ADDIS ABABA UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES

SCHOOL OF INFORMATION SCIENCE

PREDICTING THE UTILIZATION OF SKILLED BIRTH ATTENDANT AMONG ANTENATAL CARE CLIENTS

BY

YISHAK ESHETU

FEBURARY 2018

ABABA, ETHIOPIA

ADDIS ABABA UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES

SCHOOL OF INFORMATION SCIENCE

PREDICTING THE UTILIZATION OF SKILLED BIRTH ATTENDANT AMONG ANTENATAL CARE CLIENTS

A Thesis Submitted to the School of Information Science of Addis Ababa University in Partial Fulfillment of the Requirement for the Degree of

Master of Science in Information Science

By: Yishak Eshetu

Advisor: Million Meshesha (PhD)

February 2018

Addis Ababa, Ethiopia

ADDIS ABABA UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES

SCHOOL OF INFORMATION SCIENCE

PREDICTING THE UTILIZATION OF SKILLED BIRTH ATTENDANT AMONG ANTENATAL CARE CLIENTS

By: Yishak Eshetu

Name and signature of Members of the Examining Board

Million Meshesha (PhD)	____________________	__________________
Advisor	Signature	Date
Tibebe Beshah (PhD)	____________________	__________________
Examiner	Signature	Date
Wondossen Mulugeta (PhD)	____________________	__________________
Examiner	Signature	Date

DECLARATION

This thesis has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree in any university.

This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by citations giving explicit references. A list of references is appended.

Signature: ________________________

Yishak Eshetu

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

Advisor's Signature: ________________________

Million Meshesha (PhD)

Acknowledgement

First and foremost, I would like to thank the almighty God for giving me the strength, ability and opportunity to undertake this study. Without His blessings, this achievement would not have been possible.

My sincere gratitude goes to my advisor Dr. Million Meshesha for his support, encouragement, and insightful comments.

I would also like to extend my gratitude to Berhanu Sahelie, Teshager Mersha and Deslagegn Getaneh for their valuable technical assistance, comments and feedback on my thesis. I also thank you all the Medical experts who are participated and provided their feedbacks for the improvement of the constructed model for this study.

Table of contents

Table of Contents

List of Figures

List of Tables

Acronyms and Abbreviations

ANC	Antenatal care
AUC	Area under the ROC curve
CART	Classification And Regression Tree
CHAID	Chi-squared Automatic Interaction Detection
CMi	Confusion matrix of i
CRISP	Cross Industry Standard process – Data mining
CSA	Central Statistical Agency
DBS	Dried Blood spot
DM	Data Mining
EDHS	Ethiopian Demographic Health Survey
EmOC	Emergency Obstetric Care
EOC	Emergency Obstetric Care
FANC	Focus Antenatal Care
FAR	FALSE Acceptance Rate
FRR	FALSE Reject Rate
HIV/AIDS	Human immunodeficiency virus infection and acquired immune deficiency syndrome
KDD	knowledge discovery in data base
KDP	Knowledge Discovery Process
MDGs	Millennium Development Goal
MMR	Maternal Mortality rate
MOH	Ministry of Health
QMNHCP	Qualified Maternal and Newborn Health Care Professionals
ROC	Receiver Operating Characteristic
SBA	Skilled birth attendance
SBA	Skilled birth attendance
SEMMA	Sample, Explore, Modify, Model and Assess
SPSS	Statistical Package for the Social Sciences
WEKA	Waikato Environment for knowledge analysis
WHO	World Health Organization

Abstract

In Ethiopia maternal deaths are currently estimated at 412 per 100,000 live births and gets worst compared with other countries. It calls an alarm for Ethiopia to reduce such high maternal mortality rate by ensuring all birth has to be attended by skilled birth attendant (SBA). The majority of maternal deaths occur during labor, delivery, and the immediate postpartum period.

Ethiopia is one of the countries with high maternal morbidity and mortality in sub-Saharan Africa which needs more public health care effort in the country. Hence the main objective of the study is to construct a predictive model for the utilization of skilled birth attendant among Antenatal Care clients. The study is guided by experimental research and followed Hybrid Data mining process model to achieve the goal of building predictive model. Andersen's Health Behavioral Model was the basis for the assumption of this study to better understand the problem under study. The most recent data from the EDHS, 2016 is used and analyzed using data mining tools to predict the utilization of skilled birth attendant among ANC users. To build a model which can predict the utilization of skilled birth attendant among ANC users three data mining algorithms are used; J48 Decision tree, PART Rule Induction and Naïve Bayes. The data mining tool employed in this research is WEKA3.8.1.

The experimented result shows that J48 decision tree outperforms by a classification accuracy of 98.74%.

The most important rules generated from the datasets which surprised both domain and non-domain experts to predict the utilization of SBA among ANC users highly depend on the place of residence of the woman, place of region, house hold wealth index, husband highest level of education, and husband occupation. The experts suggested the applicability of the rules for both clinicians and as well as for policy makers.

From this study it is observed that data mining techniques can effectively be used in the health sectors specially to predict the utilization of skilled birth attendant to avert the high maternal mortality rate in Ethiopia.

Chapter One

Introduction

1.1. Background of the Study

1.1.1. Maternal Health

Maternal health care addresses family planning, preconception, prenatal, and postnatal care of mothers and it refers to the health of women during pregnancy, child birth and postpartum period [1]. The maternal health care has to be provided by competent qualified maternal and newborn health care professionals (QMNHCP) who are educated and regulated as per international and national standards and who work as a team within an enabling and supportive environment [2].

As Wanjira [3] suggested, time of Child birth is very critical for mothers and child's life. Therefore, during child birth, mothers should be assisted by skilled birth attendant supported by an enabling environment. Skilled birth attendant is used as an indicator to measure the achievement of Millennium Development Goal (MDG 5) which aims to improve maternal health [4].

Maternal survival has significantly improved since the adoption of the MDGs. The maternal mortality ratio dropped by 45 percent worldwide between 1990 and 2013, from 380 maternal deaths per 100,000 live births to 210. Despite this progress, every day hundreds of women die during pregnancy or from childbirth-related complications. Globally, there were an estimated 289,000 maternal deaths in 2013, equivalent to about 800 women dying each day [5]. Maternal deaths are concentrated in sub-Saharan Africa and Southern Asia, which together accounted for 86 per cent of such deaths globally in 2013 [5].

Most maternal deaths, stillbirths and neonatal deaths are preventable [4]. Access to Skilled Birth Attendance during childbirth and in the immediate post-natal period and access to Emergency Obstetric Care (EmOC) in case of obstetric complications are considered to be effective interventions to reduce the number of global maternal and newborn deaths [4].

A key strategy for reducing maternal morbidity and mortality is ensuring that every birth occurs with the assistance of skilled health personnel, meaning a medical doctor, nurse or midwife. Progress in raising the proportion of births delivered with skilled attendance has been modest over the course of the MDG time frame, reflecting lack of universal access to care [6].

Globally, the proportion of deliveries attended by skilled health personnel increased from 59 percent around 1990 to 71 per cent around 2014 that decreased the maternal mortality from 380 to 210 and contributed the major part in saving the life of both mothers and child but the goal of reducing maternal mortality remains underachieved in sub-Saharan African countries [6].

In Ethiopia, a lot of efforts have been done by Ethiopian Federal Ministry of Health together with other partners to reduce the maternal mortality rate towards Millennium development goal. However, compared with the Global MDG5 achievement; it is very far, considered as underachievement, Maternal mortality per 100,000 live births was 1,400 in 1990 and had declined to 990 in 2000. According to the EDHS, 2016 survey report, the figure declined to 412 in 2016, it was target to decrease the number by 75% at the end of 2015 [10] and therefore ,it needs further efforts to reduce the leading causes of maternal death in the country [7].

The lesson learnt from MDGs implementation, the 2030 agenda for sustainable development has been developed with the objective of reducing the global maternal mortality ratio to less than 70 per 100,000 births, with no country having a maternal mortality rate of more than twice the global average [3]. This calls a lot for Ethiopia because MMR is 412 according to EDHS, 2016 [8].

The proportion of births delivered by skilled birth attendant has been identified as the maternal health intervention indicator with the most pronounced economic related inequality [3].

Most maternal deaths are preventable as the health-care solutions for preventing or managing the complications of pregnancy and childbirth are well known. The main causes of death are postpartum hemorrhage (24%); indirect causes such as anemia, malaria, and heart disease (20%); infection (15%); unsafe abortion (13%); eclampsia (12%); obstructed labor (8%); and ectopic pregnancy, embolism, and anesthesia complications (8%) [9].

Maternal health status in Ethiopia is one of the worst in the world. The country is characterized by high maternal and child mortality. The maternal mortality rate was estimated at 412/100,000 according to the 2016 EDHS [10]. It is noted that there has been a minimal change in maternal mortality in five years from 676/100,000 in 2011 to 412/100,000 in 2016 [10, 11]. According to 2016 EDHS and 2016 world health statistics, Ethiopia maternal mortality ratio is one of the highest, 412 and 353 mothers die per 100,000 live births respectively [8,10,11].

The most recent EDHS 2016, shows that a large number of births (72%) in Ethiopia took place at home yet for 62% of the births the mothers had received antenatal care from a skilled provider. Thus, only 28% of the births were attended by a skilled provider [8].

1.1.2.Skilled Birth Attendance

Skilled birth attendance (SBA) is the process by which a woman is given adequate care during labor, delivery and the early postpartum period [12]. This requires a skilled birth attendant or skilled personnel to attend the delivery in an enabling environment, which includes adequate supplies and equipment, transport and effective communication systems [12].

Most obstetric complications could be prevented or managed if women had access to a skilled birth attendant – doctor, nurse, midwife – during childbirth. Globally coverage of skilled attendant during childbirth increased from 61% in 2000 to 78% in 2016. However, despite steady improvement globally and within regions, millions of births were not assisted by a midwife, a doctor or a trained nurse. In sub-Saharan Africa approximately only half of all live births were delivered with the assistance of skilled birth attendant in 2016 [13].

Improvements in the coverage of the proportion of births attended by skilled health personnel and their provision of care may have contributed to declines in maternal mortality between 1990 and 2015. However, the estimated coverage of births attended by skilled health personnel in 2016 shows inequality between WHO regions as only half of the births in the sub-Saharan Africa Region, where maternal mortality is highest, are attended by skilled health personal whereas in the other WHO regions over 70% to 99% of all births are attended by skilled health personnel [13].

1.1.3. Antenatal Care

Antenatal care is the process of recording medical history, assessment of individual needs, advice and guidance on pregnancy and delivery, screening tests, education on self-care during pregnancy, identification of conditions detrimental to health during pregnancy, first-line management and referral if necessary [14]. It is more effective in preventing adverse pregnancy outcomes when sought early in the pregnancy and continued through to delivery. Under normal circumstances, WHO recommends that a woman without complications should have at least four ANC visits [14].

Antenatal care (ANC) from a skilled provider is important to monitor the pregnancy, detect malformation problems and reduce the risks for mother and child during pregnancy and at delivery. ANC visits constitute one of the few times women in many resource-poor settings seek care for their own health and represent an important opportunity to help women best prepare for birth, as well as inform them about pregnancy-related complications, and the advantages of skilled birth attendance [15].

Beyond these roles of detecting malformation problems and other risk factors, Bernis et al [16] agree that in areas where skilled birth attendance remains uncommon, this educational role of ANC is far from negligible in importance, as the timely use of qualified personnel reduces the risk of death for both the mother and new-born.

1.1.4.Data Mining

Data mining can be defined as the process of finding previously unknown patterns and trends in data and using that useful information to build predictive models [25]. Alternatively, it can be defined as the process of data preparation and building models using vast data stores to uncover previously unknown patterns [26].

Data mining helps in healthcare organizations to pass quality decision in customer relationship management during diagnosis and treatment, to optimize healthcare resource management [27]. It also enables physicians identify effective treatments and best practices, as well as patients receive better and more affordable healthcare services. In medical decision support (to doctors) data mining can apply in analysis of digitized images of skin lesions to diagnose melanoma, in computer-assisted texture analysis of ultrasound images aids monitoring of tumor response to chemotherapy. Predicting the presence of brain neoplasm with magnetic resonance spectroscopy and in analysis of digital images of tissue sections to identify and quantify senile plagues for diagnosing and evaluating the severity of Alzheimer's disease [27].

Another factor is that the huge amounts of data generated by healthcare transactions are too complex and voluminous to be processed and analyzed by traditional methods. Data mining can improve decision-making by discovering patterns and trends in large amounts of complex data [28]. Such analysis has become increasingly essential as financial pressures have heightened the need for healthcare organizations to make decisions based on the analysis of clinical and financial data. Insights gained from data mining can influence cost, revenue, and operating efficiency while maintaining a high level of care [29].

Healthcare organizations that perform data mining are better positioned to meet their long-term needs [30]. Data can be a great asset to healthcare organizations, but they have to be first transformed into information. As noted by SIAM "in healthcare, data mining is becoming increasingly popular, if not increasingly essential."

Data mining can be considered a relatively recently developed methodology and technology, coming into prominence only in 1994 [31]. It aims to identify valid, novel, potentially useful, and understandable correlations and patterns in data by combing through many data sets to sniff out patterns that are too subtle or complex for humans to detect [28].

1.2. Statement of the Problem

Since 1990, the maternal mortality ratio has declined by 45 per cent worldwide, and most of the reduction has occurred since 2000 but it is a big challenge in much of Sub-Saharan Africa, with maternal death estimates still as high as live births in some countries according to the millennium development report [5]. In Ethiopia maternal deaths are currently estimated at 412 per 100,000 live births [8]. Skilled birth attendants could help reduce maternal mortality by 75% [17]. It calls an alarm for Ethiopia to reduce such high maternal mortality rate by ensuring all birth has to be attended by skilled birth attendant. According EDHS 2016, compared with skilled birth attendant, women attended ANC services are 62% but SBA attendant is only 28% [8].

There were researches carried out in the relationship between ANC and SBA, findings showed us there is a limitation on the relation between antenatal care and skilled birth attendance. Some studies only use the number of antenatal visits to study the relationship but neglect to investigate the impact of the quality of care on neither possible skilled attendance use nor the relationship linking the frequency of visits to the quality of the services [18].

Studies in developing countries have been on the determinants of maternal and child health. However, studies see the effect of antenatal care on women's decision to utilize skilled birth attendance has received less attention by researchers. Gasmsch and Campbell [19] reviewed literature on the determinants of delivery service use and found that most of the studies focused on how socioeconomic, demographic and cultural factors influenced utilization using only limited number of data sets. This study [19] further shows that residence and parity are determinant factors for pregnant mothers to use Skilled birth attendant on limited data sets using statistical tools. Other studies that have been of interest are on ANC's impact on maternal and child health outcomes like birth weight. In Ethiopia studies have not critically focused on the role of ANC in predicting SBA rather than explore determinants of maternal health care in general.

Most of the studies done in Ethiopia so far used statistical techniques such as regression analysis on a limited set of data to assess the factors which contribute to the underutilization of maternal health care service utilization. Since the analysis made by using these methods focuses on problems with much more manageable number of variables and cases than may be encountered in real world databases, they have limited capacity to discover new and unanticipated patterns and relationships that are hidden in conventional relational databases.

Therefore, the aim of this study is to apply data mining classification techniques to predict the utilization of skilled birth attendant during delivery among pregnant mother following Antenatal care by skilled birth attendant. The researcher considered selective techniques and tools which will predict the utilization of skilled birth attendant among ANC clients using EDHS 2016 dataset.

To this end the following research questions are formulated for investigation and find answer for the problem at hand:

- What are the major factors that contribute pregnant mothers with ANC follow up to end with skilled birth attendant during birth in Ethiopia?

- Which data mining classification algorithm is more appropriate to construct a model for ANC Clients pattern to predict Skill Birth Attendant?

1.3. Objective of The Study

1.3.1. General Objective

The general objective of the study is to construct a predictive model for the utilization of skilled birth attendant among Antenatal care clients.

1.3.2. Specific Objectives

In order to accomplish the general objective of this study carried out the following specific objectives:

- To conduct a literature review concerning antenatal care client and those requiring skilled birth attendant.
- To prepare good quality dataset for data mining techniques.
- To select suitable classification algorithm for predictive modeling.
- To construct a model that can predict utilization of skilled birth attendant among antenatal care clients.
- To evaluate the performance of the constructed predictive model by domain experts for its applicability.

1.4. Significance of the Study

The finding from this research will be significant to/for:

- Policy makers & executives to formulate/develop a system to minimize missed opportunity & improve skilled birth attendant rate to decrease maternal mortality rate.
- Pregnant mothers to be assisted by skilled birth attendant at health institution during delivery.
- Antenatal care health service providers to advocate the latter use of Skilled birth attendant for pregnant mothers during their ANC visit.
- A researcher for further investigation on the findings since this research was done on Secondary Data.

1.5. 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. In this research predictive model has been constructed to utilize skilled birth attendant. 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 from Ethiopia demographic health survey data but some of the variables such as distance to health facility, availability of transportation is not included. This research attempted to apply three DM techniques such as J48 decision tree, naïve Bayes, and PART rule induction in predicting factors affecting antenatal care clients to utilize skilled birth attendant. It was aimed to include large number of women throughout the country; the data used in this study was only 4,933 women instances that had ANC visits only.

1.6. Methodology

We need to have research methodology to see the overall process/steps of the research activities to be done for data mining research. It helps to ensure that sufficient checks have been done in each steps. Data mining research falls in the category of experimental research because experimental research is data-based research that helps to examine the effect of a variable or treatment which is known as experimental variable [92]. In this study, the experimental variable is utilization of skilled birth attendant to be examined through factors affecting its utilization among independent variables.

Following the experimental study, Hybrid datamining process model was followed because it combines the best features of CRISP and KDD process models to build a predictive model using data mining techniques. The study was done based on the EDHS 2016 data.

1.6.1.Research Design

To achieve the objectives of this study, the researcher has used Hybrid datamining process model to build the predictive model of utilization of skilled birth attendant among Antenatal care client's using data mining techniques.

The features of Hybrid methodology are iterative and interactive in each of its six steps.

The following section discusses in detail about the steps of hybrid knowledge discovery process in detail.

1.6.2.Understanding the Business Domain

Problem domain understanding focuses on understanding the project objectives and requirements from a business perspective, then converting this knowledge into a data mining problem definition and a preliminary plan designed to achieve the objectives [21].

The first step in the hybrid KDP process is to understand the problem domain to define the Problem under study and determine the research goals and learning about current solution to the problem.

In this study appropriate health framework is identified to better understand the problem under study. The researcher discussed models to select health behavior model which can affect utilization of individual in the utilization of health facilities. Hence these models are briefly discussed in chapter four such as Andersen Health Behavioral Model, the Health Belief Model. The Theory of Reasoned Action (TRA) and The Theory Planned Behavior (TPB), and the Trans-Theoretical model of Health Behavior Change Model then Andersen Health model was selected and used to understand the problem of this study.

Andersen's Health Behavioral Model was the basis for the assumption of this study that assumes that predisposing, enabling and need factors, including social support are important factors determining women's utilization of health facilities for delivery by skilled birth attendant [22].

According to the latest version, Andersen's [23] classify conditions that facilitate or impede utilization of health services in to Predisposing factors, Enabling factors and Need factors.

From the predisposing factors the researcher has been selected: Mother's Age at birth, Mother Education level, Partner Education level, marital status, Birth Order, Religion, Partner Occupation, and Mother Occupation. From the enabling factors the researchers have been selected: Region, Residence, Access to Media, and Wealth Index. Finally, three variables were selected from Need factors: Timing of 1st ANC visit in months, Frequency of ANC Visit, and Core component of ANC services.

Finally, the researcher briefly discussed with the domain experts on the selected framework so as to select factors which affect the utilization of skilled birth attendant as per the Andersen's Health Behavioral Model considering Ethiopia context, and all appropriate variables selected and finally translated the goals into data mining objectives and initial selection of data to be used was performed in the next step [8].

1.6.3.Understanding the Data

After selected the appropriate health belief model, the next steps are analyzing and understanding the data to be used in this study. The source of data for this research was 2016 EDHS dataset available at www.measuredhs.com website after submitting the research objectives and statement of the problem. The principal objective of the 2016 Ethiopia Demographic and Health Survey (EDHS) is to provide current and reliable 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. Among these EDHS has collected high quality data on family health, including immunization coverage among children, prevalence and treatment of diarrhea and other diseases among children under age five, and maternity care indicators, including antenatal care, assistance at delivery and postnatal care.

Administratively, Ethiopia is divided into nine geographical regions and two administrative cities. The sample for the 2016 EDHS was designed to provide estimates of key indicators for the country as a whole, for urban and rural areas separately, and for each of the nine regions and the two administrative cities and it includes all the components of maternal health care services uptake.

In this step data collected from relevant source as mentioned above, decided the data to be used to mine solution for the problem under study. Data are checked for completeness, redundancy, missing values, plausibility of attribute values, etc. Finally, 4,933 instances were selected those of who have ANC follow up.

1.6.4. Data Preparation

The third step is Data preparation comprises those techniques concerned with analyzing raw data so as to yield quality data, mainly including data collecting, data integration, data transformation, data cleaning, data reduction, and data discretization used in this phase [24].

In this phase data were cleansed such as handling missing values, followed by data integration, and transformation and finally data was converted to the necessary format for mining software

This is the most crucial phases in which the success of the entire knowledge discovery process depends. In this phase, the final dataset is also constructed from the initial raw data. Data preparation tasks were performed in iterative ways.

The major tasks include: finding out distinct value, filling missing values, and data transformation/reduction activities were also undertaken in this phase.

Data cleaning (or data cleansing) routines were applied to fill in missing values (with the most frequented or modal value), smooth out noise (by removing the record), and detect outliers (by substituting with modal values) in the data. To support the preprocessing tasks, the researcher used Microsoft Excel, 2010 and Weka tool 3.8.1 to make the data suitable to train the selected algorithms with training sample and testing case. In this study the following the following major tasks performed in the data preparation steps:

- From 5,704 attributes by referring the selected framework and consultation with domain experts 17 attributes were selected.
- Missing value is corrected using Mean for Timing of first ANC visit and Frequency of ANC visit.
- Transformed and discretized variable is Mother Age at birth
- Three variables are discretized such as Birth order, First ANC visit and Frequency of ANC visit.
- The two Transformed variables are Access to media, and Core Components of ANC Visit

1.6.5.Predictive Modeling

Given the cleaned data, intelligent methods are applied in order to extract data patterns. Patterns of interest are searched for, including classification rules or trees, regression, clustering, sequence modeling, dependency, and so forth [24].

Weka 3.8.1 machine learning software was used for analyzing the preprocessed data to build the model. J48 Decision tree, PART rule based induction and Naïve Bayes algorithm were used among the available algorithms due to their popularity in the recently published papers and due to the objective of the research.

1.6.6.Evaluation of the Discovered Knowledge

Thus after the development of the model based on the training dataset, the accuracy of the model was tested using test datasets.
For evaluating classification models a test dataset is prepared and used. Then, accuracy, recall, precision, F-measure and ROC curve are used to evaluate the performance of the classification algorithms so as to select the best predicting model for this study.

The second evaluation was done by the domain experts on the selected rules to check whether the discovered knowledge is novel and interesting even for domain experts. Interpretation of the results by domain experts is done on the constructed model and accepted patterns by domain experts are further presented for its usage in the following sections.

1.6.7. Use of the Discovered Knowledge

In order to show how to use the discovered knowledge for the domain expert, the researcher design user interface by using Vb.net in such a way that the agreed interesting rule generated from WEKA data mining tool by domain experts will be an input for a designed prototype and its importance is evaluated by the domain experts so as to roll out for maternal health centers to assist the ANC service providers. The result of this study will be presented and disseminated to different organization/bodies such as Addis Ababa University, School of Information Science and Public Health submitting original copy and every effort will be made to disseminate the results of the study through the following ways:

- Presentation for school of Information Science

- Publishing in different journals
- Presentation on different conferences/workshops for the concerned bodies
- Putting the hardcopy in the libraries for further reference.

1.7. Ethical Consideration

All the data used for the study were made available upon request from the agencies website for academic purpose. Furthermore, the data were in aggregate form and no data was collected at the individual level; hence, anonymity or confidentially issues did not arise.

1.8. Organization of the Thesis

The research is presented in six chapters.

- ✓ Chapter one discusses and presents issues related to maternal health care, existing challenges to avert maternal mortality, the role of skilled birth attendant to tackle the high mortality rate of Ethiopia, the objectives of the study, significance of the study and research methodology.
- ✓ Chapter two contains of literature review relating to factors associated with the utilization of skilled birth attendant at health facility. It reviews also data mining concepts and its application in health care environment.
- ✓ Chapter three discusses the mining techniques and methods used for creating the classification models using Decision tree, Rule induction and Naïve Bayes algorithms, and also discuss the performance evaluation of models.
- ✓ Chapter four discusses in detail about problem domain understanding, data understanding, data preparation and final product of this chapter is statistical summarized data, ready for WEKA for model building
- ✓ Chapter five presents the experiment design in detail; model generated by each of the calcification model, model evaluation, presentation and selects the best model for the problem under study.
- ✓ Chapter six presents conclusion and recommendations of the study based on the findings.

CHAPTER TWO

2. Literature Review

This chapter discusses in depth about how data mining is solving the health care which is embedded in large databases by analyzing patterns, using data mining techniques and tools, various concepts, theories and practices and it also briefly discuss the concept of maternal health care and review all the related research conducted in the maternal health sector.

2.1. Overview of Maternal Health Care

Most maternal deaths are preventable as the health-care solutions for preventing or managing the complications of pregnancy and childbirth are well known. The main causes of death are postpartum hemorrhage (24%); indirect causes such as anemia, malaria, and heart disease (20%); infection (15%); unsafe abortion (13%); eclampsia (12%); obstructed labor (8%); and ectopic pregnancy, embolism, and anesthesia complications (8%) [36].

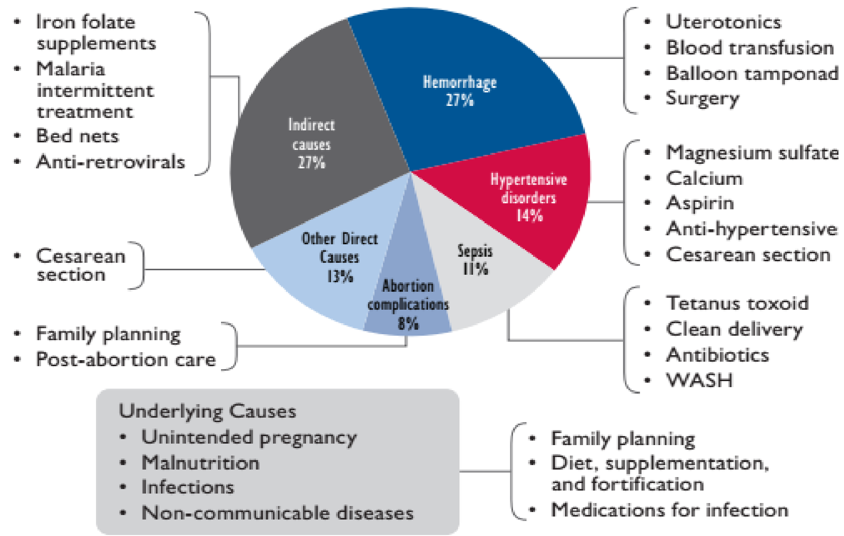

Figure 1. Global Maternal causes of death and selected key intervention [36].

