can help for better utilization of institutional delivery. Maternal health care providers should plan for more institutional delivery by extending their acceptance in the already visiting mothers [67].

Another study was also conducted by Behailu [5] to find factors affecting antenatal care utilization in Yem special woreda, southwestern Ethiopia. A cross-sectional survey was conducted in April 2008 among 651 women who have had a birth 12 months prior to the survey. Based on the study, one hundred seventy nine (28.5%) women were reported to have received antenatal care at least once during their last pregnancy [5].

Tewodros [68] also conducted another research on mothers' utilization of antenatal care services among teenagers in Ethiopia. The methodology used was stratified cluster sampling method. Both bivariate and multivariate analyses were performed to determine the differentials of ANC by explanatory variables. Most (60%) of the subjects were in the age group between 18 and 19 years. The vast majority (90%) was from the rural settings and most (87%) were in marital union. The result showed that over a quarter (27.3%) of most recent childbirths had at least one ANC service [68].

Selam [39] conducted a data mining research to predict measles outbreak in Ethiopia, on measles case based surveillance data at Addis Ababa University. She used decision tree and naive Bayes classifier data mining techniques. One of the reasons why she used decision tree was that the model is quite reasonable since it takes the form of explicit rules and this allows the evaluation of the result and the identification of key attribute in the process. Confusion matrix was used to evaluate the performance of the model. Accuracy of the classifier was measured by dividing the total correctly classified positive and negative instance by the total number of samples. Precision was calculated by dividing correctly classified instances by the total number of correctly and incorrectly classified samples. Finally she found that the scores of the Naïve Bayes classifiers for correctly classifying the true Negative were slightly higher than J48 (decision tree). However, the overall score of the Naïve Bayes model was lower than that of the decision tree model. She

also found that Naïve Bayes worked well when tested on actual datasets, particularly when combined with attribute selection procedures that eliminate attributes redundancy. According to her, dependencies between attributes inevitably reduce the power of Naïve Bayes to differentiate what is going on. Some outbreaks were misclassified by one of the models though not by the other.

She concluded that data mining is extracting meaningful patterns and rules from large quantities of data. It is clearly useful in any field where there are large quantities of data and something worth learning. In this respect, widespread use of medical information systems and explosive growth of medical databases require traditional manual data analysis to be coupled with methods for efficient computer-assisted analysis, specifically problem of predicting measles outbreak in Ethiopia could be supported by the use of data mining, in particular with decision trees technique.

Tariku [40], conducted a research on developing a predictive model for fertility preference of women of reproductive age using data mining techniques at Addis Ababa University. The data source for this study was EDHS2011 survey data acquired from CSA of Ethiopia. The 2011 Ethiopia Demographic and Health Survey (EDHS) were conducted by the Central Statistical Agency under the sponsorship of the Ministry of Health. Tool used to conduct the research was WEKA data mining tool. The objective of the study is to explore the possibility of applying data mining techniques in developing a model that can predict fertility preferences of women of reproductive age. Six-step hybrid knowledge discovery process model was adopted. Through the steps, a dataset containing 15 attributes and 16515 records of women was constructed for building models. Three data mining classification algorithms, J48, Naïve Byes and neural Network (Multilayer Perceptron), were tested using 10-fold-cross-validation. The classifiers were implemented on the dataset with all and selected features.

According to Tariku the experiment were conducted for each three classifier based on all attribute, selected attribute, and for decision tree J48 algorithm with pruning tree with all attribute or pruning tree with selected attribute. Like Selam, Tariku also used confusion matrix to evaluate the performance of the model. The result shows the error rate for J48 (not pruned) is

22.20%, the precision of the model 80% and recall 89%. Tree-pruned J48 classifier model scores 78.37% accuracy and 21.63%error rate. Naive Bayes score 74.99% accuracy and 25.1% error rate without attribute selection, and 75.92% accuracy and 24.08% error rate with attribute selection. Multilayer Perceptron model (neural network) built using all attributes correctly classified 75.03% of instances and 24.97% of instances were incorrectly classified. The second model with reduced attributes incorrectly classified 22.84% while classified with an accuracy rate of 77.17%, indicating an increment of the accuracy by more than 2 % due to attribute reduction.

Tariku concluded that based on classification accuracy, J48 model is found to be relatively the best model in predicting fertility preference of women. Furthermore, J48 model provides decision rules that can easily be understood and interpreted.

Another study conducted by Abel [41] tried to build a model which predict Heart disease. The objective of the study was to design a predictive model for heart disease detection using data mining techniques. Data used for research was Transthoracic Echocardiography Report dataset collected by International Cardiovascular Hospital from October, 2008 to March, 2011. Methodology used is Knowledge Discovery in Database which is consisting of nine iterative and interactive steps was adopted to extract significant patterns from a dataset. Abel conducted four experiments with two scenarios by selecting attribute and with all attributes. In model building using un-pruned J48 decision tree without selecting attribute it took 0.89 second to build the model with 94.29%accuracy and 5.71% error rate. Model building by using the same algorithm with selecting attribute took 0.36second with accuracy of 95.5% and 4.5%error rate. J48 algorithm pruned and with all attributes scores 95.41% accuracy and 4.5%error rate, with selected attribute 95.56% accuracy and 4.44% error rate. Naïve Bayes also used in model building and performs 92.5% accuracy and 7.5% error rate with attribute selection, and 92% accuracy and 8% error rate without attribute selection. The last experiment was on model building with neural network, with attribute selecting it achieves 94.85%accuracy and 93.83% with all attribute. Finally he concluded that the study showed that data mining techniques can be used efficiently to model and predict heart disease cases. The outcome of this study can be used

as an assistant tool by cardiologists to help them to make more consistent diagnosis of heart disease.

Santos [17] developed a data mining system for providing analytical information on brain tumors to public health decision makers. According to the authors public health decision makers requires significant amount of analytical information to manage treatment of patients and other health sector resources. The author emphasize that data mining is a technology that is used to produce analytical useful information for decision makers. The goal of the project was to present an automated data mining system that allows public health decision makers to access analytical information regarding brain tumor. Data used to conduct the study were collected and the system architecture is based on a typical data mining system architecture. It includes five components, the database data ware house store raw data, the ontology which is formal knowledge regarding brain tumor, integration the data structure from the data ware house and the knowledge from the domain ontology, mining and information. general description of the system is the public health decision makers created a list of analytical information that is necessary to manage brain tumor, the domain experts explained the ontology to the data miner, the data miner linked each concept from the ontology with related data mining attribute, the data miner create several data mining model with different data mining techniques, and finally the result obtained were evaluated by both domain expert and data miner. They concluded that the result of the test encourage the continuation of the research. It brought satisfactory data mining result for large group of decision makers and superior to excellent result for small group of users [17].

All the research work reviewed was tried to address different health care problems on different database. Some articles tried to find factors affecting antenatal care but on specific region on different databases with traditional statistical method. This research identifies factors affecting antenatal care on EDHS database in regional state of Ethiopia using data mining technology.

CHAPTER THREE

BUSINESS UNDERSTANDING AND DATA PREPROCESSING

In this study, one of the tasks of data mining which is classification method is implemented for predicting factors affecting antenatal care in Ethiopia. The objective of this study was to develop a model by training and testing different classifier using "EDHS 2011 dataset", and the model was used to predict factors affecting antenatal care in Ethiopia. Depending on the CRISP data mining process model, the first step is business understanding, which includes determining background business objective and set business criteria. To achieve this task, domain experts in the health sector were consulted, documents were analyzed, and related works were explored. In understanding the business, data mining goals and data mining success criteria will be set. The second step in CRISP data mining process model is data understanding. To achieve this initial data collection, description of the data, exploration of the data and verification of the data were done in this step. In data preprocessing phase dataset were selected, data cleaning were done, construction of the data format of the dataset were selected.

3.1. Business Understanding

Business understanding includes understanding business background, identifying business success criteria, and identifying data mining goals.

3.1.1. Background business objective

According to the dataset EDHS 2011, thirty-four percent of women who gave birth in the five years preceding the survey women received antenatal care from a skilled provider, that is, from a doctor, nurse, or midwife, for their most recent birth. The business objective of this study was to identify what makes women to attend or not to attend antenatal care during their pregnancy or determining factors affecting antenatal care in Ethiopia using EDHS 2011.

3.1.1.1. Overview of Antenatal Care

The health related millennium development goal 5 is to improve maternal health. Among all the regions of the world, Africa has the highest maternal mortality rate and the lowest child health

status. In year 2000, WHO estimated maternal mortality ratio of 920/100,000 live births for sub-Saharan Africa with a life time risk of maternal death of 1 in 16. This rate is very high when compared to the developed countries' life time risk of 1 in 2800, estimated during the same time period. Complications during pregnancy and childbirth are the leading cause of death and disability among women of reproductive age in developing countries. There are estimated 529,000 maternal deaths each year, of which 99% occur in developing countries. Millions of women in these countries lack access to adequate care during pregnancy. Inadequate access and under-utilization of modern healthcare services are major reasons for poor health in the developing countries. This inequality in the health and well being of women in the developing and the developed world is a growing concern [25] [24].

Antenatal care (ANC) is an important determinant of safe delivery and one of the basic components of maternal care, on which the lives of mothers and babies depend. The World Health Organization (WHO) defines antenatal care as a dichotomous variable with a pregnant woman having one or more visit to a trained person during the pregnancy. Although certain obstetric emergencies cannot be predicted through antenatal screening, women can be educated to recognize and act on symptoms leading to potentially serious conditions; this is one strategy for reducing maternal mortality. One of the most important functions of ANC is to offer health information and services that can significantly improve the health of women and their infants. In addition, ANC during pregnancy appears to have a positive impact on the utilization of postnatal healthcare services. Empirical evidence shows that four visits are sufficient for uncomplicated pregnancies and more are necessary only in cases of complications; hence the World Health Organization currently recommends at least four ANC visits in the course of pregnancy [24][26]. The components of antenatal care include: health promotion which involve advising pregnant women on nutrition and health care as well as counseling to alert women on signs of danger and give them a health plan for the birth, assessment, history taking, physical examination, screening test for Human Immunodeficiency Virus Infection and some other sexually transmitted diseases, chronic and hereditary diseases, early detection and management of complications where needed, prevention of malaria, hookworm, tetanus, Management of sexually transmitted diseases, anemia and other conditions. Antenatal care represents a significant opportunity to reach a large no of pregnant women [27].

3.1.1.2. Roles of Antenatal Care

The Impact of Antenatal Care on Reducing Maternal Mortality

Scientific studies have demonstrated the role of antenatal care in reducing maternal mortality by several folds. The main impact had been through the reduction in severe anemia, reduction of cases of obstructed labour, improved nutrition such as iron and folate supplementation in areas with a high prevalence of anemia, the availability of local midwives, treatment of medical conditions like the serologic screening for and treatment of syphilis, routine measurement of fundal height, malaria prevention, tetanus immunization and screening for high risk and an efficient referral system. However several barriers to the effective delivery of antenatal care in both urban and rural areas of sub-Saharan Africa have been documented. These are; lack of physical access to health care facilities which presents a fundamental hurdle to receiving care even in urban settings, poor quality of services which still ranks high as a major concern in most health systems, also high patient volume and limited resources combine to constrain service provision [28] [29] [30].

