ADDIS ABABA UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES

SCHOOL OF INFORMATION SCIENCE

DEVELOPING KNOWLEDGE BASED SYSTEM TO DETERMINE THE CHOICE OF CONTRACEPTIVE METHODS USING DATA MINING TECHNIQUE

BY
DESALEGN GETANEH

OCTOBER 2017

ADDIS ABABA, ETHIOPIA

ADDIS ABABA UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES

SCHOOL OF INFORMATION SCIENCE

DEVELOPING KNOWLEDGE BASED SYSTEM TO DETERMINE THE CHOICE OF CONTRACEPTIVE METHODS USING DATA MINING TECHNIQUE

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: Desalegn Getaneh

Advisor: Martha Yifiru (PhD)

October 2017

Addis Ababa, Ethiopia

ADDIS ABABA UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES

SCHOOL OF INFORMATION SCIENCE

DEVELOPING KNOWLEDGE BASED SYSTEM TO DETERMINE THE CHOICE OF CONTRACEPTIVE METHODS USING DATA MINING TECHNIQUE

By: Desalegn Getaneh

Name and signature of Members of the Examining Board

Martha Yifiru (PhD) Advisor	_____ Signature	_____ Date
Temtim Assefa (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: _____

Desalegn Getaneh

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

Advisor's Signature: _____

Martha Yifiru (PhD)

ACKNOWLEDGMENTS

First and foremost, I would like to thank the almighty God and his Mother Saint Marry 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. Martha Yifiru for her support, encouragement, and insightful comments.

I would also like to extend my gratitude to Teshager Mersha and Dr. Teferi Girma for their valuable comments and feedback on my thesis draft. I also thank you Mathias Tenaye and Yishak Eshetu supporting me in technical matters.

My acknowledgement would be incomplete without thanking the source of my strength, support and unconditional love, my wife, Selamawit Degefa and the whole my family. Thank you all

DEDICATION

This thesis work is dedicated to those women who are not able to use contraceptive method of their choice.

Contents

List of Figures

List of Tables

List of Abbreviation

CRISP-DM	CRoss Industry Standard Process for Data Mining
EDHS	Ethiopian Demographic and Health Survey
EFDR	Ethiopian Federal Democratic Republic
EHR	Electronic Health Record
FMOH	Federal Ministry of Health
FP	False Positive;
HIV	Human Immino Virus;
HKDD	Hybrid Knowledge Discovery in Databases
ICI:	Incorrectly Classified Instances
IUCD	Intra Uterine Contraceptive Device
KBS	Knowledge Based System
KDD	Knowledge Discovery from Data
MCH	Maternal and Child Health
mCPR	Modern Contraceptive Prevalence Rate
MEC	Medical Eligibility Criteria
MMR	Maternal Mortality Ratio
MSIE	Marie Stopes International Ethiopia
Open EMR	Open Electronic Medical Record
PID	Pelvic Inflammatory Disease;
RIPPER	Repeated Incremental Pruning to Produce Error Reduction
ROC:	Receiver Operating Characteristic;
SEMMA	Sample, Explore, Modify, Model, Assess
SPSS	Statistical Package for Social Science
SRH	Sexual and Reproductive Health
STI	Sexual transmitted diseases;
TFR	Total Fertility Rate
TP	True positive;
UNFPA	United Nation Population Fund
WEKA	Waikato Environment for Knowledge Analysis
WHO	World Health Organization

Abstract

Ethiopia is one of the most populous country in Africa with a fertility rate of 4.6 and eight percent unwanted pregnancy. This contributes for maternal mortality and child death. To make every child is wanted, family planning plays an essential role through delaying, spacing or limiting birth. But the contraceptive prevalence rate of Ethiopia is low and needs to work hard to address the demand of family planning through providing choice of contraceptive method to a woman.

Socio demographic characteristic has a contributing factor for choice of contraceptive method but which variables determine the choice of contraceptive method is a challenge. Moreover, among clients received contraceptive method, they might not get their choices. Due to this, there is 13 percent discontinuation rate though service provider uses different job aids to support the choice of contraceptive method. Thus, this research focus on identifying key variables that determine the choice of contraceptive method though applying data mining techniques and develop a knowledge based system that supports the health service provider for the choice of contraceptive methods.

Empirical research design is applied to achieve this objective which combine both experimental and non-experimental researches. Prototyping approach is followed to develop the knowledge based system. As a research method, knowledge engineering and hybrid data mining methodology was employed. Interview and document analysis was also conducted to acquire knowledge from domain experts and documents respectively.

A decision tree J48 algorithm was used to predict variables that determine the choice of contraceptive method. Thus, client age, number of children, education, residence, marital status, religion, region and contraceptive history determine the choice of contraceptive methods are key variables in the choice of contraceptive method. In addition, medical eligibility criteria and life style of a woman has a factor in the choice of contraceptive method.

Prototype knowledge based system is developed that determine choice of contraceptive methods through integrating data mining results (socio demographic variables), medical eligibility criteria (explicit knowledge) and life style of a woman (tacit knowledge).

Based on system performance evaluation and user acceptance test, 86.6 % of accuracy and 76% acceptance was scored respectively so that integrating socio demographic data, medical eligibility criteria and life style of a woman is possible and can be implemented in the domain area. Finally, further exploration has to be done to refine the knowledge base and boost the advantages of choice of contraceptive method of knowledge based system to incorporate woman who has special characteristics.

Keywords: *Family planning, Choice of Contraceptive Method, Data Mining, Hybrid Modeling, Knowledge Based System, Contraceptive method choice knowledge based system*.

Chapter One

Introduction

1.1. Background

1.1.1. Family planning

Ethiopia is one of the most populous countries in the world with more than 99 million people, of which 86% of the population live in rural areas [1] [2]. The fertility rate of Ethiopia is 4.6 in 2016, this is high number that a woman bears an average of five children in her life time [3]. According to [4], the maternal mortality in Ethiopia is 353 per 100,000 women and the proportion of maternal deaths among deaths of female reproductive age is 16.7 percent. The survey [3] also states that 8 percent of births are unwanted and 17 percent are mistimed birth (decision made to be the birth wanted is late). In addition, Ethiopian women are currently having, on average, one child more than they want. In every 10 pregnancies, one is ended with abortion [5]. Thus, family planning play a vital role in protecting women's and children's health, reducing unintended pregnancy and reducing maternal and child mortality and morbidity.

Family planning program gives to couples the freedom and ability to bear the number of children they want and to achieve the spacing of births they prefer. World Health Organization (WHO) also stated as *"Family planning saves lives of women and children and improves the quality of life for all. It is one of the best investments that can be made to help ensure the health and well-being of women, children, families and communities"* [6].

The right of women to get family planning methods is clearly stated in Federal Democratic Republic of Ethiopia constitution. In this constitution, Article 35.9, the right of women in accessing family planning is stated as *"to prevent harm arising from pregnancy and childbirth and in order to safeguard their health, women have the right to Family Planning education, information, and capacity"* [7].

As of 2016, 35 percent of married Ethiopian women of childbearing age (15–49) use any method of family planning; this is a great improvement from 2005, when only 15 percent of married women of childbearing age were using any form of contraception.

However, according to [3], 22 percent of married women do not want any more children or want to wait for two or more years before having another child but are not currently using any form of contraception though the knowledge of contraceptive use is nearly universal.

There are different types of contraceptive method that used to space or limit births for women of reproductive age group. Among these methods, oral contraceptive pills, injection Depo-Provera for three months, implant for 3 to 5 years, IUCD for 5 to 10 years and Bilateral Tuba ligation (BTL). In fact, vasectomy is also a contraceptive method for men used to limit birth [8].

EDHS 2016 [3] also report that among contraceptive users, method choice is limited very much to the shorter-term methods which are pills and injectable. This means that out of 35 percent of modern contraceptive users, short term family planning users cover 25 percent i.e. three clients out of four are short term family planning users. On the other hand, the uptake of long term [implant and IUCD] and permanent [BTL] family planning methods is only 10 percent out of 35.

1.1.2. Data Mining Concepts

Data mining is also known as Knowledge Discovery from Data (KDD) can be defined as mining knowledge from data and also a process of finding interesting patterns and hidden knowledge from vast dataset without setting prior assumption [9].
Since there is a data, data mining can be applied in any of the data i.e. it can be applied on data like text data, special data, audio data, video data, web data etc. Data mining adopts techniques from many domains like machine learning, statistics, pattern recognition, visualization, algorithms, high performance computing, applications, information retrieval, data warehouse and database systems [9].

Data mining has two main tasks i.e. prediction and description. Prediction aims to find different kinds of structures and relations in the data to generate rules and models that will be used for

decision making process. Description is identifying patterns or relations in the data by examining the property of the data itself [10].

Healthcare industry today generates large amounts of complex data about patients, hospital resources, disease diagnosis, electronic patient records, medical devices etc. Larger amounts of data are a key resource to be processed and analyzed for knowledge extraction that enables support for cost-savings and decision making. Data mining applications in healthcare can be grouped as treatment effectiveness, healthcare management, customer relationship management, fraud and abuse [11]. Through identifying the hidden pattern in health care dataset, an expert or decision maker used for better management of patients and the health system in general.

1.1.3. Knowledge Based System

Knowledge based system can be defined as "*sophisticated interactive computer programs which use high quality, specialized knowledge in some narrow problem domain to solve complex problems in that domain*" [12]. It is a software system that contains a significant amount of knowledge in an explicit and declarative form.

The knowledge based system has knowledge base, working memory and inference engine [13]. The knowledge base represents the repository of knowledge for specific and narrow domain. Usually in any knowledge base there are many facts, rules, and Meta knowledge [13]. Knowledge base constitutes the problem solving rules, facts, or intuition that a human expert might use in solving problems in a given problem domain. The knowledge base is usually stored in terms of if–then rules. In the knowledge base, the working memory represents relevant data for the current problem being solved. The inference engine is a component of the system that applies logical rules to the knowledge base to deduce new information and works in two modes: forward chaining or backward chaining. Facts and rules which are the heart of knowledge based system is extracted from data mining knowledge discovery process since the basis of data mining is a process of using tools to extract useful knowledge from large datasets [14].

1.2. Statement of the Problem

The percentage of married women currently using a modern contraceptive method has increased substantially over the past 15 years. There is a four-fold increase in the use of a method of contraception by currently married women, from 8 percent in 2000 to 35 percent in 2016. Much of this increase is attributable to the sharp increase in the use of injectable. Use of injectable increased from 3 percent in 2000 to 23 percent in 2016. In addition, although the overall use of implants continues to be low, its use has increased in the last 10 years from less than one percent (0.2 percent) in 2005 to 8 percent in 2016. Total fertility rate also decreased from 4.8 in 2011 to 4.6 in 2016. Similarly, unmet need for family planning has dropped from 26.3 percent in 2011 to 22.3 percent in 2016 [3].

Although there is remarkable achievement in increasing the contraceptive prevalence rate in the country, still the uptake of contraceptive method couldn't address the demand. In 2016, total contraceptive demand was 58% among married women ages 15 to 49 but actually received the methods are only 35 percent. Of 58 percent demand for family planning, 22 percent couldn't receive family planning even if they want 13 percent for spacing and 9 percent for limiting) [3] [18].

The Low uptake of contraceptive methods could happen because of different reason. According to EDHS depth analysis result which is conducted by UNFPA in 2012, contraceptive users are also vary by religion which is Orthodox Christians covers 35% of the users followed by Protestant 30%, Muslim 20% and other religions took the share 15%. But, this didn't show whether the religion of women affects their choice of contraceptive methods or not [19].

A survey on performance monitoring and accountability also show that one of the reasons for low uptake of family planning is poor counseling. More than half of contraceptive users (51.3%) used the method without receiving counseling on side effects. Out of ten health professionals referred to as service providers in this thesis, four of them (41.3%) couldn't give counseling on methods other than clients pre choice decision [20]. The study conducted by Marie Stopes international Ethiopia also shows that there is a great variance among service providers while they conduct counseling on modern contraceptives i.e. one third of the service providers give either too much or too little information to the client [21].

This shows that there is service provider difference in providing the appropriate information to the client in the process of choosing contraceptive method. In other words, the information provided to the user varies from provider to provider which results in the client might not receive the right and appropriate contraceptive method. Providing accurate information about side effects and gaining a full understanding of the client's reproductive intentions and preferences during counselling will impact on the likelihood of dis-continuation or switching to another method. The discontinuation rate for contraceptive method vary from method to method i.e. discontinuation rate of pills, implant/IUCD and injectable is 44%, 11% and 5% respectively [22].

Studies conducted in Ethiopia are basically focused on cross sectional study with statistical techniques for data analysis and interpretation to assess the barriers and determinant factors of low uptake of family planning service in general. There is also a paper done using data mining techniques that is applied in health care data to predict the likelihood of family planning [23]. In [23], the determinant variable was identified between user and non-user of contraceptive methods. In other words, the study lack the key determinant variables among contraceptive method choice.