Based on WHO systematic analysis on Global causes of maternal death on 417 datasets from 115 countries shows about 73% of all maternal deaths between 2003 and 2009 were due to direct obstetric causes and deaths due to indirect causes accounted for 27.5% of all deaths. Hemorrhage accounted for 27.1% hypertensive disorders 14.0% and sepsis 10.7% of maternal deaths. The rest of deaths were due to abortion 7.9% embolism 3.2% and all other direct causes of death 9.6% [37].

The direct and indirect causes of maternal death are well-known and there are effective interventions to mitigate them. Those interventions can be best delivered through quality maternity care provided by skilled health providers in facilities who are working in teams to ensure that all women can be attended throughout the antepartum, intrapartum, and postpartum periods and with backup support through referral mechanisms [38].

Only in 2015, Worldwide an estimated 303,000 maternal deaths (MMR= 216) occur of which the overall MMR in developing regions is 239 which is roughly 20 times higher than that of developed regions (MMR=12). Developing regions account for approximately 99% (302,000) of the estimated global maternal deaths in 2015, with sub-Saharan Africa alone accounting for roughly 66% maternal deaths [33].

Maternal survival has significantly improved since the adoption of the MDGs. The maternal mortality ratio dropped by 45 per cent worldwide between 1990 and 2013, from 380 maternal deaths per 100,000 live births to 210 [34]. Despite this progress, every day hundreds of women die during pregnancy or from childbirth-related complications. Maternal deaths are concentrated in sub-Saharan Africa and Southern Asia, which together accounted for 86 per cent of such deaths globally in 2013. According to 2016 world health statistics Ethiopia maternal mortality ratio is one of the highest, 353 mothers die per 100,000 live births [34, 35].

Focus Antenatal Care (FANC) is a systematic assessment and follow up of pregnant women that include education, counseling, screening and treatment to ensure the best possible health of the mother and fetus [39].

Antenatal care can also contribute to successful pregnancy outcomes by encouraging women to obtain skilled care for labor and delivery. According to WHO estimates, more than half of all women give birth without the assistance and supervision of a skilled birth attendant [94]. A study of 300 women from low- and middle-income families in urban India showed that those who received a relatively high level of antenatal care were four times more likely than those who had little or no antenatal care to deliver with a skilled attendant [95]. Antenatal care providers can help women and their families find a place to give birth, a skilled attendant, and the essential items necessary for a clean delivery. Planning for delivery should also anticipate complications and the need for referral to an appropriate medical facility with the appropriate level of good quality essential obstetric care. It may involve transport arrangements, emergency funds, a family member to accompany the woman and assist in decision making.

Antenatal screening can only be important as a maternal mortality reduction tool if the main causes of maternal mortality have detectable premorbid states but it is the first gate way to skill birth attendance. The top three priorities for the reduction of maternal mortality thought to be universal access to family planning services, skilled attendance at every birth and prompt access to emergency obstetric care when the need arises. Antenatal care services contribute immensely to newborn survival; it is for this reason that they must be strengthened access to antenatal care services will contribute in a little way but will not yield significant reductions in maternal mortality unless a mother has access to skill birth attendance and access to emergency obstetric care [44].

Skilled birth attendance (SBA) is the process by which a woman is given adequate care during labor, delivery and the early postpartum period with a skilled birth attendant (Doctor or midwife) and an enabling environment, which includes adequate supplies and equipment, transport and effective referral systems. ANC and SBA are part of WHO sexual and reproductive health Packages of Interventions to decrease maternal mortality [39].

Delivery through a SBA can help prevent infections through the practice of good hygiene during child birth and can help manage the obstetric complications if supported by a functioning health system [96]. Provision of adequate medical attention during delivery is important for the wellbeing of mother and child. Absence of such care and lack of hygienic conditions at the time of birth may lead to complications that would increase the risk of death of the mother, child or both [97]. Recent research shows that delivery by a SBA serves as an indicator of progress towards maternal mortality worldwide. Rogan and Olvena also agree that the danger of childbearing can be greatly reduced if a woman is healthy and well-nourished before becoming pregnant and if a skilled birth attendant assists the birth.

Since every pregnant, delivering, or postpartum woman is at risk for serious, life-threatening complications, an important goal of antenatal care in developing countries should be to teach women and their families to recognize signs of obstetric complications and respond promptly [98]. Signs and symptoms of pregnancy and labor complications are not always recognized as causes for concern. In rural West African communities, for example, symptoms such as swelling of the feet (a possible sign of pre-eclampsia), late-term spotting or bleeding (a sign of antepartum hemorrhage), and long labors are not viewed as potential medical emergencies [99].

According to 2015 WHO report although more than 71 per cent of births were assisted by skilled health personnel globally in 2014, in the developing regions, only 56 per cent of births in rural areas are attended by skilled health personnel, compared with 87 per cent in urban areas. Only half of pregnant women in the developing regions receive the recommended minimum of four antenatal care visits [34]. According to Ethiopian 2016 DHS, only 62 percent of women who gave birth in the five years preceding the survey received antenatal care from a skilled provider at least once for their last birth and only three in 10 women (32 percent) had four or more ANC visits for their most recent live birth. Slightly over one in 4 live births in the five years preceding the survey were delivered by a skilled provider (28 percent) or in a health facility (26 %) [40].

The majority of maternal deaths occur during labor, delivery, and the immediate postpartum period [41]. As most maternal deaths occur due to obstetric complications, most of these deaths could be prevented if women had access to high-quality maternal health care, including antenatal care, skilled assistance at delivery, and postnatal care [37]. A skilled health professional can administer interventions to prevent and manage life-threatening complications, such as heavy bleeding, or refer the patient to a higher level of care when needed [33]. By increasing the use of skilled birth attendance at delivery, antenatal care can have an indirect influence on the survival of mothers and children [35].

2.2. Overview of Data Mining Technology

With the advent of computers and means for massdigital storage we started collectingand storing all sorts ofdata, counting on the power of computers to help sortthrough this amalgam of information. This massivecollections of data stored on disparate structures very rapidlybecame overwhelming and led to the creation of structureddatabases and database management systems (DBMS).Thedatabase management systems efficiently manage largecorpus of data and effective and efficient retrieval ofparticular information from a large collection wheneverneeded and also contributes to recent massive gathering ofall sorts of information. This retrieval of data as and whenneeded contributes the technology of data mining[56]. Datamining can be viewed as a result of the natural evolution ofinformation technology. This technology provides a wideavailability of huge amounts of data and the imminent needfor turning such data into useful information and knowledge.Data mining is the extraction of interesting patterns orknowledge from huge amount of data. It can be known bydifferent names like knowledge discovery (mining) indatabases (KDD), knowledge extraction, data/patternanalysis, data archeology, data dredging, informationharvesting, business intelligence [56].

As data mining has been evolved, and continues to evolve, from the intersection of researchin such fields as databases,machine learning, pattern recognition, statistics, artificial intelligence and reasoning with uncertainty, knowledgeacquisition for expert systems, data visualization, machine discovery , scientific discovery, information retrieval, and high-performance computing[58].

Data mining is an analysis tool which may uncover an important and previously unknown and overlooked data pattern. This uncovered hidden knowledge contributes for business, scientific and medical strategy. The observed gap between data and knowledge has made to look for an improved tool in order to mine important knowledge from a large amount of data [57].

2.2.1.Knowledge Discovery Process Model

As we march into the age of digital information, theproblem of data overloadlooms ominously ahead.Our abilityto analyze andunderstand massivedatasets lags far behindour ability to gather andstore the data. A new generationof computational techniquesand tools is required to support theextraction of useful knowledge fromthe rapidly growing volumes of data.These techniques and tools are thesubject of the emerging field of knowledgediscovery in databases (KDD) anddata mining .The term “KDD” torefer to the overall process of discovering useful knowledgefrom data. Data mining is a particular step in thisprocess— application of specific algorithms for extracting patterns (models) from data. The additional steps inthe KDD process, such as data preparation, data selection, data cleaning, incorporation of appropriate priorknowledge, and proper interpretation of the results ofmining ensure that useful knowledge is derived fromthe data. KDD process id the nontrivial process of identifying valid, novel, potentiallyuseful, and ultimately understandable patterns in data [59].

2.2.2.Knowledge Discovery in Database (KDD)

Database theories and tools provide the necessary infrastructure to store, access, and manipulate data. Data warehousing, a recently popularized term, refers to the current business trend of collecting and cleaning transactional data to make them available for online analysis and decision support[58]. A popular approach foranalysis of data warehouses is called online analytical processing (OLAP). Fields concerned with inferring models from data— including statistical pattern recognition, applied statistics, machine learning, and neural networks— were the impetus for much early KDD work. KDD largely relies on methods from these fields to find patterns from data in the data mining step of the KDD process. KDD focuses on the overall process of knowledge discovery fromdata, including how the data is stored and accessed, how algorithms can be scaled to massive datasets and still run efficiently, how results can be interpreted and visualized, and how the overall human-machine interaction can be modeled and supported. KDD places a special emphasis on finding understandable patterns that can be interpreted as useful or interesting knowledge[58].

Scaling and robustness properties of modeling algorithms for large noisy datasets are also of fundamental interest.Statistics provides a language and framework for quantifying the uncertainty resulting when one tries to infer general patterns from a particular sample of an overall population. As mentioned earlier, the term data mining has had negative connotations in statistics since the 1960s, when computer based data analysis techniques were first introduced[58]. The concern arose over the fact that if one searches long enough in any dataset (even randomly generated data), one can find patterns that appear to be statistically significant but in fact are not. This issue is of fundamental importance to KDD. There has been substantial progress in understanding such issues in statistics in recent years, much directly relevant to KDD. Thus, data mining is a legitimate activity as long as one understands how to do it correctly. KDD can also be viewed as encompassing a broader view of modeling than statistics, aiming to provide tools to automate (to the degree possible) the entire process of data analysis, including the statistician’s art of hypothesis selection [58]

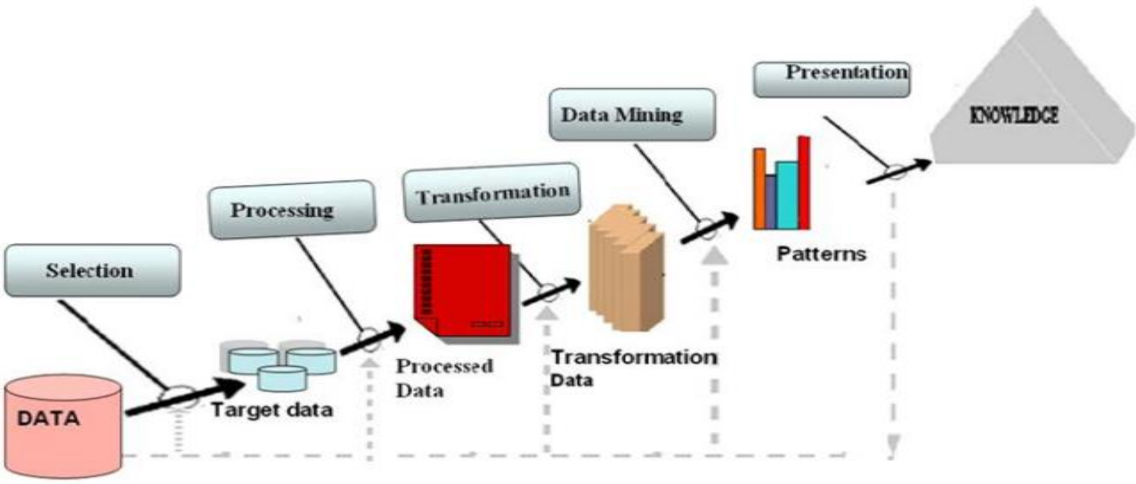

Figure 2. *Knowledge Discovery Processes.*

2.2.3.CRISP

CRISP–DM was developed in 1996 by analysts for fitting data mining into the general problem solving strategy of a business or research unit [60]. CRISP–DM is one of the most widely used methodologies in extraction of knowledge which has a life cycle consisting of six phases that involves iterative and adaptive process [60], as depicted in Figure 3.

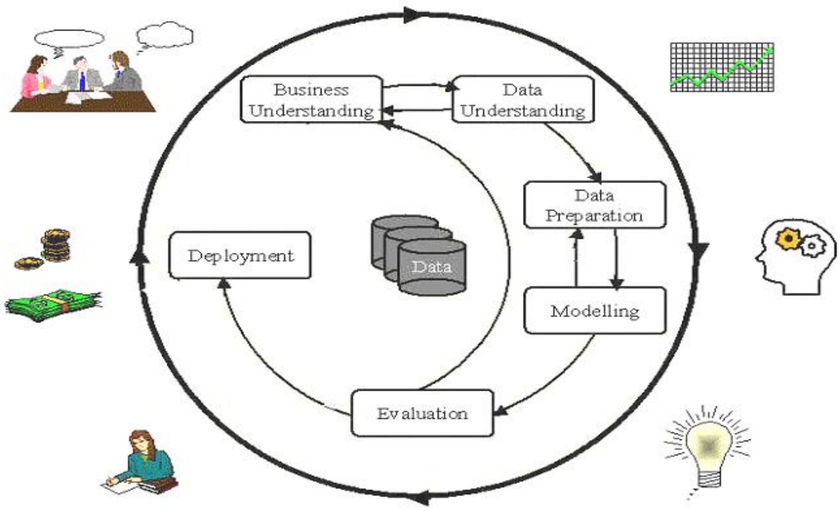

Figure 3. The CRISP-DM knowledge discovery Process Model

2.2.4.Hybrid Model

The development of academic models such as the nine-step model and eight-step model and industrial models such as five-step model and the six-step CRISP-DM model has led to the development of hybrid model that combines aspects usable for DM research. It was developed by chapman et al based on the CRISP-DM model. Hybrid process is characterized by providing more general, research oriented description of the steps. The hybrid model also encourages the application of knowledge discovered for a particular domain in other domains and it has a six step process as depicted in Figure 4.

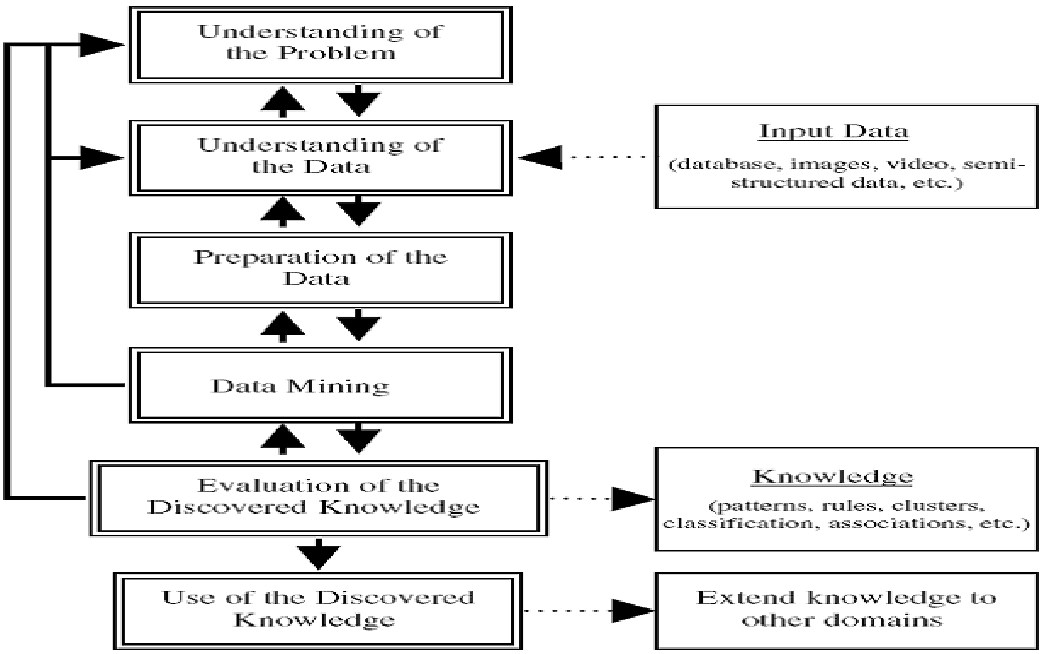

Figure 4. *Hybrid Process Models [63]*

One of the most important aspects of this model is iterative and interactive feature. The feedback loops are necessary because any changes and decisions made in one of the steps can result in changes in subsequent steps.

The initial step in the hybrid model is understanding of the problem domain to define the problem and determine the research goals and learning about current solution to the problem. It also involves learning domain – specific terminology and preparation of a description of the problem, including its restriction. Finally, the research goals will be translated in to data mining goals and initial selection of data mining tools or data to be used later in the process is performed.

This is followed by data understanding step which includes collecting sample data and deciding which data, including format and size, will be needed.

The third step is data preparation which concerns deciding about the data that is used as input for DM methods in the subsequent steps. The cleaned data may be further processed by feature selection and extraction algorithms.

The fourth step is data mining that data miner uses various data mining techniques such as classification, clustering and association rule discovery to derive hidden knowledge from processed data. This step creates Predictive and/or Descriptive models. The discovered knowledge is evaluated for understanding the result, checking whether the discovered knowledge is novel and interesting. Interpretation of the results by domain experts and checking the impact of the discovered knowledge is part of this phase.

2.3. Data Mining Tasks

Data mining tasks are used to specify the kind of patterns to be found using data mining techniques. In general, data mining tasks can be classified into two categories, Descriptive and Predictive. Descriptive mining tasks characterize the general properties of the data in the database. Predictive mining tasks perform inference on the current data in order to make predictions [57].

The descriptive model identifies the patterns or relationships in data and explores the properties of the data examined [61]. Descriptive models belong to the realm of unsupervised learning. Such models interrogate the database to identify patterns and relationships in the data. Clustering (segmentation) algorithms, pattern recognition models, visualization methods, among others, belong to this family of descriptive models [57].

Patterns found in the data to predict future values in predictive modeling. Predictive modeling consists of several types of models such as classification and regression models. Predictive models are built, or trained, using data for which the value of the response variable is already known. This kind of training is sometimes referred to as supervised learning, because calculated or estimated values are compared with the known results. Descriptive techniques are sometimes referred to as unsupervised learning because there is no already-known result to guide the algorithms [64].

2.4. Related Works

A study done to determine the medical cause, the preventability of the deaths and the type of substandard care in Morocco, three hundred and three cases of maternal deaths were analyzed for the year 2009. Direct causes accounted for 80.8%. 75.9% were considered avoidable and the three main factors were insufficient follow-up of care in 45.6% of cases, inadequate treatment in 43.9% and delay in seeking care in 41.3%. The auditors found that 54.3% of all maternal deaths could have been avoided if appropriate action had been taken at the health facilities [45].

A systematic review and meta-Analysis [46] suggests that continuous uptake of antenatal care, skilled birth attendance, and postnatal care is necessary to improve MNCH outcomes in low- and middle-income countries. Among the packages that linked antenatal care, skilled birth attendance, and postnatal care, a significant reduction was observed in combined neonatal, perinatal, and maternal mortality risks (RR 0.83; 95% CI 0.77 to 0.89, I2 79%) [46].

Promotion of universal Ante Natal Care follow-up and encouragement of mothers regarding the need for Skilled Birth Attendants during childbirth is of paramount importance. In a cross-sectional study done determinants of use of skilled birth attendance among mothers who gave birth in the Past 12 months in Raya Alamata district, North East Ethiopia, a significant association between ANC visit during last pregnancy of the women and the use of skilled assistance during delivery. Women who had ANC visit were 3.4 times higher odds of delivering with assistance by health professionals than mothers who didn’t visit ANC. [47] a similar cross sectional study in Kenya to identify determinants of use of skilled birth attendant at delivery in Makueni, Kenya: Mothers with tertiary/university education were more likely to use a skilled birth attendant during delivery, a woman whose partner had secondary education was 2.9 times more likely to seek skilled delivery. Attending ANC was equally significant, adjusted or 11.938, 95% ci, (4.086- 34.88). Living within a distance of 1-5 kilometers from a facility increased the likelihood of skilled birth attendance [48].

Although there is a general agreement on the importance of antenatal care to improve the maternal and perinatal health, little is known about its importance to improve health facility delivery in developing countries. In a systematic review with meta-analysis of Mantel–Haenszel; the pooled analysis also demonstrated that woman attending antenatal care had more than 7 times increased chance of delivering in a health facility. The comparative descriptive analysis, however, demonstrated a big gap between the proportion of antenatal care and health facility delivery by the same individuals (27%-95% vs 4%-45%) [49].

Distance to health facility and problems during pregnancy were factors positively and significantly associated with institutional delivery service utilization. In a study done to identify factors associated with institutional delivery service utilization in Ethiopia People living in urban areas (OR =13.16, CI =1.24, 3.68), with primary and above educational level of the mother and husband (OR =4.95, CI =2.3, 4. 8, and OR =4.43, CI =1.14, 3.36, respectively), who encountered problems during pregnancy (OR =2.83, CI =4.54, 7.39), showed significant association with institutional delivery service utilization [50]. Analysis of the 2011, Ethiopian Demographic and Health Survey to determinants of maternal health service utilization in Ethiopia also shows although 34% of women had ANC visits only 11.7% used skilled delivery attendants. Education of women, place of residence, ethnicity, parity, women’s autonomy and household wealth had a significant association with the use of maternal health services [51].

In Tanzania, more than 90% of all pregnant women attend antenatal care at least once and approximately 62% four times or more, yet less than five in ten receive skilled delivery care at available health units. A study identifies that the women, husbands, TBAs, and Elders interviewed agreed that the largest obstacle to receiving skilled and emergency obstetric care is failure to plan in advance for transport. Planning in advance for delivery is not part of traditional practice in the two communities where home delivery is the norm. Lack of planning for delivery are reinforced by the failure of health care providers to consistently communicate the importance of skilled delivery and immediate post-partum care for all women during routine antenatal visits [52]

Academic research conducted by Dawit [54] has been tried to predict maternal health care pattern using data mining techniques to determine the mostly seeking factors affecting women to utilize maternal health service. The study followed Hybrid methodology of Knowledge Discovery Process to achieve the goal of building predictive model using data mining techniques. The two most widely used classification techniques; J48 Decision tree and Naïve Bayes algorithms used on Ethiopia from EDHS 2011 dataset. The result of the study showed that the J48 Decision tree algorithm outperforms Naïve Bayes on the three of the outcome variables.

Tekelhaimanot [55] predict factors affecting antenatal care by applying data mining techniques. EDHS 2011 data set. In the study J48 algorithm, Naïve Bayes algorithm and neural network were used to build the models and performance of the classification algorithms.

This research reviewed the above literatures which briefly discussed the conceptual and empirical studies which were factors affecting Skilled birth attendant and yet no relevant studies was done on the role of ANC to utilize skilled birth attendant using data mining techniques on EDHS 2016 data sets. This study aims at construct predicting model for skill birth attendant among antenatal care users.

CHAPTER THREE

3. Research Design and Methods

The study proposed a model to predict skilled birth attendant using data mining techniques among ANC Clients. Hybrid methodology is followed to explore the application of data mining on EDHS 2016 dataset. WEKA 3.8.1 data mining tool, techniques and expertise are utilized as means to address the research problem.

The aim of the study is to build a model that can predict the utilization of skilled birth attendant among ANC clients using J48 Decision tree and Naïve Bayes and Rule based algorithms in Ethiopia from EDHS 2016 dataset. WEKA 3.8.1 data mining tools and techniques were employed as a means to address the research problem.

This research attempts to conduct a used predictive model using classification algorisms to model inheritances and relations between determinate variables, and the prediction using observed classification.

3.1. Architecture of the System Design

The general system design of the proposed model for discovering hidden knowledge from EDHS dataset are shown in figure 5.

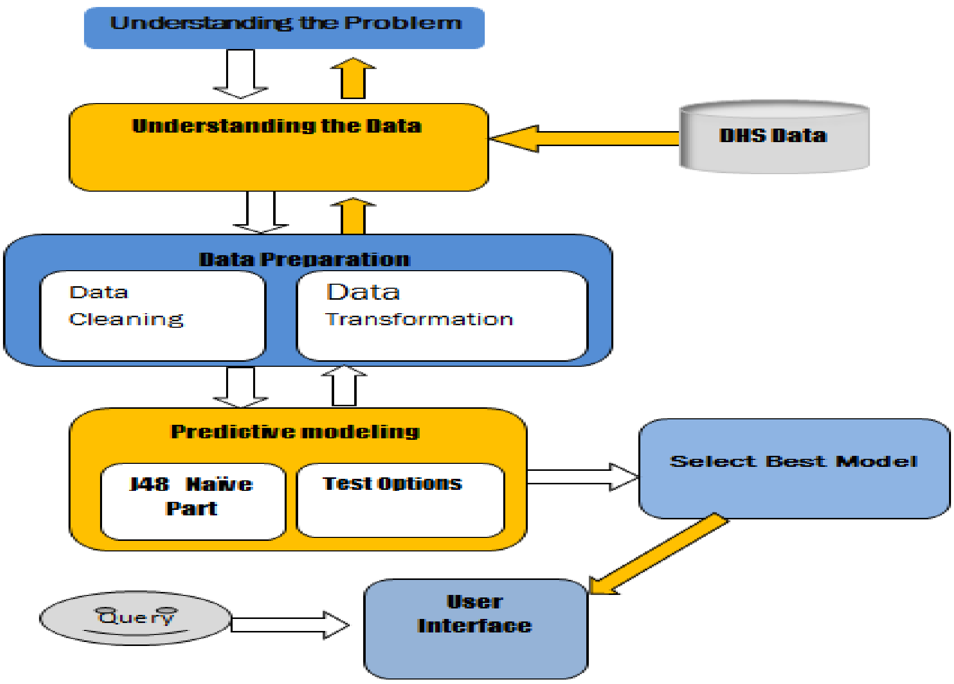

Figure 5. *The overall process and procedures of the study*

As depicted in figure 5, the first task is problem understanding during which the overall activity in the problem domain is analyzed and important attributes are selected. Then the next step is understanding of the data during which the information about the selected attributes are discussed in detailed and before the dataset passes into the data mining algorithm, preprocessing technique such as fill in missing values, smooth noisy data, identify or remove outliers, resolve inconsistencies, data reduction and data transformation are applied on the data. The preprocessing phase makes the input data suitable for the data mining algorithm. Once the preprocessing has done, the preprocessed data are used to create the classification model by training the data mining algorithms. Then after having the classification model, the next very important step is the evaluation of the model with the test data set. Finally, we measure the performance of the model created and we test our model with a real data.

3.2. Classification Algorisms

Here the data miner uses various DM methods to derive knowledge from preprocessed data. Data mining is the third step in the knowledge discovery process. It involves the use of several DM tools on data prepared. The main objective of this tool is discovering new knowledge. The process of discovering new information includes: constructing a predictive model using the chosen DM tools and following training and testing procedures.