Delivery of a Healthy Baby

Women who regularly go to antenatal care institutions and have good health as well are more likely to have live-births while women who had bad health and no antenatal care services are more prone to have dead babies and women who had good health and receive no or insufficient antenatal care services are more likely to have live-births except in a very few cases, where there are some medical problems that increase their liability to have stillborn babies [37].

3.1.2. Factors Affecting Antenatal Care Attendance

Several studies have indicated the presence of numerous factors in the initiation and the efficient use of antenatal care by women of child bearing age.

3.1.2.1. Maternal Gravida and Parity Number

Maternal experience to pregnancy and parity number has been associated with late initiation of antenatal care. High parity, indicating that the mother had experience giving birth before has

been associated with late initiation of antenatal care since these women feel that they do not need to attend as early because they already know what to expect during pregnancy and childbirth. These women may also have difficulty arranging childcare for other children in their home in order to attend antenatal care. Mothers who reported this to be their first pregnancy are also reported more likely to initiate their care late, suggesting that lack of knowledge or experience with pregnancy and childbirth is also a contributing factor to late initiation of antenatal care. These mothers simply may not have recognized that they were pregnant until later in the pregnancy. Furthermore, mothers whose pregnancies were unplanned and had a less than a very happy reaction to the pregnancy were more likely to initiate antenatal care late and attend inadequately [31].

3.1.2.2. Socioeconomic Factors

A number of socio-demographic factors are related to late initiation of antenatal care or having fewer antenatal visits. These include young maternal age, nonwhite ethnic group, low income, and high parity, low socio-economic status and unmarried status. Other financial barriers to adequate antenatal care, such as having no health insurance, are also influential on attending antenatal care in the developed world. Women living in poorest condition use antenatal services much less frequently than do those in the well to do. Women from many of the aforementioned social sub-groups associated with poor attendance for antenatal care also have an established increased risk of poor pregnancy outcomes [36].

3.1.2.3. Education Level

The utilization of antenatal care has been reported to go better with increased level of educational attainment. Majority of women who ended with stillborn babies have been reported to be absolutely illiterate or of poor education. The frequency of maternal attendance at antenatal care centers has been reported to be significantly related to maternal education. This is due to the fact that maternal education is important in deciding the preference for skilled provider; it also leads to better awareness about health facilities to get skilled professionals. This indicates out to the fact that universal primary education will go a long way to improve utilization of antenatal care [37] [34] [35].

3.2.2.4. Location of Antenatal Care Facility

Location also appears to have effect on the use of antenatal care. Women based in urban areas used antenatal care services than those based in the rural areas. This is probably due to the fact that women in urban areas are better informed and the rural women have limited access to antenatal care services as they have to depend on the primary healthcare center for maternal health care services. Thus, women based in urban areas have a lot of options to choose from [32] [33].

3.1.3. Business Success Criteria

According to health experts, antenatal care is a leading reason of death in sub-Saharan Africa including Ethiopia and the governors set a goal to minimize maternal death. The Government of Ethiopia is committed to achieve Millennium Development Goal 5 (MDG5), to improve maternal health, with a target of reducing the maternal mortality ratio (MMR) by three-quarters over the period 1990 to 2020. The business success criteria of the study were measured accordingly with its help in predicting the factor affecting antenatal care by using data mining application.

Age, Region, Residence, Highest education, Religion, Sex Of Household Head, Frequency Of Reading Newspaper, Frequency Of Listening Radio, Frequency Of Watching Television, Wealth Index, Total Children, Total Living Children, Terminated Pregnancy, Wanted Child, Marital Status, Husband Education, Job Status, and Husband Help are selected attribute important for predicting task. Data mining goal is build models with three classifier (J48, naïve Bayes, neural network).

3.1.4. Selecting the Target Dataset

The 2011 EDHS dataset used three questioners: the Household Questionnaire, the Woman's Questionnaire, and the Man's Questionnaire. These questionnaires were adapted from model survey instruments developed for the MEASURE DHS project to reflect the population and health issues relevant to Ethiopia.

From the demographic health survey data set which contains birth recode table, children recode table, couples recode table, household recode table, household member recode table, and women record table which all are in the form .dta data format individual women table were selected as target data set. Women's table has 16511 instances with 3807 attributes. Most of the attributes are redundant and irrelevant for this research. Among 16515 women at age 15-49 with a live birth in the five years preceding the survey counted 7751.

3.2. Data Understanding

3.2.1. Data Source

The 2011 Ethiopian Demographic and Health Survey (2011 EDHS) is part of the worldwide MEASURE DHS project which is funded by the United States Agency for International Development (USAID). The survey was implemented by the Central Statistical Agency (CSA) of Ethiopia. The goal of the survey was to give information about data on fertility and family planning behavior, child mortality, adult and maternal mortality, children's nutritional status, use of maternal and child health services, knowledge of HIV/AIDS, and prevalence of HIV/AIDS and anemia. In EDHS 2011 dataset 6702 households, 16515 individual women, 14110 individual men, 45540 children, 77744 household members, 6745 couples, 29812 HIV tests are surveyed.

3.2.2. Dataset exploration

The Women's table was used to collect information from all women age of 15-49. During DHS survey these women were asked questions on the following attributes: Background characteristics such as age, region, religion, ethnicity, education and media exposure, Birth history and childhood mortality, knowledge and use of antenatal care, reason didn't deliver at health facility, frequency of antenatal care, delivery and postnatal care, time spent at place of delivery, breastfeeding and infant feeding practices, vaccinations and childhood illnesses, marriage and sexual activity, women's work, husband's background characteristics, awareness and behavior regarding AIDS and other sexually transmitted infections (STIs), adult mortality, including maternal mortality [62].

The importance of data exploration in data mining process is to help to select the right tool for preprocessing or analysis. There are techniques used in data exploration. Visualization, clustering and anomaly detection are some of the techniques used. in this research all the distinct instance of the data set were visualized using WEKA tool which is Simple tools that show

histograms of the distribution of values of nominal attributes, and graphs of the values of numeric attributes (perhaps sorted or simply graphed against instance number), are very helpful.

3.2.3. Dataset description

For data preprocessing to be successful, it is essential to have an overall picture of target data set. Descriptive data summarization techniques can be used to identify the typical properties of our data and highlight which data values should be treated as noise or outliers. Each instance is characterized by the values of the attribute it has. There are many different type of attribute but the target data set has nominal and numeric attribute data type. The target data set contain social and demographical description of 16515 women instance (see table 3.1 below).

Table: 3.1 Attribute description

No	Attribute name	Description	Type
1	Respondent's current age	Age of a woman	Numeric
2	Region	Region where a woman live	Nominal
3	Residence	Current place of living	Nominal
4	Husband help	Willingness of husband in house work help	String
5	Sex of the house hold head	To identify who is responsible	Nominal
6	highest educational level	Education attainment of a women	Nominal
7	Household has: electricity	Has a house hold electricity	Nominal
8	Frequency of listening to radio	Measure media exposure	Nominal
9	Frequency of watching television	Measure media exposure	Nominal
10	Religion	Religion of a women	Nominal
11	Wanted pregnancy	Want(ready) to have a new child	Nominal
12	Women job status	Whether she has her own income or not	Nominal
13	Husband education	Education level of husband	Nominal
14	Age of household head	Age of the house Leader	Numeric
15	Marital status	Live with husband	Nominal
16	Wealth index	Level of income	Nominal
17	Age of respondent at 1st birth	Age of a woman having first child	Numeric
18	Ever had a terminated pregnancy	Ever had miscarriages before	Nominal

The above table describes all the attribute value of the women instance and the data type of the attribute.

To have an overview of the data set, it is necessary that explaining the distinct values of each attribute by using WEKA preprocessing tool. Using WEKA preprocessing tool simply observed

the distinct value of each attribute after loading the dataset to the WEKA tool. The data set has the following attribute with distinct values.

Table: 3.2 Attribute with distinct values

No	Attribute name	Distinct value
1	Wealth index	5
2	Frequency of watching television	3
3	Frequency of reading magazines	3
4	Frequency of listening radio	3
6	Sex of the household	2
7	Religion	6

Table 3.1 showed exploration result of the data set, the table indicated that missing values of the instance and no anomaly detection was observed. As mentioned above in this research we use WEKA tool and its' preprocessing functionality

Table: 3.3 Attribute with missing values

No	Attribute	Missing values	Percentage
1	Religion	8	0%
2	Frequency of listening radio	22	0%
3	Frequency of watching television	15	0%
4	Frequency of reading newspaper	25	0%
5	Terminated pregnancy	8	0%
6	Wanted child	5	0%
7	Job status	10	0%

3.3. Preprocessing

The real world data may not be complete, noisy and inconsistent and this can lead the output of knowledge discovery process irrelevant. Data preparation is the process of making the original

real world data to fit to a specific data mining method. Incomplete data occurs for many reasons, data may not be included because it was not considered important at the time of entry, and relevant data may not recorded due to misunderstanding or recording equipment malfunctioning [61]. Missing attribute values can be replaced by a fixed value (average of the non-missing values of training instance) or can be represented explicitly by adding another input for the attribute that is 1 if the attribute value is missing or replacing all missing attribute values by the same constant, using the most probable value. However, the WEKA tool replace missing values by mean and mode of the training set.

Dimensionality reduction and feature selection were applied when the data set includes more variables than could be included in the actual model building, it is necessary to select predictors from the large list of candidates. The target dataset contain redundant and very large number of attribute but with help of expert important attribute were selected. From three thousand attributes only eighteen attributes were selected. From 1615 instance of women between the age of 15-49, 7515 instance were selected because they had recent birth five year preceding the survey.

Frequently real-valued attributes are rescaled by some function that maps the value into the range 0 to 1 or −1 to 1 in a way that makes even distribution within that range but in this research most of the values are nominal and no need of rescaling. Integer-valued attributes are most often handled as if they were real-valued, age, number of living children and number of terminated pregnancy are integer value and used as if they are integer. If the number of different values is small, one of the representations used for ordinal attributes may also be appropriate. In this research all integer-valued attributes were considered as real-valued.