In order to standardize the family planning service provision among service providers and to minimize provider difference, different family planning domain experts prepare different job aids. Some of these job aids are medical eligibility criteria wheel, counseling tools and charts [24] [25]. But all these wheels, charts, and other job aids are designed based on the health condition of client. None of them includes the socio demographic data of the client in the job aid tools. Designing a tool that combine socio demographic characteristics and medical eligibility of a client would support the service provider in choosing contraceptive methods which is appropriate to the client.

Thus, this study focused on identifying the hidden patterns of the contraceptive users in case they have common characteristics and identifying the key determinant factor for using any of the contraceptive methods. Finally, rules were generated from these determinant variables, develop a KBS that provide support to service providers in suggesting contraceptive methods that match with the client lifestyle, socio demographic and medical eligibility criteria of a woman. Therefore, the main objective of this study is to develop a knowledge based system through acquiring knowledge from data mining, domain experts and documents. Hence, based on the identified possible factors that determine the choice of contraceptive methods, decision makers can design

programs which help to increase utilization of family planning. In addition, developing prototype knowledge based system to support the service provider during counselling sessions to guide their clients through the process of making a decision about which contraceptive method best fits to a woman was also another goal of this study. Using mining results and other sources acquired knowledge, KBS helps providers can easily decide the appropriate choice of options of contraceptive methods to their clients. As a result this, KBS will create standardized level of counseling between service provider so as to provide choice of contraceptive method.

1.3. Research Question

In this study, the researcher aims to answer the following research questions which are designed to attain the research objectives.

- Which classification algorithm best predicts choice of contraceptive methods?
- What are the most determinant variables for choice of contraceptive methods?
- To what extent the KBS predict choice of contraceptive methods?

1.4. Objective of the Research

1.4.1. General Objective

The general objective of this study is to develop knowledge based system that determine choice of contraceptive methods.

1.4.2. Specific Objective

The followings are the specific objectives of the study to:

- Compare different data mining classification algorithm and choose the one which predicts contraceptive methods in better way.
- Identify if there is a hidden pattern on choice of contraceptive method use in the dataset.
- Identify determinant variables of contraceptive method choice
- Model and represent knowledge acquired from data mining, documents and domain experts.
- Build prototype knowledge based system which support health provider in choice of contraceptive method.

1.5. Significance of the Study

Knowing the determinant variables and develop a knowledge based system for choice of contraceptive method will play a significant role for decision makers, service providers and clients. Decision makers can design a program that will address the gaps of the determinant variables. For example, choice of contraceptive method is determined by religion, then decision maker can design intervention that will address this issue.

Knowledge based system now-a-days becomes very hot research areas due to the fact that these systems will help the users in giving quick decision in their day to day activities. This knowledge based system reduces the provider biasness and experience difference among providers so that there will be standardized service for contraceptive users.

Moreover, a woman can also use contraceptive methods that will match with her health condition, socio demographic variables, and lifestyle so as to reduce method discontinuation rate. In general, this study will help to increase the uptake of contraceptive methods through addressing the family planning counselor biasness so as to decrease the maternal and child mortality and to make every birth wanted.

Finally, the output of this research work can also be used as a baseline for future works in this domain area to develop mobile KBS that can be used in both service providers and clients.

1.6. Scope and limitation of the Research

1.6.1 Scope of the Research

Since the objective of this study is developing knowledge based system, knowledge acquisition covers application data mining, reviewing documents and interviewing domain experts. To acquire knowledge from data mining, MSIE dataset were employed which is collected from 24

centers between year of 2015 and 2016. This dataset covers Tigray, Amhara, Oromia, SNNPR, Addis Ababa and Dire Dawa regions and city administrations.

This study also only covers the development of knowledge based system but not test the change of contraceptive use in between facilities which use this KBS system and which are not.

1.6.2. Limitation of the research

Literatures [23] showed that partner occupation and wealth index of the client are determinant variables for the uptake of contraceptive method use i.e. if partner's occupation is non agriculturalist and the wealth index is rich, then there is high likely to use contraceptive method. But MSIE dataset lacks these variables so that data mining prediction has done without them.

Literatures on application of data mining and development of KBS in family planning dataset is also very limited. Due to this, related works in this paper is short. Moreover, the developed contraceptive method match knowledge based system hasn't learning component only respond using the hard coded rules to the user.

1.7. Organization the Thesis

This research is organized in six chapters. These are:

Chapter 1: It introduces the rationale for family planning, the current status of family planning, data mining concepts and its application in health sector, knowledge based system and this helps the health providers. The chapter also states the problem statement, research question, and objective of the research, significance of the research, and scope and limitation of the research.

Chapter 2: This chapter includes literature review on data mining concepts, models and tasks. In addition, it also includes knowledge base definition, key process of knowledge base and application of knowledge base in health sector. The chapter also includes related works done in data mining and knowledge based system developments as well as current contraceptive status of the country are included.

Chapter 3: Research design and methods are discussed. Data sources, tools required, data analysis and presentation method are briefly described. Moreover, testing and evaluation has been conducted.

Chapter 4: knowledge acquisition from data mining and other sources are discussed. Knowledge modeling and representation are described.

Chapter 5: KBS Design and implementation has been illustrated. The key system design activates are discussed. Implementation components are discussed. In addition, testing and evaluation of the system both system performance and user acceptance test results are described.

Chapter 6: Conclusion and recommendation of the system has discussed.

Chapter Two

Literature Review

2.1. Contraceptive Use in Ethiopia

In Ethiopia, contraceptive prevalence increased by more than about fourfold from 8.2% to 35% in 16 years from 2000 to 2016 [3]. Most regions, rural, urban areas as well as populations in different socio-demographics have seen significant increase in contraceptive use over the last decade. Yet, the increase varies greatly by region.

Inequalities in contraceptive uptake by urban-rural residence, region, education, women's employment and religion have persisted in the country.

EDHS report [3] also showed that the method mix of short term family planning accounts 71% whereas long term and permanent family planning has a share of only 19%. In other words, the skewed of the method mix that heavily relied on Injectable is not compatible with the high proportion of women who desire to limit birth. Analysis demonstrated that a good portion of current contraceptive users are in need of better –i.e. long acting/permanent – contraceptive method in order to decisively meet their desire for limiting. Notable improvement in the use of Implant has also been recorded in the last decade, accounting for 13% of the overall contraceptive use in 2016. The increase in contraceptive use during 2000-2016 emerged as the single most important source for the recorded decline in TFR of one child per women; accounted for 16% of the decline in TFR.

According to [3], currently married women in urban areas showed increment by 54 percent over their rural counterparts to use modern contraceptive method (50 and 32 percent, respectively), Use of any contraceptive method varies by region, ranging from 56 percent in Addis Ababa to 1.5 percent in the Somali region. Use of any modern contraceptive methods is highest in Addis Ababa (50 percent) and lowest in the Somali (1.4 percent). Current contraceptive use increases with women's education. Current use of any method to a woman with no education is thirty one whereas more than secondary education uses any method are 55 percent of women. Similarly, current use of any contraceptive method increases with wealth, from 22 percent of women in the lowest quintile to 48 percent of women in the highest quintile.

Contraceptive use in Ethiopia has improved considerably in the last decade principally due to the synergy of a conducive family planning program landscape and favorable social changes. Nevertheless, Ethiopia still remains one of the countries with low contraceptive use rate. Unmet need is still high (22.3%), thirteen percent of currently married women want to space birth whereas nine percent of the currently married women do not want more children. Although it has declined in the last decade as contraceptive use has risen and about half of the women have unsatisfied demand for family planning. There is a great potential to further improve contraceptive usage in the country [3].

2.2. Knowledge based systems Concepts

Before discussing the details of knowledge based system, let us first state the definition of the knowledge base referencing different scholar and authors. *"A knowledge base used to store complex structured and unstructured information used by a computer system. The initial use of the term was in connection with expert systems which were the first knowledge-based systems"* [26].

According to [27], *"Knowledge-based systems are considered to be a major branch of artificial intelligence. They are capable of making decisions based on the knowledge residing in them, and can understand the context of the data that is being processed"*. Another scholar also define knowledge based system as:

"Knowledge based system or expert system is a software that attempts to reproduce the performance of one or more human experts, most commonly in a specific problem domain, and is a traditional application and*/or subfield of artificial intelligence"* [28].

Knowledge based systems are sophisticated interactive computer programs that solves complex problems in that domain which use high quality, specialized knowledge in some narrow problem domain [29]. Knowledge Based Systems (KBS) also name as expert systems, intelligent assistants, epistemological systems and design and analysis systems. Usually scholars also use KBS and expert systems interchangeably [30]. Knowledge based system emulates the behavior of human expert within a well-defined and narrow domain of knowledge [31]. It is a system that

draws upon the knowledge of human experts captured in a knowledge base to solve problems that normally require human expertise [32].

2.2.1. Advantages of Knowledge Based System

Knowledge based systems is more useful in many situations than the traditional computer based information systems. [33] Underlined the following advantages of knowledge based system:

- Save time: The time that will be spent doing manually will be minimized.
- Work Quality: The work quality of the service will increase due to reducing the errors happen in the center.
- Complete: unless there are implementation errors, knowledge based systems will always produce the desired result as they will not leave out any rule (consideration) in the reasoning processes.
- Replication: human experts are scarce resources. They are physically bound to their geographical locations and can only available at one place at a time but knowledge based system can be replicated and in effect to be transferred to any other locations to perform other task.
- Updating knowledge. Knowledge based system can be updated easily by editing the rule base; but human expert take to retrain.

2.2.2. Knowledge Based System Architecture

Architecture is a blue print that used to denote the structure of system. System architecture is a conceptual model that defines the structure and guidelines of the system. According to [34], figure 1 shows the architecture of knowledge based system with its components.

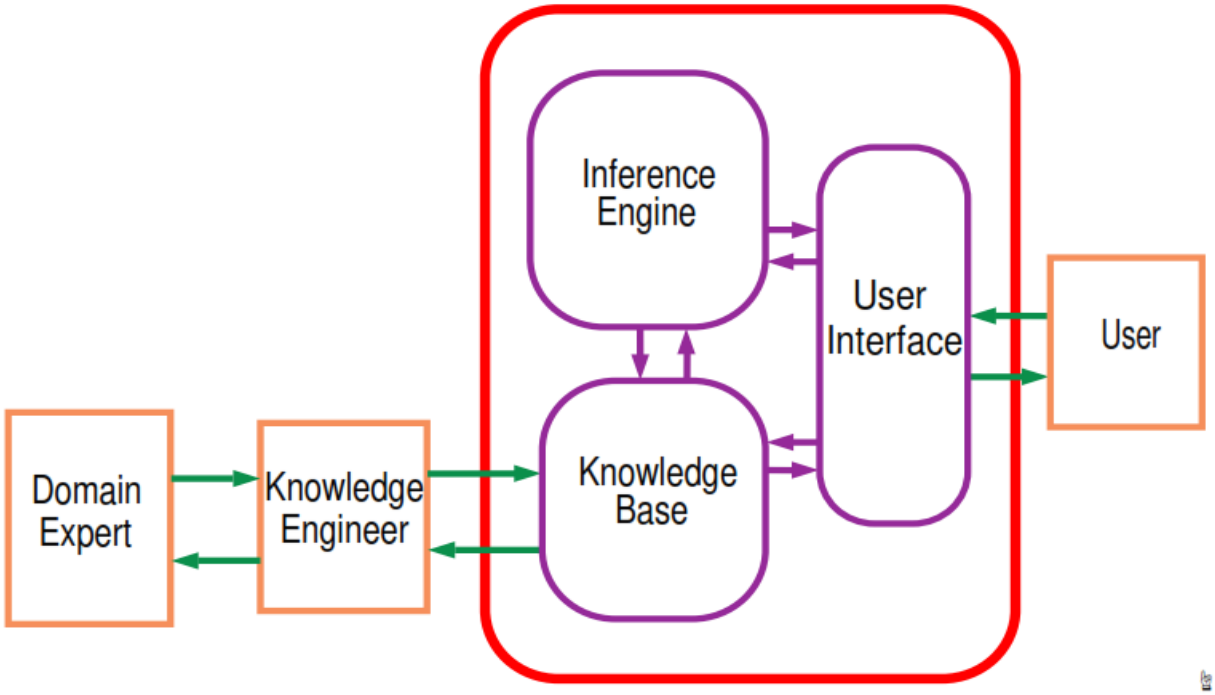


Figure 1: Knowledge Based System Architecture [34] [35]

Domain Expert: Domain Expert is a person who has knowledge and skill and able to solve problems in a specific area or domain. The expert know about the domain, but nothing about particular cases or how the system works. In the knowledge based system development, the expertise has been captured in the knowledge based system [29] [34].