In this research, the basic classification techniques for data classification are applied using the three well known supervised machine learning classifying algorithms such as J48 decision tree, Naive Bayes, and rule based induction in order to build a model for the purpose of classifying the pattern of ANC clients to use the latter skilled birth attendant.

3.2.1. Decision Tree

J48 algorithm is the Weka implementation of the C4.5 [83]. J48 implements a later and slightly improved version of which is known as C4.5 Revision, which was the last public version of this family version of algorithms before C5.0 (new version of C4.5), a commercial implementation was released [83]. J48 decision tree algorithm is a predictive machine learning model that decides the target value (dependent variable) of a new sample based on various attribute values of the available data. It can be applied on discrete data, continuous or categorical data.

J48 is the decision tree algorithm that is used in this study to classify the DHS data. The J48 decision tree can serve as a model for classification as it generates simpler rules and remove irrelevant attributes at a stage prior to tree induction. In several cases, it was seen that j48 decision trees had a higher accuracy than other algorithms [83]. J48 offer also a fast and powerful way to express structures in data.

Therefore, Weka software based J48 decision tree algorithm is used which is a greedy algorithm i.e. it constructs trees in a top-down recursive or divide-and-conquer manner and orders the class rule sets so as to minimize the number of false-positive error [84]. It uses the concept of information gain or entropy reduction to select the attribute with the highest information gain.

Suppose that we have a variable X whose K possible values have probabilities P1, P2 .Pk, the smallest number of bits, on average per symbol, needed to transmit a stream of symbols representing the values of X observed is the entropy of X. Entropy is the expected information needed to classify a tuple in X:

$$\boldsymbol{H}(\boldsymbol{X}) = -\sum \boldsymbol{pjlog2}(\boldsymbol{pj}) \qquad 3.1$$

For an event with probability p, the average amount of information in bits required to transmit the result is −log2p. For variables with several outcomes, we simply use a weighted sum of the log2 pj's, with weights equal to the outcome probabilities. Therefore, the mean information requirement can then be calculated as the weighted sum of the entropies for the individual subsets, as follows [84].

$$\text{InfoA(X)} = \sum_{j=1}^{V} \frac{|Xj|}{|X|} x \; info(Xj) \qquad 3.2$$

Information gained by branching on attribute A is:

$$Gain\ (A) = Info(X) - InfoA(X) \qquad 3.3$$

At each decision node, C4.5 uses the attribute with the maximum gain ratio as the splitting attribute and recursively visits each decision node, selecting the optimal split, until no further splits are possible. J48 also used the same concept to construct the decision tree and it supports both numeric and nominal predictors and nominal class attribute. It has the capability to handle missing values in datasets [57]. Once the tree is constructed, it is possible to generate the rule in order to apply it for new instances which are independent of the training tuples.

The J48 algorithm gives several options related to tree pruning. Pruning produces fewer, more easily interpreted results. More importantly, pruning can be used as a tool to correct for potential over fitting. J48 recursively classifies until each leaf is pure, meaning that the data has been categorized as close to perfectly as possible. When tested on new data, the rules may be less effective. Pruning always reduces the accuracy of a model on training data. This is because pruning employs various means to relax the specificity of the decision tree, hopefully improving its performance on test data.

Witten & Frank [83] emphasized the importance of understanding the variety of options during implementation of J48 algorithm. These options can make a significant difference in the quality of results. In many cases, the default settings will prove adequate, but in others, each choice may require some consideration. These options include:
Tree pruning options: The J48 algorithm gives several options related to tree pruning to produce fewer, more easily interpreted results. Pruning can be used as a tool to correct for potential over fitting. This algorithm recursively classifies until each leaf is pure, meaning that the data has been categorized as close to perfectly as possible. This process ensures maximum accuracy on the training data, but it may create excessive rules. These rules may be less effective when tested on new data. The overall concept of pruning is to gradually generalize a decision tree until it gains a balance of flexibility and accuracy.

J48 uses two pruning methods

- Sub tree replacement: The pruning method using J48 in which the nodes in a decision tree may be replaced with a leaf to reduce the number of tests along a certain path. This process starts from the leaves of the tree, and works backwards toward the root.
- Sub tree rising: The pruning method using J48, in which a node may be moved upwards towards the root of the tree, replacing other nodes along the way. Sub tree

rising often has a negligible effect on decision tree models. There is often no clear way to predict the utility of the option, though it may be advisable to try turning it off if the induction process is taking a long time. This is due to the fact that sub tree rising can be somewhat computationally complex.

Reduced-error pruning option: Error rates are used to make actual decisions about which parts of the tree to replace. There are multiple ways to do this. The simplest is to reserve a portion of the training data to test on the decision tree. The reserved portion can then be used as test data for the decision tree, helping to overcome potential over fitting.
This approach is known as reduced-error pruning. This method reduces the overall amount of data available for training the model.

Confidence factor option: Confidence factor is the approach that seeks to forecast the natural variance of the data, and to account for that variance in the decision tree. This approach requires a confidence threshold, which by default is set to 25 percent. This option is important for determining how specific or general the model should be. If the training data is expected to conform fairly closely to the data we would like to test the model on, this value of confidence factor can be lowered. The reverse is true if the model performs poorly on new data. Try decreasing the rate in order to produce a more pruned or more generalized tree.
Minimum number of instances per leaf option: This is the lowest number of instances that can constitute a leaf. The higher the number, the more general the tree will be. Lowering the number will produce more specific trees, as the leaves become more granular.
Binary split option: The binary split option is used with numerical data. If turned on, this option will take any numeric attribute and split it into two ranges using an inequality. This greatly limits the number of possible decision points. Rather than allowing for multiple splits based on numeric ranges, this option effectively treats the data as a nominal value. Turning this encourages more generalized trees.
Laplace smoothing option: This option is used to prevent probabilities from ever being calculated as zero. This is mainly to avoid possible complications that can arise from zero probabilities.

Witten & Frank [83] have also stated that the most basic parameter is the tree pruning options. If the data mining researcher decides to employ tree pruning, it is advisable to consider these options. Depending on how the training and test data have been defined, the performance of an unpruned tree may superficially appear better than a pruned one.

This can be a result of over fitting. Hence, it is important to repeatedly experiment with models by intelligently adjusting these parameters to obtain the best set of options.

Stopping Criteria

The growing phase continues until a stopping criterion is triggered. The following conditions are common stopping rules:

1. All instances in the training set belong to a single value of y.
2. The maximum tree depth has been reached.
3. The number of cases in the terminal node is less than the minimum number of cases for parent nodes.
4. If the node were split, the number of cases in one or more child nodes would be less than the minimum number of cases for child nodes.
5. The best splitting criterion is not greater than a certain threshold.

```
    Tree Growing (S,A,y,SplitCriterion,StoppingCriterion)
Where:
S - Training Set
A - Input Feature Set
y - Target Feature
SplitCriterion - the method for evaluating a certain split
StoppingCriterion - the criteria to stop the growing process
Create a new tree T with a single root node.
IF StoppingCriterion(S) THEN
        Mark T as a leaf with the most common value of y in S as a label.
ELSE
        ∀a_i ∈ A find a that obtain the best SplitCriterion (ai, S).
        Label t with a
        FOR each outcome v_i of a:
                Set Subtreei= TreeGrowing (σa = viS, A, y).
                Connect the root node of t_T to Subtree_i with an edge that is labeled as v_i
        END FOR
END IF
RETURN TreePruning (S, T, y)
TreePruning (S, T, y)
Where:
S - Training Set
y - Target Feature
T - The tree to be pruned
DO
        Select a node t in T such that pruning it maximally improve some evaluation criteria
        IF t ≠ Ø THEN T = pruned (T, t)
UNTIL t = Ø
RETURN T
```

Figure 6. *Top-Down Algorithmic Framework for Decision Trees algorithms.*

3.2.2.Rule Induction

Rule induction is one of the techniques most used to extract knowledge from data, since their presentation of knowledge as if/then rules is very intuitive and easily understandable by problem-domain experts [61]. It is an area of machine learning in which formal rules are extracted from a set of observations. The extracted rules may represent a full model of the data or represent local patterns in the data (in the form of individual rules) [85].

Hence, regularities hidden in the data are frequently expressed in terms of rules; rule induction is one of the fundamental tools of data mining at the same time. Usually rules are expressions of the form

if (attribute − 1; value − 1) and (attribute − 2; value − 2) and ……and (attribute − n; value − n) then (decision; value):

Rule induction on a database is a process undertaking using intelligent software where all possible patterns are systematically pulled out of the data. In this process the accuracy is added to them that tell the user how strong the pattern is and how likely it is to occur again [86]. It has been widely used to represent knowledge in expert system and they have the advantage of being easily interpreted by human experts because of their modularity [87]. Rule induction systems are highly automated and are probably the best of data mining techniques for exposing all possible predictive patterns in a database. They can be used in prediction problem but algorithm for combining evidence from a variety of rules comes from practical experience [86].

3.2.2.1. Rule Induction Algorithms

Rule induction seeks to go from the bottom up and collect all possible patterns that are interesting and then later use those patterns for some prediction target. It also retains all possible patterns even if they are redundant or do not aid in predictive accuracy. Hence, in a rule induction system if there were two columns of data that were highly correlated (or in fact just simple transformations of each other) they would result in two rules [86].

Rule induction is also known as Separate-And-Conquer method. This method applies an iterative process consisting of first generating a rule that covers a subset of the training examples and then removing all examples covered by the rule from the training set. This process is repeated iteratively until there are no examples left to cover. The final rule set is the collection of the rules discovered at every iteration of the process [88].

Written and Frank [83] Combined these two approaches in an algorithm called PART (for partial decision trees) in order to circumvent problems that can arise with both these techniques. Rules induced from decision trees are computationally expensive and this expense can grow alarmingly in the presence of noise, while separate-and-conquer methods suffer from a form of over pruning called “hasty generalization”. PART works by building a rule and removing its cover, as in the separate-and-conquer technique, repeatedly until all the instances are covered. The rule construction stage differs from standard separate-and-conquer methods because a partial pruned decision tree is built for a set of instances, the leaf with the largest coverage is made into a rule, and the tree is discarded. The pruned decision tree helps to avoid the over pruning problem of methods that immediately prune an individual rule after construction. Also, the expensive rule optimization stages associated with decision tree rule learning are not performed. Results on standard data sets show smaller rule sizes with no loss in accuracy when compared with the decision tree learner C4.5 and greater accuracy when compared with the separate-and-conquer rule learner RIPPER. In this paper we adapt the basic procedure of PART to continuous class prediction to examine whether similar results can be obtained, namely smaller rule sets with no loss in accuracy. Some examples of these kinds of systems which are supported by WEKA software are discussed below:

1. OneR: OneR or “One Rule” is a simple algorithm proposed by Holt. The OneR builds one rule for each attribute in the training data and then selects the rule with the smallest error rate as its ‘one rule’. To create a rule for an attribute, the most frequent class for each attribute value must be determined. The most frequent class is simply the class that appears most often for that attribute value. A rule is simply a set of attribute values bound to their majority class.

   OneR selects the rule with the lowest error rate. In the event that two or more rules have the same error rate, the rule is chosen at random [89].
2. PART: PART is a separate-and-conquer rule learner proposed by [79]. The algorithm producing sets of rules called ‘decision lists’ which are ordered set of rules. Anew data is compared to each rule in the list in turn, and the item is assigned the category of the first matching rule (a default is applied if no rule successfully matches). PART

builds a partial C4.5 decision tree in each iteration and makes the "best" leaf into a rule [78].

3.2.3. Naïve Bayes Classifier

Naïve Bayes classifier is statistical processing based on the Bayes decision theory is a fundamental technique for pattern recognition and classification. It is based on the assumption that the classification of patterns (the decision problem) is expressed in probabilistic terms. The statistical characteristics of patterns are expressed as known probability values that describe the random nature of patterns and their features. These probabilistic characteristics are mostly concerned with a priori probability and conditional probability densities of patterns and classes. The Bayes decision theory provides a framework for statistical methods for classifying patterns into classes based on probabilities of patterns and their features [60].

Bayes theorem provides a direct method for calculating the best probabilities [90]. More precisely, Bayes theorem provides a way to calculate the probability of a hypothesis based on its prior probability which may reflect any background knowledge about the chance that h is a correct hypothesis before we have observed the training data, P(h); the probability of observing various data given the hypothesis, P(D/h); and the observed data itself given no knowledge about which hypothesis holds, P(D). If we have no prior knowledge, then we might simply assign the same prior probability to each candidate hypothesis, P (h). The formula for Bayes Theorem is given:

$$P(h/D) = P(D/h)P(h)/P(D) \quad \dots\dots \qquad 3.4$$

P(h/D) increases with P(h) and with P(D/h). P(h/D) decreases as P(D) increases, because the more probable it is that D will be observed independently of h, the less evidence D provides in support of h.

In machine learning problems we are interested in the probability p(h/D) that h holds given the observed training data D, which is called the posterior probability of h, because it reflects our confidence that h holds after we have seen the training data D, in contrast to the prior probability p(h), which is independent of D.

In many learning scenarios, the learner considers some set of candidate hypotheses H and is interested in finding the most probable hypothesis $h \epsilon H$ given the observed data D (or at least one of the maximally probable if there are several). Any such maximally probable hypothesis is called a maximum a posteriori (MAP) hypothesis. It is possible to determine the MAP hypotheses by using Bayes theorem to calculate the posterior probability of each candidate hypothesis, i.e., h MAP is a MAP hypothesis provided

$$hMAP = argmax_{h \epsilon H} \, \mathrm{P}\left(\frac{\mathrm{h}}{\mathrm{D}}\right)$$

$$= argmax_{h \epsilon H} \, P(D/h) \; P(h)/P(D)$$

$$= argmax_{h \epsilon H} \; P(D/h)P(h) \quad \ldots \qquad 3.5$$

We dropped the term P (D) because it is a constant independent of h. In some cases, it is assumed that every hypothesis in H are equally probable a priori (P(hi and h j in H). Therefore, we need only to consider the term P (D/h) to find the most probable hypothesis. P(D/h) is often called the likelihood of the data D given h, and any hypothesis that maximizes P(D/h) is called a maximum likelihood (ML) hypothesis,

$$hML = argmax_{h \epsilon H} \; P(D/h) \ldots \qquad 3.6$$

In machine learning problems the data D is considered as training examples of some target function H as the space of candidate target functions. Since Bayes theorem provides a principled way to calculate the posterior probability of each hypothesis given the training data, we can use it as the basis for a straightforward learning algorithm that calculates the probability of each possible hypothesis, and then outputs the most probable [90].

Naive Bayes classifiers are among the most successful known algorithms for learning to classify text documents, long a favorite punching bag of new classification techniques. It has had a long history as a simple, yet powerful classification technique. It has recently emerged as a focus of research itself in machine learning. Naïve Bayes probabilistic classifiers are commonly studied in machine learning. Machine learning researchers tend to be aware of the large pattern recognition literature on naive Bayes, but may be less aware of an equally large information retrieval literature dating back almost forty years. In fact, naive Bayes methods, along with prototype formation methods, accounted for most applications of supervised learning to information retrieval until quite recently [90].

In addition, naïve Bayesian classifier learning is robust to noise and irrelevant attributes. In general, with Bayesian learning formulations and in particular with a naïve Bayes classifier, there is no problem with missing values at all. The calculation would simply omit the missing attribute – it is not included in the frequency counts, and the probability ratios are based on the number of values that actually occur rather than on the total number of instances [57].

3.2.3.1. Naïve Bayesian Classification Algorithm

One highly practical Bayesian learning method is the naive Bayes learner, often called the naive Bayes classifier. In some domains, its performance has been shown to be comparable to that of neural network and decision tree learning. This section introduces the naive Bayes classifier. The naive Bayes classifier applies to learning tasks where each instance x is described by a conjunction of attribute values and where the target function f (x) can take on any value from some finite set V. A set of training examples of the target function is provided, and a new instance is presented, described by the tuple of attribute values (al, a2...an). The learner is asked to predict the target value, or classification, for this new instance.

The Bayesian approach to classifying the new instance is to assign the most probable target value, VMAP, given the attribute values (al, a2 . . . an) that describe the instance.

$$VMAP = argmax_{vj\varepsilon v}\ \ P(Vj/a1, a1, \ldots an) \tag{3.7}$$

We can use Bayes theorem to rewrite this expression as

$$VMAP = \frac{argmax_{vj\varepsilon v}\, P(a1, a1 \ldots an/vj)P(vj)}{P(a1, a2 \ldots an)}$$
$$= argmax_{vj\varepsilon v}\ \ P(a1, a1 \ldots an/vj)P(vj)\ \ \ldots \tag{3.8}$$

Now we could attempt to estimate the two terms in Equation (3.4) based on the training data. It is easy to estimate each of the P(vj) simply by counting the frequency with which each target value vj occurs in the training data. However, estimating the different P (al, a2...an/vj) terms in this fashion is not feasible unless we have a very, very large set of training data. The problem is that the number of these terms is equal to the number of possible instances times the number of possible target values. Therefore, we need to see every instance in the instance space many times in order to obtain reliable estimates. The naive Bayes classifier is based on the simplifying assumption that the attribute values are conditionally independent given the target value. In other words, the assumption is that given the target value of the instance, the probability of observing the conjunction al, a2...a, is just the product of the probabilities for the individual attributes:

$$P\left(a1, a2 \dots \frac{an}{vi}\right) = \prod_i \boldsymbol{P(ai/vi)} \qquad 3.9$$

Substituting this into Equation (3.4) [90], we have the approach used by the naïve Bayes classifier.

$$\text{VMAP} = argmax_{vj\varepsilon v}\ \text{P(vj)} \prod_i \text{P(ai/vj)} \ \dots \qquad 3.10$$

Where, VNB denotes the target value output by the naive Bayes classifier. Notice that in a naive Bayes classifier the number of distinct P(ai/vj) terms that must be estimated from the training data is just the number of distinct attribute values times the number of distinct target values-a much smaller number than if we were to estimate the P(a1,a2…an/vj) terms as first contemplated.

To summarize, the naive Bayes learning method involves a learning step in which the various P(vj) and P(ai/vj) terms are estimated, based on their frequencies over the training data. The set of these estimates corresponds to the learned hypothesis. This hypothesis is then used to classify each new instance by applying the rule in Equation (3.5) [90]. Whenever the naive Bayes assumption of conditional independence is satisfied, this naive Bayes classification VNB is identical to the MAP classification.

Due to its apparent over-simplified assumptions, the naïve Bayes classifiers often work much better in many complex real-world situations than one might expect. The naïve Bayes classifiers have been reported to perform surprisingly well for many real world classification applications under some specific conditions [91].

An advantage of the naïve Bayes classifier is that it requires a small amount of training data to estimate the parameters necessary for classification. Bayesian classification approach arrives at the correct classification as long as the correct category is more probable than the others. The category's probabilities do not have to be estimated very well. In other words, the overall classifier is robust enough to ignore serious deficiencies in its underlying naïve probability model.

3.3. Data Preparation and Preprocessing Techniques

Data preparation concerns deciding which data will be used as input for DM methods in the subsequent step. It involves sampling, running correlation and significance tests, and data cleaning, which includes checking the completeness of data records, removing or correcting for noise and missing values, etc. The cleaned data may be further processed by feature selection and extraction algorithms (to reduce dimensionality), by derivation of new attributes (say, by discretization), and by summarization of data (data granularization). The end results are data that meet the specific input requirements for the DM tools selected in problem domain understanding [34]. The researcher does data preparation to prepare quality data accordingly.

Preprocessing is to help improve the quality of the data and, consequently, its aim is to improve the efficiency and ease of the mining process. Databases are highly susceptible to noisy, missing, and inconsistent data due to their typically huge size (often several gigabytes or more) and their likely origin from multiple, heterogeneous sources [41].

There are a number of data preprocessing techniques. Data cleaning can be applied to remove noise and correct inconsistencies in the data. Data integration merges data from multiple sources into a coherent data store, such as a data warehouse. Data transformations, such as normalization, may be applied. For example, normalization may improve the accuracy and efficiency of mining algorithms involving distance measurements; data reduction for instance. These techniques are not mutually exclusive; they may work together. For example, data cleaning can involve transformations to correct wrong data, such as by transforming all entries for a date field to a common format. Data processing techniques, when applied before mining, can substantially improve the overall quality of the patterns mined and/or the time required for the actual mining [41]. Therefore, the researcher will have done the following tasks on EDHS data set.

3.3.1.Data Cleaning

Real-world data tend to be incomplete, noisy, and inconsistent. Data cleaning (or data cleansing) routines attempt to fill in missing values, smooth out noise while identifying outliers, and correct inconsistencies in the data [41]. Data cleaning is a technique for handling missing data and for smoothing data, by avoiding missing values, noise, and inconsistencies that contribute to inaccurate data. So, handling errors and formatting data are applied on the data for better performance of the algorithm.

3.3.2.Data Integration and Transformation

It is likely that data analysis task will involve data integration, which combines data from multiple sources into a coherent data store, as in data warehousing. These sources may include multiple databases, data cubes, or flat files. Data mining often requires data integration—the merging of data from multiple data stores [41]. In this research the DHS data is collected from one source, CSA in SSPS file format there is no need of applying data integration task.

The data may also need to be transformed into forms appropriate for mining. Smoothing, normalization and discretization are applied for our dataset. Smoothing which works to remove noise from the data. Such techniques include binning, regression, and clustering. Normalization is just scaled to fall within a smaller, specified range of values. Discretization used to reduce data size by dividing the range of a continuous attribute into intervals and the interval labels can then be used to replace actual data values.

3.3.3.Reduction

Complex data analysis and mining on huge amounts of data can take a long time, making such analysis impractical or infeasible. Data reduction techniques can be applied to obtain a reduced representation of the data set that is much smaller in volume, yet closely maintains the integrity of the original data. That is, mining on the reduced data set should be more efficient yet produce the same (or almost the same) analytical results [41]. In this research dimensionality reduction and discretization methods are applied for our dataset.

3.4. Data Mining Tools

In order to select the best tools for dealing with the classification task that helps in decision making, this study explored a number of free available data mining and knowledge discovery tools and software packages.

Many open data mining tools are available for free on the Web such as Waikato Environment for Knowledge Analysis (WEKA), Tanagra, the Konstanz Information Miner (KNIME), and Orange Canvas.

In the following section all the above free and open data mining tools are briefly described ant the one which fits for this study will be selected.

3.4.1. Waikato Environment for Knowledge Analysis (WEKA)

WEKA was developed at the University of Waikato in New Zealand, and the name stands for Waikato Environment for Knowledge Analysis. The system is written in Java and distributed under the terms of the GNU General Public License. It runs on almost any platform and has been tested under Linux, Windows, and Macintosh operating systems—and even on a personal digital assistant. It provides a uniform interface to many different learning algorithms, along with methods for pre- and post-processing and for evaluating the result of learning schemes on any given dataset [39].

WEKA is a state-of-the-art facility for developing machine learning (ML) techniques and their application to real-world data mining problems. It is a collection of machine learning algorithms for data mining tasks. The algorithms are applied directly to a dataset. WEKA implements algorithms for data preprocessing, classification, regression, clustering, association rules; it also includes visualization tools [40].

Weka provides implementations of learning algorithms that you can easily apply to your dataset. It also includes a variety of tools for transforming datasets, such as the algorithms for discretization. You can preprocess a dataset, feed it into a learning scheme, and analyze the resulting classifier and its performance—all without writing any program code at all.

The workbench includes methods for all the standard data mining problems: regression, classification, clustering, association rule mining, and attribute selection. Getting to know the data is an integral part of the work, and many data visualization facilities and data preprocessing tools are provided. All algorithms take their input in the form of a single relational table in the ARFF format described which can be read from a file or generated by a database query. One way of using weka is to apply a learning method to a dataset and analyze its output to learn more about the data. Another is to use learned models to generate predictions on new instances. A third is to apply several different learners and compare their performance in order to choose one for prediction. The learning methods are called classifiers, and in the interactive weka interface you select the one you want from a menu. Many classifiers have tunable parameters, which you access through a property sheet or object editor. A common evaluation module is used to measure the performance of all classifiers. Implementations of actual learning schemes are the most valuable resource that weka provides. But tools for preprocessing the data, called filters, come a close second. Like classifiers, you select filters from a menu and tailor them to your requirements. We will show how different filters can be used, list the filtering algorithms, and describe their parameters. Weka also includes implementations of algorithms for learning association rules, clustering data for which no class value is specified, and selecting relevant attributes in the data [39].

3.4.2. Tanagra

Tanagra is free data mining software for academic and research purposes. It offers several data mining methods like exploratory data analysis, statistical learning and machine learning. The first purpose of the Tanagra project is to give researchers and students easy-to-use data mining software. The second purpose of TANAGRA is to propose to researchers an architecture allowing them to easily add their own data mining methods, to compare their performances. The third and last purpose is that novice developers should take advantage of the free access to source code, to look how this sort of software was built, the problems to avoid, the main steps of the project, and which tools and code libraries to use for. In this way, Tanagra can be considered as a pedagogical tool for learning programming techniques as well [100].

3.4.3. KNIME (Konstanz Information Miner)

KNIME is a user-friendly and comprehensive open-source data integration, processing, analysis, and exploration platform. KNIVEM is the leading open solution for data-driven innovation, helping you discover the potential hidden in your data, mine for fresh insights, or predict new futures and it is very fast to deploy, easy to scale, and intuitive to learn. It is a modular data exploration platform that enables the user to visually create data flows (often referred to as pipelines), selectively execute some or all analysis steps, and later investigate the results through interactive views on data and models [101].

3.4.4. Orange –Canvas

Orange is a library of C++ core objects and routines that includes a large variety of standard and not-so-standard machine learning and data mining algorithms, plus routines for data input and manipulation. This includes a variety of tasks such as pretty-print of decision trees, attribute subset, bagging and boosting, and alike. Orange also includes a set of graphical widgets that use methods from core library and Orange modules. Through visual programming, widgets can be assembled together into an application by a visual programming tool called Orange Canvas. All these together make the Orange tool, a comprehensive, component-based framework for machine learning and data mining, intended for both experienced users and researchers in machine learning who want to develop and test their own algorithms while reusing as much of the code as possible, and for those just entering who can enjoy in powerful while easy-to-use visual programming environment [102].

The research done by Abdullah H. Wahbeh et al [93] on comparison between four data mining toolkits for classification purposes, nine different data sets were used to judge the four toolkits tested using six classification algorithms namely; Naïve Bayes (NB), Decision Tree (C4.5), Support Vector Machine (SVM), K Nearest Neighbor (KNN), One Rule (OneR), and Zero Rule (ZeroR).

The study concluded that no tool is better than the other if used for a classification task, since the classification task itself is affected by the type of dataset and the way the classifier was implemented within the toolkit. However; in terms of classifiers' applicability, the study concluded that the WEKA toolkit was the best tool in terms of the ability to run the selected classifier followed by Orange, Tanagra, and finally KNIME respectively [93].

Finally, the study concluded that; WEKA toolkit has achieved the highest performance improvements when moving from the Percentage Split test mode to the Cross Validation test mode followed by Orange, KNIME, and then Tanagra Respectively [93].