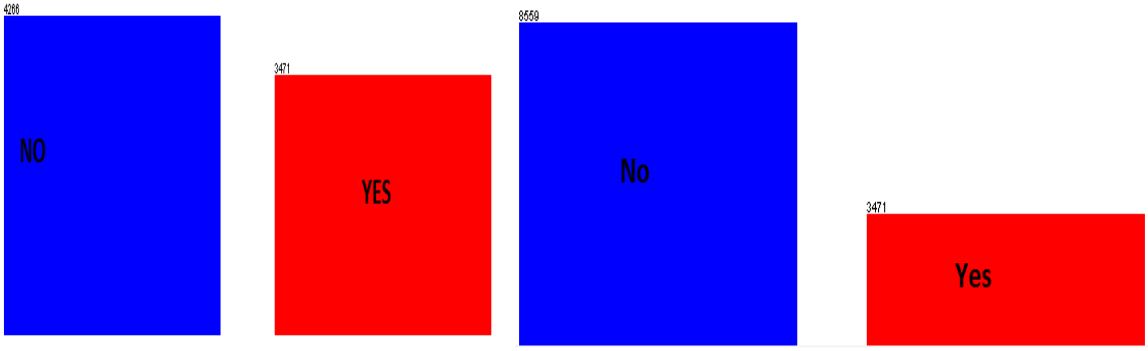


Figure 3.1.: Effect of SMOTE

3.3.1. Data Transformation

Data transformation is the process of converting data into appropriate forms for mining. Data analysis and mining on EDHS2011which is huge may take very long time and can give biased result output. Data reduction techniques have been helpful in analyzing reduced representation of data yet producing quality of knowledge. Data reduction techniques such as data cube aggregation attribute subset selection, dimensionality reduction, numerosity reduction, and discretization can be used to obtain a reduced representation of the data while minimizing the loss of information content [61]. Data reduction specifically dimensionality reduction were applied in the data set. The total number of attribute before reduction was 3807 and reduced to 18 attribute after expert advice. Discretization was also applied in the research. It is a technique that can be used to reduce the number of values for a given continues attribute by dividing the range of attribute into interval. In this research discretization were applied in the age of the women, living children, and age of the household by using WEKA tool (filter.unsupervised.attribute.discritize). Concept hierarchy also were used to transform five distinct values of wealth index(poorest, poorer, middle, richest, richer) to three distinct values (poor, middle, rich) and marital status of a women was described first by five distinctive values(never married, living with partner, married, divorced, widowed) were changed to two distinctive values (living with husband and not living with husband).

3.3.2. Data format conversion

The EDHS database was first in .dta format and was opened with SPSS, and then this file was converted into comma separated file format. Comma separated file format file were simply converted to ARFF file format which will be used in WEKA. Variables identification recode to be understood in the process vo11 for date of birth, v024 for region, v025 for residence, v103 for childhood residence and etc. the data type also converted from numeric to nominal for easy understandability. Residence was first numeric converted to nominal(1= urban, 2 = rural), data type of terminated pregnancy was numeric changed to nominal(0=no, 1=yes), data type of Religion was numeric changed to nominal(1=orthodox, 2=catholic, 3=protestant, 4=Muslim, 5=traditional, 96=other, 99 to missing value), data type of region was numeric changed to nominal (1= Tigray, 2= Affar, 3=Amahara, 4= Oromia, 5=Somalia, 6= Benshingual,7=Snnp,

12=Gambela, 13=Hararie, 14= Addis Ababa ,15= Deridewa), education attainment was in numeric data format changed to nominal(0=no education, 1= primary, 2= secondary 3= higher),marital status was in numeric converted to nominal(0=no,1=yes), husband education was in numeric converted to nominal(0=no education,1=primary, 2=secondary,3= higher),working status of the respondent was in numeric converted to nominal(0=no,1=yes) wealth index was also in numeric data type changed to nomial (1=poorest,2=poorer,3=middle,4=richer, 5=richest)and sex of the house hold converted 1 to male and 2 to female.

3.3.2.1. Attribute-Relation File Format (ARFF) in WEKA

The data set was first found in .dta SPSS data format however, most of the time WEKA uses ARFF file format. ARFF file format consist of a list of instance and attribute value for the instance separated by comma. Sample data in ARFF format is shown in Figure 3.2.

```
@relation 'lastindividal228-weka.filters.unsupervised.attribute.Remove-R1-
2,7-18'

@attribute Residence {Rural,Urban}
@attribute HighestEducationLevel {HasEducation,NoEducation}
@attribute Religon {christian,Muslim,other,Traditional}
@attribute SexOfHouseholdHead {Male,Female}
@attribute ANCvisit {no,yes}

@data
Rural,HasEducation,christian,Male,no
Rural,NoEducation,christian,Female,no
Rural,NoEducation,christian,Male,no
Rural,NoEducation,christian,Female,no
Rural,HasEducation,christian,Female,no
Rural,NoEducation,christian,Male,no
```

Figure 3.2. Attribute-Relation File Format (ARFF)

Attribute definitions is @data line that shows the start of the instances in the dataset. Instances are written one per line, with values for each attribute in turn, separated by commas. If a value is missing it is represented by a single question mark. ARFF files accommodate data types, such as nominal, numeric, date and string.

3.4. Performance investigation

Once a model is built using training data one would be curious as to how the model will perform in the future data or compare multiple model type for the same prediction problem to decide to use for the real world decision making [4] to do this we have to measure a performance and

accuracy of the model. In this research costs insensitive (precision, recall and F-measure) and ROC curve which is cost sensitive evaluation method were used. The amount of training percentage determined to have accurate model also used to measure model performance. In this study classifier were tested for 66 %(the default), 50% training percentage and for 80% training percentage and the remaining for testing.

3.4.1. Confusion matrix

Confusion matrix contains information about actual and predicted attribute classification [54]. Confusion matrix is a useful tool for analyzing how well your classifier can recognize instance of different classes [61]. If the outcome from a prediction is p and the actual value is also p, then it is called a true positive (TP), but the actual value is different from the predicted it is said to be false positive [54].

		predicted class	
		positive	Negative
True class	positive	True positive	False negative
	negative	False positive	True negative

Figure 3.3.: Confusion matrix

In classification problem the primary source of performance measurement is confusion matrix. After we draw diagonal line from upper left to lower right represent the correct decision made and the number outside this diagonal represent the error from the above confusion matrix table correctly classified instance are true positive and true negative and false negative and false positive incorrectly classified instances [61].

3.4.2. Precision and recall

Precision is the fraction of true positive from the predicted instance, while recall is the fraction of relevant instances that are retrieved or the fraction of true positive predicted from all true positive instance. Both precision and recall measure of relevance. Precision can be seen as a measure of exactness or quality, whereas recall is a measure of completeness or quantity [54].

Error rate or misclassification rate of a classifier M is (1-ACC (M)) where ACC (M) is accuracy of classifier M. The Mean Absolute Error measures the average magnitude of the errors [57], and is defined as the quantity used to measure how close predictions or forecasts are to the eventual outcomes [59]. The root mean squared error is the square root of the average quadratic loss and defined as frequently used measure of the differences between values predicted by a model or an estimator and the values actually observed [57][59]. The relative absolute error is very similar to the relative squared error in the sense that it is also relative to a simple predictor, which is just the average of the actual values [57].

$$\text{Precision} = \frac{tp}{tp+fp}$$

$$\text{Recall} = \frac{tp}{tp+fn}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{F-measure} = \frac{2}{1+\frac{1}{\text{Precision}}\ \frac{1}{\text{recall}}}$$

The ROC curves are two-dimensional graphs which are very powerful tools for measuring classifiers' accuracy in binary-class problems like the problem on hand (followed antenatal care and not followed antenatal care). ROC curves have been also used as a performance measure like above mentioned precision recall and accuracy which emphasized about positive predictions that are correct or incorrect [58]. The above mentioned performance evolution methods aren't cost sensitive do not take into account the cost of making wrong decisions, wrong classifications but ROC curves measure not only correctly classified instance or the accuracy of the classifier but it also measure the error or the cost of the classifier. In WEKA tool ROC curves were built by assigning x as false positive and y axis true positive [60].

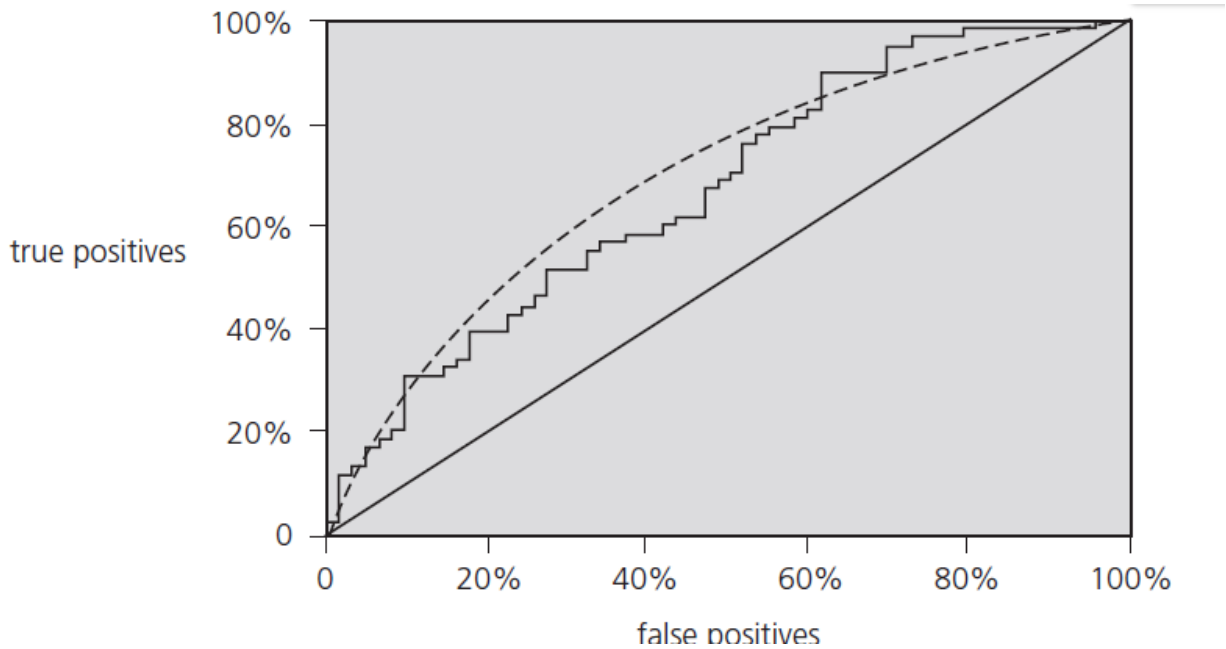


Figure 3.4.: Area under Roc Curve

From the above figure classifier has best accuracy if the curve goes closer to the upper left corner (close to y = 100%) [60].

Kappa is a chance-corrected measure of agreement between the classifications and the true classes. It's calculated by taking the agreement expected by chance away from the observed agreement and dividing by the maximum possible agreement. A value greater than 0 means that classifier is doing better than chance [57].

3.4.3. K-fold cross validation

In order to minimize the problems associated with the random sampling of training and testing data in comparing the predictive accuracy of two or more methods, we can use k-fold cross validation. This model minimizes the effect of the assumption that all subset of data sample are same kind. In *k*-fold cross validation the complete data set is randomly split into k mutually exclusive subsets of approximately equal size. The classification model is trained and tested k times. The first step in k-fold cross validation randomly divide the data set into k disjoint subset or fold. With each containing approximately the same number of records. For each fold, a classifier is constructed using all records except the ones in the current fold that means in the

first iteration the first partition is reserved as a test set and the remaining others as training set. Then the classifier is tested on the current fold to obtain a cross-validation estimate of its error rate. The result is recorded. After repeating the steps for all k folds, the kth cross validation Estimates are averaged to provide the aggregated classification accuracy estimate of each model type [60].