Knowledge Engineer: design, build, and debug the knowledge base in consultation with domain experts. The knowledge engineer interviews the expert to provoke his or her knowledge; the knowledge engineer encodes the elicited knowledge for the knowledge base; Interviews the domain expert to find out how a particular problem is solved; Establishes what reasoning methods the expert uses to handle facts and rules and decides how to represent them in the knowledge based system; Chooses some development software or knowledge based system shell, or looks at programming languages for encoding the knowledge [34] [36].

Knowledge Base: It represents the repository of knowledge for specific domain. In any knowledge base, there are many facts, rules, and meta knowledge. Knowledge base constitutes the problem solving rules, facts, or intuition that a human expert might use in solving problems in a given problem domain. If-then rules are the usual storage mechanism of knowledge base [37] [29].

Inference Engine: "*refers to the part of knowledge based system that specifies the logical process by which new facts and belief are derived from known facts and beliefs. It also contains the control*

strategy that orders the search for an inferential solution. Inference engine is a generic control mechanism that applies the axiomatic knowledge in the knowledge base to the task specific data to arrive at some solution or conclusion" [29].

User Interface: the interaction between the user and the knowledge based system i.e. it simulates the communications with the environment unit of the functional model of human system. User interface is allowing the user to input information in response to questions generated by the system [38].

Explanation and justification: involves the design and programming of explanation capability. For example, the ability to answer questions such as why a specific piece of information is needed by the computer or how a certain conclusion was derived by the computer [38].

2.2.3. Phases of Knowledge Based System Development

According to [39], the development of knowledge based system has different stages. The main phases are discussed below.

Knowledge Acquisition: Knowledge acquisition is the accumulation, transfer, and transformation of problem solving expertise from experts or documented knowledge sources to a computer program for constructing or expanding the knowledge base [40]. It also concerned with the eliciting and representing the knowledge of human experts. It is the most important stage in the development of KBS [41]. In fact, knowledge acquisition is the most difficult task in developing KBS due to that the expert has insufficient knowledge about programming and expert system techniques and the expert find it difficult to describe his knowledge completely and correctly [41]. Knowledge acquisition is a two-step process: knowledge elicitations and structuring and can be described as follows.

Knowledge elicitation: is the process of explicating domain specific knowledge underlying human performance. The elicitation knowledge from experts can be extracted either manually or with the help of computers. Manual methods knowledge elicitation methods are interview, observation, process tracking, protocol analysis etc. Computer aided are usually knowledge discovery process using data mining technique [42].

Knowledge Structure: Elicited knowledge from domain experts should be structured in well-organized manner. Knowledge modeling, representation of knowledge using symbols to represent new pieces and to create new knowledge, is used to structure the process, acquire and store the knowledge [43].
In general, knowledge acquisition is useful to increase the productivity of knowledge engineering, to reduce the need of domain expert, and to increase the quality of acquired knowledge [37].

Knowledge Modeling: Knowledge modeling is a cross disciplinary approach to capture and model knowledge. Knowledge models view the knowledge based system using diagram and other structured representations such as trees, maps, and KBS construction methods typically provide tools for knowledge analysis in the form of conceptual models of knowledge. So, knowledge model provides an implementation independent specification of knowledge in an application domain [44] [45]. Knowledge model can be constructed by ontology, entity relational data model, UML class diagrams UML activity diagram, business process model [46].

Knowledge Representation: Once knowledge is acquired, it is the time of knowledge representation. Preparation of a knowledge map and encoding of the knowledge in the knowledge base are the main role of this phase. In this phase, human knowledge is encoded into machine or equivalent language. Knowledge can be represented by production rules, decision trees, frames, objects etc. [42]. It is responsibility of the knowledge engineer to select appropriate knowledge representation scheme that is natural, efficient, and transparent and developer friendly [39]. Knowledge representation cannot be defined in pure epistemological terms. So, good knowledge representation can be measured by [47]:

- *Support to efficient reasoning*
- *Expressivity – how expressive the knowledge is*
- *Adequacy – is the represented knowledge adequate*
- *Satisfiability – role of knowledge which satisfies the goal*
- *Quality – quality of knowledge within the knowledge representation*
- *Uncertainty – how much certain the expressed knowledge is*
- *Consistency – how much consistent the knowledge is*.

Knowledge representation is a difficult task and has challenges. Some of the challenges are to cope with dynamism in the world knowledge, to preserve consistency of information across domains, to accept belief revisions in the knowledge covered under representation, to represent beliefs in an easy and resourceful manner, to facilitate better reasoning and inferences over the represented knowledge, to enable changes from the knowledge definition perspective, to adapt to addition of necessary information with change in the specification or conceptualization [47].

Common knowledge representation mechanisms are discussed below.

Rule based system: is the most common and popular knowledge representation methods in KBS. A rule means a structure which has if component and a then component.

if < conditions >

then < conclusion >

where <conditions> represents the conditions of a rule, whereas <conclusion > represents its conclusion.

A rule-based system, therefore, either identifies a pattern and draws conclusions about what it means, **OR** identifies a pattern and advises what should be done about it, **OR** identifies a pattern and takes appropriate action [48] [29].

There are two main kinds of reasoning strategies employed in rule based systems which include forward and backward chaining mechanism [49].

Forward Chaining: Forward chaining also called data driven approach, starts from the facts. It also apply rules to find all possible conclusions. It is data driven not goal oriented. This is an initial data and uses inference rules. It helps in extracting more data until a goal is reached. An inference engine using forward chaining searches the inference rules until it finds one. Here the antecedent is known to be true [49].

Backward Chaining: Backward chaining start with the desired conclusion(s) and work backwards to find supporting facts. It is also goal-directed. An inference engine using “Backward Chaining” would search the inference rules until it finds one which has a “THEN clause” that matches a desired goal [49].

According to [49], table 1 shows the comparison between the forward and backward chaining.

Table 1: Comparison between forward and backward chaining in Rule based system *[49]*

Forward Chaining	**Backward Chaining**
It starts with new data.	It starts with some goal or hypothesis.
It asks few questions.	It asks many questions.
It examines all rules.	It examines some rules.
Slow approach.	Fast approach.
Gather larger information from small amount of data.	It produce small amount of information from available data.
Forward Chaining is primarily data driven.	Backward Chaining is primarily Goal Driven.
It uses its input. It searches rules for answers.	It proves the considered hypothesis.
It is a form of Top-Down reasoning.	It is a form of bottom up reasoning.
Works forward to find conclusions from facts.	Works backward to find facts that support the hypothesis.
It tends to breath – first.	It tends to depth – first.
Forward Chaining is suitable for problems that start from data collection; e.g. planning, monitoring and control.	Backward Chaining is suitable for problems that start from hypothesis, e.g. diagnosis.
This type of chaining is non-focused because it infers all conclusions, may answer unrelated questions.	This type of chaining is focused to prove the goal and search as only the part of knowledge base that is related to the problem.
Explanation is not facilitated in Forward Chaining.	Explanation is facilitated in Backward Chaining
All data is available.	Data must be acquired interactively (i.e. on demand)
It deals with less number of initial states and many results.	It deals with less starting goals and many facts.
Forming a goal is difficult in case of Forward Chaining.	Forming a goal is easy in case of Backward Chaining.

Case based Representation: Case-based reasoning means understand solve new problems using old experiences. Previous situation is remembered similar to the current one and uses that to solve the new problem. It works through adapting new situations and solve the old one. Solutions to meet new demands; using old cases to explain new situations; using old cases to critique new solutions; or reasoning from precedents to interpret a new situation or create an equitable solution to a new problem. In case-based reasoning, a reasoner remembers previous situations similar to the current one and uses them to help solve the new problem [50].

The quality of a case-based reasoner's solutions depends on four things:

- the experiences it's had,
- its ability to understand new situations in terms of those old experiences,
- its adeptness at adaptation, and
- Its adeptness at evaluation [50].

Frame: A frame consists of a collection of slots which can be filled by values or pointers to other frames. A Frame is a collection of questions to be asked about a hypothetical situation: it specifies

issues to be raised and methods to be used in dealing with them. They are a natural extension of Semantic Networks. A frame includes all the knowledge about an object. The knowledge in a frame is divided into slots. They consist of sets of slots filled by values, procedures for calculating values, or pointers to other frames. Frames provide visual context to guide scene interpretation. It tells the program what to look for and where to look for it. A Frame is composed of a set of "slots" and "methods".

A slot is a named place holder for a point. Slots point to other frames that represent entities that are described (or interpreted) by the frames [51] [52].

Script: A script is a remembered precedent, consisting of tightly coupled, expectation suggesting primitive action and state change frames. It is a structured representation describing a stereotyped sequence of events in a particular context. Scripts are data structure used to represent a sequence of events. They are used for interpreting stories. They are frame based systems that describe stereotyped sequences of events and actions that enable an intelligent agent to perform appropriately in a particular context. Scripts have been used to interpret, understand and reason about stories; understand and reason about observed events; reason about observed actions; plan actions to accomplish tasks [52].

Semantic Nets: A semantic network is widely used knowledge representation technique. As the name semantic network, it represents the connection between objects or class of objects. It is composed of nodes and links (arcs) that show hierarchical relationships between objects. A semantic network is often used as a form of knowledge representation [29]. It is a directed graph consisting of vertices which represent concepts and edges which represent semantic relations between the concepts. A semantic network represents knowledge as a set of labeled nodes and links. Properties of semantic nets is:

- Allows to structure the knowledge to reflect the structure of that part of the universe which is being represented.
- Default values (e.g. height of a baseball player to be 195cm), very strong representation facilities by procedural attachment
- Notion explained up to now quite general, for a useful tool, must be much refined. In particular clear syntax, but clear semantics has to be worked out [52].

Knowledge Validation: Knowledge validation (or verification) involves validating and verifying the knowledge (e.g., by using test cases) until its quality is acceptable. Testing results are usually shown to a domain expert(s) to verify the accuracy of the knowledge based system [53].

2.2.4. Knowledge Based System in health sector

The role of knowledge based system in health sector was studied by the Bristol group identified that Knowledge Based Systems (KBS) will play in health care over the next decade. In particular, it seeks to identify the key clinical areas that will require computerized decision support, and examines the way in which KBS technology may prove to be the key enabling technology [54]. There is a long tradition now in developing clinical support systems that focus on diagnostic assistance for clinicians. Faced with growing evidence both that diagnosis is only one of many problems in clinical medicine, as well as the manifest failure in the adoption of diagnostic systems into routine practice.

2.3. Data Mining Concepts

Now days vast amounts of data are collected daily in the world we live [9]. Due to this too much data with the phenomenal rate of growth of data, users expect more sophisticated useful and valuable information. Because of this, users are interested to extract useful information which is usually knowledge from this massive data. Without a support of tool to extract the knowledge, accumulating huge data would be non-sense and becomes meaningless. So, the concept of data mining is used to extract interesting patters.

According to [9], data mining is the process of discovering insightful, interesting, and novel patterns, as well as descriptive, understandable, and predictive models from large-scale data. Jeffery and Anand also define data mining as the discovery of models for data [55]. Data mining is also known as Knowledge Discovery from Data (KDD) can be defined as a process of finding interesting patterns and hidden knowledge from vast dataset without setting prior assumption. If we look gold mining: mining gold from sand or rock is not saying sand or rock mining. Likewise, mining from data is knowledge discovery [9].

Data mining can be applied in any of the data as long as the data is available so that it can be applied in any data like text data, special data, audio data, video data, web data etc. Data mining adopts techniques from many domains like machine learning, statistics, pattern recognition,

visualization, algorithms, high performance computing, applications, information retrieval, data warehouse and database systems [9].

In today's word data mining is applied in different fields like health, business, education, agriculture etc. to mine the hidden knowledge which is essential for competitive advantage [56].

2.4. Data mining Process Model

The objective of data mining technology is construct/build a model going through serious of steps [57]. Different process models from academic and industrial areas with their level of steps discussed below. These are Knowledge Discovery in Databases (KDD), Cross-Industry Standard Process for Data Mining (CRISP-DM), Sample, Explore, Modify, Model, Access (SEMMA) and Hybrid Model [58].

2.4.1. Knowledge Discovery in Databases (KDD)

As presented by [59] KDD is the process of using DM methods to extract what is believed knowledge according to the specification of measures and thresholds, using a database along with any required preprocessing, sub sampling, and transformation of the database. KDD is also the process of discovery knowledge to databases. It is also defined as a non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data [60].

KDD is one of the known academic research data mining model applied in different works [59]. The reason to drive this model was that knowledge considered as the end product of data driven discovery process.

KDD has five phases. These are selection, prepossessing, transforming, data mining and interpretation. The following diagram shows phases of this model.