3.5. Performance Evaluation for Predictive Model

In the final phase of the data-mining process, when the model is obtained using one or more inductive-learning techniques, one important question is how does one verify and validate the model. The data-mining results are validated and verified by the testing process. Model testing is demonstrating that inaccuracies exist or revealing the existence of errors in the model. We subject the model to test data or test cases to see if it functions properly. "Test failed" implies the failure of the model, not the test. Some tests are devised to evaluate the behavioral accuracy of the model (i.e., validity), and some tests are intended to judge the accuracy of data transformation into the model (i.e., verification) [67].

The objective of a model obtained through the data-mining process is to classify/predict new instances correctly. The commonly used measure of a model's quality is predictive accuracy. Since new instances are supposed not to be seen by the model in its learning phase, we need to estimate its predictive accuracy using the true error rate. The true error rate is statistically defined as the error rate of the model on an asymptotically large number of new cases, where this number converge to the actual population distribution. In practice, the true error rate of a data-mining model must be estimated from all the available samples, which are usually split into training and testing sets. The model is first designed using training samples, and then it is evaluated based on its performance on the test samples. In order for this error estimate to be reliable in predicting future model performance, not only should the training and the testing sets be sufficiently large, they must also be independent. This requirement of independent training and test samples is still often overlooked in practice [67].

The researcher had tacitly assumed that the goal of the performance evaluation was to maximize the success rate of the predictive model for DHS dataset. Therefore, predictive models are evaluated in terms of performance, and applicability.

As Berry Michael described [67], after a predictive model is developed for determining the skilled birth attendance on DHS dataset, the model should be evaluated as to how it will perform for the future data which, it has not seen during the model building process.

For purposes of this study 10-fold cross validation and 66% percentage split test options are used because of which are relatively low bias and variations [80].

In 10-fold cross validation; the data were divided into 10 folds where 9 folds were used as training data whereas the remaining one folds as test data. In the percentage split method, where 66% of the data was used as training and the remaining 34% was used as test data. Accuracy, Precision, specificity, ROC curve, mean absolute error, Recall and confusion matrix standard metrics were also used for evaluation of the results.

3.6. Evaluation of the discovered knowledge

Evaluation includes understanding the results, checking whether the discovered knowledge is novel and interesting, interpretation of the results by domain experts, and checking the impact of the discovered knowledge. Only approved models are retained, and the entire process is revisited to identify which alternative actions could have been taken to improve the results. A list of errors made in the process is prepared [34].

In data mining evaluation has two primary functions. The first one helps to predict how well the final model will work in the future. Secondly, evaluation is an integral part of many learning methods and helps to explore the model that best represents the training data.

For evaluating classification models a test dataset is prepared and used. Then, accuracy is calculated to measure the overall performance of classification algorithm. Accuracy is the proportion of correctly classified instances out of the total instances. In addition, effectiveness and efficiency of the model by class is computed in terms of recall and precision. Recall is the proportion of correctly classified instances of a given class; whereas, precision is the proportion of instance that are correctly classified. F measure is the harmonic relationship between precision & recall; it is the point to conclude that the precision and recall of the model are significantly balanced. ROC curve is useful visual tool for comparing classification models. It shows the trade-off between the true positive rate (proportion of positive tuples that are correctly identified) and the false-positive rate (proportion of negative tuples that are incorrectly identified as positive) for a given model.

The accepted rules by domain experts are used as an input to design a porotype to assist ANC service providers to predict the utilization of skilled birth attendant during child birth using

VB.net language and then the designed user interface has been tested by the users for its applicability and usability.

3.6.1.Confusion Matrix

The confusion matrix is a useful tool for analyzing how well the researcher's classifier can recognize tuples of different classes [57].

In many data-mining applications, the assumption that all errors have the same weight is unacceptable. So, the differences between various errors should be recorded, and the final measure of the error rate will take into account these differences. When different types of errors are associated with different weights, we need to multiply every error type with the given weight factor cij. If the error elements in the confusion matrix are eij, then the total cost function C (which replaces the number of errors in the accuracy computation) can be calculated as: [67]

$C = \sum m\ i = 1 \sum m\ \ j = 1\ cijeij$ 3.10

In many data-mining applications, it is not adequate to characterize the performance of a model by a single number that measures the overall error rate. More complex and global measures are necessary to describe the quality of the model. Consider a classification problem where all samples have to be labeled with one of two possible classes. A typical example is a diagnostic process in medicine, where it is necessary to classify the patient as being with or without disease. For these types of problems, two different yet related error rates are of interest. The False Acceptance Rate (FAR) is the ratio of the number of test cases that are incorrectly "accepted" by a given model to the total number of training cases. For example, in medical diagnostics, these are the cases in which the patient is wrongly predicted as having a disease. On the other hand, the False Reject Rate (FRR) is the ratio of the number of test cases that are incorrectly "rejected" by a given model to the total number of cases. In the previous medical example, these are the cases of test patients who are wrongly classified as healthy, without any diagnosed disease [67].

For the most of the available data-mining methodologies, a classification model can be tuned by setting an appropriate threshold value to operate at a desired value of FAR. If we try to decrease the FAR parameter of the model, however, it would increase the FRR and vice versa. To analyze both characteristics at the same time, a new parameter was developed, the

Receiver Operating Characteristic (ROC) Curve. It is a plot of FAR versus FRR for different threshold values in the model. This curve permits one to assess the performance of the model at various operating points (thresholds in a decision process using the available model) and the performance of the model as a whole (using as a parameter the area below the ROC curve) [67]. The ROC curve is especially useful for a comparison of the performances of two models obtained by using different data-mining methodologies, in this research it is discussed in detail in section 3.6.2.

To confirm the model performance evaluation for the results of the predicted model of the skilled birth attendants among ANC Users in Ethiopia, the researcher implemented rules and procedures on EDHS area.

Given M classes; a confusion matrix is a table of at least size M by M. An entry, CMi,j in the first M rows and M columns indicates the number of tuples of class i that were labeled by the classifier as class j. For a classifier to have good accuracy, ideally most of the tuples would be represented along the diagonal of the confusion matrix, from entry CM1, 1 to entry CMm,m, with the rest of the entries being close to zero [57].

Predicted class

Actual class

	C1	C2
C1	True Positive (TP	False Negative (FN)
C2	False Positive (FP)	True Negative(TN)

Table 1. Confusion matrix for positive and negative tuples [57].

In this research performance evaluation computational techniques on confusion matrix are used.

Given two classes, we can talk in terms of positive tuples (tuples of the main class of interest, True positives refer to the positive tuples that were correctly labeled by the classifier, while true negatives are the negative tuples that were correctly labeled by the classifier. False positives are the negative tuples that were incorrectly labeled. Similarly, false negatives are the positive tuples that were incorrectly labeled. There are alternatives to the accuracy measure, that are sensitivity and specificity measures can be used, respectively, for this purpose. Sensitivity is also referred to as the true positive (recognition) rate (that is, the proportion of positive tuples that are correctly identified), while specificity is the true negative rate (that is, the proportion of negative tuples that are correctly identified) [57].

$sensitivity\ = TP/\,(TP + FN)$ 3.11

$specificity\ = TN/(FN + TN)$ 3.12

$precision\ = TP/(TP +\ FP)$ 3.13

Accuracy is a function of sensitivity and specificity:

$Accuracy\ = (TP + TN)/(TP + TN + FP + FN)$ 3.14

The true positives, true negatives, false positives, and false negatives are also useful in assessing the costs and benefits (or risks and gains) associated with a classification model. The cost associated with a false negative is far greater than that of a false positive. In such cases, we can outweigh one type of error over another by assigning a different cost to each [57].

3.6.2.Area under the ROC Curve

In many machine learning applications, the ranking quality of a classifier is critical. For example, the ordering of the list of relevant documents returned by a search engine or a document classification system is essential. The criterion widely used to measure the ranking quality of a classification algorithm is the area under an ROC curve (AUC). But, to measure and report the AUC properly, it is crucial to determine an interval of confidence for its value as it is customary for the error rate and other measures. It is also important to make the computation of the confidence interval practical by relying only on a small and simple number of parameters. In the case of the error rate, such intervals are often derived from just the sample size N [71].

A Receiver Operating Characteristics (ROC) curve is a technique for visualizing, organizing and selecting classifiers based on their performance. In essence, it is another performance evaluation technique for classification models and also useful tool for comparing two or more classification models. ROC curves have long been used in signal detection theory to depict the tradeoff between hit rates and false alarm rates of classifiers. The use of ROC analysis has been extended into visualizing and analyzing the behavior of diagnostic systems. Recently, the medical decision making community has developed an extensive literature on the use of ROC curves as one of the primary methods for diagnostic testing [71].

ROC curve shows the trade-off between the true positive rate or sensitivity (proportion of positive tuples that are correctly identified) and the false positive rate (proportion of negative tuples that are incorrectly identified as positive) for a given model. That is, given a two-class problem, it allows us to visualize the trade-off between the rate at which the model can accurately recognize 'yes' cases versus the rate at which it mistakenly identifies 'no' cases as 'yes' for different "portions" of the test set. Any increase in the true positive rate occurs at the cost of an increase in the false-positive rate. The area under the ROC curve is a measure of the accuracy of the model.

In order to plot an ROC curve for a given classification model, true positive (TP) rate is plotted on the Y axis and false positive (FP) rate is plotted on the X axis [57].

As Han and Kamber described [57], the process of drawing Roc curve we start at the bottom left-hand corner (where the true positive rate and false-positive rate are both 0), we check the actual class label of the tuple at the top of the list. If we have a true positive (that is, a positive tuple that was correctly classified), then on the ROC curve, we move up and plot a point. If, instead, the tuple really belongs to the ‘no’ class, we have a false positive. On the ROC curve, we move right and plot a point. This process is repeated for each of the test tuples, each time moving up on the curve for a true positive or toward the right for a false positive.

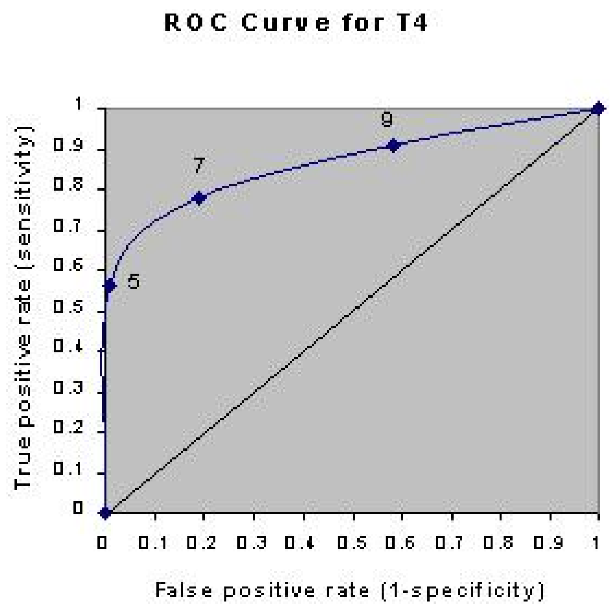

Figure 7. A Sample ROC curves

To compare classifiers or to judge the fitness of a single classifier one may want to reduce the ROC measures to a single scalar value representing the expected performance. A common method to perform such task is to calculate the area under the ROC curve. Area under the Curve (AUC) is a portion of the area of the unit square, and its value will always be between 0 and 1.0. A perfect accuracy gets a value of 1.0. The diagonal line y = x represents the strategy of randomly guessing a class. If a classifier randomly guesses the positive class half the time (much like flipping a coin), it can be expected to get half the positives and half the negatives correct; this yields the point (0:5; 0:5) in ROC space, which in turn translates into area under the ROC curve value of 0.5. No classifier that has any classification power should have an AUC less than 0.5 [71].

The area under the curve helps to observe the accuracy of a model, we measure which is a portion of the area of the unit square and its value is ranged from 0-1. It is assumed that increasing numbers on the scale represents that the subject belongs to one category while decreasing numbers on the scale represent the increasing belief that the subject belongs to the other category. Thus, from the ROC curve, the closer the ROC curve of a model is to the diagonal line, the less accurate the model is closer to the area of 0.5.

No.	ROC Area	Performance
1	0.9-1.0	Excellent
2	0.8-0.9	Good
3	0.7-0.8	Fair
4	0.6-0.7	Poor
5	0.5-0.6	Fail

Table 2. Performance Measures of ROC Area No.

As [57] described, the model with perfect accuracy will have an area of 1.0; the larger the area, the better performance of the model or the larger values of the test result variable indicate the stronger evidence for a positive actual state that is 1.00.

In figure 7, classifier 1 performs better than 2. Therefore, we can indicate that predictors to find the one with optimal characteristics and their associated cut-points using ROC analysis.

The researcher considered the accuracy, sensitivity, specificity and ROC area when the classifier performance is evaluated to select the best model.

3.7. The prototype development approaches to use the discovered Knowledge of the study

In this study, the software development approach was examined and the one which fits with this study is selected based on the goals of this project. According to Floyd, prototype is developed by selecting appropriate software development approaches that is based on one of three goals - exploration, experiment and evolution [103].

An exploratory prototyping approach is mainly undertaken to elicit or clarify user requirements. Exploratory prototyping helps developers gain insight into the users work tasks and problems, and helps to crystallize hazy user perceptions and needs into the requirements for an initial system.

Experimental prototyping provides users with alternative possibilities, helps assess the feasibility of the future system in terms of performance and other technical aspects, when requirements are known.

Evolutionary prototyping is an approach that allows flexibility in the software development process so that it can adapt to changing organizational contexts. Here, software development process is not seen as an isolated self-contained project but as something that continuously evolves.

Since this study trying to solve the problem under study by deploying data mining techniques and tools to construct a model and the constructed model should be tested by user for feasibility of the system, experimental approach is selected.

In this approach, prototypes are built so that the feasibility of proposed solutions to problems can be examined by means of experimental use. The prototype could be a partial functional simulation demonstrating certain aspects of the system, a full functional simulation which makes available all the functions of the proposed system, a mock-up interface or a skeletal program showing the structure of the system [104].

The prototype is developed through the process of functional selection, construction, evaluation and further use. The functions in this study are the selected important rules which were generated from the data mining tools and consulted with domain experts and a prototype is constructed in such a way that all these important rules generated from data mining tools are inputted to the system, the developed system is evaluated again by the domain experts for its feasibility and the prototype further will be used by outlining the problem of this study .The development of this prototype has been passed through the following steps:

1. user's basic needs are identified; the basic user needs for this system is a system required for ANC service providers to assist the providers in such a way that whether the ANC client will utilize skilled birth attendant or not and together with the reasons or factors which most likely hinder the client not to utilize the service and then according to the clients specific factor ,the health service provider will provide appropriate counseling and action to convince the clients for utilization of skilled birth attendant based on her specific factors.
2. A working prototype is developed; the most important rules or models generated are fed into the system and then the prototype developed and passed through user acceptance test.
3. The working prototype is then implemented and used; The working prototype is piloted in one of the Non-governmental maternal health center to see its applicability for further implementation in the other maternal health centers and the providers' comments and feedbacks are presented in chapter five.
4. the prototyping system is revised and enhanced; the most valuable comments from the health professionals were forwarded and integrated in the next version of the prototype for further strengthen of a system.

Therefore, this prototype development process has been undergoing with several iterations and steps until the user accepts the system that will assist the providers by providing appropriate information to counsel clients by providing specific factors that either enable or hinders client to utilize the service.

Chapter Four

4. Data Preparation and Pre-processing

This chapter briefly discuss about the Health models which can deals with factors that can mostly affect the individual's behaviors to utilize the health service at the health institution and these models helps us to understand the problem under study, the data mining model adapted for this study; hybrid model and final discuss data source used for the study and selection, data cleaning and data transformation or reduction.

4.1. Theoretical framework for Describing the problem

The utilization of health services can be viewed as a type of individual behavior. In general, the behavioral sciences have attempted to explain individual behavior as a function of characteristics of the individual, characteristics of the environment in which he/she lives, and/or some interaction of these individual and societal forces. There are several ways in which behavior is conceptualized and defined. The largest number of studies focuses squarely on the individual as the locus of behavior [72].

A theoretical framework for viewing health services utilization is presented, emphasizing the importance of the (1) characteristics of the health services delivery system, (2) changes in medical technology and social norms relating to the definition and treatment of illness, and (3) individual determinants of utilization. These three factors are specified within the context of their impact on the health care system [72].

These theories posit a greater or lesser impact by external factors such as society, but each hold behavior to be an outcome of competing influences balanced and decided upon by the individual.

The Framework Societal determinants of utilization are shown to affect the individual determinants both directly and through the health services system. Various types of individual determinants then influence health services used by the individual. In subsequent sections, these determinants will be defined and data will be presented to illustrate some of the suggested relationships. Prior to such discussion the nature of the utilization variable itself needs to be considered [72].

4.1.1.Andersen Health Behavioral Model (1968 -present)

This model is among the most widely acknowledged behavioral model that has been applied in a broad range of health services topics and diseases [73]. Ronald Andersen, a US medical sociologist and researcher of health services, developed the model in 1968 in his attempt to predict and explain "how" and "why" families use formal health services. The model he developed has been influential in explaining an individual's use of health care services, especially physician care [73].

The model is a valuable tool to explore determinants and identify relevant variables in the process of health service use. It asserts that health service is a function of individual traits, population characteristics, and the environment [74]. Recently, Scholars have expanded and modified the application of this model to address efficiency, equity, effectiveness concepts in health care and broader issues including policy, health systems factors, external environment, and patient satisfaction [74].

When the model is used as an explanatory tool for particular conditions, which is the case in this research, it provides a better comprehensive and informative analysis of a problem which in this case is the under-utilization of skilled birth attendants [72]. The application of Andersen's model mostly relies on individual level factors that affect health services utilization [74]. The model has undergone four phases of development over the years. According to the latest version the model classifies conditions that facilitate or impede utilization of health services as follows:
Predisposing factors: These factors are antecedents to behavior that provides the rationale or motivation for behaviors. These can be either individual or contextual predisposing factors. Demographic factors/biological imperatives, social factors, skill, and mental f actors in terms of health belief, attitude, values and knowledge on health and health services belong to individual predisposing factors. On the other hand, contextual factors include demographic and social composition of communities; collective and organizational values, cultures, norms, and political perspectives.

Enabling factors: These are individual (for example income and health care insurance), organizational (including health service facilities and personnel, transportation means, travel time to and waiting time for care), and contextual (resources available in the community that enable the utilization of health services).

Need factors: These antecedents to behavior enable the behavior to be realized. These factors also can be either individual or contextual need factors. At individual level, the view and experience of individuals towards their health perceived need and evaluated need for health services, which is a professional assessment of individual's health status and need for medical care makes up individual level need factors. The contextual need factors include the health related condition of the environment and the overall measures of community health.
Many models of health behavior are also relevant to the model chosen for this study. The reasons why Andersen's Health Behavioral Model was selected the next section discuss other models commonly used by researchers in the health and behavioral sciences.

4.1.2.The Health Belief Model(HBM)

The HBM is a health specific social cognition, the key components and constructs (that is, complex theoretical components) of which are:

- Perceived susceptibility. The subjective perception of the risk the individual is at from a state or condition.
- Perceived severity. Subjective evaluation of the seriousness of the consequences associated with the state or condition.
- Perceived threat, the product/sum of severity and susceptibility. This combined quantum might be seen as indicative of the level of motivation an individual has to act to avoid a particular outcome.
- Perceived benefits. The subjectively understood positive benefits of taking a health action to offset a perceived threat. This perception will be influenced not only by specific proximal factors, but an individual’s overall ‘health motivation’.
- Perceived barriers. The perceived negatively valued aspects of taking the action, or overcoming anticipated barriers to taking it.

- Self-efficacy. This component has been added to the HBM on many occasions since the late 1970s, when Bandura first introduced this concept of act or task specific self-confidence, i.e. belief in one’s ability to execute a given behavior
- Expectations, which are the product/sum of perceived benefits, barriers and self-efficacy. This may be seen as indicative of the extent to which the individual will try to take a given action.
- Cues to action. Reminders or prompts to take actions consistent with an intention, ranging from advertising to personal communications from health professionals, family members and/or peers.
- Demographic and socio-economic variables. These may include age, race, ethnicity (cultural identity), education and income.

Social, economic and environmental factors integration applied in a systematic way the full set of model components described above (to which may on occasions be added a general health perception variable) would have the potential to provide a relatively comprehensive understanding of the influence of social, economic and environmental factors on health behaviors, in addition to that of cognitive factors contained in the psycho-social equation at the heart of the HBM. However, the use of this model has in practice focused largely on measurements and analyses of susceptibility, severity, benefit and barrier perception components alone. The research literature analyzed during this review did not provide evidence that applications of the HBM have enabled the influence of social, economic or other environmental factors (including variables such as low income, exposure to racial prejudice, cultural exclusion, low health valuations as cultural norms or inconvenient service access arrangements) to be better understood by researchers, practitioners or policy makers.

The HBM is characterized by a lack of adequate combinatorial rules and inconsistent application. Its main components have weak effect sizes, and its predictive capacity is limited as compared to that of other social cognition models [75].

4.1.3.The Theory of Reasoned Action (TRA) and the Theory Planned Behavior (TPB)

The Theory of Reasoned Action was formulated towards the end of the 1 960s, and in some respects may be seen as refining and taking forward approaches embodied in the HBM. At that time psychologists were concluding that attitudes (at least in the form of uni-dimensional phenomena) have very limited validity as predictors of future behavior. As expressed in its final form, the TRA combines two sets of belief variables, described under the headings of 'behavioral attitudes' and 'the subjective norm'. The Theory of Planned Behavior built further on this framework. Its design and dissemination followed Bandura's work on self-efficacy and the publication of his Social Cognitive Theory in 1986. It is differentiated from the TRA, by the additional dimension of perceived behavioral control.

Both the TRA and the TPB assume that the immediate cognitive precursors to behaviors are not attitudes but behavioral intentions. These are in essence defined as complex amalgams of prior beliefs. Hence the shared components of the TRA and the TPB are:

- Behavioral beliefs, salient to a) the likelihood that an action might promote or negate a given outcome and b) evaluating outcomes achieved or avoided, in terms of their desirable and negative consequences.
- Behavioral attitudes, defined as the multiplicative sum of the individual's relevant likelihood and valuation/severity related behavioral beliefs. However, such attitudes may also be independently measured.
- Normative beliefs, including a) referent beliefs about what behaviors others expect and b) the degree to which the individual wants to comply with others' expectations.
- Subjective norms, which (like behavioral attitudes) are defined as the multiplicative sum of the two sets of normative beliefs, although these may also be independently assessed.
- Behavioral intentions, derived from the combination of the behavioral attitude and the subjective norm. Intents rather than attitudes are, as noted above, regarded as the main proximal cognitive precursors to acting.

The Theory of Reasoned Action has been criticized because it is said to ignore the social nature of human action. Behavioral and normative beliefs are derived from individuals' perceptions of the social world they inhabit, and are hence likely to reflect the ways in which economic or other external factors shape behavioral choices. Yet there is a compelling logical case to the effect that the model is inherently biased towards individualistic, rationalistic, interpretations of human behavior. Its focus on subjective perception does not necessarily permit it to take meaningful account of social realities.

The general theoretical frameworks of the TRA and the TPB have allowed them to be very widely used in the retrospective analysis of health behaviors and to a lesser extent in predictive investigations and the design of health interventions. There is evidence that the Theory of Reasoned Action and the Theory of Planned behavior can both be used to predict health related behavior with greater effect than the Health Belief Model. There is also evidence that the predictive power of the TPB exceeds that of the TRA and across a wide range of health behaviors the TPB can explain 20 per cent or more of observed behavioral variance. However, there is also evidence that TPB based research is infrequently used directly to inform behavioral change interventions, and when this has been the case the additional health benefits gained appear to have been relatively limited [76].

4.1.4.The Trans-Theoretical Model of Health Behavior Change

In order to link together concepts drawn from a variety of theories it uses a temporal dimension, the stages of change (SoC) construct, as a basic framework around which other model components relating to the promotion of behavioral change (that is, the processes of change components) and its monitoring and support are located. The TTM therefore differs significantly from the other models because it is designed to be of direct value in the delivery of desired behavioral change in individuals and populations. Nevertheless, some of the elements it includes are similar or identical to those utilized in other social cognition based models of health behavior change. The precise format of the TTM and its central stages of change construct has varied over time. The five (or six) stages of change (SoC). These are pre-contemplation (in which the individual has no intention of changing his or her behavior in the foreseeable future); contemplation, in which the individual is considering changing his or her behavior in the next six months; preparation, in which change is planned within the coming month; action, in which stage the individual has made the behavior change within the last six months; and maintenance, in which the health behavior has been sustained for at least six months. A final stage, termination, is included in some versions of the TTM. In this stage the new behavior is seen as being fully established, after a period of five or more years. The progress of individuals between stages is not seen as linear, but as ‘a spiral staircase’ upon which subjects may on occasions ‘jump’ either up or down.

As with other social cognition models the TTM does not normally include objective – defined here as external fact based – measures of health related social, economic and environmental variables. Although it could be used in conjunction with such measures, and so might be able to support action relevant to the reduction of health inequalities, it is not primarily designed to facilitate such approaches. The TTM has been extensively used in health behavior change programs in this country and elsewhere. Regardless of their relative efficacy, such programs appear to have contributed to achieving intermediate health outcomes such as (for example) smoking cessation. The evidence available is also strongly supportive of the view that in the case of smoking cessation improved health outcomes will have in time resulted from such interventions, and that the average cost per quality adjusted life year (QALY) gained is likely to have been modest [77].

The researcher adapted hybrid model to select data for this study but prior to data understanding, hybrid model suggest that business understanding has to be carefully examined in order to solve the problem under discussion, in our case poor utilization of skilled birth attendant given high number of ANC utilization. Therefore, in the study Conceptual frame work from Andersen's Health behavioral model is adapted to understand the problem of the study. The conceptual framework of the use of skilled delivery care services used in this study was based on Andersen's Behavioral Model of Health Services [53]. This model has been widely used to understand the factors that determine an individual's use of health care services in this case delivery service uptake from health institution [53].

Predisposing factors are the combination of demographic characteristics, social structure, and health beliefs. Demographic characteristics are the tendency of the individual to use services which include: age, gender, family size, number of previous pregnancies and marital status. Social structure such as education, occupation and religion or ethnicity will measure the coping ability of the individual with the problem and availability of the resources. Health beliefs are the knowledge about health and health care system; for example, attitudes towards disease and medical care [53].

Enabling factors are the thing which make individual to get access to the health care services such as income, health insurance, availability of health care providers [53].

Need factors are factors which the most direct cause of health service that are the perception of one's own health status and expectation of benefit from the health treatment [53].

For this study, the selected independent variables were in the categories of predisposing, enabling factors and need based factors.

4.2. Data Source and Selection

All the data used for the study were made available upon request from the agencies website for academic purpose. The data sets were main sources to describe key health indicators measuring level of morbidity, mortality and socio-economic progress. Thus, data on skilled delivery with relevant socio-economic characteristics were extracted for this study.

This study is conducted based on the national EDHS 2016 data which were carried out under the guidance of the Federal Ministry of Health (FMOH) and were implemented by the Central Statistical Agency (CSA). The surveys were part of the worldwide Demographic and Health Survey (DHS) program, which was designed to provide information on population, family planning, and health.