Different percentage split also the other classifier evaluation method to train the classifier with determined percentage of the date and test with the remaining. From this it is simple to identify the learning curve and compare the training percentage which gives the highest accuracy.

CHAPTER FOUR

EXPERIMENTATION

The goal of this research is to find pattern of factors affecting antenatal care using classification techniques. Three different supervised techniques are used; Decision tree classifier, Bayesian classifier and neural network classifier on WEKA (Waikato Environment for Knowledge Analysis) data mining tool.

4.1. Experimental Design

Different experiments were conducted in this research and all experiments with two situations, decision tree with pruning and without pruning, naïve Bayes with all attributes and with selected attribute, and neural network with all attributes and selected attribute. The tests were done with the use of test set and 10-fold cross validation and on different test split percentages. The more data is used in training phase the better the generated model.

To check the quality of the classifier the accuracy of the algorithm is measured. The accuracy of classifier measured by true positive rate, true negative rate, F-measure, Recall, Precision, and additionally by the AUC (Area under ROC Curve). The research also measures the error rate of each classifier Mean absolute error, Root mean squared error, Relative absolute error, Root relative squared error are measured. WEKA version 3.6, is the tool which is chosen in experiment to analyze algorithms output. Experimenter and Explorer are two mainly used interfaces during this experiment WEKA can perform data preprocessing and modeling technique, and easy to use for the research with its easy to use graphical user interface. WEKA handles different data mining tasks which are, data preprocessing, clustering, classification, regression, visualization, and feature selection.

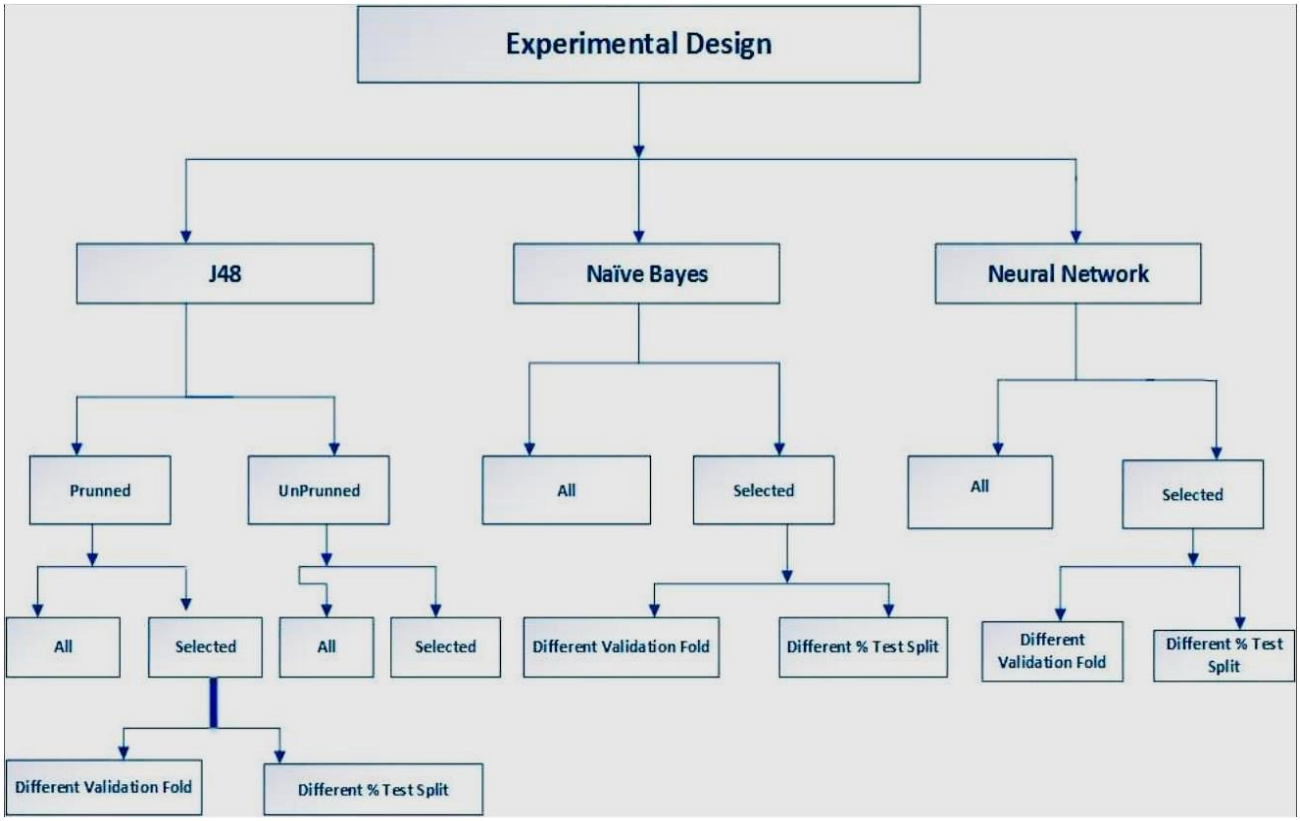


Figure 4.1: Flow of experiments

4.2. Feature Selection

All data set attributes may not be relevant for specific data mining task. Feature selection or attribute selection in search of the best attribute subset to have unbiased output is the first task in data mining. The key benefits of performing feature selection before model building is to avoid redundant data with this effect decrease over fitting, less data means training time for the algorithm minimized, and mainly increase the accuracy of the model. Attribute selection were conducted using WEKA data mining tool (filtering algorithm). There are two kinds of filter algorithm unsupervised and supervised algorithm and further classified to attribute filter which work on variables in the data set and record filter which work on instances in the data set. WEKA provides an attribute selection tool which are attribute evaluator; attribute subset are assessed to find best attribute subset with the information gain with respect to the class, and search method ranker which ranks attribute by their individual ranks.

```
=== Attribute Selection on all input data ===

Search Method:
      Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 19 ANCvisit):
      Information Gain Ranking Filter
      Numeric attributes are just binarized

Ranked attributes:
 0.0791852   2 Region
 0.0753274  10 WealthIndex
 0.0727035   4 HighestEducationLevel
 0.0702516   3 Residence
 0.0613003   9 FrequencyOfWatchingTelevision
 0.0590346  16 husbsndeducation
 0.0392878   5 Religon
 0.0285398   7 FrequencyOfReadingnewespaper
 0.0251781  11 TotalChildrens
 0.0192193   8 FrequencyOfListineningRadio
 0.0188706  12 livingchildren
 0.0052668   1 Age
 0.0052663   6 SexOfHouseholdHead
 0.0044116  13 TerminatedPregnancy
 0.0043382  14 Wantedchild
 0.0016211  15 maritalstatus
 0.0000537  17 husbandoccupation
 0.0000346  18 Respondentjobstatus

Selected attributes: 2,10,4,3,9,16,5,7,11,8,12,1,6,13,14,15,17,18 : 18
```

Figure 4.2.: Output of WEKA after unsupervised attribute filter

The output showed ranked attribute with their corresponding information gain value. WEKA calculate information gain using Cloud Shannon information theory which use entropy as measure of information gain. The result of attribute selection shows that region has gained the highest information gain with 0.071852 information gain and the lowest for respondent job status. J48 decision tree also select region as a root node because of highest information gain to split the dataset to binary class.

4.3. Model Building Using J48

Experiment 1

Thinking that the goal of this research is to predict factors patterns which affect antenatal care using classification techniques, the first experiment was conducted with unpruned J48 classifier with all attributes. In analyzing effectiveness and accuracy of data mining techniques it is necessary to measure the accuracy, error rate and true positive rate of J48 classifier. In addition to the above mentioned method AUC were analyzed to measure performance of a model.

The first situation the algorithm was run on the full data set with all attribute. It took 0.51 second to build the model and the model generated a tree with a size of 3134 and 4185 leaves.

Table 4.1.: Comparison of unpruned J48 on attribute selection

Mode	True positive rate	False positive rate	Precision	recall	F-measure	Roc
Unpruned J48 with all attributes	76%	40%	75%	76%	75%	70%
Unpruned J48 with selected attribute	77.8%	42.7%	76.5%	77.8%	76%	73%

Table 4.1 shows result of the first experiment which was performed by J48 classifier unpruned with all attributes. The classifier correctly classified 76.33% of the instances and the remaining 23.67% of the instances were incorrectly classified. False positive rate on the table 4.1 indicates the portion of negative case in the actual case but classified as positive in the experiment 1062 instance are identified false positive. True negative rate of the experiment is the portion negative instance in actual case and were classified as negative in the experiment which amounts 62.3% of negative instance (percentage of women who said no to antenatal care during the interview and identified as no to ante natal care during prediction). Recall or the true positive rate is the proportion of positive cases that were correctly classified as positive. In this experiment recall of the classifier is 71%. Precision of the classifier is the portion of the true classified cases which were correct; the first experiment scored 79% precision. The combined measure of precision and recall or the F measure 72%. The kappa statistics of the classifier interpreted as fair agreement because it is in the range of 0.2 to 0.4 which is fair agreement.

Table: 4.2 confusion matrix of unpruned J48

		Prediction	
		No	Yes
Actual	no	7497	1062
	yes	1786	1685

Confusion matrix shows the number of correct and incorrect classifications made by the classification model; in this experiment it is two by two matrixes because the classification is

binary class problem. True positive rate shows the portion of positive instance correctly identified as positive.1685 women who said yes to antenatal care during pregnancy correctly identified by the model, but 1786 of them were incorrectly identified as they said no to antenatal care even though they said yes to antenatal care. True negative indicate the portion of women who actually said no to antennal care and were classified as they said no to antenatal care during pregnancy 7497 of the instance. False negative of the output indicate the portion of women who actually followed antenatal care but incorrectly classified as if they did not follow antenatal care during pregnancy. False positive in the output indicate the portion of women who said no to antenatal care in the actual case but classified as followed antenatal during pregnancy. Accuracy of the model out of the total women predicted as yes to antenatal who actually followed antennal care and predicted as no to antenatal care who actually not followed antenatal care.

For unpruned J48 classifier when run with all attribute, instead of smote synthetic balancing with over sampling the minority class replaced with resampling, output of the classifier improve true positive rate. But accuracy of classifier decrease compared to balancing with smote.

Confusion matrix of unpruned J48 after resampling

		Prediction	
		No	Yes
Actual	no	3517	744
	yes	872	2110

Experiment was also done with unpruned J48 algorithm with attribute selected with WEKA tool in attribute selection step. The classifier correctly classified 77.79% instance and 22.21% instances are incorrectly identified, from this we understand that the classifier slightly decrease accuracy with slight decrease in computational time when attribute were selected.