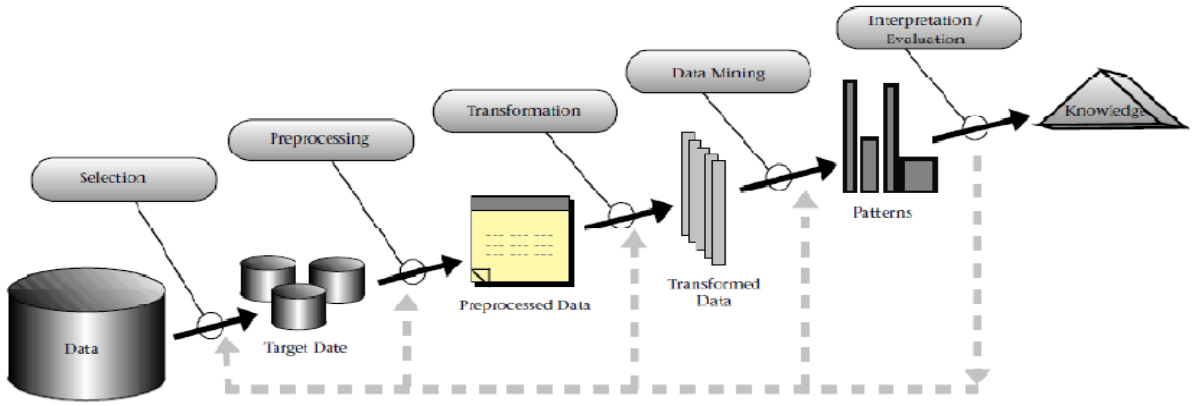


Figure 2: An Overview of the Steps That Compose the KDD Process *[59]*

"The KDD process is interactive and iterative, involving numerous steps with many decisions made by the user" said [59]. Each process steps are discussed below.

Selection: In this stage selecting raw data for discovery is taken place. Unnecessary attributes will be reduced and only few and important attributes are selected. As a result target dataset will be selected.

Preprocessing: In this process, missing values will be handled, noise data and duplicate data are managed.

Transformation: data is transformed or consolidated and make ready for data mining. Smoothing, aggregation, generalization, normalization and feature selection are the main activity of this stage.

Data Mining: Analysis will be done and decision will be made which algorithm achieves the target. Interesting patterns also discovered.

Interpretation: In this final stage the mined result will be interpreted. The results will be interpreted to user form for better understanding. Finally evaluation the result will be done whether the result is accepted by the user or not. [59].

2.4.2. CRoss Industry Standard Process for Data Mining (CRISP-DM)

The CRISP-DM process was developed by the means of a consortium initially composed with Daimler Chryrler, SPSS and NCR. CRISP-DM stands for Cross Industry Standard Process for Data Mining [61].

CRISP-DM is a product neutral data mining model developed by a consortium of several companies. As the name indicates, CRISP-DM is widely applicable in industry areas [60].

CRISP-DM has six sequence of steps which are business understanding, data understanding, data preparation, modeling, evaluation, deployment. Figure 3 shows life cycle of the process model.

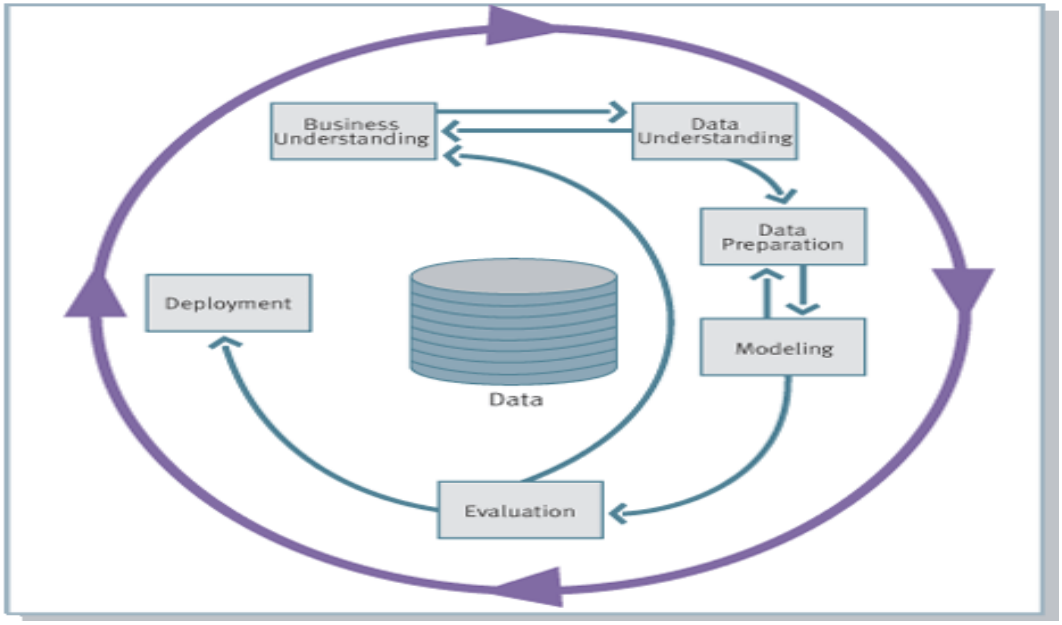


Figure 3: Phases of the CRISP-DM Process Model *[61]*

- **Business understanding** – This initial phase 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.
- **Data understanding** – The data understanding phase starts with an initial data collection and proceeds with activities in order to get familiar with the data, to identify data quality problems, to discover first insights into the data or to detect interesting subsets to form hypotheses for hidden information.
- **Data preparation** – The data preparation phase covers all activities to construct the final dataset from the initial raw data.
- **Modeling** – In this phase, various modeling techniques are selected and applied and their parameters are calibrated to optimal values.
- **Evaluation** – At this stage the model (or models) obtained are more thoroughly evaluated and the steps executed to construct the model are reviewed to be certain it properly achieves the business objectives.

- **Deployment** – Creation of the model is generally not the end of the project. Even if the purpose of the model is to increase knowledge of the data, the knowledge gained will need to be organized and presented in a way that the customer can use it.

2.4.3. The SEMMA Process

SAS institute was develop this SEMMA process. The acronyms stands for **S**ample, **E**xplore, **M**odify, **M**odel, **A**ssess, and refers to the process of conducting a data mining project [61]. In this stage, data mining can be viewed as a process rather than a tool [62].

The SAS Institute considers a cycle with 5 stages for the process:

Sample – This stage consists on sampling the data by extracting a portion of a large data set big enough to contain the significant information, yet small enough to manipulate quickly. This stage is pointed out as being optional.

Explore – This stage consists on the exploration of the data by searching for unanticipated trends and anomalies in order to gain understanding and ideas.

Modify – This stage consists on the modification of the data by creating, selecting, and transforming the variables to focus the model selection process.

Model – This stage consists on modeling the data by allowing the software to search automatically for a combination of data that reliably predicts a desired outcome.

Assess – This stage used to assess the data by evaluating the usefulness and reliability of the findings from the data mining process and estimate how well it performs [61] [62] [63]. The following diagram shows the life cycle of SEMMA process.

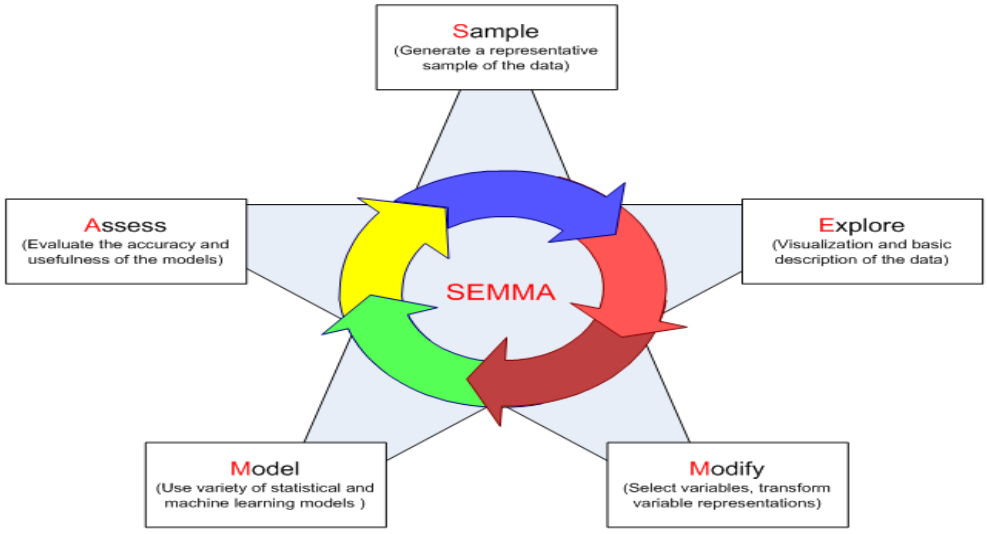


Figure 4: SEMMA Process Model *[62]*

In SEMMA process model, iterative is possible and one can proceed back to the exploration phase for additional refinement of the data.

2.4.4. Hybrid Model

A combination of two or more data mining techniques together built hybrid model due the fact that to use the strength of different classifiers and to improve the performance of the classifiers [64]. According to [65], the development of academic and industrial models has led to the development of hybrid models, i.e., models that combine aspects of both. One such model is a six-step KDP model. It was developed based on the CRISP-DM model by adopting it to academic research. A difference from CRISP-DM is that this model has a general research description steps, includes data mining stage instead of modeling, and has several feedback mechanism. Figure 5 shows the steps of the model. These steps are:

Understanding of the problem domain: In this stage, the highly engaged with the domain experts to define the project goals, to identify key peoples and to learn a solution given to the existing problem. It also involves learning domain-specific terminology. A description of the problem, including its restrictions, is prepared. Finally, project goals are translated into DM goals, and the initial selection of DM tools to be used later in the process is performed.

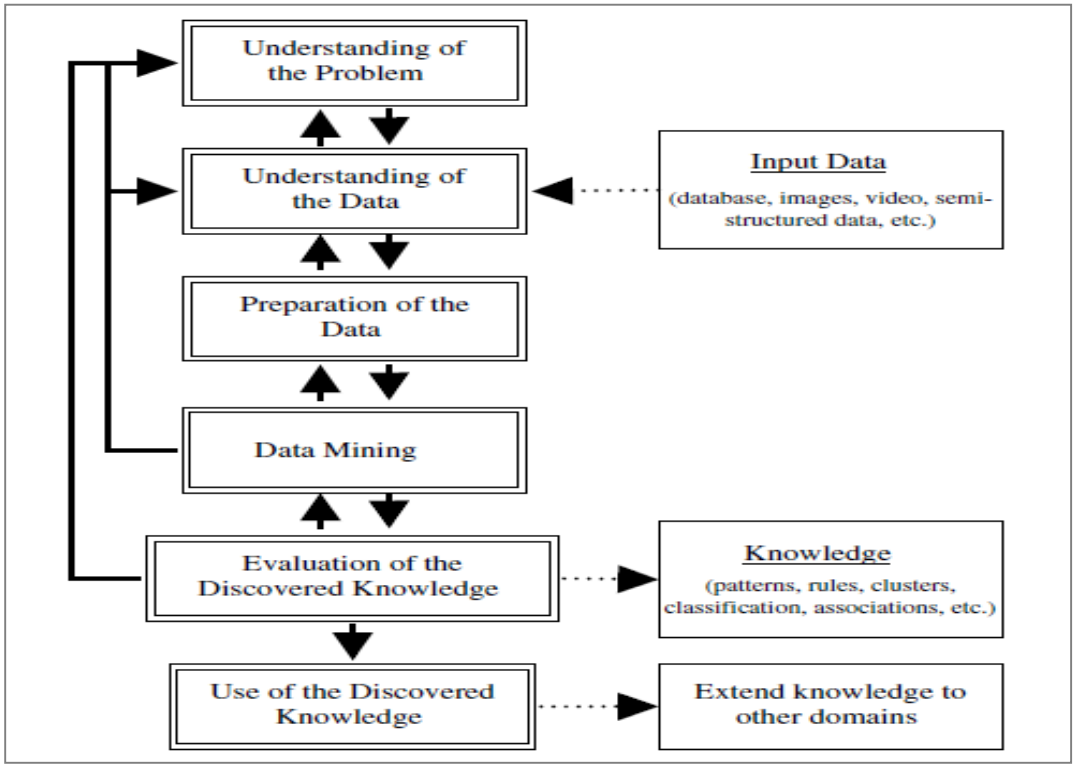


Figure 5: The six steps of Hybrid model *[65]*

Understanding of the data: Sample data collected, decide which data including format and size will be needed. Data completeness, redundancy, missing values, plausibility of attribute values, will be checked. Lastly verification will be done whether the data with respect to DM goals is useful or not.

Preparation of the data: This step focuses 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. The end results are data that meet the specific input requirements for the DM tools selected in the first step.

Data mining: one of the important key steps in the knowledge discovery process is data mining. The main objective of this tool is discovering new knowledge. The process of discovering new information includes: the data model was constructed using one of the chosen DM tools and training and testing procedures are designed. Then generated data model was verified by using testing procedures.

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.

Use of the discovered knowledge: This final step consists of planning where and how to use the discovered knowledge. The application area in the current domain may be extended to other domains. A plan to monitor the implementation of the discovered knowledge is created and the entire project documented. Finally, the discovered knowledge is deployed [62] [65].