In the EDHS 2016 used three questionnaires: the household Questionnaire, the Women's Questionnaire, and the Men'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.

The components of maternal health care covered in the survey included antenatal care, delivery and postnatal care. Women aged 15– 49 who gave birth within five years preceding the survey were asked information on utilization of skilled delivery care services. If the woman had more than one child in the five years preceding the survey, information on the use of delivery assistance was collected for the last birth.

The EDHS is basically a descriptive cross-sectional survey which employed quantitative research methods. The sample for the EDHS were designed to provide population and health indicators at the national (urban and rural) and regional levels. The sample design allowed for specific indicators to be calculated for each of Ethiopia's 11 geographic/administrative regions (the nine regional states and the two city administrations).

The 2016 Ethiopia Demographic and Health Survey (EDHS) is the fourth Demographic and Health Survey conducted in Ethiopia. It was implemented by the Central Statistical Agency

(CSA) at the request of the Federal Ministry of Health (FMoH). Data collection took place from January 18, 2016, to June 27, 2016.

The primary objective of the 2016 EDHS is to provide up-to-date estimates of key demographic and health indicators and hence maternal health care indicators were also included in the survey which is the main area of study for this research. The EDHS provides a comprehensive overview of population, maternal, and child health issues in Ethiopia. More specifically, the 2016 EDHS:

- ✓ Collected data at the national level that allowed calculation of key demographic indicators, particularly fertility and under-5 and adult mortality rates
- ✓ Explored the direct and indirect factors that determine levels and trends of fertility and child mortality
- ✓ Measured levels of contraceptive knowledge and practice
- ✓ Collected data on key aspects of family health, including immunization coverage among children, prevalence and treatment of diarrhea and other diseases among children under age 5, and maternity care indicators such as antenatal visits and assistance at delivery
- ✓ Obtained data on child feeding practices, including breastfeeding
- ✓ Collected anthropometric measures to assess the nutritional status of children under age 5, women age 15-49, and men age 15-59
- ✓ Conducted hemoglobin testing on eligible children age 6-59 months, women age 15-49, and men age 15-59 to provide information on the prevalence of anemia in these groups
- ✓ Collected data on knowledge and attitudes of women and men about sexually transmitted diseases and HIV/AIDS and evaluated potential exposure to the risk of HIV infection by exploring high-risk behaviors and condom use
- ✓ Conducted HIV testing of dried blood spot (DBS) samples collected from women age 15-49 and men age 15-59 to provide information on the prevalence of HIV among adults of reproductive age
- ✓ Collected data on the prevalence of injuries and accidents among all household members

The 2016 EDHS collected information on ANC, assistance during delivery and postnatal care coverage from responses of women who had a live birth in the five years preceding the survey. For women with two or more live births during the five-year period, the EDHS data refer to the most recent birth. Therefore, a total of 4,933 respondents in the five years preceding the survey for their last birth who had ANC follow up reported for 2016 EDHS were used for this particular study.

The 2016 EDHS shows that 62% of women who had a live birth in the 5 years before the survey received ANC from a skilled provider at least once for their last birth and 28% of births were delivered by a skilled provider

4.3. Variables Selected on Maternal Health

Important variables are selected based on Andersen's behavioral model in consultation with domain expert for the study. Five domain experts; two Gynecologists, two midwives and one center head who are involved in this study selected from maternal health centers among 18 health professionals who are providing both ANC service and Delivery. Due to the fact that if we couldn't select appropriate variables which can predict the problem under study it may take too much time to build a model, if could enroll all available variables and also may lead into wrong models [20].

Figure 8, summarizes the selected preprocessed final variables used in the analyses. From the predisposing factors the researcher has been selected: Mother's Age at birth, Mother Education level, Partner Education level, marital status, Birth Order, Religion, Partner Occupation, and Mother Occupation.

From the enabling factors the researchers have been selected: Region, Residence, Access to Media, and Wealth Index. Finally, three variables were selected from Need factors: Timing of 1st ANC visit in months, Frequency of ANC Visit, and Core component of ANC services.

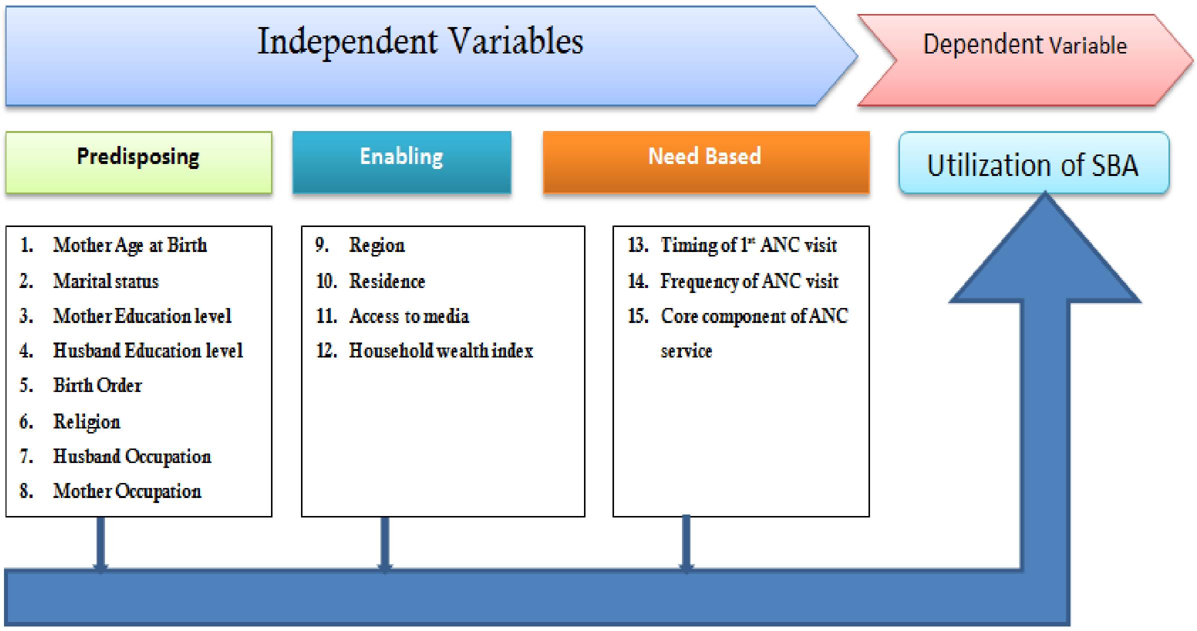

Figure 8. Selected Variables based on Anderson Health Framework

4.4. Description of the Selected Attributes

The description of the selected attributes with their data type, values, and percentage of missing values are described and presented in Annex 1. Final the selected attributes were preprocessed before feeding into data mining tool, WEKA to develop the best model which will suited for the problem under study. Data Pre-Processing

Most of the raw data contained in databases has to be passed through data cleaning and data transformation process because most database contained un-preprocessed, incomplete, and noisy data. The Databases may contain missing values, outliers, inconsistent and duplicate data which needs to be corrected before applying data mining algorithm for creating models [78]. The longlisted selected attributes have to be passed through data pre-processing to get shortlisted variables with accurate values to mine model for the problem under study and the final attribute will be presented with statistical summary on table. Data mining often deals with data that hasn't been looked at for years, so that much of the data contains field values that have expired, are no longer relevant, or are simply missing. The overriding objective is to minimize GIGO: to minimize the "garbage" that gets into our model so that we can minimize the amount of garbage that our models give out [78].

4.4.1.Handling Missing Data

Missing data is a problem that continues to plague data analysis methods. Even as our analysis methods gain sophistication, we continue to encounter missing values in fields, especially in databases with a large number of fields. The absence of information is rarely beneficial. All things being equal, more data is almost always better. Therefore, we should think carefully about how we handle the thorny issue of missing data [78].

Missing data is a common problem that can affect the data analysis process. Missed values in the data set will minimize the accuracy of classification and affect the models to be generated from the data mining algorithm.

The first step in Data mining and knowledge discovery before data analysis is Solving the problem of missing data. Handling the missed values using appropriate method doesn't affect the rule to be generated from the data [79].

A common method of handling missing values is simply to omit from the analysis the records or fields with missing values. However, this may be dangerous, since the pattern of missing values may in fact be systematic, and simply deleting records with missing values would lead to a biased subset of the data. Further, it seems like a waste to omit the information in all the other fields, just because one field value is missing. Therefore, data analysts have turned to methods that would replace the missing value with a value substituted according to various criteria.

Insightful miner offers a choice of replacement values for missing data:

1. Replace the missing value with some constant, specified by the analyst.

2. Replace the missing value with the field mean (for numerical variables) or the mode (for categorical variables).

3. Replace the missing values with a value generated at random from the variable distribution observed [78].

Therefore, in this research study the researcher replaced the missing values with the field mean for numerical attributes.

S/No.	Attribute	Data Type	Frequency of missing Value	% of Missing Values	Replaced Value
1	First ANC visit by Mothers during pregnancy in month	Numeric	27	1	4.1
2	Total number of ANC visit by mother during pregnancy	Numeric	19	1	4

Table 3. Summarized numeric valued Attributes with missing values replaced by mean

Therefore, the two missed values of the dataset were handled are presented in table 3 for numeric valued attributes.

4.4.2.Data Transformation and Reduction

Data transformation is the process of converting data into appropriate forms for mining. Data analysis and mining on EDHS2016 which is huge may take very long time and can give biased result output. The process of data transformation might include smoothing (e.g. Using bin means to replace data errors), Normalization, where the attribute data are scaled so as to fall within a small specified range (eg. Mother's age at birth), and Attribute construction, where new attributes are constructed and added from the given set of attributes to help the mining process [78]. Data reduction Obtain a reduced representation of the data set that is much smaller in volume but yet produces the same (or almost the same) analytical results.

Data reduction specifically dimensionality reduction was applied in the data set. The total number of attribute before reduction was 5,704 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. From the original dataset the age attribute for mothers was in continuous values and calculated as follows:

Mother's age at birth is calculated from the CMC of the date of last birth and the CMC of the date of birth of the Mother and discretized

$$\text{Mother's age at birth} = (Last\ birth's\ date\ \ in\ CMC\ - Mother's\ date\ \ in\ CMC\ /12 \quad 4.5$$

And has to be changed to integer. After the calculation we found that the mother's age in continuous values. Therefore, for this study the mother's age at birth is discretized by equal width and presented in the following table.

Mother's Age at birth	Represented Value	Count
15-19	1	588
20-24	2	1310
25-29	3	1305
30-34	4	836
35-39	5	483
40-44	6	150
45-49	7	12

Table 4. Discretized Mother Age in to seven categories.

The attribute birth order is a continuous variable in which discretization was performed to convert into four distinct values as shown in table 5 below:

Birth Order of the child	Represented Value	Count
1	First	1147
2-3	2nd or 3rd	1578
4-5	Fourth or Fifth	993
6+	Sixth or Higher	966

Table 5. Discretized Birth Order.

The attribute number of Timing of first antenatal visit in month is a continuous variable which discretization was performed to convert attributed values into four distinct values as shown in table 6 below:

Timing of 1st ANC Visit in Month	Represented Value	Count
1-3	less than Four Month	1807
4-5	Four to Five	1897
6-7	Six to Seven	842
8+	Eight Plus	138

Table 5. *Discretized number attributed Timing of 1st ANC Visit in the month.*

The attribute frequency of ANC visit is also a continuous variable which discretization performed and presented in Table 7.

Frequency of ANC Visit	Represented Value	Count
1	Only Once	336
2-3	Two/three times	1739
4+	Four/above	2609

Table 6. *Discretized number of ANC visit.*

Other attributes which need transformation are Frequency of reading news, Frequency of listening to radio, and Frequency of watching television each of them has three values (Not at all, Less than once a week, and At least once a week). The three attributes are combined and a new attribute named Media Exposure is created (see Table 8). Highest values from the three attributes taken to get values for the new Attribute-Media Exposure (Not at all, Less than a week, and at least once a week).

Frequency of Public Media access(Media Exposure)	Represented Value	Count
0	Not at all,	2547
1	Less than once a week	742
2	At least once in a week	1395

Table 7. Transformed variable, Media Exposure.

The other attributes transformed is core ANC services expected to be taken during ANC visits such as Blood pressure measured, Urine sample taken, blood sample taken, nutrition counseling provided. The values for this new attribute (Number of Core component of ANC services) are No Service, only one of the core service, two, three and four (see table 9).

Number of Core ANC services taken during ANC	Represented Value	Count
No Service	0	342
Only One core service	1	370
Two	2	524
Three	3	1130
Four	4	2318

Table 8. Transformed variable, core ANC services.

Therefore, in this research discretization were applied in the Mother's age at birth, birth order, Timing of 1st ANC visit in month and Frequency of ANC Visit during pregnancy for continuous values variables.

Finally, after pre-processing the original dataset assumed to be relevant to the target variable, which consists of 16 variables 15 Predictors (Independent) and 1 Outcome (Dependent) variables were selected and 4,933 instances was used for constructing the model.

4.5. Statistical Summary of the Selected Attributes

The summary of the selected attributes used for model building are statistically described in detail in Table 10. This statistical summary of the attributes is helpful for understanding of the data set for DM model building phase.

No.	Variable	Data Type	Represented values	Frequency	Percent(%)
I	Predisposing factors: Independent				
1	Mothers Age at birth	Nominal	4933		
	15-19		1	588	12.6
	20-24		2	1,310	28.0
	25-29		3	1,305	27.9
	30-34		4	836	17.8
	35-39		5	483	10.3
	40-44		6	150	3.2
	45-49		7	12	0.3
	Total			4,933	100.0
2	Mother' Highest level Education	Nominal			
	No Education		0	2,321	49.6
	Primary Education		1	1,538	32.8
	Secondary Education		2	518	11.1
	Higher Education		3	307	6.6
	Total			4,933	100.0
3	Husband/Partner Highest level Education	Nominal			
	No Education		0	1,994	42.6

	Primary Education		1	1,560	33.3
	Secondary Education		2	634	13.5
	Higher Education		3	496	10.6
	Total			4,933	100.0
4	Marital Status	Nominal			
	Never Married		0	43	0.9
	Married		1	4,271	91.2
	Living together with partner		2	57	1.2
	Widowed		3	58	1.2
	Divorced		4	191	4.1
	Not living together		5	64	1.4
	Total			4,933	100.0
5	Birth Order	Nominal			
	One		1	1,147	24.5
	Two or three		2	1,578	33.7
	Four or Five		3	993	21.2
	Six Plus		4	966	20.6
	Total			4,933	100.0
6	Religion	Nominal			
	Orthodox		1	1,843	39.3
	Catholic		2	28	0.6
	Protestant		3	852	18.2
	Muslim		4	1,924	41.1
	Traditional		5	16	0.3

	Others		6	21	0.4
	Total			4,933	100.0
7	Mother's Occupation	Nominal			
	Don't work		0	2,469	52.7
	Agricultural employee		1	901	19.2
	Professonial/technical Manager's and Higher position		2	140	3.0
	Non-Manager's		3	1,174	25.1
	Total			4,933	100.0
8	Husband/Partner Occupation	Nominal			
	Don't work		0	331	7.1
	Agricultural employee		1	2,455	52.4
	Professonial/technical Manager's and Higher position		2	392	8.4
	Non-Manager's		3	1,506	32.2
	Total			4,933	100.0
II	Enabling factors: Independent				
9	Region	Nominal			
	Tigray		1	690	14.7
	Afar		2	286	6.1
	Amhara		3	502	10.7
	Oromiya		4	522	11.1
	Somali		5	347	7.4
	Benshangul-Gumuz		6	387	8.3

	SNNP		7	623	13.3
	Gambela		8	317	6.8
	Harari		9	316	6.7
	Addis Ababa		10	363	7.7
	Dire Dawa		11	331	7.1
	Total			4,933	100.0
10	Residence	Nominal			
	Urban		1	1,391	29.7
	Rural		2	3,293	70.3
	Total			4,933	100.0
11	Access to Media	Nominal			
	Not at all		0	2,547	54.4
	Less than once a week		1	742	15.8
	At least once a week		2	1,395	29.8
	Total			4,933	100.0
12	Household Wealth Index	Nominal			
	Poorest		1	1,059	22.6
	Poorer		2	754	16.1
	Middle		3	704	15.0
	Richer		4	667	14.2
	Richest		5	1,500	32.0
	Total			4,933	100.0
III	Need factors: Independent				
13	Timing of 1st antenatal check (months)	Nominal			

	less than four month		1	1,807	38.6
	Four to five month		2	1,897	40.5
	Six to Seven		3	842	18.0
	Eight plus		4	138	2.9
	Total			4,933	100.0
14	Number of ANC Visit during pregnancy	Nominal			
	Only Once		1	336	7.2
	Two or three		2	1,739	37.1
	Four plus		3	2,609	55.7
	Total			4,933	100.0
15	Core compenents of ANC service taken during ANC Visit	Nominal			
	No Service		0	342	7.3
	Only One		1	370	7.9
	Two		2	524	11.2
	Three		3	1,130	24.1
	Four/All service		4	2,318	49.5
	Total			4,933	100.0
IV	Dependent Variable				
16	Utilization of Skilled birth attendant	Nominal			0.0
	No		No	2,215	47.3
	Yes		Yes	2,539	52.7
	Total			4,933	100.0

Table 9. Statistical summary of the selected variables from EDHS,2016 data.

Further descriptions and coding of the dependent and independent variables which were employed in this study are briefly described in the following sections.

Dependent Variable: Utilization of Skilled birth attendant:

In the 2016 EDHS, the respondents were asked, with respect to the last birth occurring in the five years preceding the survey, who assisted with the delivery. From this specific question, dichotomous variable was created for this study. It was coded as 1 if the woman received assistance at delivery from SBAs including: Births delivered with the assistance of doctors, nurse/midwives, health officers, and health extension workers and 0 otherwise.

Independent variables selected under this study were from three components according to Andersen's behavioral model of the use of health services, and hence 18 independent variables were selected and presented in table 13 above for this study. Nine variables were from predisposing factors and five from enabling factors and 5 from Need factors.

From the predisposing factors 8 variables such as Mother's Age at birth, Mother Education level, Partner Education level, marital status, Birth Order, Religion, Partner Occupation, and Mother Occupation were selected and presented as follows:

Mothers 'Age at birth: In the EDHS, 2016 survey, women's age was ranging from 15-49. In this study, women were classified into seven age groups. These were preprocessed and coded as 1, 2, 3,4,5,6 and 7 for the age groups of 15-19, 20-24,25-29,30-34, 35-39,40-44, and 45-49 respectively.
Accordingly, the highest frequencies are 1,310(28%) and 1,305(28%) for the age group of 20-24 and 25-29 respectively and the least frequent age group is from 45-49, which only 12(0.3%).

Mother's Education level: This referred to the highest levels of education of the women. In the survey, women were classified into four levels of education including:

no education, primary, secondary, and higher education. In this study, women were categorized into the same four categories and were coded as 0, 1, and 2, 3 for the level of education of no education, primary education, and higher education, respectively. According to table 13, the highest education level attained by women is no education which is 2,321(50%) and least scored education level of women is Higher education which is 307(7%).

Husband/Partner Education level: Similar to the level of education of women. The highest number of education level attained by Husband is No education which is 1,994 and the least scored is Higher education attained by Husband/partners of women is 496(11%).

Religion: This referred to the type of religion that the women had at the time of the survey. In the survey, women were classified into six different religions namely Orthodox, Catholic, Protestant, Muslim, Traditional and Others. These were coded as 1, 2, 3, 4, 5, and 6 for the classification of Orthodox, Catholic, Protestant, Muslim, Traditional and Others respectively. As shown in the table 13 the highest and least numbers are 1,924(41%) and 16(0.3%) for Muslim and traditional respectively.

Mother's occupation: This was based on the question of the type of work that the respondent's mother does primarily. In this study, mother's occupations were categorized into four categories including: Don't work, agriculturally employed, Professional/technical managers, and Non-Managers were coded as 0, 1, 2, and 3, respectively. According to table 13, 2,539(53%) of mothers categorized into don't have work and only 140(3%) of mothers were employed at Higher post and categorized as Professional and technical managers group.

Husband/Partner's occupation: This similar to mother's occupation category and accordingly a total of 2,455(52%) of husband/partners of mothers were employed as agricultural and only about 8.4%(392) husbands were employed as professional or technical managers.

Marital Status: The respondents were asked about their current marital status during the survey. In this study, the Marital status of Mothers were classified into six categories including: Never married, Married, living together with partner, widowed, divorced and not living with partner and were coded as 0,1,2,3,4, and 5 respectively. Almost all or 4,271(91%) of mothers were categorized into Married and the least Marital status category of women was widowed, which is 58(1.2%).

Birth Order: The respondents were asked about their last birth order during the survey. In this study, the last Birth Order of mother were categorized into first, second/third, fourth/fifth and sixth plus and were coded as,1,2,3, and 4 respectively. About 34% of Mothers birth order was categorized into second/third time's delivery and the last categorized order was sixth plus which is around 20% of the respondents.

As shown in table 13, four attributes were selected from enabling factors 3such as region, residence, Access to media and Wealth Index and described in the following section as follows.

Region: In the 2016 EDHS, all regions of the country were included in the survey. In this study, the region of the country was categorized as: Tigray, Afar, Amhara, Oromia, Somali, Benshangul Gmuz, SNNP, Gambella, Harari, Addis Ababa and Dire Dawa and coded as 1,2,3,4,5,6,7,8,9,10, and 11 respectively. As shown in the table 13 above, all regions were representatively included in the survey but Tigray region was taking the first rank,14% of respondents were from Tigray region and the least represented region was Afar which is around 6% of the respondents included from the Afar region.

Residence: In the 2016 EDHS, respondents were involved from Rural and Urban. In this study, the residence of the mother was categorized as: Urban and Rural, and coded as 1 and 2 respectively.70% of respondents were categorized as Rural and the remaining 30% included from Urban.

Wealth Index: In the 2016 EDHS, wealth index was constructed from data on household possession. This was based on the questions about whether a household had items such as radios, televisions, and bicycles, and facilities such as type of floor, piped water, toilets, and electricity. Each asset was assigned a weight, and each household was then assigned a score for each asset, and the scores were summed for the particular household. Individuals were then ranked according to the total score. The higher the score, the higher the economic status of the household. This variable was coded as 1, 2, 3, 4, and 5 for poorest, poorer, middle, richer, and richest, respectively. As shown in the table 13 above, the highest economic status of mothers was categorized into richest group is about 32% and the least scored percentage of the economic status was categorized as richer which is about 14.2%.

Access to Media: In the 2016 EDHS, respondents were asked whether thy had access to media. In this study, Access to media were categorized as: Not at all, less than once a week and at least once a week coded as 0,1, and 2 respectively. More than half around 54% of respondents were categorized as Not at all and the least access to media category according to table 13 was less than once a week, which accounts for 16%.

The third health behavior factors according Andersen Model is Need factors. In this study four attributes were selected under Need factors, such as Timing of 1st antenatal check (months), Number of ANC Visit during pregnancy, and Core components of ANC service taken during ANC Visit.
These attributes will be briefly described in the following section as follows:
Core Components of antenatal care: In the 2016 EDHS, Pregnant women were asked about the core components of ANC, such as blood pressure measured, blood sample taken, urine sample and nutritional counselling. In this study, core components of ANC were categorized as
No service, only one service, two service, three service and Four/all services and coded as 0,1,2,3 and 4.50% of women categorized as Four/all, that means almost half of the women received all core components of ANC service during ANC visit and 7.3% of Mothers categorized as No service.

Timing of 1st antenatal check (months): In 2016 EDHS, pregnant women were asked when thy made the first antenatal check during pregnancy. In this study, this attribute categorized as less than four months, four to five months, six to seven months, and eight plus months coded as 1,2,3 and 4 respectively. Table 13 shown as 41% of mothers had their first antenatal check within four to five months and only 3% of mothers had their first antenatal visit in the last category eight plus months.

CHAPTER FIVE

5. Experimentation and Knowledge Discovery

This research work is aimed to find pattern of factors affecting utilization of skilled birth attendant among ANC users using classification techniques. Three different supervised techniques are used; Decision tree classifier, Naïve Bayesian classifier and PART rule induction classifier using WEKA data mining tool.

In the hybrid KDP methodology selected for this study, model building is an iterative process. Therefore, it is important to conduct different experiments using various data mining methods to derive knowledge from the preprocessed data to predict the factors which affects the use of skilled birth attendant among ANC users.

Totally thirty-six (36) experiments were conducted using the three algorithms to find the pattern which mostly affect the utilization of Skilled birth attendant during birth among ANC users: J48 Decision Tree, Rule Induction and Naïve Bayes using Weka 3.6.18 tools. All experiments were evaluated to find the best performing methods for model construction.

The class is somehow not balanced so that Resampling is used to balance the class. The following figure shows the total instances after Resampling. From the total of 4,933 instances 2,490 are "No", Not assisted by Skilled Health providers during delivery and 2,443 "yes", assisted by skilled professionals during birth.

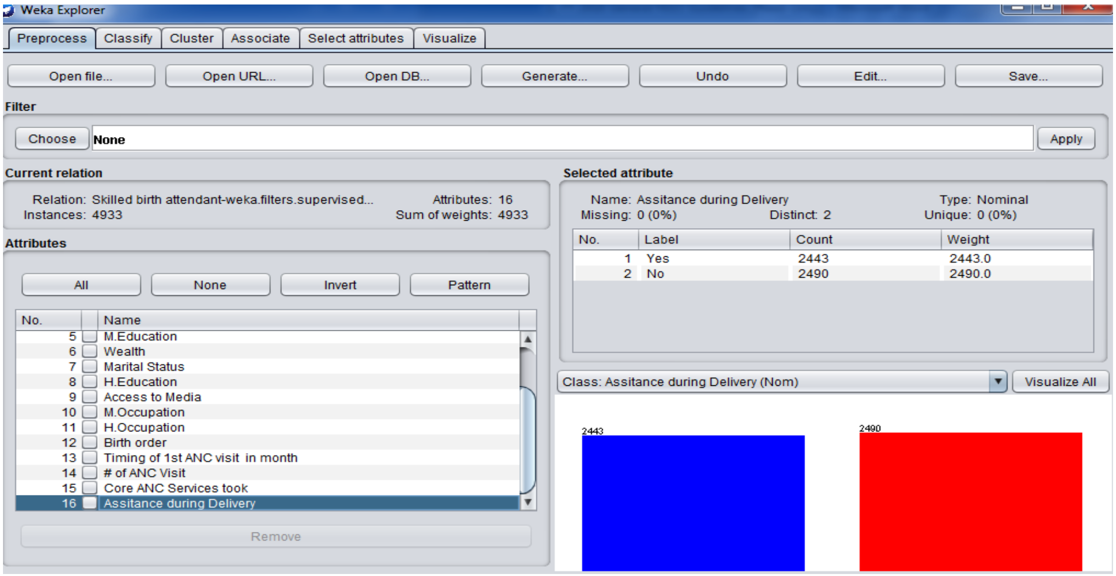

Figure 9. WEKA 3.8.1 Explorer Windows Showing the Number of Attribute and Instances with SMOTE-Resampling.