Experiment two

In this experiment J48 algorithm with pruning was used to build the model and similarly to the first experiment the classifier was evaluated with the use of same performance evaluation matrix. In the experiment confidence factor was 0.25 which made the algorithm more pruning. Lowering the confidence factor decreases the amount of post-pruning, minNumObj is parameter of J48 in WEKA which means the minimum number of instances per leaf, the default two were used in the

experiment, numFolds is another parameter which describes the amount of data used for c4.5 error prunning the default three is used one fold is used for pruning, the rest two for growing the tree, reduced-error pruning is used instead of C.4.5 pruning, sub tree raising operation were not done when pruning.

Tree pruning is a step undertaken by WEKA tool to optimize time and computational complexity as well as accuracy of a decision tree algorithm. As mentioned above reduced error pruning which is implanted after fully induced decision tree, during the process statistically insignificant nodes were removed. The algorithm evaluates the decision error (estimated percent misclassifications) at each decision node and propagate this error up the tree.

Table 4.3.: Comparison of pruned J48 on attribute selection

Mode	Time taken	True positive rate	False positive rate	Precision	Recall	F-measure	Roc
Pruned J48 with all attributes	0.59 second	0.785	0.448	0.777	0.785	0.761	0.686
Pruned J48 with selected attribute	0.21 second	0.7827	0.466	0.777	0.783	0.755	0.69

The above table shows the effect of pruning in J48 algorithm with selected attribute and all attributes. From the result it is visible that the size of tree and computational time is less than the algorithm without pruning. Time taken to build the model was smaller for selected attribute with a slight decrease in true positive rate compared to tree building with all attribute. From the confusion matrix accuracy of the classifier decrease from 78.85% to 78.27% after attributes selection.

Additionally for this experiment J48 pruned with selected attribute was tested with different folds by tuning cross validation and different percentage of the training set, with the goal to get a model with better accuracy and classifier complexity.

Table 4.4.: Output comparison on different folds and different test split percentage of pruned J48

	Training set	5-fold cross validation	10-fold cross validation	15-fold cross validation	40% split	50% split	66% split	80%split
Correctly Classified Instances	80.51%	78.06%	78.27%	78.42%	77.88 %	78.28 %	78.72 %	79.85%
Incorrectly classified	19.49%	21.94%	21.73%	21.58%	22.12 %	21.72 %	21.28 %	20.15%
TP rate	80.5%	78.1%	78%	78.4%	77.9%	78.3%	78.7%	79.9%
FP rate	43.4%	46.6%	46.6%	46.4%	49.6%	47%	44.3%	44%
Precision	81%	77.4%	77.8%	78%	77.8%	77.5%	77.8%	79.3%
Recall	80.5%	78.1%	78.3%	78.4%	77.9%	78.3%	78.7%	79.4%
F-measure	78%	75.3%	75.5%	75.7%	74.4%	75.6%	76.6%	77.6%
Roc	72%	69.1%	69%	68.5%	68.7%	68.8%	68.5%	70.1%

The experiment showed that the highest number of correctly classified instance is gained for training Set of dataset (over 80.51% of instance). The worst correctly classified instance value is obtained for 40% split of dataset (77.88% of instance). When it comes to the True positive rate the best results were obtained for training set with slight difference of 80% split.

From table 4.4 results also showed that slight increase of correctly identified instance occurred with increase of cross validation folds and increase of split percentage. Precision, recall, and F-measure increase in the same trend like correctly identified instance with increase of cross validation folds and increase in percentage split.

From the experiment it is identified that in increasing the percentage split from 66% to 80% correctly identified instance increase to 79.85% instance with computational time trade off (increase from 0.23 second to 0.32 second) and the size of the tree keeps constant for all split percentage. Accuracy of the classifier increased from 78.06% in 5-fold cross validation to 78.43% in 15-fold cross validation.

4.4. Model Building Using Naïve Bayes Classifier

Experiment 3

The second algorithm used in this experiment was naïve Bayes. Similarly to the previous algorithm naïve Bayes was also tested with the same situation (with all attribute and selected attribute) and evaluation matrix with the goal of finding better classification algorithm for the data set. The parameter which is used to manipulate the algorithm is small when compared to neural network and J48. In the first case, the algorithm was run using the entire dataset with 19 attributes and it took 0.03 second to build the model which is very small compared to the above classifier and 10 fold cross validation were used.

```
time taken to build model: 0.03 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        9253               76.916 %
Incorrectly Classified Instances      2777               23.084 %
Kappa statistic                          0.4002
Mean absolute error                      0.2583
Root mean squared error                  0.4268
Relative absolute error                 62.9222 %
Root relative squared error             94.1894 %
Total Number of Instances            12030
Ignored Class Unknown Instances             27

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.88      0.505      0.811      0.88      0.844      0.773    no
                 0.495     0.12       0.627      0.495     0.553      0.772    yes
Weighted Avg.    0.769     0.394      0.758      0.769     0.76       0.773

=== Confusion Matrix ===

   a    b   <-- classified as
 7536 1023 |   a = no
 1754 1717 |   b = yes
```

Figure 4.3.: Output of Naïve Bayes Classifier with all attribute

From the experiment output, the accuracy of the classifier is 76.92% with error rate 23.08%. The classifier correctly predicted 7536 instance which said no to ante natal care classified to no class and 1717 instance said yes to antenatal care during pregnancy and classified to yes class. 1023 instance incorrectly classified as yes class although they said no to ante natal care follow-up during pregnancy and again 1754 instances were incorrectly classified as no class even though they followed antenatal care during pregnancy.

Table: 4.5 Output comparison of Naïve Bayes Classifier on all attributes and selected attribute

Model	Time taken	True positive rate	False positive rate	Precision	Recall	F-measure	Roc
Naïve bayes with all attributes	0.03	0.769	0.394	0.758	0.769	0.76	0.773
Naïve bayes with selected attribute	0.02	0.768	0.399	0.756	0.768	0.759	0.767

The above table shows the output of naïve bayes run with selected attribute and all attribute. From the output computational time showed slight decrease from 0.03 second to 0.02 second for selected attribute and True positive rate, precision, recall, and f-measure showed slight decrease when the algorithm run with selected attribute.

Table: 4.6 Confusion matrices for Bayesian algorithm run with all attribute

		Prediction	
		No	Yes
Actual	No	7548	1011
	Yes	1779	1692

From this confusion matrix it can be identified that 7548 women who haven't followed ante natal care during pregnancy correctly classified by the algorithm to no class (true negative). It is also showed that 1692 women who actually followed antenatal care during pregnancy correctly classified as yes class (true positive). The proportion of the sum of two classes with all instance showed the precision of the classifier.

Confusion matrices for Bayesian algorithm run with all attribute after resampling

		Prediction	
		No	Yes
Actual	no	3493	768
	yes	1576	1906

For naïve bays classifier when run with all attribute, instead of smote synthetic balancing with over sampling the minority class replaced with resampling, output of the classifier improve true positive rate. But accuracy of classifier decrease compared to balancing with smote.

Table: 4.7 Output of Naïve Bayes Classifier on different folds and test split percentage

	Training set	5fold cross validation	10fold cross validation	15fold cross validation	50%split	66%split	80%split
Correctly Classified Instances	76.90%	76.80%	76.81%	76.80%	77.38%	77.52%	77.11%
Incorrectly classified	23.10%	23.20%	23.19%	23.20%	22.62%	22.47%	22.89%
TP rate	76.9%	76.8%	76.8%	76.8%	77.4%	77.5%	77.1%
FP rate	39.8%	39.9%	39.9%	39.9%	39%	38%	38.6%
Precision	75.7%	75.6%	75.6%	75.6%	76.3%	76.6%	76.2%
Recall	76.9%	76.8%	76.8%	76.8%	77.4%	77.5%	77.1%
F-measure	75.9%	75.9%	75.9%	75.9%	76.6%	76.8%	76.5%
Roc	76.9%	76.7%	76.7%	76.7%	77.3%	77.7%	77.7%

The above Table 4.7 shows that the highest number of correctly classified instance was gained for training Set of dataset (over 76.90% of instance). The worst correctly classified instance value was obtained for 5fold and 15fod cross validation (76.80% of instance for both situations). Concerning to the True positive rate the best results were obtained for 66%percentage split (77.5%). From the experiment it is identified that in increase or decrease of cross validation folds and test split percentage doesn't show any proportional increase or decrease of the accuracy of naïve Bayes output like J48 algorithm.

4.5. Multilayer Perceptron

Experiment four

The last algorithm used to classify and discover patterns of factors affecting antenatal care was the Multilayer Perceptron. Similarly to the previous experiments also in case of the Multilayer Perceptron the different testing configurations were tested. Evaluation of effectiveness and accuracy of data mining methods also measured with the same model evaluation methods. Multilayer perceptron in WEKA tool has many parameters which can affect the output performance. Decay a parameter which will cause learning rate to decrease which helps to stop the network from diverging from the target output. The other parameter hidden layers defines the hidden layer of the output. Learning rate is also a parameter which defines the amount the weight is updated. Training time is the number of epoch to train through. There is also a parameter which filters and normalizes attributes during the process.

```
Time taken to build model: 520.24 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        9278               77.1239 %
Incorrectly Classified Instances      2752               22.8761 %
Kappa statistic                          0.4088
Mean absolute error                      0.2451
Root mean squared error                  0.4456
Relative absolute error                 59.6917 %
Root relative squared error             98.3423 %
Total Number of Instances            12030
Ignored Class Unknown Instances             27

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.878    0.493      0.815     0.878     0.845      0.747    no
                 0.507    0.122      0.628     0.507     0.561      0.746    yes
Avg.             0.771    0.386      0.761     0.771     0.763      0.747

=== Confusion Matrix ===

   a    b   <-- classified as
 7519 1040 |    a = no
 1712 1759 |    b = yes
```

Figure 4.4.: Output of neural network

The first experiment was done on multilayer preceptron using the entire attribute with cross validation techniques with default parameters. The classifier took 520.24 second which is the worst computational time compared to the above two algorithms. The classifier correctly classified 9278 instance (77.12% of the instances) and the true positive rate was 77.1%. From the confusion matrix it is identified that multilayer perceptron correctly classified 7519 instance which said no to antenatal care during pregnancy correctly classified as no class during classification and the classifier also correctly classified 1759 instance which said yes to antenatal care during pregnancy classified as yes class.

Table: 4.8 Multilayer perceptron output on different folds and test split percentage

	Training Set	5-fold cross validation	10-fold cross validation	15-fold cross validation	50% split	66% split	80% split
Correctly Classified Instances	81.44%	78.%	77.3%	78.42%	77.18%	77.94%	78.98 %
Incorrectly classified	18.56%	22%	22.7%	21.58%	22.82%	22.06%	21.02 %
TP rate	81.4%	78%	77.3%	78.4%	77.2%	77.9%	79%
FP rate	33.4%	43.9%	41.7%	46.4%	44.1%	43.9%	40.7%
Precision	80.7%	76.9%	76%	78%	75.7%	76.7%	77.8%
Recall	81.4%	78%	77.3%	78.4%	77.2%	77.9%	78%
F-measure	80.7%	76%	75.9%	75.7%	75.5%	76.1%	77.6%
Roc	82.6%	74.6%	75.3%	68.5%	71.9%	74.2%	76.6%

Multilayer perceptron output showed that the highest number of correctly classified instance is gained for training Set configuration of WEKA (over 80.44% of instance). The worst correctly classified instance value is obtained for 50% split of dataset for training and same amount for testing (77.18% of instance). Concerned to the True positive rate the best results were obtained for training set with slight difference of 80% split.