Among these data mining models, the researcher used hybrid model for this study. The reason why this model is selected is that hybrid model is a combination of both academic and industry. Other models are applied either in academic/research area or industry area so that peoples from both areas couldn’t understand equally.

2.5. Data Mining Tasks

In data mining, the central idea is that what kind of patterns can be mined. There are two kinds of functions involved in data mining which are descriptive data mining task and predictive data mining task [66].

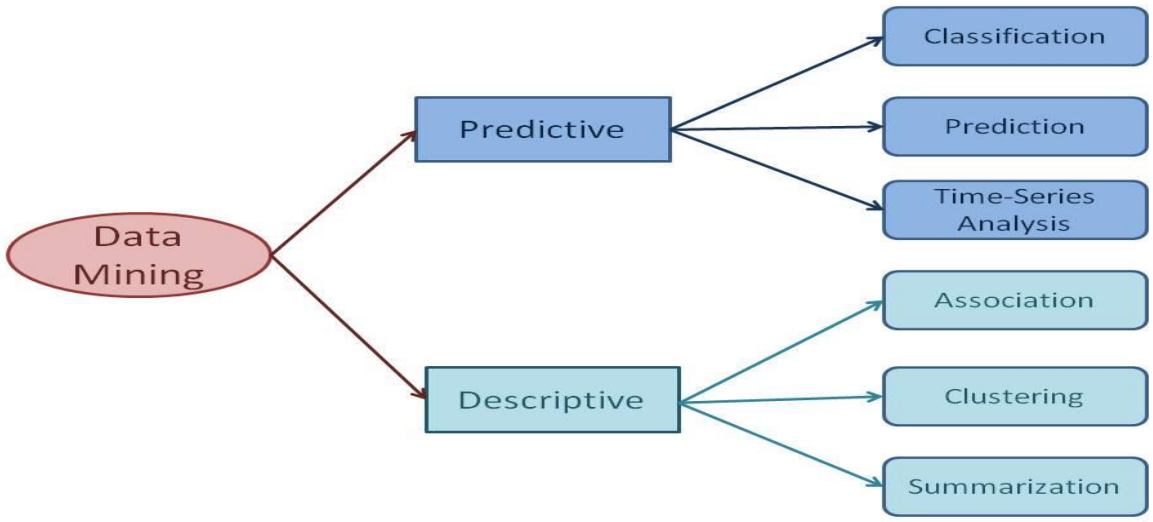


Figure 6: Data mining task and models [67]

[9] Also said that the task of data mining can be grouped as prediction and description. The details of these tasks are discussed below.

Predictive data mining tasks come up with a model from the available data set that is helpful in predicting unknown or future values of another data set of interest. A medical practitioner trying to diagnose a disease based on the medical test results of a patient can be considered as a predictive data mining task [67]. This model also makes a prediction about values of data using known results found from different data and its goal is to identify strong links between variables of a data table. Classification, prediction and time serious analysis are considered as predictive data mining task.

Descriptive data mining tasks usually finds data describing patterns and comes up with new, significant information from the available data set. For example, a retailer trying to identify products that are purchased together can be considered as a descriptive data mining task. In addition, it targets to find different kinds of structures and relations in the data to generate rules

and models that will be used for decision making process. Description is identifying patterns or relations in the data by examine the property of the data itself (characterize the properties of the data in a target dataset). The task of data mining could be developing model, is a high-level description summarizing a large collection of data and describing its important features [67] [66].

In general the major difference between these two models is that a predictive model has the specific objective of allowing us to predict the value of some target characteristic of an object on the basis of observed values of other characteristics of the object. On the other hand, descriptive model serves as a way to explore the properties of the data examined, not to predict new properties [55] [9] [62].

2.5.1. Predictive Model

The aim of predictive modeling is to estimate a function from the training data set that can predict a value y given input variables X_i. The predicted variable is called the response variable and the input variable called explanatory variable. Predictive modeling task can be classification, regression, time serious and prediction analysis based on the data that will be processed.

2.5.1.1. Classification

It predicts the class of objects whose class label is unknown. Its objective is to find a derived model that describes and distinguishes data classes or concepts. The derived model is based on the analysis set of training data i.e. the data object whose class label is well known. Classification is a form of data analysis that extracts models describing important data classes. Such models, called classifiers, predict categorical (discrete, unordered) class labels. For example, we can build a classification model to categorize bank loan applications as either safe or risky. Such analysis can help provide us with a better understanding of the data at large. Classification has numerous applications, including fraud detection, target marketing, performance prediction, manufacturing, and medical diagnosis. Data classification is a two-step process, consisting of a learning step (where a classification model is constructed) and a classification step (where the model is used to predict class labels for given data) [23] [67] [57].

2.5.1.2. Prediction

In this task, based on the available data, prediction involves developing a model and this model is used in predicting future values of a new data set of interest. E.g. a model can predict the income of an employee based on education, experience and other demographic factors and also applicable in health sectors [67] [66].

2.5.1.3. Time Series Analysis

According to [67], time series is a sequence of events where the next event is determined by one or more of the preceding events. Time series reflects the process being measured and there are certain components that affect the behavior of a process. Time series analysis includes methods to analyze time-series data in order to extract useful patterns, trends, rules and statistics. Stock market prediction is an important application of time- series analysis.

Predicative modeling uses different algorithm for its purpose and the most common one are decision tree, decision rules, Naïve Bayes, support vector machine, artificial neural network etc. The brief discussion on this algorithm is explained below.

Decision Tree: is one of the most common algorithms used for data mining classification. Most decision tree methods learn the decision trees by a top down approach, beginning with the question "which attribute should be used to define a partition?" To answer this question, each attribute is evaluated using the information gained to determine how well it alone classifies the training examples. The entire process is then repeated using the training examples associated with each descendant node to select the best attribute to partition with at that point in the tree. This forms a greedy search for a decision tree, in which the algorithm never backtracks to reconsider earlier choices.

A decision tree is a flowchart-like tree structure that has three types of nodes.

- Root node: it has no incoming edges and zero or more outgoing edges.
- Internal nodes: it has one incoming edge and two or more outgoing edges.
- Leaf/terminal nodes: it has no outgoing edges and exactly one edge incoming.

Each leaf node is assigned a class label. The internal node and root nodes contains attribute test conditions to separate records that has different characteristics [68].

There are different decision tree algorithm and the common decision was ID3 (Iterative Dichotomiser) which was developed by J. Ross Quinlan in early 1980s. Later on J. Ross Quinlan developed a successor of ID3 called C4.5. These algorithm adopt a greedy approach in which decision trees are constructed in top down recursive in dive and conquer manner. J48 algorithm is a good example that applies c4.5. In decision tree, over-fitting is the common problem which is happed because of the expected error of the model on previously unseen records. Overfitting occurs when the model does not fit to the future states which is caused by the use of small size and unbalanced training database. In other words, a classification algorithm is said to overfit to the training data if the model is strongly dependent on particularly on a certain feature of the dataset. So, pruning the tree helps to minimize the over fitting problem. Tree pruning is an important point in order to diminish the size of the tree for the reduction of cost complexity and this is done in two techniques pre tree pruning and post tree pruning [9].

REPTree: This also another decision tree algorithm which is fast decision tree learner. It builds a decision/regression tree using information gain/variance and prunes it using reduced-error pruning. Only sorts values for numeric attributes once. Missing values are dealt with by splitting the corresponding instances into pieces like C4.5. Reduced Error Pruning results in a more accurate and simple classification tree, even in cases with large amount of training and testing data [69].

JRIP: JRip uses an algorithm called Repeated Incremental Pruning to Produce Error Reduction. (RIPPER) is one of the basic and most popular algorithms. Classes are examined in growing size and an initial set of rules for the class is generated using incremental reduced error. JRip (RIPPER) proceeds by treating all the examples of a particular decision in the training data as a class, and finding a set of rules that cover all the members of that class. Thereafter it proceeds to the next class and does the same, repeating this until all classes have been covered [70].

RIPPER is especially more efficient on large noisy datasets .There are two kinds of loop in Ripper algorithm. This algorithm was designed by Cohen in 1995 namely, Outer loop and Inner loop. Outer loop adds one rule at a time to the rule base and Inner loop adds one condition at a time to

the current rule. The information gain measure is maximized by adding the conditions to the rule. This process is continued until it covers no negative example [71].

PART: is also another decision rules that uses separate-and-conquer. It builds a partial C4.5 decision tree in each iteration and makes the "best" leaf into a rule. The algorithm producing sets of rules called „decision lists" which are planned set of rules. A new data is compared to each rule in the list in turn, and the item is assigned the class of the first matching rule [71].

Naïve Bayes: Naïve Bayesian Classification is unsupervised classification technique which is based on the theory of Bayes. The Naive Bayes classifier is a simple probabilistic classifier based on applying Bayes Theorem with strong independence assumptions which assumes all of the features are equally independent. It uses a Bayesian algorithm for the total probability procedure, the principle is according to the probability that the text belongs to a category of prior probability, and the text would be assigned to the category of posterior probability. In simple terms, a naïve Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature. Naïve Bayesian classifiers assume that the effect of an attribute value on a given class is independent of the values of the other attributes and this assumption is called class conditional independence [71].

2.5.2. Descriptive Model

Descriptive mining tasks describe properties of the data in a target data set. Description model creates a concise and convenient representation of a given data set, which is called descriptive model. Descriptive modeling helps us to look into the important aspect of the data. It describes a given set of task-relevant data in a concise and summative manner, presenting interesting general properties of the data.

These descriptions can be derived by the following two ways:

- Data Characterization − this refers to summarizing data of class under study. This class under study is called as Target Class. It describes a given set of task-relevant data in a

concise and summative manner, presenting interesting general properties of the data. Concept (or class) description consists of characterization and comparison (or discrimination). The former summarizes and describes a data collection, called the target class, whereas the latter summarizes and distinguishes one data collection, called the target class, from other data collection(s), collectively called the contrasting class [10] [9].

- Data Discrimination − It refers to the mapping or classification of a class with some predefined group or class [10] [9].

Clustering, association rule, summarization and sequence discovery are examples of descriptive model.

Clustering: is the process of partitioning a set of data objects (or observations) into subsets. Each subset is a cluster, such that objects in a cluster are similar to one another, yet dissimilar to objects in other clusters. The set of clusters resulting from a cluster analysis can be referred to as a clustering.

Clustering consists in identifying groups for entities that have characteristics in common and are cohesive and separated from each other. Interest in clustering has increased due to several applications in distinct knowledge areas. Highlighting the search for grouping of customers and products in massive datasets, document analysis in Web usage data, gene expression from microarrays and image analysis where clustering is used for segmentation. In this context, different clustering methods may generate different clustering on the same data set. The partitioning is not performed by humans, but by the clustering algorithm. Clustering analyzes data objects without consulting class labels unlike of classification. It can be used to generate class labels for a group of data. The objects are clustered or grouped based on the principle of maximizing the intra-class similarity and minimizing the interclass similarity.

Association Rules: is the form of unsupervised learning. The main objective of the association rule is to check if there are common relationships between objects. It also applied in both categorical and numerical data. Association rule mining consists of first finding frequent item sets (sets of items, such as A and B, satisfying a minimum support threshold, or percentage of the task-relevant tuples), from which strong association rules in the form of A=>B are generated. These

rules also satisfy a minimum confidence threshold (a pre- specified probability of satisfying B under the condition that A is satisfied). Associations can be further analyzed to uncover correlation rules, which convey statistical correlations between item sets A and B.

The objective of the data mining that the researcher used is to predict the choice of contraceptive method, the researcher used predictive data mining task to classify it using classification algorithm.

2.6. Related Works

The researcher reviewed different paper those employed data mining in family planning dataset. To the knowledge of the researcher, there are very limited paper studied on the application of data mining in family planning. The following are papers reviewed for the purpose of this study.

Abrham Gebregiorgis [23] has done his research on application of data mining to predict the likelihood of contraceptive users and non-users of women on EDHS 2005 dataset. He has used decision tree J48 algorithm and Naïve Bayes algorithm to predict the likelihood of contraceptive use among reproductive age group (15-49). Classifying contraceptive user and non-contraceptive user group on the basis of variable such as region, type of place, religion, visited by family planning worker, exposure to family planning message, Education, Number of living children, Knowledge of family planning, Marital status, Partner's Education, Partner's Occupation, respondent's Occupation, wealth index, Respondent's age and current contraceptive use. He conducted 8 different experiments through applying J48 default parameters, minNumObj, percentage split, pruned and unpruned decision tree, first balancing data by SMOTE, Naïve Bayer default and naïve Bayes percentage test [23].

[23] Result showed that among eight experiments, he proved that J48 with pruned tree algorithm is best due to its accuracy, number of leaves, and size of the tree. He extracted 13 rules that predict the likelihood of family planning users among reproductive age group. He conclude that the variables knowledge of any method, partner occupation, partner's education level, current marital

status, wealth index, type of place, family planning massage, number of living children, religion, education level were found to be the most determinant attributes of contraceptive method use.