5.1. Experimental Design

Model building is one of the tasks in data mining process and the model is prepared by providing the processed data to the selected data mining algorithms with different parameters to have a model with better performance. There are a number of tasks involved during model building like selecting the modeling technique, experimental setup or design, building the model and evaluating the model.

In any data mining process before building a model, there is a need to generate a procedure or mechanism to test the model's quality and validity. For instance, in supervised data mining tasks such as classification, it is common to use classification accuracy measure or error rates as quality measures for data mining models.

In this study after doing the preprocessing task 4,933 datasets with 15 independent attributes and one dependent variable were used for training and testing. WEKA 3.8.1 tool has been used to measure the quality, validity and test of the selected model, since it is open source and widely used software. For purposes of this study 10-folds cross validation and 66% split test options are used because of its relatively low bias and variations [80].

In 10-fold cross validation; the data were divided into 10 folds where 9 folds were used as training data whereas the remaining one folds as test data. In the percentage split method, where 66% of the data was used as training and the remaining 34% was used as test data. Accuracy, Precision, specificity, ROC curve, mean absolute error, Recall and confusion matrix standard metrics were also used for evaluation of the results.

In order to get a better predictive model three classification algorithms namely J48 decision tree, PART induction rule, and Naive Bayes algorithms were selected and experimented through Weka machine learning software. These algorithms were selected; because as discussed clearly in the literature reviews that these models were widely used and applied in health care sectors.

To select the best predictive model for this study the following three scenarios has been conducted with different parameter values of the classification algorithms using WEKA 3.8.1 tool.

Scenario 1: J48 Decision tree with pruning.

Scenario 2: PART Rule Induction with pruning.

Scenario 3: Naive Bayes.

In the first scenario sixteen experiments will be done using J48 Decision tree with pruning with 10-fold cross validation and 66 percentage split tests with its default and modified parameters and also with PART Rule induction with pruning a total of 16 experiments done with its default and modified parameters with 10-fold cross validation and 66 percentage splits. In third scenario a total of four experiments done with Naïve Bayes classification algorithms with its default values with 10-fold cross validation and 66 percentage splits. The detail parameter for each of the classification algorithms will be presented in the later sections.

5.1.1.Feature Selection

The feature selection algorithm removes the irrelevant and redundant features from the original dataset to improve the classification accuracy. The feature selections also reduce the dimensionality of the dataset; increase the learning accuracy, improving result comprehensibility [82].

The feature selection avoids over fitting of data [82]. The feature selection also known as attributes selection which is used for best partitioning the data into individual class. 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 is 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.

```
Ranked attributes:
 0.1662    2 Region
 0.1569    6 Wealth
 0.1566    5 M.Education
 0.1396    3 Residence
 0.1048    8 H.Education
 0.0885    9 Access to Media
 0.0791   11 H.Occupation
 0.0778   15 Core ANC Services took
 0.0763   10 M.Occupation
 0.076    12 Birth order
 0.0483   14 # of ANC Visit
 0.042    13 Timing of 1st ANC visit  in month
 0.0316    4 Religion
 0.0283    7 Marital Status
 0.0123    1 M.Age at birth
```

Figure 10. Attribute Selection on all input data using WEKA.

The above output showed ranked attribute with their corresponding information gain value. WEKA evaluates the worth of an attribute by measuring the information gain with respect to the class. The results show us the first ranked attribute is Region with information gain value of 0.1662 and the last ranked attribute is mother's age at birth with information gain value of 0.0123. Therefore, the tree will be constructed in such a way that the root of the tree is region and going down to leaves until ends with the values of the class either yes or no.

5.2. Model Building Using J48 Decision Tree

In this study J48 decision tree classifier is selected due to it is predictive ability and potential produce accurate and easily interpretable patterns/tree that helps to achieve the research objectives. The J48 decision tree algorithm builds decision trees from a set of predefined training dataset using the concept of information entropy and attribute ordering. It uses the fact that each attribute of the data was used to make a decision by splitting the data into smaller subsets. The following experiments will be done and Table 14 presents lists of parameter of the experiments to be done. In this study pruning decision tree is used because pruning a decision tree is a fundamental step in optimizing the computational efficiency as well as classification accuracy. Pruning usually results in reducing size of tree, avoids unnecessary complexity, and to avoid overfitting of the data sets when classifying new data [81].

The synopsis of the selected J48 classifier parameters with description as follows: [82]

- binarySplits -- Whether to use binary splits on nominal attributes when building the trees. Its default value is set by WEKA to false, if this value change true it enforces the model to generate binary tree than generalized tree. Therefore, in this this study default value is used to generate generalized tree.
- confidenceFactor -- The confidence factor used for pruning (smaller values incur more pruning). The default value for confidence factor is 0.25 which made the algorithm more pruning and 0.50 are used in the study.

- debug -- If set to true, classifier may output additional info to the console. Default value is used.
- minNumObj -- The minimum number of instances per leaf. The default value of minimum number of instances per leaf is set 2 and 5 is used in this study.
- numFolds -- Determines the amount of data used for reduced-error pruning. One fold is used for pruning, the rest for growing the tree and default value is used.
- reducedErrorPruning -- Whether reduced-error pruning is used instead of C.4.5 pruning.
- saveInstanceData -- Whether to save the training data for visualization.
- seed -- The seed used for randomizing the data when reduced-error pruning is used.
- subtreeRaising -- Whether to consider the subtree raising operation when pruning.

Therefore, a total of sixteen experiments are built with the default parameter values of most of its parameters but only changed to see the effects of changing the default of confidence factor and minNumbobj for some of the experiments for J48 Decision tree algorithm and the details of the parameters used in this experiments are presented in table 11.

Experiments	Pruned	Confidence factor	(min Numobj)	Test option	Attribute
1	True	0.25	2	10 fold cross validation	All
2	True	0.25	5	10 fold cross validation	All
3	True	0.5	2	10 fold cross validation	All
4	True	0.5	5	10 fold cross validation	All
5	True	0.25	2	10 fold cross validation	Selected
6	True	0.25	5	10 fold cross validation	Selected
7	True	0.5	2	10 fold cross validation	Selected
8	True	0.5	5	10 fold cross validation	Selected
9	True	0.25	2	66% percentage split	All
10	True	0.25	5	66% percentage split	All
11	True	0.5	2	66% percentage split	All
12	True	0.5	5	66% percentage split	All
13	True	0.25	2	66% percentage split	Selected

14	True	0.25	5	66% percentage split	Selected
15	True	0.5	2	66% percentage split	Selected
16	True	0.5	5	66% percentage split	Selected

Table 10. *Values of parameters used for J48 Decision tree experiments.*

The evaluation of the performance of the model generated from J48 decision trees are presented in terms of Accuracy, precision, TP rate, FP rate of the above sixteen experiments are described in the Table 12.

Experiments	Accuracy (%)	TP rate	FP rate	Precision	Recall	F-Measure	ROC area
1	98.49	0.985	0.015	0.985	0.985	0.985	0.994
2	97.81	0.987	0.022	0.978	0.978	0.978	0.992
3	98.74	0.987	0.013	0.987	0.987	0.987	0.995
4	97.97	0.980	0.02	0.98	0.98	0.98	0.993
5	98.39	0.984	0.016	0.984	0.984	0.984	0.995
6	97.20	0.972	0.028	0.972	0.972	0.972	0.993
7	98.74	0.987	0.013	0.987	0.987	0.987	0.996
8	97.28	0.973	0.027	0.973	0.973	0.973	0.993
9	98.03	0.98	0.02	0.988	0.988	0.988	0.994
10	97.07	0.971	0.029	0.971	0.971	0.971	0.99
11	98.74	0.986	0.014	0.986	0.986	0.986	0.995
12	97.79	0.978	0.022	0.978	0.978	0.978	0.994
13	97.79	0.978	0.022	0.978	0.978	0.978	0.994
14	96.42	0.964	0.036	0.965	0.964	0.964	0.985
15	98.27	0.983	0.018	0.983	0.983	0.983	0.991
16	96.89	0.969	0.32	0.97	0.969	0.969	0.987

Table 11. Experimental result of J48 Algorithms.

As described in Table 12, most of the experiments performed nearly equally but Experiment 7 performed best than the other experiments with slightly higher accuracy of 98.74 and ROC area of 99.6% with test option of 10-fold cross validation with selected attributes.

Therefore, the best performing experiment selected is Experiment 7 of J48 decision tree with test option of 10-fold cross validation with selected attributes to be compared with others classification algorithms so as to select the best algorithms to build the model.

5.3. Model Building PART Rule Induction

In this study, the second classifier algorithm selected is PART rule indication because of its ability and potential to produce accurate and easily interpretable patterns that helps to predict the utilization of skilled birth attendant among ANC clients at health institution. The same experiments were used as J48 decision tree for comparison. Table 13 presents lists of experiments prepared to conduct experiments with default and modified values.

Experiments	Pruned	Confidence factor	(min Numobj)	Test option	Attribute
17	True	0.25	2	10 fold cross validation	All
18	True	0.25	5	10 fold cross validation	All
19	True	0.5	2	10 fold cross validation	All
20	True	0.5	5	10 fold cross validation	All
21	True	0.25	2	10 fold cross validation	Selected
22	True	0.25	5	10 fold cross validation	Selected
23	True	0.5	2	10 fold cross validation	Selected
24	True	0.5	5	10 fold cross validation	Selected
25	True	0.25	2	66% percentage split	All
26	True	0.25	5	66% percentage split	All
27	True	0.5	2	66% percentage split	All
28	True	0.5	5	66% percentage split	All
29	True	0.25	2	66% percentage split	Selected
30	True	0.25	5	66% percentage split	Selected
31	True	0.5	2	66% percentage split	Selected
32	True	0.5	5	66% percentage split	Selected

Table 12. *Values of parameters used for PART Rule Induction experiments.*

The evaluation results of all the sixteen experiments of PART rule Induction are presented in table 14 for analysis.

Experiments	Accuracy (%)	TP rate	FP rate	Precision	Recall	F-Measure	ROC area
17	98.4796	0.985	0.015	0.985	0.985	0.985	0.993
18	97.6485	0.976	0.024	0.977	0.976	0.976	0.992
19	98.2707	0.983	0.017	0.983	0.983	0.983	0.99
20	98.0539	0.981	0.02	0.981	0.981	0.981	0.994
21	98.4796	0.985	0.015	0.985	0.985	0.985	0.995
22	97.2228	0.972	0.028	0.972	0.972	0.972	0.99
23	98.4796	0.985	0.015	0.985	0.985	0.985	0.994
24	97.2228	0.972	0.028	0.972	0.972	0.972	0.99
25	98.0918	0.981	0.019	0.981	0.981	0.981	0.994
26	96.8992	0.969	0.031	0.969	0.969	0.969	0.988
27	97.5552	0.976	0.025	0.976	0.976	0.976	0.994
28	97.0185	0.97	0.03	0.97	0.97	0.97	0.991
29	98.2707	0.983	0.017	0.983	0.983	0.983	0.99
30	96.9589	0.97	0.03	0.97	0.97	0.97	0.989
31	97.7937	0.978	0.022	0.978	0.978	0.978	0.99
32	96.3626	0.964	0.037	0.964	0.964	0.964	0.984

Table 13. Experimental result of PART Rule Induction Algorithms

As described in Table 14, experiment 21 performed best with accuracy of 98.48 % and ROC area of 99.5% with 10-fold cross validation test option with selected attributes.

Therefore, the researcher selected pruned PART Rule Induction with 10-fold cross validation test option with selected attributes to compare it with other classification algorithms.

5.4. Model Building Using Naïve Bayes classifier

The Bayesian Learning Algorithms combine training data with a priori knowledge to get the posteriori probability of a hypothesis. So, it is possible to figure out the most probable hypothesis according to the training data.

The researcher selected Naïve Bayes for the reason that Naïve Bayes has the ability and potential to produce accurate that helps to achieve the research objectives. To build the model, WEKA software package and the same number of datasets was used as an input. The experiments were divided into two test options that are 10-fold cross validation and percentage split. The following lists of experiments were prepared for Naïve Bayes algorithm with its default values of its parameters as bellow:

Experiments	Debug	Test option	Attribute
33	False	10 fold cross validation	All
34	False	10 fold cross validation	Selected
35	False	66% percentage split	All
36	False	66% percentage split	Selected

Table 14. Values of parameters used for Naïve Bayes algorithms.

The performance measures in terms of accuracy, precision, recall and other effectiveness measures of the above six experiments of Naïve Bayes algorithms are organized in Table 16.

Experiments	Accuracy (%)	TP rate	FP rate	Precision	Recall	F-Measure	ROC area
33	76.0592	0.761	0.242	0.778	0.761	0.756	0.86
34	74.5379	0.745	0.251	0.766	0.745	0.74	0.85
35	73.9915	0.740	0.263	0.757	0.740	0.735	0.851
36	72.3912	0.724	0.272	0.741	0.724	0.72	0.836

Table 15. Experimental result Naïve Bayes algorithms.

As presented in Table 16 Naïve Bayes classifier which was implemented on experiment 33 (Naive Bayes with 10-fold cross validation with all attributes) achieved the highest accuracy of 76.06% and ROC area of 86.0%.

Therefore, the researcher selected Naive Bayes with 10-fold cross validation to compare it with the other two selected classification algorithms in this study.

5.5. Model Comparison

After selecting the best algorithms from each classification techniques experimented the next step is comparing those algorithms for selecting a best classification technique to the utilization of skilled birth attendant among ANC clients.

Accordingly, the three selected classification models J48, PART, and Naive Bayes with their respective best performance metrics presented in the table 17 below:

Classifier Algorithm	Accuracy (%)	TP rate	FP rate	Precision	Recall	F-Measure	ROC area
Experiment 7 - Pruned J48	98.74	0.987	0.013	0.987	0.987	0.987	0.996
Experiment 5 –Pruned PART	98.4796	0.985	0.015	0.985	0.985	0.985	0.995
Naive Bayes	76.0592	0.761	0.242	0.778	0.761	0.756	0.86

Table 16. *Performance comparison of selected best algorithms.*

As shown on Table 17 Pruned J48 decision tree with 10-fold cross validation with selected attributes performed highest with accuracy of 98.74% with modified values it's parameters while with slight difference pruned PART Rule Induction with 10-fold cross validation with selected attributes came to the second position with classification accuracy of 98.48%. Naive Bayes with 10-fold cross validation with all attributes scored the last classification algorithms with accuracy which is 76.05%.

The models built from Pruned J48 decision tree scored ROC Area of 0.996 is the highest among the three classification algorithms used in this study and fall under excellent category.

The experimental results shown that, Pruned J48 decision tree algorithm performed best than the other two classifiers in the domain of predicting utilization of skilled birth attendant taking all measurement tools used in this study such as accuracy, TP rate, FP rate, precision, recall, F-measure, and ROC area curve Therefore, it is reasonable to conclude that the Pruned J48 decision tree is more appropriate for the specified problem domain than the other algorithms used for experimentation based on its better performance results.

5.6. Analysis of the Selected Model

It can be observed from the results obtained J48 decision tree algorithms, selecting attributes provided with better result than with all attributes. Therefore, the performance of pruned J48 decision tree with selected attributes is selected for predicting the utilization of skilled birth attendant among ANC clients.

Table 18 shows the confusion matrix of the selected J48 decision tree for our predictive model

		Predicted Class	
		Yes	No
Actual Class	Yes	2,416	27
	No	35	2,455

Table 17. *Confusion matrix of the selected I48 decision model*.

From the confusion matrix of the selected model 98.74% of total instances are correctly classified.

From the selected model 99.8% instances from the Yes class are correctly classified as Yes and 98.59% instances from the No class correctly classified as No. The model performed in classifying the instances as True Positive class (Sensitivity) of 98.9% and True Negative class (Specificity) of 98.6% with 98.74% of predictive performance. Thus, it is possible to conclude that the model is in a good performance to classify instances into yes and no classes.

From the precision score of the model, the precision of this model for “Yes” class is 98.6%, and the precision of “No” class is 98.9% with an average precision of 98.7 % instances labeled as belonging for each class yes and no. From harmonic mean of precision and recall which is F-score, with value of 0.987, so it can be concluded that the precision and recall of the model are most significantly balanced.

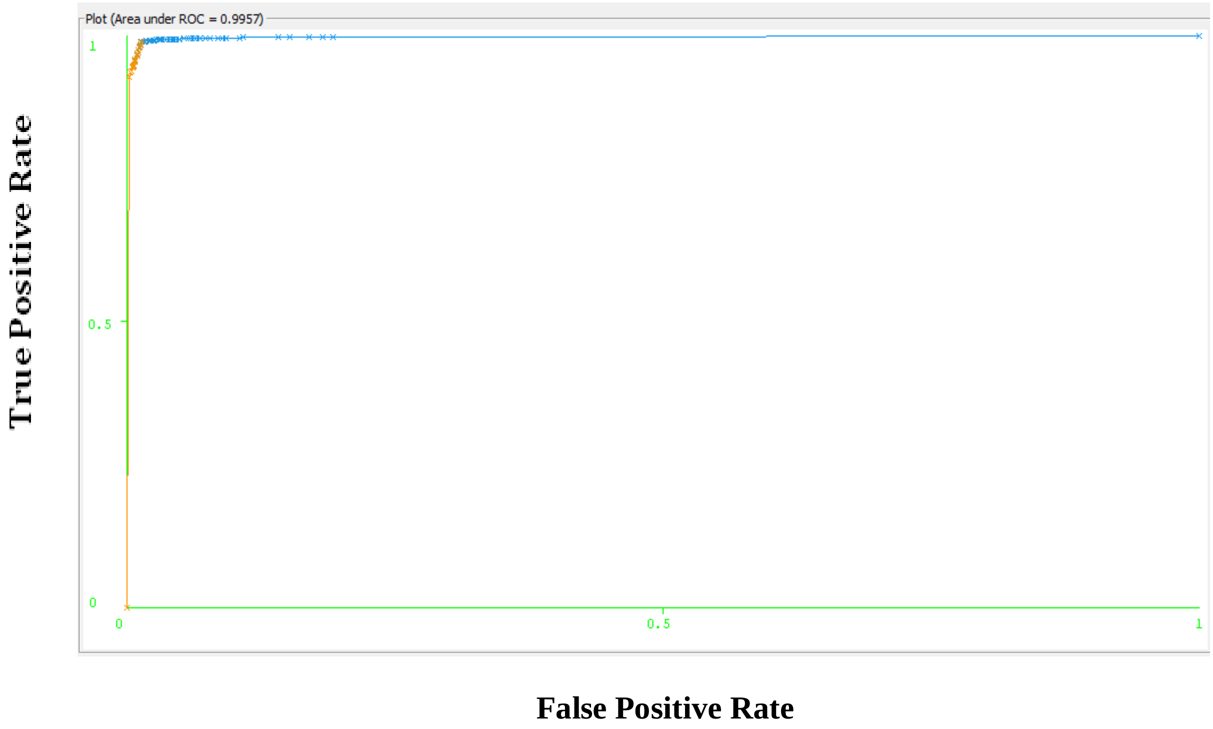

Figure 11. Area under ROC curve.

As it can be observed from the fig. 11 the area under the ROC curve shows better test, which categorized it to excellent ,99.6% which fall under excellent category, therefore the model can best predict the utilization of skilled birth attendant.

5.7. Generated rules from the Selected Model

Pruned J48 decision tree with selected attribute is selected as the best model for rule generation. This is because it achieved better Predictive performance compared with other classifier algorithms and the rules provided by Pruned J48 decision tree like decision tree models can be easily assimilated by human without any difficulty. A total of 358 Leaves and 463 trees produced. However, the researcher selected interesting rules from the generated leaves that cover the problem under study and then consulted with domain experts for validation.

5.8. Discovered Knowledge

Knowledge/rules were generated from the model and discussed with Domain experts for validity whether the generated rules are acceptable by domain experts to better utilize the model for the problem under study, high maternal mortality ratio due to very low practice of skilled birth attendant at health institution during birth by mothers among ANC clients.

Therefore, the researcher selects and discusses a rule that satisfy the assumptions of the domain expert which can describe the problem and based on this assumption, some of the best rules/patterns are extracted from J48 decision tree for the utilization of skilled birth attendant. The following rules indicate that the possible conditions in which a woman could be classified in skilled birth attendant during birth among ANC follow up clients.

The following rules indicate that the possible conditions in which a woman could utilize the later skilled birth attendant or not during birth among ANC follow up clients.

Rule 1:

IF residence =” Rural” AND Region =” Tigray” AND Household wealth index/economic status of the family =” Richer” AND Husband highest education level = “none”: Then utilization of skilled birth attendant during delivery is NO (17).

Rule 2:

IF residence =” Rural” AND Region =” Tigray” AND household wealth index/economic status of a family =” poorest” AND Highest education level of a woman =” none” AND Husband highest education level = “none” AND Birth order=” First”: Then utilization of skilled birth attendant during delivery is Yes (14).

Rule 3:

IF residence =” Rural” AND Region =” Tigray” AND household wealth index/economic status of a family =” poorest” AND Highest education level of a woman =” Primary” AND Husband highest education level = “Primary”: Then utilization of skilled birth attendant during delivery is Yes (21).

The most determining variables for a woman who is living in rural area of Tigray region are husband highest level of education and Birth order. If a husband’s education level is at least primary and higher then a woman will most likely utilize skilled birth attendant and a woman with first birth order, then she will most likely utilize skilled birth attendant.

Rule 4:

IF residence =” Rural” AND Region =” Afar” AND Women Occupation=” Non-Professional/technical Manger” AND Access to media =” at least once per week”: Then utilization of Skilled birth attendant during delivery No (6).

The media communication channel has to be further investigated because a woman who has good access to media most likely will not utilize skilled birth attendant during delivery for a woman who is living in rural area of Afar region.

Rule 5:

IF residence =” Rural” AND Region =” Amhara” AND Husband’s highest level of education=” Primary” AND Access to media= “none” AND household wealth index/economic status of the family =” Richest”: Then utilization of Skilled birth attendant during delivery No (32).

A woman who lives in rural area of Amhara region with her husband’s highest level of education is at primary level and with no media access then the variable which determine the utilization of skilled birth attendant is the wealth status of a family and it is inversely related with the utilization of skilled birth attendant; if household wealth status is categorized as “richest” then a woman most likely will not utilize skilled birth attendant during delivery.

Rule 6:

IF residence =” Rural” AND Region =” Oromia” AND highest education level of a woman= “none” AND Access to media = “at least once a week” AND Core component of ANC =” four/all” AND # of ANC visit =” more than four times”: Then utilization of Skilled birth attendant during delivery No (22).

The interesting pattern observed for a woman who lives in rural area of Oromia region with no education but having daily access to media, took all required services during ANC visit and frequently visited or more than four times visited health center during ANC (highest ANC visit) then, she will not take skilled birth attendant during delivery.

Rule 7:

IF residence =” Rural” AND Region =” Somali” AND woman occupation =” Don’t work” AND Core components of ANC service =” four/all” AND Husband’s highest level of education = “none”: Then utilization of Skilled birth attendant during delivery No (28).

The most determining variable for a woman who lives in rural Somali region is depends on her Husband’s highest level of education; if her husband’s education level is none then a woman will most likely may not take skilled birth attendant during delivery even if she will take all core components of ANC service during her ANC visit.

Rule 8:

IF residence =” Rural” AND Region =” Benshangul-Gumuz” AND Husband Occupation =” Agricultural employee” AND Birth order =” Fourth/Fifth” AND Access to media=” daily”: Then utilization of Skilled birth attendant during delivery No (18).

Rule 9:

IF residence =” Rural” AND Region =” Benshangul-Gumuz” AND Husband Occupation=” Agricultural employee” AND Birth order =” Sixth or higher” AND Access to media=” daily”: Then Skilled birth attendant during delivery No (5).

For a rural woman who lives in Benshangul-Gumuz region, the utilization of skilled birth attendant will depend on woman husband occupation even if she will have daily access to media. If her husband’s occupation is agricultural employee, then she will not utilize skilled birth attendant.

Rule 10:

IF residence =” Rural” AND Region =” SNNP” AND Core components ANC Service =” none”: Then utilization of Skilled birth attendant during delivery NO (79).

Rule 11:

IF residence=” Rural” AND Region=” SNNP”AND Husband occupation=” Agricultural employee” AND Core components ANC Service =” four/all” AND Household Wealth Index =” Richer” AND Highest education level of a woman= “Primary”: Then utilization of Skilled birth attendant during delivery No (22).

The probability of skilled birth attendant for a woman who lives in rural regions of southern nation nationalists & people (SNNP) depends on two variables: -

1. If a woman will not take any core components of ANC service, then utilization of skilled birth will be “No”.
2. If her husband is agricultural employee, then she will not utilize skilled birth attendant even if she will be very rich in economic and will take all core components of ANC service and completed primary level education.

Rule 12:

IF residence =” Rural” AND Region =” Gambella” AND Household wealth index=” Richest”: Then utilization of Skilled birth attendant during delivery No (5).

Rule 13:

IF residence =” Rural” AND Region =” Harrari” AND Household wealth index=” Richer” AND # of ANC service =” four/all”: Then utilization of Skilled birth attendant during delivery No (13).

Rule 14:

IF residence =” Rural” AND Region =” Diredawa”AND Household Wealth Index = “Richest”: Then utilization of Skilled birth attendant during delivery No (3).

Rule 15:

IF residence =” Rural” AND Region =” Diredawa”AND Household Wealth Index = “poorest” AND Core components of ANC service=” two”: Then utilization of Skilled birth attendant during delivery yes (39).

The only factor which affects the utilization of skilled birth attendant for a rural woman who lives either in Harrari, Gambella of Diredawa region is household wealth index. If her economic status falls in either in richer/richest category, then she will have highest chance not to give birth with the assistance of skilled health personnel.

Rule 16:

IF residence =” Urban” AND region =” Tigray” AND highest level of education of a woman =” none” AND Access to Media =” none” AND Husband highest level of education =none”: Then utilization of Skilled birth attendant during delivery yes (4).

Rule 17:

IF residence =” Urban” AND region =” Tigray” AND highest level of education of a woman =” none” AND Access to Media =” none” AND Husband highest level of education =” primary”: Then Skilled birth attendant during delivery No (25).

For a woman who lives in urban area of Tigray region with no education, the probability of the utilization of skilled birth attendant will depend on Husband highest level of education; if her husband’s education level is none then she will have highest chance to have skilled birth attendant during delivery.

Rule 18:

IF residence =” Urban” AND region =” Amhara” AND highest level of education of a woman =” none” AND Access to Media =”daily”: Then utilization of Skilled birth attendant during delivery No (17).

For a woman who lives in urban area of Amhara region with no education and access to media is daily then she will have highest chance not to have skilled birth attendant during delivery.

Rule 19:

IF residence =” Urban” AND highest level of education of a woman =” Primary” AND Birth Order = “3” AND Husband occupation =” Agricultural employee”: Then skilled birth attendant during delivery No (55).

For an urban woman with primary level of education and her decision for the utilization of skilled birth attendant will depend on her husband’s occupation; if her husband’s occupation is agricultural employee then a woman most likely not utilize skilled birth attendant during delivery.