From table 4.8 results showed that slight increase of correctly identified instance occurred with increase of cross validation folds and increase of split percentage. Precision, recall, and F-measure increase in the same trend like correctly identified instance with increase of cross validation folds and increase in percentage split.

4.6. Discussion

To build a model that can identify feature patterns which affect antenatal care follow-up of women J48, Naïve Bayes and Multilayer Perceptron algorithms were used. Experiments were done on all feature and selected feature of the data set, with different validation folds and with different test split percentage. On the first experiment J48 decision tree algorithm tested with all attribute with pruning and without pruning, selected attribute with pruning and without pruning. On selected attribute the algorithm run on different cross validation folds and test split percentage. Naïve Bayes algorithm was run on all attribute and selected attribute again on different cross validation folds and test split percentage. Neural network (multi-layer preceptron) algorithm was also run on different cross validation folds and test split percentage.

4.6.1. Effect of attribute selection on j48, naïve Bayes, and neural network

Two situation were considered for each classifier to develop models .the first situation consider all attribute of an instance to develop models and the other situation consider selected attribute using information gain attribute evaluator and ranker search method. Again with selected attribute the algorithm run on different test split percentages and cross validation folds. Generally the classifier with selected attribute run on training test set produced improved result.

Table 4.3 shows the improvement of accuracy after attribute selection. From the output of J48 algorithm it was observed that unpruned j48 algorithm showed an improved accuracy, precision, and recall result when it was run on selected attribute. For pruned J48 algorithm the result was reverse of unprunned j48 algorithm. Accuracy, precision, and recall decreased for selected attribute with improvement of computational time from 0.59 second to 0.21 second. It was also observed that because decision tree size decreased computational space also decreased.

From the result of naïve Bayes it was observed improvement of computational time from (0.03 second to 0.02 second) for selected attribute which is the smallest of all classifier in this study. But Accuracy precision, and recall showed slight decrease for selected attribute, finally effect of attribute selection was observed on multilayer perceptron from the output accuracy showed improvement from 77.12%to 77.3% and recall also improved from 77.1% to 77.9% but precision showed no change.

4.6.2. Effect of pruning on J48 algorithm

Effect of Pruning on J48 algorithm when run on all the attributes improve the true positive rate from 76% to79%, pruning also improve precision and recall from 75% to 79% and from 76% to 79% consecutively. Computational time increase from 0.51 second to 0.59 second and decrease of computational space indicated by the size of the tree was also the other effect of tree pruning in J48 decision tree. It was also visible that with the cost of slight increase in computational time, pruning achieved accuracy and computational space improvement. Other situation tree pruning effect observed was running J48 algorithm with selected attribute, from the output it was observed that true positive rate almost keeps constant and slight improvement on precision and recall (precision improved from76.5%to77.7%and recall from 77.8%to78.5%).from this the effect of pruning is slightly increase when algorithm run on all attribute relative to selected attribute

4.6.3. Effect of test split and K-fold cross validation on J48, naïve Bayes and multi-layer perceptron

J48 pruned with selected attribute was tested using training set (test on the set classifier trained), different cross validation folds and percentage split (train the classifier on the percentage identified and test with the reaming percentage). Result showed that the highest accuracy was achieved on use training set configuration with slight difference 80% percentage split test configuration (80.51% and 79.85% accuracy respectively).Concerning on precision the highest amount was observed on training test set with slight difference 80%split percentage the same like accuracy. Precision increase with cross validation folds increase (when increase from 5-fold to 15-fold precision increase from 77.4% to 78%). The same trend observed on split percentage test, when the percentage split increase from 40%to80%precision also increase from77.8% to 79.3%.concerrning on recall the same pattern observed like precision even though increasing rate is decreased, recall increase with cross validation fold increase (from 5-fold to 15-fold recall increase from 78.1% to 78.4%) and recall increase from 77.9% to 79.4% with increase of test split percentage from 40% to 80%.

Naïve Bayes algorithm also tested use training set configuration (test on the set classifier trained), different cross validation folds and percentage split (train the classifier on the

percentage identified and test with the reaming percentage) from the output it was observed that the highest accuracy was achieved on 66% test split (77.9%). precision of the classifier was not changed with cross validation fold changes (keeps constant to 75.6% when the folds increased from 5-fold to 15-fold)precision showed slight variation with change split percentage test. Concerning on recall like precision it was not changed with change of cross validation folds it was keep constant to 76.8% for every fold.

Cross validation change in neural network from 5-fold to 15-fold decreased accuracy from 78% to 77.3% but change from 10-fold to 15-fold showed increment of accuracy from 77.3% to 78.42%. On test split variation increment from 50% split to 80%split accuracy also increased from 77.18% to 78.98% proportionally.

4.7. Evaluation and comparison of data mining algorithms

For this research J48 classifier, naïve Bayes classifier, and multi-layer perceptron were used, in this section this algorithm were evaluated and compared against each other. Here the comparison of the algorithms in terms of performance is done. From the above experiments Use training test 10-fold cross validation and the 80% split have a slight advantage over the rest of the testing configurations. The result of the comparison of the algorithm is presented as follow.

Table: 4.9 performance comparison of J48, naïve Bayes and multilayer perceptron

Model		Accuracy	False positive	Precision	Recall	F-measure
J48	Unpruned with all attribute 10fold cross validation	76%	40%	75%	76%	75%
	Unpruned with selected attribute 10fold cross validation	77.8%	42.7%	76.5%	77.8%	76%
	Pruned with all attribute 10fold cross validation	78.5%	44.8%	77.7%	78.5%	76%
	Pruned with selected attribute 10fold cross validation	78.27%	46.6%	77.7%	78.3%	75.5%
	Pruned with selected attribute 66% percentage split	78.72%	44.3%	77.8%	78.7%	76.6%
	Pruned with 80%percentage split	79.85%	44%	79.3%	79.4%	77.6%
	Pruned with selected attribute use training set	80.51%	43.4%	81%	80.5%	78%
Naïve Bayes	With all attribute 10fold cross validation	76.9%	39.4%	75.8%	76.9%	76%
	With selected attribute 10fold cross validation	76.8%	39.9%	75.6%	76.8%	75.9%
	With selected attribute Use test set	76.9%	39.8%	75.7%	76.9%	75.9%
	With selected attribute 66%percentage split	77.5%	38%	76.6%	77.3%	76.8%

	With selected attribute 80%percntage split	77.1%	38.6%	76.2%	77.1%	76.5%
Multi-layer perceptron	With all attribute 10fold cross validation	77.12%	38.6%	76.1%	77.1%	76.3%
	With selected attribute 10fold cross validation	78.42%	41.7%	76%	77.3%	75.9%
	With selected attribute 66%percentage split	77.94%	43.9%	76.7%	77.9%	76.1%
	With selected attribute 80%percentage split	78.98%	40.7%	77.8%	78%	77.6%

From Table 4.9 it can be observed that pruned j48 classifier with selected attribute on training set testing configuration achieved highest accuracy level (80.51%). The next situation high accuracy achieved was again pruned j48 algorithm with selected attribute but on 80%percentage split test configuration (79.85%). The lowest accuracy observed on unpruned J48 algorithm with selected attribute. J48 algorithm showed the highest precision with pruning and attribute selection the same like accuracy (86.1%). The next highest accuracy percentage was achieved by pruned J48 algorithm with selected attribute on 80%percentage split test configuration (79.3%). The lowest precision was observed for unpruned J48 algorithm when it was run on the entire attribute. Concerning on recall the highest percentage was observed the same like the upper cases for pruned j48 algorithm with selected attribute on use training test set. The next highest also the same the upper situation it was pruned J48 algorithm with selected attribute on 80% percentage split test. It is clear that accuracy, recall, and precision showed same result for J48 algorithm with decreasing and increasing of cross validation folds and test split percentage.

From the output of the experiment it was observed that the time spent for multilayer perceptron with all attribute on 10fold cross validation was the longest while, classifying with naïve Bayes with selected attribute on 10fold cross validation was the shortest. Pruned J48 classifier with selected attribute also showed better computational time than J48 unpruned with all attribute.

Accuracy, true positive rate, precision, recall, and f-measure don't consider the cost of making wrong decision and wrong classification that means they are not considered cost sensitive analysis. To make cost sensitive analysis area under ROC curve were implemented false positive on the x axes and true positive in the y axes. The main goal is to have the curve more close to the upper left corner (close to y = 1). WEKA environment allows graphically evaluating of data mining algorithms' performance using ROC curves. From the above experiments naïve Bayes

algorithm with selected attribute showed better area Under ROC curve 77.3% and multi-layer perceptron and J48 showed 74.7% and 70.3% respectively showed.

Information of every fold for 10fold cross validation especially error rate of each fold for J48 algorithm on WEKA experimenter environment was observed. The values of error rate describe whether the classification algorithm is over-fitted or not by checking every cross validation folds.

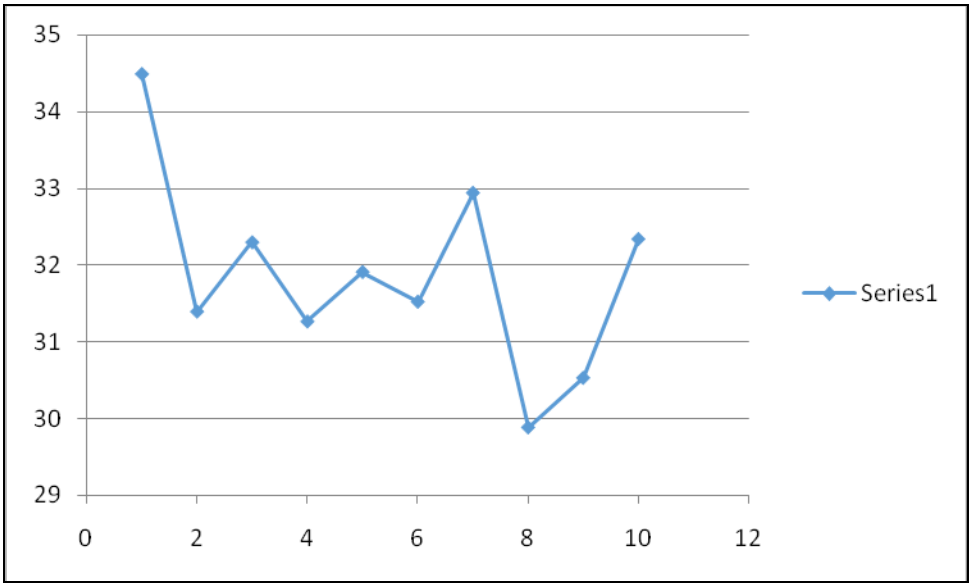


Figure 4.5.: *Error rate of each fold for J48 algorithm*

From the output it is observed that error rate range from 29% to 35% for J48 algorithm unpruned with all attribute for each 10 folds the highest error rate was observed for the first fold and most of the folds failed between the error of 31% and 33%.