The result also describes that a woman who doesn't know any contraceptive method hasn't any chance of using any method. Knowledge about a family planning couldn't be the only reason which makes women to use any method of contraception. The other point is that a rural or urban woman who are not married, have no child and their partner is agriculturalist are less likely to use family planning. On the other point, a woman aged 24-34 who are rich and whose partner are a non-agriculturalist and living together have a high probability of using contraception. Poor women have low likelihood of utilizing any contraceptive method; i.e. may be because poor woman have low exposure to FP message and have low education compared with the woman who are wealthy [23].

According to Vincent Lemaire et al. [72] has used Naïve Bayes technique on Indonesia data set to predict the contraceptive method choice (no contraceptive method, short term contraceptive method or long term contraceptive method) using the explanatory variables; age, education, husband's education, number of children ever born, religion, working or not, husband's occupation, working or not, standard of living index, FP message). By selecting variables which was considered as possible targets for policies experimented to increase the probability of one class i.e. to increase the probability of using a short term contraceptive and the probability of using a long term contraceptive.

Applying the method to increase the probability of using a long term contraceptive showed that the most significant variable is the education level. Out of 1473 instances, 577 instances were already at a higher education level. Out of the remaining 895 instances, 99 were predicted to switch from no contraceptive to a long term contraceptive. If the education level was changed from whatever value (low or middle) to a high value, and 230 instances were predicted to switch from short term contraceptive to long term contraceptive with the same change in education level. Media exposure could not make significant impact (only 2 instances changed to long term contraceptive, by changing the media exposure to good media exposure). Applying the method to increase the probability of using a short term contraceptive, 157 instances were predicted to switch

from no contraceptive to short term contraceptive with a higher education, and 18 with change to good media exposure.

There was also a paper done in India entitled with “Contraceptive use in India: A data mining approach” [73]. In this paper, the researcher uses data mining approach to analyze patterns of contraceptive use in India by comparing contraceptive use among groups of women with distinct demographic, economic, cultural, and social characteristics. PART rule predicts the best of contraceptive users’ dataset. The analysis suggests that currently married, non-pregnant women aged 15–49 years in India can be classified into 13 mutually exclusive groups on the basis of six characteristics of women—surviving children, household standard of living, religion, women’s years of schooling, husbands’ education, and residence. Contraceptive use pattern in these 13 groups is essentially different and reflects the orientation of family planning efforts, especially, official family planning efforts in the country [73].

As observed in the above literatures, none of them addresses choice of contraceptive method rather predicts between user and non-user of contraceptive method. These studies result end with identifying best classification algorithm and determinant variables of the dependent class. In addition, none of them couldn’t integrate with knowledge based system.

Abdulkerim [37] has done his thesis work through integrating data mining with knowledge based system, the case of network intrusion detection. According to this thesis work, knowledge discovery process approach is employed. JRIP model was best predictor and generate the rule so as to ingrate with the knowledge based system automatically. Rule based knowledge based system method was used to develop the KBS. Though this research is not directly related with my study, it tells us how data mining results are integrated with knowledge based system. It also tells us the extent to which data mining results are converted into KBS and will support our day to day activities.

Table 2 showed summary of related works reviewed by the researcher.

Table 2: Summary of related works

No	Researcher	Objective	Model used	Output of the result	Remark
1	Abraham	Predicting likelihood of contraceptive use	J48	knowledge of any method, partner occupation, partner's education level, current marital status, wealth index, type of place, family planning massage, number of living children, religion, education level	
2	Vincent Lemaire et al.	predict the contraceptive method choice (no method, short term or long term contraceptive method)	Naïve Bayes	Education level, media exposure	
3	Aalok Ranjan Chaurasia	Contraceptive use in India: A data mining approach	PART	surviving children, household standard of living, religion, women's years of schooling, husbands' education, and residence	
4	Abdulkerim	Integrating data mining with KBS: Network intrusion detection	JRIP	Network behavers like duration, repeated logins, destination host name, guest login, protocol type classify the network as normal, Probe, U2R, R2L	

To the knowledge of researcher, application of data mining in contraceptive use and knowledge based system for it is very limited. So, the researcher reviewed the above related works literature.

Chapter Three

Research Design and Methods

In this chapter, the research design and methodology used to achieve the research goals of this study discussed. The research goal of this paper is to develop knowledge based system that determine choice of contraceptive method match knowledge based system through acquiring knowledge from life style, medical eligibility criteria and application of data mining techniques. Here below, the research design and methods are discussed in detail.

3.1. Research Design

A research design is a plan used as a guide in collecting and analyzing research data for the study to be conducted. It describes the methods used to collect and analyze the data that helps to answer the research question [74]. For this research, among the different research designs, empirical research design is used to explain the research and answer the research questions. This empirical study combines both experimental and non-experimental researches. Unlike statistical studies which has both experiment and control group, this study is data mining approach which experiment the predicting models which classifies best of the choice of contraceptive method. Non experimental study was also followed to acquire knowledge from documents and domain experts.

3.2. Research Methods

Prototyping approach is followed to develop the knowledge based system. Prototyping is an iterative process of quickly building an experimental system, for demonstration and evaluation so that users can dynamically determine their information requirements and explore and test the design of the system [75]. Prototyping allows participating users and domain experts for evaluating systems performance and efficiency. Since domain experts were highly involved during problem understanding, data understanding, rule selection, system evaluation. Hence the researcher used this prototyping approach.

As research process model, knowledge engineering is followed in the process of knowledge based system development. Knowledge Engineering can be defined as it is an engineering discipline that involves integrating knowledge into computer systems in order to solve complex problems normally requiring a high level of human experts [76]. Knowledge Engineering steps to follow in order to accomplish successful knowledge based system are discussed below.

3.2.1. Knowledge acquisition

For the purpose of this study, knowledge is acquired from data mining, domain experts and documents. Data mining techniques, primary and secondary data collection methods were employed to acquire knowledge. Acquired knowledge from this sources are discussed as follows

3.2.1.1. Knowledge acquisition using data mining techniques

To acquire knowledge from data mining, data mining knowledge process model was employed. In data mining there are different activities done for the purpose of optimal and desired outcome. Since the objective is finding a hidden pattern which is a knowledge discovery, the researcher employed the Hybrid Knowledge Discovery in Databases (HKDD) process which includes the understanding the problem, understanding the data, preparation the data, mining the data, evaluation of the discovered knowledge and use of the discovered knowledge. HKDD is the process model to find useful information and patterns in the database which starts from business understanding and data understanding. The reason why the researcher select HKDD is that HKDD is the combination of both the academic and industrial knowledge process models [59]. This is to mean that, both sectors can understand the result of this study and will be easy to implement in the real world. This Hybrid KDD also gives a chance to the researcher to know the business and data understanding of the organization that as key role during the development of the knowledge based system.

Source of data mining dataset: for this study, the researcher was used dataset from Marie Stopes International Ethiopia (MSIE) data warehouse. The reason why MSIE's data preferred is that:

- MSIE works on family planning and sexual reproductive health services in the last 26 years in the country.

- It is a pioneer international organization providing family planning and sexual reproductive health (SRH) services in Ethiopia more than a quarter of century so that able to get huge instances.
- MSIE record clients' data electronically and has a central warehouse.
- Its intervention area is throughout Ethiopia except Somali, Gambella, Benishangul-Gumuz and Harari region
- The current method mix of MSIE is higher than the national average.

This dataset is a serious of data recorded between 2015 to 2016 that receive contraceptive methods from 24 MSIE clinics which are located in Tigray, Amhara, Oromia, South Nations Nationalities Peoples Representative (SNNPR) regions and Addis Ababa and Dire Dawa city administrations.

Using this dataset, the hybrid knowledge discovery in databases process model steps has done to discover novel/interesting pattern that determine choice of contraceptive methods. Activities done in each step has discussed below.

- **Understanding the problem**: learn the organization business and work closely with the domain experts to define the problem was done. In addition, determining the central idea of the problem will lead to the data mining goal so that identification of data mining tools and techniques which will be applied later also the important aspect of this stage. To understand the problem, the researcher reviews literature how the current uptake of contraceptive use looks like. In addition, the researcher conduced open discussion with domain expert of MSIE to define the problem. The researcher learnt that the current practice of counseling and providing contraceptive method is using charts and wheels which focus on the medical eligibility criteria of a woman. This luck the socio demographic characteristic as well as lifestyle of a woman.
- **Data Understanding**: Once the problem is defined, the researcher with domain experts, open discussion on each attribute has been taken place to understand the values, the purpose of the data element etc. The data was given to me in Microsoft excel comma separated values (CSV) file format. In order to understand the landscape of the data, descriptive statistics was employed using Microsoft excel. In addition, from all attributes in the dataset, only relevant data related to this study was selected with the help of domain experts and reviewed literatures.

- **Data Preprocessing**: the extracted data from the database might get inconsistent, incorrect, incomplete values. The data preparation is the main activity to be done in data mining activity which covers 80 percent of the total work. Different techniques were employed based on the data that lacks. In this stage, the researcher has done data cleansing and data transformation and feature extraction as well as data reduction.
- **Data Mining:** Once data preprocessing has done, the researcher converted the dataset from CSV file format into ARFF file format using Waikato Environment for Knowledge Analysis (WEKA) software. Weka is a collection of machine learning algorithms implemented in java so that experiment has done using this software. The reason why the researcher select weka for data mining application is that Weka is open source and widely applicable software in data mining [77]. Five scenarios was experimented to predict choice of contraceptive methods using different classification algorithms of J48, REPTree, PART, JRIP and Naïve Bayes. .
- **Evaluation of the Discovered Knowledge:** Evaluation of models were conducted using performance measurement criteria (accuracy, true positive rate, false postive rate, time taken, ROC curve, recall, precision) As a result, best predictor algorithm was identified. Finally, evaluating the discovered knowledge is novel or interesting was also done in this stage.
- **Use the discovered knowledge:** this novel discovered knowledge was acquired and rules were generated. The researcher selected interesting rules using the accuracy of predicted class node (e.g. IUCD (290/10), the prediction accuracy of this rule is 290/300=97%) . Finally, selected rules has taken a discussion with domain experts, how match they fit with the real world.

3.2.1.2. Knowledge acquisition from documents

As literatures showed that choice of contraceptive method is not only determined by socio demographic variables but also medical eligibility criteria [78] and life style of a woman. Since family planning dataset contains only socio demographic attributes, document analysis was also carried out to acquire explicit knowledge from different source of knowledge on the area of family planning method choice preference. The researcher adopt the World Health Organization

(WHO) medical eligibility criteria manual to identify variables that determine choice of contraceptive use [78].The reason why the researcher adopt this WHO manual is that health facilities use this manual during their counseling session as well as at the time of contraceptive method provision.

3.2.1.3. Knowledge acquisition from domain experts

The acquired knowledge from data mining and document analysis still not sufficient to determine choice of contraceptive method. Thus, acquire knowledge from domain experts was required. Structured interview was conducted to elicit implicit knowledge from domain experts.

To elicit the required knowledge from these domain experts, purposive sampling technique was employed. Purposive sampling would be the best way to elicit the views of persons who have expertise and knowledge about specific domain [79]. Accordingly, employing purposive sampling in this study is its appropriateness to capture demonstrable experience and expertise of the experts. Thus domain experts from Marie Stopes international Ethiopia Quality and Clinical Training department team were selected purposively according to their working experience in family planning and level of education, current roles in MSIE irrespective of age and sex.

Interview was carried out with face to face communication. During face to face communication the information obtained from experts was recorded manually.

3.2.2. Knowledge Modeling and Representation

The acquired knowledge need to be modeled since it helps the knowledge engineer ease understanding of the acquired knowledge. Knowledge acquired from data mining is modeled by the data mining tool, weka. Since the selected model was J48, acquired knowledge was modeled by decision tree. Likewise, the researcher used decision tree to model knowledge which are acquired from domain experts and documents. Since decision tree is easy to understand and to make it similar with data mining result modeling.

Once the required knowledge is acquired and modeled from the data mining and other sources, knowledge representation has done. Among many representation, rule based representation method was implemented to represent expert and document knowledge.

The rule based approach is particularly useful because in many domains much of an expert's knowledge is acquiescent to expression in if-then rules, many of which are empirical. In addition, data mining results are also generated in the form of rules. The rule-based knowledge representation has many advantages compared with other knowledge representation methodologies. The reason for the choice of rule based reasoning is that this method is common one and it can be gratifyingly powerful from the perspective of building useful applications. Moreover, the experience and knowledge of domain experts are captured in the form of IF-THEN rules [48] [80].

3.2.3. Knowledge codification

Rules represented with IF-Then rules were converted into machine understandable form. The researcher used Microsoft visual basic.Net programming langue to codify the represented rules. The researcher used this language is that the researcher is familiar with this language. Finally KBS system has developed which has user interface, knowledge base, inference engine and explanation module.