The health professional will be benefited from the knowledge generated from this study to advocate the utilization of skilled birth attendant based on client's specific factor obtained from the system. All the above generated rules/knowledge are imbedded into a system to explicitly represent the knowledge generated and will enable explanation for each client about the utilization of skilled birth attendant.

5.9. Use of Discovered Knowledge

Following the generated knowledge, user interface(Prototype) was designed in a such a way that the generated and accepted rules by domain experts will be an input for the user interface to further assist the domain experts who are currently working at maternal health centers developed using VB.net programming language

The designed user interface was presented and tested by the domain experts to see the acceptance of the system by the health professional who are currently working at maternal health center. The results of the acceptance test made by the users are presented in the following section.

This user interface will assist the health service providers while providing ANC service for ANC clients in such a way that, first it will predict whether the client will take the later utilization of skilled birth attendant or not then it will provide more information about the client's specific factor either enabled or hinder the utilization of skilled birth attendant that will make the health professional to provide appropriate counseling based on the information he/she provided from the system.

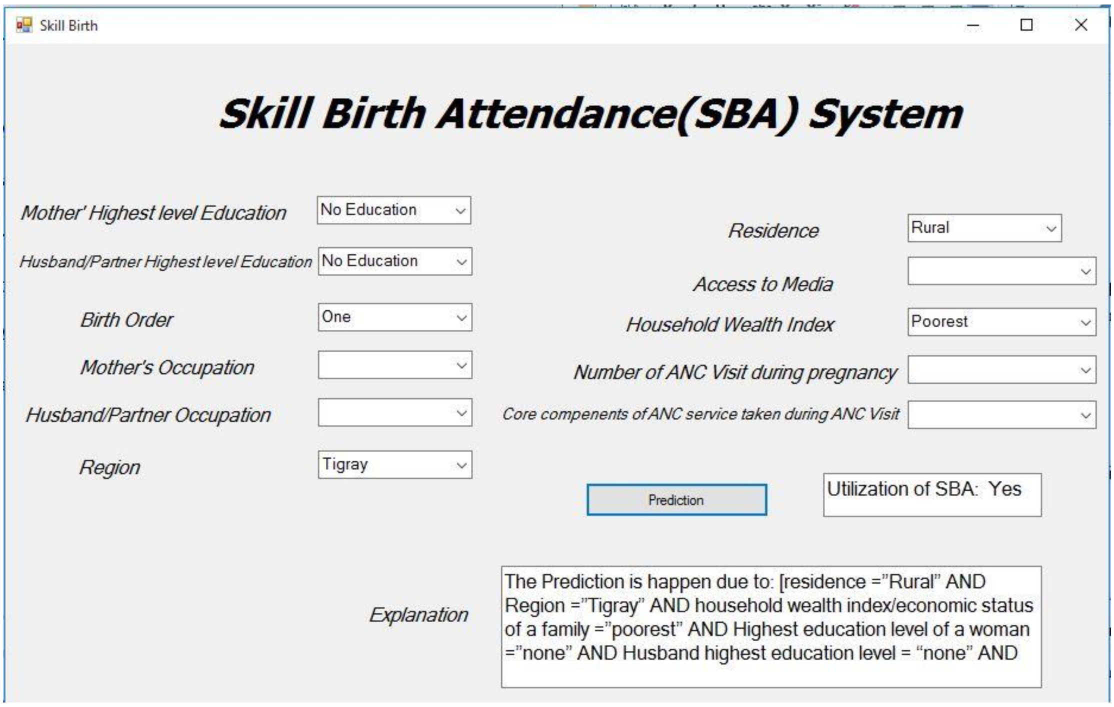

Figure 12. *The skilled birth attendance System for most likely users*

As presented in figure 12 above, the system will assist the users or domain experts first by predict the possibility of a pregnant woman for the utilization of skilled birth attendant during her ANC visit and it will provide more information on the factors which most likely enable or hinder the utilization of the service specific to each clients based on their predisposing, enabling and need factors. Some of the rules imbedded in the system will be populated for each client based on their predisposing, enabling and need factors and will assist the user's /domain experts that the likelihood of a woman to use the service with driving factors. After filling all the required information the user expected to click on predict button, then the system will pop up the message in the screen for the user of the system as depicted in the figure 12 that will tell him/her on the probability of the mother to use skilled birth attendant and the critical variables hindering a woman not to utilize skilled birth attendant that will help them to work more to provide counseling for the woman based on her specific variable from their side during her ANC visit at maternal health center.

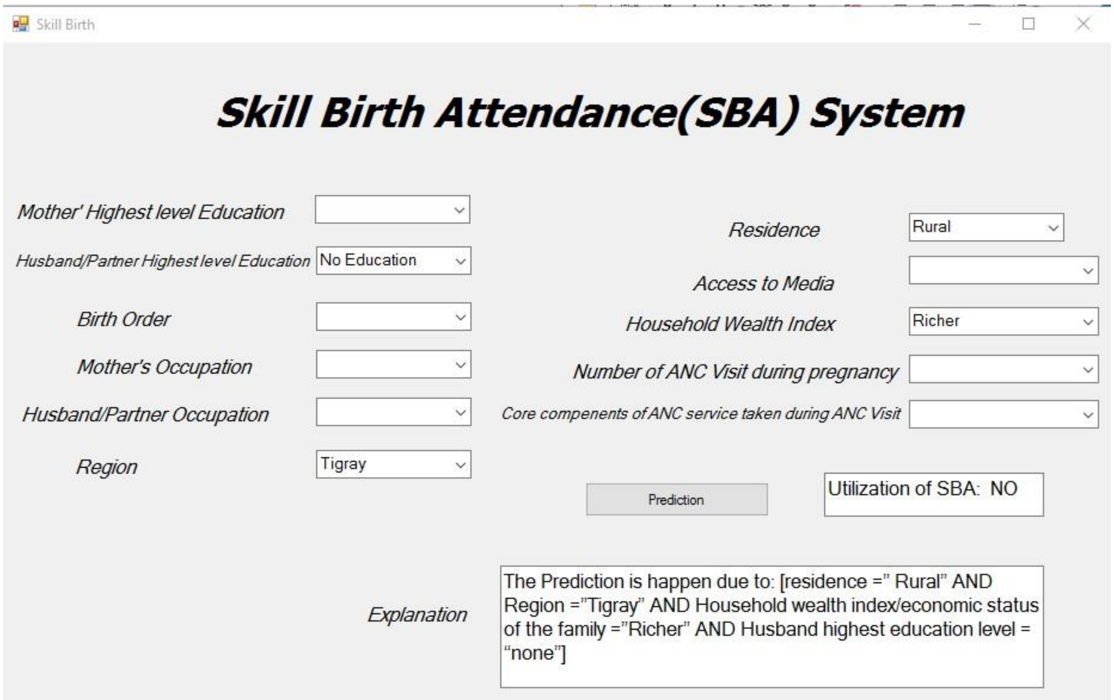

Figure 13. *The skilled birth attendance System for most unlikely users.*

The figure 13 shows that the message for the user of the system that a woman currently under ANC follow up examination will not most likely use skilled birth attendant then the health professional should have to advise her more based on a woman clinical findings on examination and based on her demographic status.

A total of 12 domain experts were tested and evaluated the system on its usability, understandability, relevance and further improvement.

The evaluators filled in the questionnaire as Excellent, very good, Good, Satisfactory and poor for questions 1 up to 8 and at the end they putted their overall feedbacks and recommendations. The researcher assigned values for the answers to make suitable to estimate the performance of the system as an Excellent = 5, V/good = 4, Good = 3, Satisfactory = 2 and poor = 0.

The evaluator's response is summarized in table 19 below.

	Evaluation Criteria	Poor	Satisfactory	Good	V/Good	Excellent
1	The system is usable for our day to day activity.	-	-	1	**8**	**3**
2	The system is easy to use.	-	-		**7**	**5**
3	The prototype provides fast response.	-	-	**1**	**7**	**4**
4	The prototype result is accurate and reliable.	-	-		**10**	**2**
5	The system gives the appropriate prediction for the users.	-	-	2	**5**	**5**
6	Health institution benefits from such prediction Model.				10	2
7	Working with this system is relevant.	-	-		**8**	**4**
8	Using the system would improve the current uptake of skilled birth attendant	-	-	1	7	4

Table 18. The result of user acceptance evaluation

As depicted in table 19, all respondents agreed with simplicity and usability of the system. In the same way, all users agreed with the prediction capability of the system. It can be concluded from the quantitative findings from the users that the system is more reliable and can be used and easily integrated in to the existing electronic management system for any maternal health centers, if they may have such automated system.

In the qualitative discussion with those experts, they pointed out that for further improvements on the system in presented as the following:

- The system has to recommend for the users critically things to be observed from the findings.
- The system has to separate probably the findings in to the three categories; what is expected from the health professional to convince her to use skilled birth attendant, what is expected from a woman to use skilled birth attendant and lastly what is expected from other concerned bodies then this will help the clinical works to more focus on their parts.
- The system should be used before a woman physical and laboratory examination and the latter counseling session to give more time on counseling session on what has to be done to ensure the usage of skilled birth attendant.
- The same user acceptance test has to be continued for regions because their regional differences from the findings for each region.
- Health center officials who are working on maternal health should be shared the finding from this study and their feedbacks also very important to further strengthen the system.
- The system has to be designed for non-automated maternal health centers especially for rural areas.

CHAPTER SIX

6. Conclusion and Recommendations

6.1. Conclusion

The objective of this study is to develop a predictive model for Antenatal care clients seeking pattern to predict skilled birth attendant using data mining techniques among women aged 15-49 in Ethiopia from the most recent EDHS 2016 datasets. A total of 16 experiments were conducted using J48, PART Rule Induction and Naïve Bayes classification algorithms with all and selected attributes. The most appropriate attributes were selected for J48 decision tree algorithm using Weka's built in CfsSubsetEval function.

The researcher has used Hybrid methodology of Knowledge discovery process to build the predictive model of Antenatal care client's pattern to use skilled birth attendant using data mining techniques. Andersen Health Behavioral Model is used to better understand the problem under study because it assumes that predisposing, enabling and need factors determine women utilization of health facilities for delivery by skilled birth attendant.

The data source used for this study was 2016 EDHS dataset. The EDHS 2016 contains all the variables required by the business under study. Accordingly, the variables selected for this study were for each components of Andersen Health Behavioral Model which were determine the utilization of skilled birth attendant of Antenatal care clients.

The longlisted selected attributes have passed through data pre-processing for the purpose of data cleaning, data transformation and dimension reduction. The missing values replaced by field mean and modal values for categorical variables. Data discretization was also used transform numeric values in to nominal.

In this study, the widely used classification algorithms such as J48 decision tree, PART rule induction and Naïve Bayes are used to build the model for predicting Skilled birth attendant among Antenatal care clients using EDHS 2016 dataset.

The performance of the models was evaluated using the standard metrics of accuracy, precision, recall and F-measure, and ROC. Pruned J48 decision tree with 66 percent split with selected attribute is selected with a classification accuracy of 98.74%.

The most important rules generated from the datasets which surprised both domain and non-domain experts to predict the utilization of SBA among ANC users and hence the knowledge generated from WEKA to predict the utilization of Skilled birth attendant among ANC users highly depend on the place of residence of the woman, place of region, Access to media, house hold wealth index, husband highest level of education, and husband occupation. The experts suggested the applicability of the rules for both clinicians and as well as for policy makers.

From this study it is observed that data mining techniques can effectively be used in the health sectors specially to predict the utilization of skilled birth attendant to avert the high maternal mortality rate in Ethiopia. At the end user interface is designed to further assist the decision makers both at clinic levels and at higher level (policy makers).

The surprising results generated from the EDHS 2016 will enable decision makers and researchers to increase the utilization of skilled birth attendant by ANC clients.

In this study secondary data obtained from EDHS 2016 is used to extract the interesting rules but we have to wait for five years to get the EDHS data, therefore, all concerned body should have to exert their efforts to captures all the necessary data at clinic level in the electronic format to further to dig the knowledge within the data to make evidence based decision on timely manner so as to decrease the high maternal mortality rate by closely working to ensure all ANC follow-up clients should have to end with skilled birth attendant.

6.2. Recommendations

Based on the findings from this study the researcher suggests both for public health policies makers and maternal health clinic professionals aimed to reduce maternal mortality rate by increasing the utilization of Skilled birth attendant in Ethiopia and very interesting model generated from the study so as to increase the uptake of skilled birth attendant by using the generated knowledge user interface at each maternal health centers and at higher decision makers to know more specific factors for each regions and residences and should have to look in the following strategies made by the researcher:

- During ANC visit a woman should come to the center with her husband to council both of them about the benefits of utilization of skilled birth attendant and birth preparedness.
- The benefits of giving birth at health institution should be continuously promoted through all the available public media.
- Increase the access of education for all women in the country especial for those who are living at urban areas.
- Higher official/Ministry of education should strongly work on promotion of skilled birth attendant through all available medias for those who are not educated will benefit more to make decision to utilize skilled birth attendant.
- There should be a system to capture all the required details of mothers during ANC follow up at health facilities to further analysis on the data to strengthen the findings of this study for the future study and other variables such as distance from the health facility, transportation availability and information about birth preparedness has to be included.
- The developed knowledge based system based on the findings from this study should be integrated at the health facilities as an input to provide counseling for pregnant mothers to utilize the later skilled birth attendant.

- However, EDHS 2016 is secondary data and the same analyzes has to be researched on the primary data which is captured by the health institution to further strengthen the study at country level and so as to avert the high maternal mortality rate of the country.

References

1. Say L, Inoue M. Maternal mortality in 2005: estimates developed by WHO, UNICEF and UNFPA. Geneva, 2007.
2. Competent attendance in maternal and newborn health: the definition of the competent health care provider in maternal and newborn health. A joint statement by WHO, UNFPA, UNICEF, ICM, ICN, FIGO and IPA , 2017
3. Global, regional, and national levels and trends in maternal mortality between 1990 and 2015, with scenario-based projections to 2030: a systematic analysis by the UN Maternal Mortality Estimation Inter-Agency Group.
4. Wanjira, et al, Practices and Associated factors among Mothers Seeking Child welfare Services in selected Health Facilities in Nyandarua South District, Kenya. BMC, Public Health, 2011.
5. The Millennium Development Goals Report 2015, United nations New York, 2015
6. Waage J, Banerji R, Campbell O, Chirwa E, Collender G, Dieltiens V, et al. The millennium development goals: a cross-sectoral analysis and principles for goal setting after 2015.
7. WHO Maternal mortality in 1990–2013. http://www.who.int/gho/countries/eth/country_proiles/en/. Accessed June, 2014.]
8. Central Statistical Agency and ICF. Ethiopia Demographic and Health Survey 2016: Key Indicators Report. Addis Ababa, Ethiopia, and Rockville, Maryland, USA. CSA and ICF, 2016.
9. Ending Preventable Maternal Mortality: USAID Maternal Health Vision for Action Evidence for Strategic Approaches ,January 2015
10. The Millennium Development Goals Report 2015, United nation New York ,2015.

11. World health statistics 2016: monitoring health for the SDGs, sustainable development goals. WHO, 2016
12. Adegoke et al, Skilled Birth Attendance- Lessons learnt. International Journal of Obstetrics and Gynecology, 2009.
13. http://www.who.int/gho/maternal_health/skilled_care/skilled_birth_attendance_text/en/.
14. WHO defines Health Services Coverage Statistics ,2013.
15. Carroli, G., Rooney C., & Villar, J. (2001). How effective is antenatal care in preventing maternal mortality and serious morbidity?,2001.
16. Tangam de Bernis and et al, Skilled attendants for pregnancy, childbirth and postnatal care,2003.
17. UNFPA and Guttmacher Institute, the Costs and Benefits of Investing in Family Planning and Maternal and Newborn Health. New York, 2009.
18. Bloom and et al, Does antenatal care make a difference to safe delivery? A study in urban Uttar Pradesh, India. Health Policy and Planning, 199.
19. Gabrysch, S., & Campbell,Still too far to walk: literature review of the determinants of delivery service use,2009.
20. Eyeresualem Dagne , Role of socio-demographic factors on utilization of maternal health care services in Ethiopia,2010.
21. Han J and Kamber M. Data Mining: Concepts and Techniques. New York. USA: Morgan Kaufmann; 2001
22. Babitsch, B, Gohi, D & Lengerke, TV. Re-visiting Andersen's behavioral model of health services use: a systematic review of studies from 1998-2011,2012.
23. Andersen, RM. 1995. Revisiting the behavioral model and access to medical care: Does it matter? Journal of Health and Social Behavior 36(1) ,1995.
24. T. Larose, Data mining methods and models. New Jersey: John wily and Sons Inc., 2006.
25. Kincade, K. Data Mining: digging for healthcare gold, Insurance & Technology, 23(2), IM2-I, 1998..

26. Milley, A. Healthcare and Data Mining. Health Management Technology, 21(8), 44-4,2000.
27. SIAM. Data mining for healthcare management, 2011.
28. Kreuze, D. Debugging Hospitals:Technology Review, 104(2), 32,2001.
29. Silver, et al : how to apply data mining techniques in a healthcare datawarehouse. Journal of Healthcare Information Management, 15(2), 155-164, 2001.
30. Benko, A. & Wilson, B. Online decision support gives plans an edge: Managed Healthcare Executive, 13(5), 20, 2003.
31. Trybula, W.J. Data mining and knowledge discovery: Annual Review of Information Science and Technology, 32, 197-229, 1997.
33. Trends in maternal mortality: 1990 to 2015: estimates by WHO, UNICEF, UNFPA, World Bank Group and the United Nations Population Division, World health organization, 2015.
34. The Millennium Development Goals Report2015, United nation New York, 2015.
35. World health statistics 2016: monitoring health for the SDGs, sustainable development goals WHO, 2016
36. Ending Preventable Maternal Mortality: USAID Maternal Health Vision for Action Evidence for Strategic Approaches, January 2015.
37. Say L et.al, Global causes of maternal death: a WHO systematic analysis ,2014
38. Campbell OM et.al, Graham WJ. Lancet Maternal Survival Series steering group. Strategies for reducing maternal mortality: Getting on with what works,2006 .
39. Every women every child, Packages of Interventions for Family Planning, Safe Abortion care, Maternal, Newborn and Child Health, Global strategy for women's and children's health,WHO,2010
40. Central Statistical Agency and ICF. Ethiopia Demographic and Health Survey 2016: Key Indicators Report. Addis Ababa, Ethiopia, and Rockville, Maryland, USA. CSA and ICF, 2016.

41. Wanjira, et.al, Delivery Practices and Associated Factors among Mothers Seeking Child Welfare Services in Selected Health Facilities in Nyandarua South District, Kenya,2011.
43. Montgomery AL ,et.al ,The Million Death Study Collaborators (2014) Maternal Mortality in India: Causes and Healthcare Service Use Based on a Nationally Representative Survey,2014.
44. Oyerinde K, Can Antenatal Care Result in Significant Maternal Mortality Reduction in Developing Countries?, 2013.
45. S. Abouchadi et al. Preventable maternal mortality in Morocco, Tropical Medicine and International Health, 2013.
46. Kikuchi K, et al. ,Effective Linkages of Continuum of Care for Improving Neonatal, Perinatal, and Maternal Mortality: A Systematic Review and Meta-Analysis,2015.
47. Tadese F, Ali A,Determinants of Use of Skilled Birth Attendance among Mothers Who Gave Birth in the Past 12 months in Raya Alamata District, North East Ethiopia,2014.
48. Gitimu et al. ,Determinants of use of skilled birth attendant at delivery in Makueni, Kenya,2015.
49. Berhan Y, Berhan A. Antenatal care as a means of increasing birth in the health facility and reducing maternal mortality: A SYSTEMATIC REVIEW, Ethiop J Health Sci. , 2014
50. Kebede A., Hassen K. Nigussie A. Factors associated with institutional delivery service utilization in Ethiopia; International Journal of Women's Health ,2016
51. Mulu S., Sue L. And Giedraitis V. Determinants of maternal health service utilization in Ethiopia: analysis of the 2011 Ethiopian Demographic and Health Survey: BMC Pregnancy and Childbirth, 2014.
52. Magoma M., Requejo J., Campbel O., Cousens S. and Filippi V. High ANC coverage and low skilled attendance in a rural Tanzanian district: a case for implementing a birth plan intervention; 2010
53. Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? , 1995.

54. Dawit A., Predicting maternal health care seeking pattern using data mining techniques in Ethiopia ,2013
55. Tekelehaimanot,Predicting antenatal care follow up of women during pregnancy in ethiopia using data mining technology , 2015.
56. Taylor & Francis, DATA PREPARATION FOR DATA MINING, 2003.
57. Han and Kamber. Data Mining: Concepts and Techniques, 2006.
58. International Journal of Computer Applications, (0975 – 8887) Volume 138 – No.1, March 2016.
59. Usama Fayyad , The KDD Process for Extracting Useful Knowledge from Volumes of Data, 1996.
60. Cios Krzyszt,et al, A. Data Mining: A knowledge Discovery approach. New York: Springer-Verlag Science Business Media LLC; 2007.
61. M. L. WONG, LEUNG, K. S. , Data Mining using Grammar Based Genetic Programming and application: Springer, 2000.
62. Pete Chapman, et al, CRISP-DM 1.0 Step-by-step data mining guide.
63. Obenshain, Mary K. Application of data mining techniques to healthcare data: Infection control and hospital epidemiology. The official journal of the Society of Hospital Epidemiologists of America 25 (8): 690-5; 2004.
64. Two Crows Corporation. Introduction to Data Mining and Knowledge Discovery., USA,2005.
65. Weka Machine Learning Project, http://www.cs.waikato.ac.nz/~ml/index.html
66. Mohamed Kantardzic J.B. Data Mining-Concepts, Models, Methods, and Algorithms. USA: John Wiley & Sons Publication Inc, 2003
67. Berry Michael W, Browne K M., Lecture Notes in Data Mining. USA: World Scientific Publishing Co. Pte. Ltd Inc. Rosewood Drive, Danvers, MA, 2006.
68. Ruben D and Canlas Jr., Data Mining in Healthcare: Current Application and Issues. Thesis, Australia: Carnegie Mellon University, 2009
69. K. Ming Leung, Naive Bayesian Classifier, POLYTECHNIC UNIVERSITY Department of Computer Science / Finance and Risk Engineering, 2007.
70. DL Olson, D Delen - Advanced data mining techniques, Springer,2008 .
71. Moore, Wilbert E. Social Structure and Behavior, The Handbook of Social Psychology, Volume 4, 2nd edition. Reading, Massachusetts: Addison-Wesley 1969.
72. Babitsch, B, Gohi, D & Lengerke, Re-visiting Andersen's behavioral model of health services use: a systematic review of studies from 1998-2011 .

73. Andersen, RM & Davidson PL. Improving access to care in America: individual and contextual indicators ,2001.
74. Ajzen, I. The theory of planned behavior. Organizational Behavior and Human Decision Processes, 50(2), 179-211 1991.
75. Hardeman, W., Griffin, S., Johnston, M., Kinmonth, A.L. and Wareham, N.J.: A systematic review of psychological models and behavior change methods 2000.
76. Harrison JA , Mullen PD, and Green LW. A meta-analysis of studies of the Health Belief Model with adults. Health Educ Res. 1992.
77. Daniel T. Larose. Discovering Knowledge in Data: An Introduction to Data Mining. John Wiley & Sons, Hoboken, New Jersey; 2005
78. Data Mining: Concepts, Models, Methods, and Algorithms, 2nd Edition Data Mining: Concepts, Models, Methods, and Algorithms, 2nd Edition by Mehmed Kantardzic September 2011
79. I. H. F. WITTEN, E., "Data Mining: Practical machine learning tools and techniques," Morgan Kaufmann, 2005.
80. Max Bramer," Pre-pruning Classification Trees to Reduce Overfitting in Noisy Domains", Faculty of Technology, University of Portsmouth, UK,2009
81. Ross Quinlan . C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo, CA, 1993.
82. Y. H. LEI, L., "Feature selection for high-dimensional data: A fast correlation-based filter solution," in Proceedings of the Twentieth International Conference on Machine Learning, Washington DC, USA, 2003.
83. I. W. a. E. Frank, Data mining: practical machine learning tools and techniques with java implementations, 2nd ed. San Francisco: Morgan kaufmann publishers, 2000.
84. T. C. Corporation, Introduction to Data Mining and Knowledge Discovery. USA: Potomac, 2005.
85. P. a. N. Clark, T. , "The CN2 induction algorithm," Machine Learning, vol. 3(4), pp. 261- 283, 1989.
86. A. B. a. S. S. THERLING K. (2010, 20 September). An Overview of Data Mining Techniques. Available: http://www.thearling.com/text/dmtechniques/dmtechniques.htm
87. H. M. MOSHKOVICH, MECHITOV,ALEXANDER I., AND OLSON,DAVID L., "Rule induction in data mining," Ltd. Technical University of Crete, vol. 22, pp. p.303-311, 2002.
88. MDMW. (2008). My Data Mining Weblog Rule learner (or Rule Induction). Available: http://mydatamine.com/?tag=induction-algorithms

89. R. C. Holte, "Very simple classification rules perform well on most commonly used datasets," Machine learning, vol. 11, pp. 63-90, 1993.
90. T. M. MITCHELL, "Machine Learning," McGraw-Hill Science, 1997.
91. B. PANG, LEE, L. & VAITHYANATHAN, S., "Sentiment classification using machine learning techniques," presented at the Proceedings of the ACL-02 conference on Empirical methods in natural language processing, 2002.
92. C.R.KOTAHRI, Research methodology methods and techniques ,MAY 1990
93. Abdullah H. et al, A Comparison Study between Data Mining Tools over some Classification Methods
94. World Health Organization (WHO). *Coverage of Maternity Care: A Listing of Available Information.* 4th edition. 1997.
95. Bloom SS, Lippeveld T, Wypij D. 1999. Does antenatal care make a difference to safe delivery? A study in urban Uttar Pradesh, India. *Health Policy and Planning* 14(1):38–48.
96. *Akhter, S & Wohab, A (2008). Millennium Development Goal of Improving Maternal health in Bangladesh: A review of Concepts. BRAC University Journal 5(1): pp 99-112*
97. *Rogan, S. B. & Olveῇa, M. R. (2004). Factors Affecting Maternal health utilization in the Phillipines*
98. *Akalin MZ, Maine D. 1995. Strategy of risk approach in antenatal care: evaluation of the referral compliance. Social Science and Medicine 41(4):595–596.*
99. *The Prevention of Maternal Mortality Network. 1992. Barriers to treatment of obstetric emergencies in rural communities of West Africa. Studies in Family Planning 23(5):279– 291.*
100. *Tanagra –a Free Data Mining Software for Teaching and Research, Available at: http://eric.univ-lyon2.fr/~ricco/tanagra/en/tanagra.html, (Accessed 20 April 2018).*
101. *KNIME (Konstanz Information Miner), Available at: http://www.knime.org/, (Accessed 20 April 2018).*
102. *Orange– Data Mining Fruitful and Fun, Available at: https://www.researchgate.net/publication/220698829_Orange_From_Experimental_Machine_Learning_to_Interactive_Data_Mining_White_Paper/, (Accessed 20 April 2018).*
103. *Floyd, C. A Systematic Look at Prototyping, in, 1984.*
104. *Tate, G.) Prototyping: Helping to Build the Right Software, Information and Software Technology, 42, 4, 237-244, 1990.*

Annex 1: The description of the selected attributes from 2016 EDHS Data Sets.