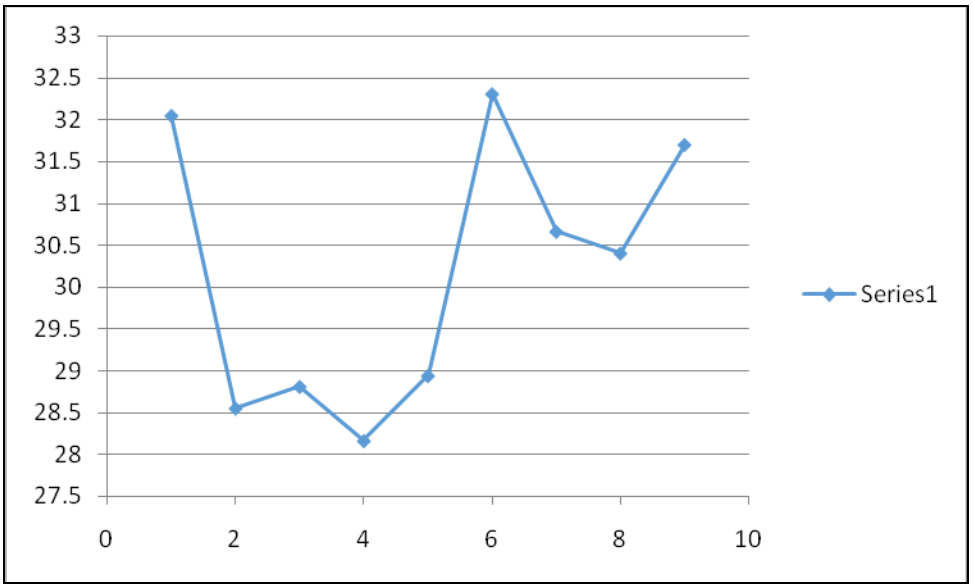


Figure 4.6.: *Error rate of each fold for naïve Bayesian algorithm*

For naïve Bayesian with all attribute the error rate for each fold described by above the graph. It showed that the error rate ranges from 28% to 32.5%. The graph doesn't show liner relation between error rate and cross validation fold change. Higher error rate was observed first fold, sixth fold and the ninth fold and the lowest error rate was observed for the second, third, fourth and fifth folds.

Some specific rules

In this study, J48 classifier has achieved relatively the highest classification accuracy as compared to Naïve Bayes and Multilayer Perceptron. The following rules are some of the rules extracted from pruned J48 algorithm with selected attribute

Rule1. If residence = Rural and highest education level = higher then yes to antenatal care

This rule shows the effect of women education attainment on antenatal care follow-up.

Rule2. If residence = Rural and highest education level =has education and frequency Of Listening Radio =At Least Once A week then yes to antenatal care

Rule3. If Residence = Urban and Frequency of Watching Television = Not At All and Region = Amhara and Highest Education Level = No Education and Frequency of Listening Radio = Not At All then no to ante natal care

Rule4. If Residence = Urban and Frequency of Watching Television = Not At All and Region = Amhara and Highest Education Level = No Education and frequency Of Listening Radio = less Than Once A week then yes to ante natal care

Rule 2, rule 3and4 describe the effect media on antenatal care

Rule5. If residence=Rural and highest education level=has education and Frequency of Listening Radio = Not At All and region Gambela and Religion Muslim then yes to antenatal care

Rule 6 If residence Rural and highest education level = has education and Frequency of Listening Radio = Not At All and region =Gambela and Religion = other then no to antenatal care

Rule 5 and 6 show effects social and region effect on antennal care

Rule 7 Residence = urban and frequency of watching television = not at all region = Amhara husband education = secondary then no to antenatal care

Rule 8 Residence = urban and frequency of watching television = less than once a week and region = Amhara and husband education = secondary then yes to antenatal care

Rule 9 Residence = urban and frequency of watching television = not at all and region = Addis Ababa and husband education = secondary then yes to antenatal care

Rule 10 Residence = urban and frequency of watching television = at least once a week and region = Addis Ababa husband education = no education then yes to antenatal care

Rule 11 Residence = urban and frequency of watching television = not at all and region = Addis Ababa husband education = primary then yes to antenatal care

From the above rules it is observed the effect of education, residence, and media exposure on antenatal care follow up.

Rule 12 residence = urban region = Hararie husband education secondary then yes to antenatal care

Rule 13 residence = urban region = Hararie husband education no education then no to antenatal care

CHAPTER FIVE

CONCLUSION AND RECOMMENDATIONS

5.1. Conclusion

According to health expert and EDHS 2011 survey one of the leading factor of maternal death in sub-Saharan Africa including Ethiopia is traditional birth attendance instead of proper antenatal care follow up with skilled provider that is from doctor, nurse, or midwife. According to the survey about six in every ten Ethiopian women (57 percent) did not receive any antenatal care for their last birth in the five years preceding the survey. The survey include women instance with corresponding socio-economic and cultural attribute describe social, economic and personal background.

With the current rapid increase in the amount data generated and collected in health care and the wide spread availability of cheap computing equipment, applying data mining in health care significantly enhance health care quality . The complexity in volume and nature of health care data has made necessary the application technique beyond the traditional statistics. From the literature reviewed although its late data mining applications are started gaining importance in the health care.

Some issues constrained application of data mining in health care in Ethiopia. Health care data usually collected manually and distributed between hospitals and health centers, collected data are not available in a centralized repository in national level and regional level, and accessibility of data because of privacy issue are limitations to apply data mining in health sector

The main goal of the research was to observe applicability of data mining algorithm and identify the most common data mining algorithms, implemented in EDHS 2011 data set, and evaluate their performance on Different test situation.

This research applied three different data mining algorithm for classifying the target dataset. From the process the algorithm develop feature pattern what makes a woman to follow or not to follow antenatal care during pregnancy. Data mining models were developed using three data

mining classification modeling techniques. The classifier extracts hidden knowledge from EDHS 2011 women data base. The models are trained and validated against a test dataset. Cost sensitive evaluation method ROC curve of the classifier and cost insensitive evaluation method precision recall and f-measure methods were used to evaluate the effectiveness of the models. All three models are able to extract patterns in response to the predictable state. The most effective model to predict women history of antenatal care follow-up appears to be Decision Trees followed by Naïve Bayes and Neural Network. From this the research question "What data mining algorithms are currently used in health care data mining?'' was answered.

In the research tree pruning effect on performance of J48 algorithm on two situation was observed (selected and all attribute) tree pruning showed performance improvement when run on all attribute. In selected attribute it shows computational space improvement with decrease of accuracy. Attribute selection showed improved result for unpruned J48 than pruned J48 and concerned on multi-layer perceptron it showed improved performance when run with selected attribute.

It was difficult to find the best testing configuration for all of the algorithms. Each metric had similar values for all of the configurations. However, there was a slight tendency favoring use training set and 80%split. On the other hand, in most of the experiments the worst configuration was the 5fold cross validation.

Generally it is difficult to make a decision that one classification algorithm is better than another because it may work well in a certain dataset but not in others. Evaluation on performance of a classification algorithm is measured on its accuracy. However, other factors such as computational time or space are also considered to have a full picture of each algorithm. But in this study, pruned J48 classifier with selected attribute on 80% split test has achieved relatively the highest classification accuracy as compared to Naïve Bayes and Multilayer Perceptron. From this research question "Which data mining algorithm implemented give the most accurate result in the target dataset?" was answered.

The objective of the study was identifying factors affecting antenatal care follow up in Ethiopia on EDHS 2011 dataset by using J48, naïve Bayes, and neural network classification techniques.

It is observed that the three classifier successfully extracted rules that are helpful for health sector stakeholders. Although promising results were achieved by pruned J48 classifier some records were classified incorrectly. Hence future research directions are recommend to come up with solution misclassification.

5.2. Recommendations

This research attempt to apply data mining in health sector, specifically women antennal care follow-up on EDHS 2011 using WEKA tool. The main objective of this study was finding factor patterns affecting ante natal care which was achieved. Based on the findings of this study, it is recommended that the following issues need to be addressed in future studies.

- To enhance the performance of the algorithm it is recommended to use algorithm boosting mechanism and class balancing mechanism.
- The need to develop knowledge base for health centers, extracted rules from EDHS 2011 dataset by applying data mining classifier algorithm so as to support decision maker and health expert if the extracted rules accessible and deposited to centralized repository.
- Develop decision support system for policy maker; the rules extracted from dataset using data mining classifier describe patterns of social and economic factors affecting antenatal care.
- Develop interface for output user, it is difficult for domain experts to extract and observe classifier output. Incorporate WEKA tool with java will made rule extraction simple.
- Develop database for all health center to identify social, economic and other factor to kebele level, which help decision maker to have specific information and better decision
- For future work it will be better to include evaluation of different algorithm for different health data set to solve different problem.

References

[1]. Bahilu T., Abebe G. and Yohannes D. (2009). Factors affecting antenatal care utilization in Yem Special Woreda. *Ethiopian Journal of Health Sciences*, 19:46-49.

[2]. Ethiopian Society of Population Studies. (2008). Maternal health care seeking behavior in Ethiopia: findings from EDHS 2005. UNFPA. Addis Ababa, Ethiopia.

[3]. Ruben D.C.Jr (2009). Data Mining in Healthcare: Current Applications and Issues. August 5.

[4].Witten, I. and Frank, E. (2005). Data mining: Practical Machine learning tools and techniques. 2nd edition. Morgan Kaufman publishers, San Francisco, California, USA.

[5]. Koh H. C. and Tan G. (2005). Data mining application in health care. *Journal of Healthcare Information Management,* 19(2): 64-72

[6]. Alemayehu T., Haidar J. and Habte D. (2011). Utilization of antenatal care service among teenagers in Ethiopia: A cross sectional study. *Ethiopian Journal of Health Development*, 24(3).

[7]. Bekri F. E., Govardhan A. (2011). Association of Data Mining and Healthcare Domain: Issues and Current State of the Art. *Global Journal of Computer Science and Technology*, 11(21):1-9.

[8]. WHO (2005) The World Health Report: Make Every Mother and Child Count. World Health Organization, Geneva.

[9]. Yoo I., Alafaireet P., Marinov M., Pena-Hernandez K., Gopidi R., Chang J.F., and Hua L. (2011). Data Mining in Healthcare and Biomedicine: A Survey of the Literature. *J Med Syst.*, 36(4): 2431-48.

[10]. Tremblay M.C., Dutta K. and Vandermeer D. (2010). Using data mining to discover bias patterns in missing data. *Journal of Data and Information Quality*, 2(1):23-55.

[11]. Wirth R. and Hipp J. (2000). CRISP-DM: Towards a Standard Process Model for Data Mining. Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining, 29-39, Manchester, UK.

[13]. Fayyad U., Piatetsky-Shapiro G., and Smyth P. (1996). From data mining to knowledge discovery in databases. *Artificial Intelligence Magazine,* 17 (3):37-53.