.

3.3. Testing and Evaluation

Performance evaluation of the data mining algorithm was measured using the precision, recall and accuracy, ROC Area, F-measure, true positive rate and false positive rate. Using this measurement, the best model was selected and rules were generated from this model [81].

In order to select best models in data mining, performance evaluation parameters result are calculated as follows. Assuming that the model has 2x2 confusion matrix, confusion matrix is a table used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known [81].

Table 3: A 2x2 Confusion matrix

	Predicted class		
Actual Class	C(i,j)	a(Yes)	B(No)
	A (Yes)	TP	FN
	B (No)	FP	TN

Keys: TP: True Positive; FP: False Positive; TN: True Negative; FN: False Negative

- TP rate (Sensitivity/Recall) = TP/ (TP+FN): These are cases in which predicted yes and actually it is yes. When it's actually yes, how often does it predict yes can be considered as TP rate.
- TN rate (Specificity) = TN/ (TN+FP): The prediction is no and actually don't have it. When it is actually no, how often it predicts no can be considered as TN rate.
- Precision = TP/ (TP+FP): When it predicts yes, how often is it correct?
- FP rate = FP/ (FP+TN) = 1-TN: When it's actually no, how often does it predict yes?
- FN rate = FN/ (FN+TP) =1-TP: predicted as no but actually as yes group.
- Accuracy = (TP+TN)/ (TP+TN+FP+FN): overall, how often the classification is correct.
- F-measure=2*Precision*Recall / (Precision + Recall): s a measure of a test's accuracy and is defined as the weighted harmonic mean of the precision and recall of the test.
- ROC Area= 100-specificity: is a commonly used graph that summarizes the performance of a classifier over all possible thresholds.

In order to assure that the system achieves its objective, evaluating the KBS is an important task. So, system performance measurement and user acceptance test has done to test and evaluate the designed knowledge based system.

To test the system performance, the researcher prepared test cases from the dataset and other sources and provide to domain experts to fill the dependent class. At the same time, these test cases are provided to the developed system. Finally, confusion matrix has developed to evaluate the performance of the KBS.

At the same time, questionnaire was developed to test the user acceptance. Descriptive statistics has done to analyze the questionnaire result.

Chapter Four

Knowledge Acquisition Using Data Mining and Other Sources

Knowledge is acquired either from peoples (implicit knowledge) or from other external sources like documents which are usually codified knowledge (explicit knowledge). In addition, knowledge is also acquired from large data set using data mining techniques. In this paper, the researcher designed and developed knowledge based system that supports the family planning service provider to choose contraceptive method to the client. To develop the KBS, the researcher describes step by step process how knowledge is acquired from these sources.

4.1. Knowledge Acquisition Using Data mining

Since the researcher used the hybrid data mining model to discover the hidden knowledge, a step by step process what the researcher has done is discussed below.

4.1.1. Understanding the Problem:

Understanding the problem is the first valid step in the data mining process to understand the problem that we solve. In other words, the goal of the data mining process depends on the type of problem to be solved using data mining technology. In this step, the researcher talked about the institution that data was collected and the problem domain of the research area.

Marie Stopes International Ethiopia (MSI Ethiopia), which is part of MSI, is registered as a 'Foreign Charity' with the Charities and Society's Agency of the Federal Democratic Republic of Ethiopia. Established in 1990, it has been providing family planning (FP) and other SRH services for the last 26 years. MSI Ethiopia has operation area of a network of 24 clinics, 10 outreach teams, 400 franchise clinics throughout Ethiopia these services [82]. In MSIE, nearly 1 million women got family planning services in 2016 through these operation area. Clients visited MSIE clinics are recorded using electronic software called Open Electronic Medical Records (OpenEMR) and Electronic Health Record (EHR). Service providers at MSIE use the medical eligibility criteria to counsel and choose a contraceptive method. These clients socio demographic data are recorded electronically. Hence, the researcher used these electronic data to discover the hidden knowledge.

According to EDHS 2016, the total fertility rate is 4.6 children per woman in Ethiopia and 37% of the currently married woman don't want more children. Likewise, 35.3 percent of married women want space at least for two years for another baby. Eight percent of birth is unwanted. Among currently married women, modern contraceptive users are 35 percent. Among this 35 percent contraceptive users, 23 percent uses injectable and 8 percent are implant users [3]. This high fertility rate and low CPR as well as skewed method, implying further rapid population growth in the years ahead which requires quite a streamlined activity to increase the CPR of the country and also avail a method mix. So, identifying key determinant factors on uptake of contraceptive methods and translate the data mining rules into KBS system was the problem domain of this research.

So, identifying key determinant factors on choice of contraceptive methods and translate the data mining rules into KBS system was the problem domain of this research.

4.1.2. Data Understanding:

Data understanding phase mainly focuses on creating a target dataset with selected sets of variables that is relevant to the discovery process. The more the size of the data and the more multiple and heterogeneous source, the less the predictive performance of a model [64]. So, understanding the data well play a vital role in the knowledge discovery process.

The data source of this research is taken from Marie Stopes International Ethiopia data warehouse. All these services were provided in MSIE clinics which are located in 17 major towns of the country (in four major regions and 2 city administrations). Years of 2015, 2016 dataset were considered for this research due to that MSIE has rollout electronic system to all clinics in 2015. The data is extracted from the data warehouse were Microsoft excel format. MSIE uses two electronic client data recording system for Sexual and Reproductive Health (SRH) Clinics (these clinics are provide only FP and other sexual reproductive health services) and Maternal and Child Health Clinics (MCH, these clinics provide FP services as well as any gynecological services including delivery). These are open Electronic Medical Record (openEMR) client recording system and Electronic Health Record (EHR) system. From these two systems, a total of 105,206 instances were found.

The dataset has been described and visualized using Microsoft Excel file format to examine the properties of the dataset relative to the whole records. Simple statistical analysis has been performed to verify the quality of the dataset such as missing values, error values and to obtain high level information regarding the data mining questions.

MSIE family planning client record data set contains 32 attributes. Before selecting relevant attributes for this study the researcher lists all attributes. As a result, table 4 showed all attributes and which will be refined during discussion with domain experts.

Table 4: Attributes before data preparation and preprocessing

Attribute	***Description***
Age	Age of Client who when she receives modern family planning
Education	Highest Education level of the client
Sex	Biological sex of the client
Marital Status	Marital Status of the client
Religion	Religion of the Client
Region	From which region the client visit MSIE facility
Residence	Current living place of client
Number of Children	Number of living children of the client
FTEU	First Time Ever User Client : If the client has an experience on Family planning or not
Last Method Used	Any modern contraceptive client used before received today's FP method
Current Method Used	Family planning method used by the client during her visit
Visit date	Date of service provided
Clinic Name	Name of Clinic where client receive the service
Clinic code	Unique code of the clinic
Psuedomized Client ID	For client confidentiality, actual unique ID of the client is changed into another number
Service Group description	Services provided in clinics are grouped within their respective domain
Service type group	Describes the service provided as service and product
FP provided?	Does the client receive family planning service?

Service Billing Status	Describes the clients visit status, as open or closed.
Service Free / Paid ?	Describes whether the client receive family planning paid or free
Woreda	From which woreda/district the client comes from
Sex	Biological sex of the client
Age	
Year	Year when client receive the contraceptive method
Month	Month when client receive contraceptive method
Visit Category	Describes the client is new to MSIE or not
How did you hear about us ?	Describes source of information where the client heard about MSIE services
Follow up description	Describes family planning checkups and follow ups
Referred by Description	Describes who is referring for the service if the client is referral
CAF Reference	Reference number for referral clients by CAF group
Call Center Reference	Reference number who are referred from call center
Area Office	Describes clients location based on MSIE geographic classification
Filed office	Describes also the sub location on MSIE classify their operation areas

The researcher conducted a discussion with domain experts and reviewed literatures to identify relevant attributes which have contributing factor for the choice contraceptive methods. Based on these, relevant attributes used for model buildings are described in details below.

Age: This attribute states that the age of the client at the state of receiving family planning and the value is Numeric

Education: this attribute describes the highest Education level of the client at the state of receiving family planning: The value is nominal {none/Some primary, comprehensive primary completed, secondary school level, Tertiary/higher education}

Marital Status: This also states that the marital status of the client while receiving family planning method. The value of the attribute is nominal {Single, Married, Divorced, and Widowed}

Religion: it describes the religion of the client at the state of receiving family planning. The value is nominal {Christian, Muslim, and Other}

Region: This states that from which region the client comes from. The value is nominal {Addis Ababa, Amhara, Diredawa, Oromiya, SNNPR, and Tigray}

Residence: this attribute states that current residence of client. The value is nominal {Rural, Urban}

Number of living Children: this attribute showed that the number of children of the client at the state of receiving family planning. The value is Numeric.

FTEU: This attribute showed that whether the client has an experience on using family planning before or not. The value is numeric {if the client has no experience before 1 and if not 0}

Last Method Used: This attribute states that any modern family planning used before receiving today's method. The value is nominal {IUCD, Implant, Injectable, Pills, Has not currently using but used before, has never used before}

Current Method Used: This attribute describes that the method received by the client at the state of her visit. The value is nominal ad {IUCD Coper T 386,IUCD LNG, Implant implanon, Implant-Sino, Implant-Jadelle, Injection-depo, Contraceptive Pills, Tuba-ligation }

4.1.3. Data Preparation and preprocessing:

Data preparation is the crucial stage of data mining process and time consuming stage of data mining process. Data cleaning, transformation, reduction etc. has been done in this stage.

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 [9]. Hence, handling errors will increase the efficiency of data mining algorithm. Table 5 shows incorrect value of instance which are considered as key activities done in this stage.

Table 5: Incorrect Values in the dataset to be cleaned.

Sr. No	Name of Attribute	Reason Incorrect instance	Number of incorrect instance	Percent of incorrect instance
1	Current Method Used	IUCD client as Male	10	0.07%
2	Current Method Used	Injectable Client as Male	16	0.05%
3	Current Method Used	Implant Client as Male	11	0.03%
4	Age	124 not real age	1	0.002%

In order to make correct the above table, all sex related issues are manually changed into female since all contraceptives except vasectomy are fabricated for females. In the meantime, I was asked the MIS officer of MSIE why 124 age was happened. This was typing error, rather than inputting year as 1992, it was typed as 1892. So, the researcher changed as 24 years old.

In addition to these incorrect values, there was also missing values. Missing values mean the value of one or more instances doesn't exist in a given attribute. Here are also missing values in the dataset.

Table 6: Handling missing values of an attribute

Sr. No	Name of Attribute	# of missed instance	Percent of missed instance	Expected value of instance	Technique applied
1	Education	10	0.01%	None/some primary -2 Com-primary 2 Secondary-3 Tertiary-3	Most frequent instance (mode) by grouping the value of other attributes
2	Marital Status	3	0.003%	Married 2, single 1	
3	Region	3	0.003%	Oromiya 2 and Addis Ababa 1	

Using mode, the most frequent instance can be taken as the value of the attribute.

4.1.4. Data Transformation and Reduction

One of the process in data transformation is discretization. Discretization can be defined as where methods are used to reduce the number of values for a given continuous attribute by dividing the range of the attribute into intervals [83]. Hence, the data are transformed into forms appropriate for mining by performing aggregation operations.

In order to make the analysis manageable and cost efficient, it is better to reduce the data. Data reduction techniques include a data discretization technique which is used to reduce the number of values for a given continuous attribute by dividing the range of the attribute into intervals. The researcher had done data transformation and reduction in 3 attribute values of dataset. Table 7 showed list of attributes with their discretized and transformed value.

Table 7: Discretized attributes with values

Attribute Name	Old Value	Discretized Value	Transformed Value
Age	Continues value from 15 to 49	15-19 20-24 25-29	Adolescent Early Youth Youth

		30-49	Adult
Current Method Used	IUCD Coper T 386 IUCD LNG Implant implanon Implant-Sino Implant-Jadelle Injection-depo Contraceptive Pills Tuba-ligation	IUCD Implant Injection-depo Pills Tuba ligation	
Number Living Children	values from 0 to 10	0 1-2 3-4 5 and above	No Children Low family size Medium Family size High family size

Source for discretized value is EDHS 2016

In table 7, client's age and number of living children are discretized into four groups and transformed into nominal values. To discretize the attributes, the researcher adapted EFDR client socio demographic variables categorization [84] [3]. Adapting this categorization will help the developed KBS more acceptable in real word. Moreover, the current method used nominal values are discretized into five nominal values. These values are changed with the consultation of domain expert from MSIE.

4.2. Data Format conversion for Weka Software

Like most application soft wares, weka software also understands its own extension. The default file extension accepted by Weka software is ARFF (Attribute Relation File Format). In addition, weka also accept an input with Comma Separated Values (CSV) file format. The data which is cleansed and transformed was in excel format. The researcher also saved this xlsx file format into CSV file format. Since weka accepts this CSV file and converted into an ARFF file format automatically. Finally, the researcher used 10 attributes with 105,205 instances that are ready for experimentation process. Table 8 below shows the final processed attributes with their data values which are ready for experiments.