No.	Name of Attribute	Description		Values	Count of Missing Values
1	Mother's Date birth(CMC)	Birth date of Mother (in CMC format)	Numerical	Continues	0.00%
2	Birth date of Last-born child(CMC)	Birth date of Last-born child(in CMC format)	Numerical	Continues	0.00%
3	ANC Service Provided by Doctor	Prenatal service: Doctor	Categorical	[0,1]	0.00%
4	ANC Service Provided by Nurse	Prenatal service: Nurse	Categorical	[0,1]	0.00%
5	ANC Service Provided by Midwife	Prenatal service: Midwife	Categorical	[0,1]	0.00%
6	ANC Service Provided by Health officer	Prenatal service: Health officer	Categorical	[0,1]	0.00%

7	ANC Service Provided by Health Extension Worker	Prenatal service: HEW	Categorical	[0,1]	0.00%
8	Delivery Assisted by Doctor	Skilled birth attendant: Doctor	Categorical	[0,1]	0.00%
9	Delivery Assisted by Nurse	Skilled birth attendant: Nurse	Categorical	[0,1]	0.00%
10	Delivery Assisted by Midwife	Skilled birth attendant: Midwife	Categorical	[0,1]	0.00%
11	Blood pressure Measured during ANC Visit	Component of ANC Service: BP Measured	Categorical	[0,1]	0.00%
12	Urine Sample taken during ANC Visit	Component of ANC Service: Urine Sample	Categorical	[0,1]	0.00%
13	Blood Sample taken during ANC Visit	Component of ANC Service: Blood Sample	Categorical	[0,1]	0.00%
14	Nutrition Counseling Provided during ANC Visit	Component of ANC Service: Nutritional Counseling	Categorical	[0,1]	0.00%

15	Region	Administrative regions of Ethiopia	Categorical	1.Tigray, 2. Afar, 3.Amhara, 4.Oromiya, 5.Somali, 6.Benishangul Gumz, 7.Southern National Nationality People SNNP), 12.Gambela, 13.Harari, 14.Addis Abeba, 15.Dire Dawa	0.00%

16	Residence	Type of place of residence	Categorical	1.Urban 2.Rural	0.00%
17	Women's Education Level	Highest level of education attained by women	Categorical	0.No Educuation 1.Primary Education 2.Secondary Education 3.Higher Education	0.00%
18	Household Wealth Index	Living standard of the household	Categorical	1.Poorest, 2.Poorer, 3.Middle, 4.Richer, 5.Richest	0.00%

19	Marital Status	Mother's marital status during birth	Categorical	0.Never married, 1. Married, 2.Living together, 3. Widowed, 4.Divorced, 5. Not living together	0.00%
20	Husband/Partner Education level	Highest level of education attained by Husband/Partner	Categorical	0.No Educuation 1.Primary Education 2.Secondary Education 3.Higher Education	356(7.6%)

21	Frequency of reading news paper	How often reading news paper	Categorical	0.Not at all, 1.Less than once a week 2.At least once in a week 3.Almost every day	0.00%
22	Frequency of listening radio	How often listening radio	Categorical	0.Not at all, 1.Less than once a week 2.At least once in a week 3.Almost every day	0.00%

23	Frequency of watching Television	How often watching television	Categorical	0.Not at all, 1.Less than once a week 2.At least once in a week 3.Almost every day	0.00%
24	Mother's occupation	Occupation of Mothers	Categorical	0 .Don't work 1. Agricultural Employed 2. Professional/Technical/managerial 3.Non-	0.00%

				Managerial 4.Missing	
25	Husband/Partner Occupation	Occupation of Husband/Partner	Categorical	0 .Don't work 1. Agricultural Employed 2. Professional/Technical /managerial 3.Non-Managerial 4.Missing	395(8.4%)
26	Birth Order	Birth order of the last birth	Categorical	1,2,3,….	0.00%
27	Timing first Antenatal check(months)	First ANC visit by Mothers during pregnancy in month	Categorical	1,2,3,…9 or 98 do not know	27(1%)
28	Number of ANC Visit during	Total number of ANC visit by mother	Categorical	0.No	19(0.5%)

	pregnancy	during pregnancy		antenatal visit 98.Donot know	
29	Religion	Religion of the mother	Categorical	1.Orthodox 2.Catholic 3.Protestant 4.Muslim 5.Traditional 6.Other	0.00%

Annex 2: Outputs of the Selected Classifier (J48 Decision tress) with generated rules

=== Run information ===

Scheme:weka.classifiers.trees.J48 -C 0.5 -M 2

Relation: Skilled birth attendant-weka.filters.supervised.instance.SMOTE-C0-K5-P11.25-S1-weka.filters.supervised.instance.Resample-B1.0-S1-Z100.0-

Instances: 4933

Attributes: 12

Region

Residence

M.Education

Wealth

H.Education

Access to Media

M.Occupation

H.Occupation

Birth order

# of ANC Visit

Core ANC Services took

Assitance during Delivery

Test mode:10-fold cross-validation

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

J48 pruned tree

Residence = 1

| M.Education = 0

| | Access to Media = 0

| | | Region = 1

| | | | H.Education = 0: Yes (4.0)

| | | | H.Education = 1: No (25.0)

| | | | H.Education = 2: No (0.0)

| | | | H.Education = 3: No (0.0)

| | | Region = 2: No (31.0)

| | | Region = 3: Yes (7.0)

| | | Region = 4: No (0.0)

| | | Region = 5: No (45.0)

| | | Region = 6: Yes (14.0)

| | | Region = 7: Yes (5.0)

| | | Region = 8: No (0.0)

| | | Region = 9: Yes (27.0)

| | | Region = 10: Yes (22.0)

| | | Region = 11: No (21.0)

| | Access to Media = 1

| | | M.Occupation = 0: No (9.0)

| | | M.Occupation = 1: No (0.0)

| | | M.Occupation = 2: No (0.0)

| | | M.Occupation = 3: Yes (4.0)

| | Access to Media = 2

| | | Region = 1: Yes (24.0)

| | | Region = 2: Yes (0.0)

| | | Region = 3: No (17.0)

| | | Region = 4: Yes (0.0)

| | | Region = 5: Yes (29.0)

| | | Region = 6: Yes (0.0)

| | | Region = 7: Yes (0.0)

| | | Region = 8: Yes (0.0)

| | | Region = 9: Yes (2.0)

| | | Region = 10: Yes (24.0)

| | | Region = 11: Yes (24.0)

| M.Education = 1

| | Birth order = 1

| | | Core ANC Services took = 0: Yes (0.0)

| | | Core ANC Services took = 1: Yes (0.0)

| | | Core ANC Services took = 2

| | | | Access to Media = 0: Yes (24.0)

| | | | Access to Media = 1: Yes (0.0)

| | | | Access to Media = 2: No (4.0)

| | | Core ANC Services took = 3

| | | | H.Occupation = 0: Yes (0.0)

| | | | H.Occupation = 1: Yes (10.0)

| | | | H.Occupation = 2: Yes (9.0)

| | | | H.Occupation = 3: No (4.0/1.0)

| | | Core ANC Services took = 4: Yes (92.0)

| | Birth order = 2

| | | Region = 1: Yes (4.0)

| | | Region = 2: Yes (0.0)

| | | Region = 3: Yes (0.0)

| | | Region = 4: Yes (25.0)

| | | Region = 5: No (1.0)

| | | Region = 6: Yes (0.0)

| | | Region = 7: Yes (0.0)

| | | Region = 8

| | | | Access to Media = 0: Yes (22.0)

| | | | Access to Media = 1: No (9.0)

| | | | Access to Media = 2: Yes (0.0)

| | | Region = 9: Yes (6.0)

| | | Region = 10: Yes (67.0/12.0)

| | | Region = 11: Yes (34.0)

| | Birth order = 3

| | | H.Occupation = 0: No (0.0)

| | | H.Occupation = 1: No (55.0)

| | | H.Occupation = 2: No (0.0)

| | | H.Occupation = 3

| | | | Region = 1: Yes (0.0)

| | | | Region = 2: Yes (0.0)

| | | | Region = 3: Yes (0.0)

| | | | Region = 4: No (5.0)

| | | | Region = 5: Yes (0.0)

| | | | Region = 6: Yes (0.0)

| | | | Region = 7: Yes (0.0)

| | | | Region = 8: Yes (23.0)

| | | | Region = 9: Yes (0.0)

| | | | Region = 10: Yes (0.0)

| | | | Region = 11: Yes (0.0)

| | Birth order = 4: Yes (7.0)

| M.Education = 2

| | Wealth = 1: Yes (0.0)

| | Wealth = 2: Yes (0.0)

| | Wealth = 3: No (7.0)

| | Wealth = 4: Yes (0.0)

| | Wealth = 5: Yes (400.0)

| M.Education = 3: Yes (286.0)

Residence = 2

| Region = 1

| | Wealth = 1

| | | M.Education = 0

| | | | H.Education = 0

| | | | | Birth order = 1: Yes (14.0)

| | | | | Birth order = 2: Yes (3.0)

| | | | | Birth order = 3

| | | | | | M.Occupation = 0: Yes (10.0)

| | | | | | M.Occupation = 1: No (19.0)

| | | | | | M.Occupation = 2: No (0.0)

| | | | | | M.Occupation = 3: No (0.0)

| | | | | Birth order = 4: No (28.0)

| | | | H.Education = 1: No (43.0)

| | | | H.Education = 2: No (0.0)

| | | | H.Education = 3: No (0.0)

| | | M.Education = 1

| | | | H.Education = 0: No (3.0)

| | | | H.Education = 1: Yes (21.0)

| | | | H.Education = 2: Yes (0.0)

| | | | H.Education = 3: Yes (0.0)

| | | M.Education = 2: No (0.0)

| | | M.Education = 3: No (0.0)

| | Wealth = 2

| | | # of ANC Visit = 1: Yes (0.0)

| | | # of ANC Visit = 2

| | | | Birth order = 1: No (4.0)

| | | | Birth order = 2: Yes (23.0)

| | | | Birth order = 3: Yes (23.0)

| | | | Birth order = 4: Yes (1.0)

| | | # of ANC Visit = 3: Yes (125.0)

| | Wealth = 3

| | | M.Occupation = 0: No (13.0)

| | | M.Occupation = 1: No (23.0)

| | | M.Occupation = 2: No (0.0)

| | | M.Occupation = 3

| | | | Birth order = 1: No (1.0)

| | | | Birth order = 2: Yes (17.0)

| | | | Birth order = 3: No (2.0)

| | | | Birth order = 4: Yes (19.0)

| | Wealth = 4

| | | H.Education = 0: No (17.0)

| | | H.Education = 1: Yes (18.0)

| | | H.Education = 2: Yes (27.0)

| | | H.Education = 3: Yes (0.0)

| | Wealth = 5: Yes (52.0)

```
|  Region = 2
|  |  M.Occupation = 0: No (122.0)
|  |  M.Occupation = 1: No (1.0)
|  |  M.Occupation = 2: No (0.0)
|  |  M.Occupation = 3
|  |  |  Access to Media  = 0: Yes (3.0)
|  |  |  Access to Media  = 1: No (6.0)
|  |  |  Access to Media  = 2: No (0.0)
|  Region = 3
|  |  H.Education = 0
|  |  |  M.Education = 0: No (269.0/2.0)
|  |  |  M.Education = 1
|  |  |  |  Birth order = 1: Yes (15.0)
|  |  |  |  Birth order = 2: Yes (0.0)
|  |  |  |  Birth order = 3: No (9.0)
|  |  |  |  Birth order = 4: Yes (0.0)
|  |  |  M.Education = 2: No (0.0)
|  |  |  M.Education = 3: No (0.0)
|  |  H.Education = 1
|  |  |  Access to Media  = 0
|  |  |  |  Wealth = 1: Yes (12.0)
|  |  |  |  Wealth = 2: No (7.0)
|  |  |  |  Wealth = 3: No (20.0)
|  |  |  |  Wealth = 4: No (47.0/1.0)
|  |  |  |  Wealth = 5: No (32.0)
|  |  |  Access to Media  = 1
|  |  |  |  # of ANC Visit = 1: Yes (0.0)
```

| | | | # of ANC Visit = 2: No (3.0)

| | | | # of ANC Visit = 3

| | | | | Wealth = 1: Yes (0.0)

| | | | | Wealth = 2: No (4.0)

| | | | | Wealth = 3: Yes (29.0)

| | | | | Wealth = 4: Yes (22.0)

| | | | | Wealth = 5: Yes (0.0)

| | | Access to Media = 2: No (0.0)

| | H.Education = 2

| | | M.Education = 0: Yes (20.0)

| | | M.Education = 1: No (2.0)

| | | M.Education = 2: Yes (0.0)

| | | M.Education = 3: Yes (0.0)

| | H.Education = 3: No (0.0)

| Region = 4

| | M.Education = 0

| | | Access to Media = 0

| | | | Core ANC Services took = 0: No (55.0)

| | | | Core ANC Services took = 1: No (89.0)

| | | | Core ANC Services took = 2: No (54.0)

| | | | Core ANC Services took = 3: No (34.0)

| | | | Core ANC Services took = 4

| | | | | Birth order = 1: No (0.0)

| | | | | Birth order = 2: No (26.0)

| | | | | Birth order = 3

| | | | | | Wealth = 1: No (0.0)

| | | | | | Wealth = 2: No (10.0)

| | | | | | Wealth = 3: Yes (4.0)

| | | | | | Wealth = 4: No (0.0)

| | | | | | Wealth = 5: No (0.0)

| | | | | Birth order = 4: Yes (23.0)

| | | Access to Media = 1

| | | | Core ANC Services took = 0: Yes (35.0)

| | | | Core ANC Services took = 1: No (25.0)

| | | | Core ANC Services took = 2: No (0.0)

| | | | Core ANC Services took = 3

| | | | | Wealth = 1: Yes (10.0)

| | | | | Wealth = 2: Yes (0.0)

| | | | | Wealth = 3: No (16.0)

| | | | | Wealth = 4: Yes (11.0)

| | | | | Wealth = 5: Yes (0.0)

| | | | Core ANC Services took = 4: No (16.0)

| | | Access to Media = 2

| | | | # of ANC Visit = 1: No (0.0)

| | | | # of ANC Visit = 2: No (57.0/2.0)

| | | | # of ANC Visit = 3

| | | | | Core ANC Services took = 0: No (0.0)

| | | | | Core ANC Services took = 1: No (0.0)

| | | | | Core ANC Services took = 2: No (0.0)

| | | | | Core ANC Services took = 3: No (22.0)

| | | | | Core ANC Services took = 4: Yes (11.0)

| | M.Education = 1

| | | Birth order = 1

| | | | Core ANC Services took = 0: Yes (0.0)

| | | | Core ANC Services took = 1: No (1.0)

| | | | Core ANC Services took = 2: No (20.0)

| | | | Core ANC Services took = 3: Yes (38.0)

| | | | Core ANC Services took = 4: Yes (12.0)

| | | Birth order = 2: No (37.0)

| | | Birth order = 3

| | | | # of ANC Visit = 1: Yes (0.0)

| | | | # of ANC Visit = 2: No (11.0)

| | | | # of ANC Visit = 3: Yes (13.0)

| | | Birth order = 4: No (0.0)

| | M.Education = 2: Yes (19.0)

| | M.Education = 3: No (0.0)

| Region = 5

| | M.Occupation = 0

| | | Core ANC Services took = 0

| | | | Birth order = 1: Yes (2.0)

| | | | Birth order = 2: No (0.0)

| | | | Birth order = 3: No (0.0)

| | | | Birth order = 4: No (40.0)

| | | Core ANC Services took = 1: No (8.0)

| | | Core ANC Services took = 2

| | | | Birth order = 1: No (3.0)

| | | | Birth order = 2

| | | | | # of ANC Visit = 1: No (0.0)

| | | | | # of ANC Visit = 2: Yes (6.0)

| | | | | # of ANC Visit = 3: No (20.0)

| | | | Birth order = 3: No (44.0)

| | | | Birth order = 4: No (22.0)

| | | Core ANC Services took = 3

| | | | Access to Media = 0

| | | | | H.Occupation = 0: No (15.0)

| | | | | H.Occupation = 1: No (14.0)

| | | | | H.Occupation = 2: No (2.0)

| | | | | H.Occupation = 3

| | | | | | M.Education = 0: Yes (6.0)

| | | | | | M.Education = 1: No (4.0)

| | | | | | M.Education = 2: Yes (0.0)

| | | | | | M.Education = 3: Yes (0.0)

| | | | Access to Media = 1: No (0.0)

| | | | Access to Media = 2: Yes (5.0)

| | | Core ANC Services took = 4

| | | | H.Education = 0: No (28.0)

| | | | H.Education = 1: No (0.0)

| | | | H.Education = 2: No (0.0)

| | | | H.Education = 3: Yes (12.0/1.0)

| | M.Occupation = 1: No (19.0)

| | M.Occupation = 2: Yes (11.0)

| | M.Occupation = 3

| | | M.Education = 0: Yes (17.0)

| | | M.Education = 1: No (4.0)

| | | M.Education = 2: Yes (15.0)

| | | M.Education = 3: Yes (0.0)

| Region = 6

| | H.Occupation = 0

| | | H.Education = 0: No (0.0)

| | | H.Education = 1: No (30.0)

| | | H.Education = 2: Yes (12.0)

| | | H.Education = 3: No (0.0)

| | H.Occupation = 1

| | | Birth order = 1: No (15.0)

| | | Birth order = 2

| | | | Core ANC Services took = 0: No (19.0)

| | | | Core ANC Services took = 1: No (3.0)

| | | | Core ANC Services took = 2

| | | | | Wealth = 1: No (3.0)

| | | | | Wealth = 2: No (5.0)

| | | | | Wealth = 3: Yes (0.0)

| | | | | Wealth = 4: Yes (14.0)

| | | | | Wealth = 5: Yes (0.0)

| | | | Core ANC Services took = 3

| | | | | M.Occupation = 0: No (0.0)

| | | | | M.Occupation = 1

| | | | | | M.Education = 0: No (5.0)

| | | | | | M.Education = 1: No (19.0)

| | | | | | M.Education = 2: Yes (4.0)

| | | | | | M.Education = 3: No (0.0)

| | | | | M.Occupation = 2: No (0.0)

| | | | | M.Occupation = 3: Yes (11.0)

| | | | Core ANC Services took = 4

| | | | | Wealth = 1: No (5.0)

| | | | | Wealth = 2: Yes (39.0)

| | | | | Wealth = 3: Yes (11.0)

| | | | | Wealth = 4

| | | | | | M.Education = 0: No (3.0)

| | | | | | M.Education = 1: Yes (3.0)

| | | | | | M.Education = 2: Yes (0.0)

| | | | | | M.Education = 3: Yes (0.0)

| | | | | Wealth = 5: Yes (0.0)

| | | Birth order = 3

| | | | Access to Media = 0

| | | | | Wealth = 1: No (4.0)

| | | | | Wealth = 2: No (12.0)

| | | | | Wealth = 3: No (0.0)

| | | | | Wealth = 4: Yes (11.0)

| | | | | Wealth = 5: No (0.0)

| | | | Access to Media = 1: No (0.0)

| | | | Access to Media = 2: No (18.0)

| | | Birth order = 4

| | | | Access to Media = 0

| | | | | H.Education = 0

| | | | | | Core ANC Services took = 0: No (0.0)

| | | | | | Core ANC Services took = 1: No (0.0)

| | | | | | Core ANC Services took = 2: No (0.0)

| | | | | | Core ANC Services took = 3: Yes (7.0)

| | | | | | Core ANC Services took = 4: No (29.0)

| | | | | H.Education = 1: No (44.0)

| | | | | H.Education = 2: No (0.0)

| | | | | H.Education = 3: No (0.0)

| | | | Access to Media = 1

| | | | | Core ANC Services took = 0: Yes (0.0)

| | | | | Core ANC Services took = 1: Yes (0.0)

| | | | | Core ANC Services took = 2: Yes (0.0)

| | | | | Core ANC Services took = 3: No (5.0)

| | | | | Core ANC Services took = 4: Yes (7.0/1.0)

| | | | Access to Media = 2: No (5.0)

| | H.Occupation = 2: Yes (9.0)

| | H.Occupation = 3: Yes (37.0)

| Region = 7

| | Core ANC Services took = 0: No (79.0)

| | Core ANC Services took = 1: No (24.0)

| | Core ANC Services took = 2

| | | M.Education = 0

| | | | Access to Media = 0: No (65.0)

| | | | Access to Media = 1: No (41.0)

| | | | Access to Media = 2: Yes (2.0)

| | | M.Education = 1

| | | | Wealth = 1: No (22.0)

| | | | Wealth = 2: No (28.0)

| | | | Wealth = 3

| | | | | M.Occupation = 0: Yes (41.0)

| | | | | M.Occupation = 1: Yes (0.0)

| | | | | M.Occupation = 2: Yes (0.0)

| | | | | M.Occupation = 3: No (9.0)

| | | | Wealth = 4

| | | | | M.Occupation = 0: No (0.0)

| | | | | M.Occupation = 1: Yes (4.0)

| | | | | M.Occupation = 2: No (0.0)

| | | | | M.Occupation = 3: No (20.0)

| | | | Wealth = 5: No (0.0)

| | | M.Education = 2: Yes (25.0)

| | | M.Education = 3: No (0.0)

| | Core ANC Services took = 3

| | | H.Occupation = 0: No (0.0)

| | | H.Occupation = 1: No (58.0)

| | | H.Occupation = 2: No (0.0)

| | | H.Occupation = 3

| | | | M.Occupation = 0: No (4.0)

| | | | M.Occupation = 1: Yes (0.0)

| | | | M.Occupation = 2: Yes (0.0)

| | | | M.Occupation = 3: Yes (31.0)

| | Core ANC Services took = 4

| | | H.Occupation = 0: No (12.0)

| | | H.Occupation = 1

| | | | Wealth = 1: No (13.0)

| | | | Wealth = 2

| | | | | M.Education = 0: No (4.0)

| | | | | M.Education = 1: Yes (5.0)

| | | | | M.Education = 2: Yes (0.0)

| | | | | M.Education = 3: Yes (0.0)

| | | | Wealth = 3

| | | | | Birth order = 1: Yes (0.0)

| | | | | Birth order = 2: Yes (0.0)

| | | | | Birth order = 3

| | | | | | M.Education = 0: No (4.0)

| | | | | | M.Education = 1: Yes (12.0)

| | | | | | M.Education = 2: Yes (0.0)

| | | | | | M.Education = 3: Yes (0.0)

| | | | | Birth order = 4: Yes (26.0)

| | | | Wealth = 4

| | | | | M.Education = 0: No (0.0)

| | | | | M.Education = 1: No (22.0)

| | | | | M.Education = 2: Yes (13.0)

| | | | | M.Education = 3: No (0.0)

| | | | Wealth = 5: Yes (0.0)

| | | H.Occupation = 2: Yes (2.0)

| | | H.Occupation = 3

| | | | M.Occupation = 0

| | | | | Wealth = 1

| | | | | | Birth order = 1: Yes (0.0)

| | | | | | Birth order = 2: Yes (23.0)

| | | | | | Birth order = 3: Yes (0.0)

| | | | | | Birth order = 4: No (4.0)

| | | | | Wealth = 2: Yes (0.0)

| | | | | Wealth = 3: Yes (22.0)

| | | | | Wealth = 4: Yes (10.0)

| | | | | Wealth = 5: Yes (0.0)

| | | | M.Occupation = 1: Yes (0.0)

| | | | M.Occupation = 2: Yes (0.0)

| | | | M.Occupation = 3: No (2.0)

| Region = 8

| | Wealth = 1

| | | Core ANC Services took = 0: No (0.0)

| | | Core ANC Services took = 1

| | | | # of ANC Visit = 1: No (0.0)

| | | | # of ANC Visit = 2: No (32.0)

| | | | # of ANC Visit = 3: Yes (3.0)

| | | Core ANC Services took = 2: Yes (3.0)

| | | Core ANC Services took = 3: Yes (1.0)

| | | Core ANC Services took = 4: No (48.0)

| | Wealth = 2

| | | M.Education = 0: Yes (0.0)

| | | M.Education = 1

| | | | # of ANC Visit = 1: No (0.0)

| | | | # of ANC Visit = 2: No (6.0)

| | | | # of ANC Visit = 3

| | | | | H.Occupation = 0: No (2.0)

| | | | | H.Occupation = 1: No (1.0)

| | | | | H.Occupation = 2: No (0.0)

| | | | | H.Occupation = 3: Yes (2.0)

| | | M.Education = 2: Yes (13.0)

| | | M.Education = 3: Yes (0.0)

| | Wealth = 3

| | | Birth order = 1: No (19.0)

| | | Birth order = 2: No (7.0)

| | | Birth order = 3: Yes (4.0)

| | | Birth order = 4: No (0.0)

| | Wealth = 4: Yes (3.0)

| | Wealth = 5: No (5.0)

| Region = 9

| | Wealth = 1: No (18.0)

| | Wealth = 2: No (20.0/2.0)

| | Wealth = 3: No (23.0)

| | Wealth = 4

| | | # of ANC Visit = 1: Yes (2.0)

| | | # of ANC Visit = 2: No (13.0)

| | | # of ANC Visit = 3: Yes (18.0)

| | Wealth = 5

| | | H.Occupation = 0: Yes (0.0)

| | | H.Occupation = 1: Yes (18.0)

| | | H.Occupation = 2: Yes (0.0)

| | | H.Occupation = 3

| | | | M.Education = 0: Yes (5.0)

| | | | M.Education = 1: No (2.0)

| | | | M.Education = 2: Yes (0.0)

| | | | M.Education = 3: Yes (0.0)

| Region = 10: No (0.0)

| Region = 11

| | Wealth = 1

| | | Core ANC Services took = 0: Yes (11.0)

| | | Core ANC Services took = 1: Yes (0.0)

| | | Core ANC Services took = 2: Yes (39.0)

| | | Core ANC Services took = 3: No (4.0)

| | | Core ANC Services took = 4: No (17.0)

| | Wealth = 2: No (26.0/2.0)

| | Wealth = 3: Yes (9.0/1.0)

| | Wealth = 4: Yes (37.0)

| | Wealth = 5: No (3.0)

Number of Leaves : 358

Size of the tree : 463

Time taken to build model: 0.09 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	4871	98.7432 %
Incorrectly Classified Instances	62	1.2568 %
Kappa statistic	0.9749	
Mean absolute error	0.0169	
Root mean squared error	0.1079	
Relative absolute error	3.3781 %	
Root relative squared error	21.5862 %	
Total Number of Instances	4933	

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure ROC Area Class

	0.989	0.014	0.986	0.989	0.987	0.996	Yes
	0.986	0.011	0.989	0.986	0.988	0.996	No
Weighted Avg.	0.987	0.013	0.987	0.987	0.987	0.996	

=== Confusion Matrix ===

a b <-- classified as

2416 27 | a = Yes

35 2455 | b = No