[14]. Ngai E.W.T., Hu Y., Wong Y.H., Chen Y. and Sun X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems,* 50: 559–569.

[15]. Zaki M. J. (2003). Data Mining Techniques. WSPC Lecture Notes Series, 12:10.

[16]. Wu X., Kumar V., Ross Quinlan J., Ghosh J., Yang Q., Motoda H., McLachlan G. J., Ng A., Liu B., Yu P. S., Zhou Z., Steinbach M., Hand D. J., Steinberg D. (2008). Top 10 algorithms in data mining. *Knowledge Information System*, 14:1–37.

[17]. Santos R.S., Malheiros S.M.F., Cavalheiro S. and Parente de Oliveira J.M. (2013). A data mining system for providing analytical information on brain tumors to public health decision makers. *Computer methods and programs in biomedicine,* 109(3): 269-282.

[18]. Fu Y. (1997). Data mining: tasks, techniques and applications. *IEEE Potentials*, 16(4):18-20.

[19]. Milovic B., Milovic M. (2012). Prediction and decision making in health care using data mining. *International journal of public health science*, 1(2): 69-78.

[20]. Ruben D. Canlas Jr. (2009). Data mining in healthcare: Current applications and issues. Carnegie Mellon University Australia.

[21]. Mary K., Obenshain, MAT. (2004): Application of Data Mining Techniques to Healthcare Data. *Statistics for Hospital Epidemiology*, 25(8): 690-695.

[22]. Kaur H. and Wasan S.K. (2006). Empirical study on application of data mining techniques in healthcare. *Journal of Computer Science,* 2(2):194-200.

[23]. Daciuk T. and Jou S. (2010). An introduction to data mining and predictive analytics. *CASCON*: 345-346.

[24]. Simkhada B., Van Teijlingen E.R., Porter M. &Simkhada P. (2008). Factors affecting the utilization of antenatal care in developing countries: systematic review of the literature. *Journal of Advanced Nursing*, 61(3): 244–260.

[25]. Shearer C. (2000). The CRISP-DM Model: The new blueprint journal of data mining. *Journal of Data Warehousing*, 5(4):3-15

[26].WHO & UNICEF (2003) Antenatal Care in Developing Countries: Promises, Achievements and Missed Opportunities: An Analysis of Trends, Levels, and Differentials: 1990–2001. WHO & UNICEF, Geneva, New York.

[27]. Adewoye K.R, Musa IO, Atoyebi O.A., Babatunde O.A. (2013). Knowledge and Utilization of Antenatal Care Services by Women of Child Bearing Age in Ilorin-East Local Government Area, North Central Nigeria. *International Journal of Science and Technology*, 3(3):188-193.

[28]. Mcdonagh M. (1996). Is antenatal care effective in reducing maternal morbidity and mortality? *Health Policy and Planning*, 11(1): 1-15.

[29]. Akazili J., Doctor H.V., Abokyi L., Hodgson A. and Phillips J.F. (2011). Is there any relationship between antenatal care and place of delivery? Findings from rural northern Ghana. *Afr J Health Sci.*, 18:62-73.

[30]. Van Eijk A. M., Bles H.M., Odhiambo F., Ayisi J.G., Blokland I.E., Rosen D.H., Adazu K., Slutsker L. and Lindblade K.A. (2006). Use of antenatal services and delivery care among women in rural western Kenya: a community based survey. *Reproductive Health, 3:2.*

[31]. Low P., Paterson J., Wouldes T., Carter S., Williams M. and Percival T. (2005). Factors affecting antenatal care attendance by mothers of Pacific infants living in New Zealand. *The New Zealand Medical Journal*, 118(1216):1-10.

[32]. Rahman M.M., Islam M.R. and Islam A.Z. (2008). Rural-Urban differentials of utilization of ante-natal health care services in Bangladesh. *Health policy and development journal*, 6(3): 117-125.

[33]. Dairo M.D. and Owoyokun K.E. (2010). Factors affecting the utilization of antenatal care services in Ibadan, Nigeria. *Benin Journal of Postgraduate Medicine*, 12(1):3-16.

[34]. Ajayi IO, Osakinle DC.(2013). Socio Demographic Factors Determining the Adequacy of Antenatal Care among Pregnant Women Visiting Ekiti State Primary Health Centers. *Online J Health Allied Scs*.,12(2):4.

[35]. Worku A.G., Yalew A.W. and Afework M. F. (2013): Factors affecting utilization of skilled maternal care in Northwest Ethiopia: a multilevel analysis. *BMC International Health and Human Rights*, 13(20):1-11.

[36]. Rowe R.E. and Garcia J. (2003). Social class, ethnicity and attendance for antenatal care in the United Kingdom: a systematic review. *Journal of Public Health Medicine*, 25(2):113–119.

[37]. Yousif E.M. and Abdul Hafeez A.R. (2006). The Effect of Antenatal care on the probability of neonatal survival at birth, Wad Medni teaching hospital, Sudan. *Sudanese Journal of Public Health*, 1(4): 293-297.

[38]. Sibte S. and Abidi R. (2001). Knowledge management in healthcare: towards 'knowledge-driven' decision-support services. *International Journal of Medical Informatics*, 63: 5–18.

[38]. Assamnew S. (2011). Predicting the occurrence of measles outbreak in Ethiopia using data mining technology. MSc Thesis, AAU.

[39]. Debela T. (2013). Developing a predictive model for fertility preference of women of reproductive age using data mining technology. MSc Thesis, AAU.

[40]. Dametew A. (2011). Designing a predictive model for heart disease detection using data mining techniques. MSc Thesis, AAU.

[41]. Padhy N., Mishra P., and Panigrahi R. (2012). The survey of data mining applications and feature scope. *Asian Journal of Computer Science and Information Technology*, 2(4): 68-77.

[42]. Maimon O. and Rokach L. (2010). Data Mining and Knowledge Discovery Handbook. 2nd Edition .pP. 1-15.

[43]. Srinivas K., Kavihta Rani B., Govrdhan A. (2010). Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks. *International Journal on Computer Science and Engineering*, 2(2): 250-255.

[44]. Manjusha K. K, Sankaranarayanan K., Seena P. (2014). Prediction of Different Dermatological Conditions Using Naïve Bayesian Classification. International journal of advanced research in computer science and software engineering, 4(1): 864-868.

[45]. Mingers, J. (1989). An Empirical Comparison of Pruning Methods for Decision Tree Induction. Machine Learning, 4:227-243.

[46]. Koh H. C. and Tan G. (2010). Data mining application In health care. Journal of Healthcare Information Management, 19(2): 64-72.

[47]. Pandey M. (2011).Applications of Data Mining in Healthcare System. International Journal of Technical Teachers. Accessed from: www.ittonline.co.in

[48]. Singh Y., Chauhan A.S. (2009). Neural networks in Data mining. Journal of Theoretical and Applied Information Technology. Accessed from: www.jatit.org.

[49]. Cilimkovic M. (2010). Neural Networks and Back Propagation Algorithm, Institute of Technology Blanchardstown, Blanchardstown Road North Dublin 15, Ireland.

[50]. Mathur N., Kumar S., Kumar S., and Jindal R. (2012). The Base Strategy for ID3 Algorithm of Data Mining Using Havrda and Charvat Entropy Based on Decision Tree. International Journal of Information and Electronics Engineering, 2(2): 253-258.

[51]. Bhardwaj R. and Vatta S. (2013). Implementation of ID3 Algorithm. International Journal of Advanced Research in Computer Science and Software Engineering, 3(6):2277–81.

[52]. Verma S. and Jain N. Implementation of ID3–Decision Tree Algorithm.

[53]. Kaur G. and Chabra A. (2014). Improved J48 Classification Algorithm for the Prediction of Diabetes. International Journal of Computer Applications, 98(22): 13-17.

[54]. Goyal A. and Mehta R. (2012). Performance Comparison of Naïve Bayes and J48 Classification Algorithms. International Journal of Applied Engineering Research, 7(11), 1-5.

[55]. Srivastava1 A., Han E.H., Kumar V. and Singh V. Parallel Formulations of Decision-Tree Classication Algorithms. Kluwer academic publishers.

[56]. Kumar A. Artificial neural network for data mining. I.A.S.R.I., Library Avenue, Pusa, New Delhi.

[57]. Raiwani Y.P., and Panwar S.S. (2015). Data Reduction and Neural Networking Algorithms to Improve Intrusion Detection System with NSL-KDD Data set. *International Journal of Emerging Trends & Technology in Computer Science , 4(1):219-225.*

[58]. Powers D. M. W. (2012). The Problem with Kappa: in Proceedings of the 13th conference of the European Chapter of the Association for Computational Linguistics. pP 345-355.

[59]. Vijayarani S. and Muthulakshmi M. (2013). Comparative Study on Classification Meta Algorithms. *International Journal of Innovative Research in Computer and Communication Engineering,* 1(8): 1768-1774.

[60] Olson D. and Delen D. (2008). *Advanced data mining techniques*. Berlin: Springer.

[61] Han J. and Kamber M. (2006). *Data mining*. Amsterdam: Elsevier.

[62]. Ethiopian Demographic and Health Survey. (2011). Central Statistical Agency Addis Ababa, Ethiopia.

[63]. Wahbeh A.H., Al-Radaideh Q.A., Al-Kabi M.N., and Al-Shawakfa E.M. A comparison study between data mining tools over some classification methods. *International Journal of Advanced Computer Science and Applications, 18-25. Accessed from, www.ijacsa.thesai.org.*

[64]. http://www.systems-thinking.org/dikw/dikw.htm

[65]. Cios, K Witold, P Roman, S and Kurgan A. (2007). Data Mining: A Knowledge Discovery Approach, Springer, New York: USA.

[66]. Yalem Tsegay Assfaw (2010). Determinants of Antenatal Care, Institutional Delivery and Skilled Birth Attendant Utilization in Samre Saharti District, Tigray, Ethiopia. MSc Thesis, Umea University.

[67]. Bereket Y., Mulat T., Wondimagegn P. (2013). Mothers‟ Utilization Of Antenatal Care And Their Satisfaction With Delivery Services In Selected Public Health Facilities Of Wolaita Zone, Southern Ethiopia. *IJSTR*, 2(2): 74-85.

[68]. Tewodros A., Jemal H., Dereje H. (2010). Utilization of antenatal care services among teenagers in Ethiopia: A cross sectional study. *Ethiopia Journal of Health Development*, 24(3):221-225.

[69]. Olson D. L. and Delen D. (2008). Advanced Data Mining Techniques. SPRINGER.

[70]. Siraj F. and Abdoulha M.A. (2011). Mining Enrolment Data Using Predictive and Descriptive Approaches. Accessed from www.intechopen.com.

[71] Introduction to KDD and data mining guyen Hung Son

[72] Deshpande S. P. and Thakare V. M. (2010). Data Mining System and Applications: A Review. *International Journal of Distributed and Parallel systems (IJDPS)*,1(1): 32-44.

[73] Introduction to Data Mining by Tan, Steinbach, Kumar.