Table 8: Summary of Preprocessed attributes

Attribute	***Description***	***Attribute Type***
Age	Age of Client who when she receives modern family planning	Nominal {Adolescent, Early youth, Youth, Adult}
Education	Highest Education level of the client	Nominal {None/Some Primary, Comp primary, Comp Secondary, Tertiary/Higher}

Marital Status	Marital Status of the client	Nominal {Single, Married, Divorced, Widowed}
Religion	Religion of the Client	Nominal {Christian, Muslim, Other}
Region	From which region the client visit MSIE facility	{Tigray, Amhara, Oromia, SNNPR, Addis Ababa, Dire Dawa}
Residence	Residence of the client	Nominal {Urban, Rural}
Number of Children	Number of living children of the client	Nominal {No Children, Low Family Size, Medium Family Size, High Family size}
FTEU	First Time Ever User Client : If the client has an experience on Family planning or not	Nominal {Yes, No}
Last Contraceptive Method Used	Any modern contraceptive client used before received today's FP method	Nominal {Has Never Used before, Not currently using but used before, Long term user, Short term user}
Current Method Used	Family planning method used by the client during her visit	Nominal {BTL, IUCD, Implant, Injection-Depo, Pills}

Here is below also the sample ARFF file format that the data mining algorithm is used for classifying.

Table 9: Sample ARFF used for classification

@relation 'Family Planning User'
@attribute Region {Tigray,Oromia,'Addis Ababa',Amhara,SNNPR,'Dire Dawa'}
@attribute 'Last Method Used' {'Has Never Used Before','Not currently using but has before','Short term user','Long term User'}
@attribute FTEU {Yes,No}
@attribute Residence {Urban,Rural}
@attribute 'Age group' {'Early Youth',Youth,Adolecent,Adult}
@attribute Education {'None/some prim',Tertiary/higher,'Comp primary','Comp secondary'}
@attribute religion {Chirstian,Muslim,Other}
@attribute 'Marital Status' {Married,Single,Divorced,Widowed,single}
@attribute 'Number of Children' {'Medium Size Family','High Family Size','Low Family Size','No Children'}
@attribute 'Current Method Used' {Implant,'Injection - Depo',IUCD,Pills,BTL}
@data
Tigray,'Has Never Used Before',Yes,Urban,'Early Youth','None/some prim',Chirstian,Married,'Medium Size Family',Implant
Oromia,'Has Never Used Before',Yes,Urban,Youth,'None/some prim',Chirstian,Married,'High Family Size',Implant
'Addis Ababa','Not currently using but has before',No,Urban,Youth,'None/some prim',Muslim,Married,'High Family Size','Injection - Depo'

Amhara,'Not currently using but has before',No,Urban,'Early Youth','None/some prim',Chirstian,Single,'Medium Size Family',Implant
SNNPR,'Not currently using but has before',No,Urban,'Early Youth',Tertiary/higher,Chirstian,Married,'Medium Size Family',IUCD
Oromia,'Not currently using but has before',No,Rural,Youth,'None/some prim',Muslim,Married,'Medium Size Family','Injection - Depo'
Oromia,'Has Never Used Before',Yes,Rural,Adolecent,'None/some prim',Muslim,Married,'High Family Size','Injection - Depo'
'Addis Ababa','Not currently using but has before',No,Urban,'Early Youth','Comp primary',Muslim,Married,'Medium Size Family','Injection - Depo'
'Addis Ababa','Not currently using but has before',No,Urban,'Early Youth','Comp primary',Chirstian,Married,'High Family Size',Implant
SNNPR,'Not currently using but has before',No,Rural,'Early Youth','Comp primary',Chirstian,Married,'High Family Size',Implant
SNNPR,'Not currently using but has before',No,Rural,'Early Youth','Comp primary',Muslim,Single,'Medium Size Family','Injection - Depo'
Amhara,'Has Never Used Before',Yes,Rural,Adolecent,'Comp primary',Chirstian,Single,'Medium Size Family',Implant

Now the data is preprocessed and ready for data mining process to conduct the experiment using selected classifiers.

From the total 105,205 FP users (instances), 37,607 were implant, 32369 were Injectable-Depo, 14094 were IUCD, 20917 Pills and 218 were BTL. Figure 8 showed that the number of instances for selected attributes of a class.

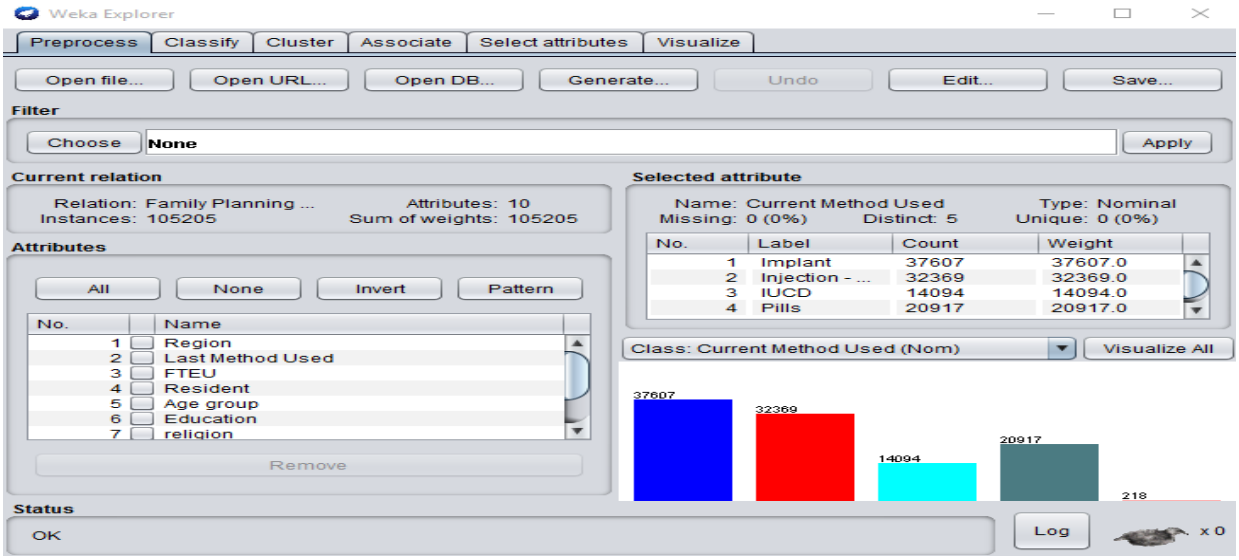


Figure 7: Class Imbalance problem

Looking this data shows class imbalance, is the problem in machine learning where the total number of a class of data (positive) is far less than the total number of another class of data (negative). In order to balance the instance, class balancer, reweights the instances in the data so that each class has the same total weight and the total sum of weights across all instances will be

maintained, technique is applied. Hence, the researcher applied this technique in weka. As a result, the balanced data using class balancer after reweight is shown as follows. Weka 3.8.1. has class balancer feature for this purpose.

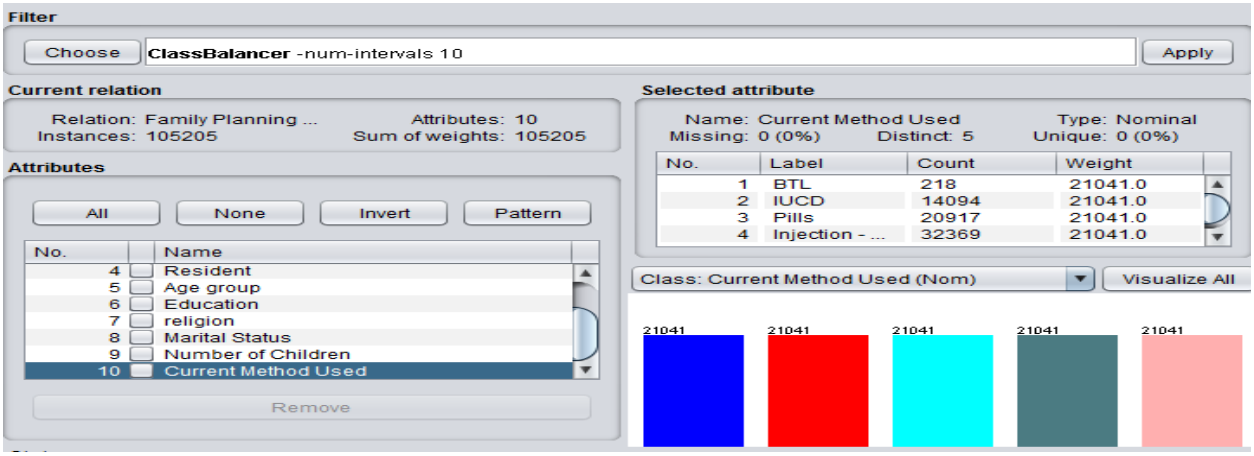


Figure 8: Balanced class using Class Balancer technique

4.2.4. Feature Selection

One of the factors that affect the success of data mining algorithms on a given task is the quality of the data. The data is irrelevant and repetitive, the quality of the data mining algorithm will decrease. In order to remove such problems, selecting relevant and valid attributes is important. Attribute selection also called feature selection, is the process of selecting a subset of relevant features (variables, predictors) for use in model construction [85].

Weka 3.8.1 has feature selection feature which determine relevance of attributes. The common method attribute selection methods are ranker method and wrapper method.

The researcher applied BestFirst method to select attributes in weka 3.8.1 since all attributes algorithm is predicted using information gain. As a result selected attributes using BestFirst method are residence, education, religion, marital status and number of children were selected.

4.3. Experimentation

After balancing the imbalanced class of targeted attribute (Current method used), experiment has been taken place using selected classifier on family planning user data set which has 105,205 instances with 9 independent variable(all relevant attributes except current method used) and 1 dependent variable (current method used). Weka 3.8.1, open source and widely used, tool was

used to measure the best model. In this study, k-fold (10-folds) cross validation and percentage split test options are used because of its relatively low bias and variations [86]. In 10-fold cross validation, the data is divided randomly into 10 parts in which the class is represented in approximately the same proportions as in the full dataset. Each part is held out in turn and the learning scheme trained on the remaining nine-tenths; then its error rate is calculated on the holdout set. Thus the learning procedure is executed a total of 10 times on different training sets (each of which have a lot in common). Finally, the 10 error estimates are averaged to yield an overall error estimate where as 66 percentage of the data is used as training and the remaining 34% is used as test data. Accuracy, Precision, specificity, Recall, number of leaves, rules time taken to build the model and confusion matrix standard metrics were also used for evaluation of the results.

In order to get a better predictive model of choice of contraceptive method, five classification algorithms namely J48 decision tree, JRIP rule, REPTree, PART, and Naïve Bayes were selected and experimented through Weka machine learning software. Those algorithms were selected, since most of the related works that are discussed in this study were experimented using two or more of the above algorithms and also they are easy for model building, interpretation and understanding.
The researcher applied different scenarios on the selected classification algorithm.

Until this stage, data were preprocessed and make ready for mining task. The researcher set five scenarios in order to select the best classifiers.

- Scenario1: J48 decision tree with all attributes and selected attributes
- Scenario 2: PART Rule with all attributes and selected attributes
- Scenario 3: JRip Rule with all attributes and selected attributes
- Scenario 4: REPTree Rule with all attributes and selected attributes
- Scenario 5: Naïve Bayes with all attributes and selected attributes

Each scenario are discussed below:

Scenario 1: J48 decision tree with all attributes and selected attributes

The researcher used J48 decision tree algorithm to classify family planning users. The reason why the researcher preferred this algorithm due to decision tree J48 is easy to understand, performs better [87] in addition to widely use in healthcare data mining task. Pruned J48 algorithm was tested by changing the values of minNumobj and confidenceFactor. Pruning simplifies a classifier by merging disjuncts that are adjacent in instance space and its goal is to generate fewer branches and good accuracy. MinNumobj is minimum number of instances per leaf (the value that the researcher used is 2 (default value) and 5(to increase the efficiency of the algorithm). Confidence factor used to test the effectiveness of post pruning. Lowering the confidence factor decrease the amount of post pruning [88]. The values of the confidence factor selected for this research is 0.25 and 0.5. According to literatures [23] [62], unpruned tree is not effective as compared to pruned tree experiment. Hence, unpruned tree is not applied in the experiment.

Table 11 showed experiments done using different J48 classifier parameters between selected attributes and all attributes.

Table 10: J48 decision tree experiment parameters

Experiments	Attributes used	Testing parameters		
		Confidence Factor	minNumobj	Test Option
Experment1	all	0.25	2	10 fold cross validation
Experment2	all	0.25	5	10 fold cross validation
Experment3	all	0.5	2	10 fold cross validation
Experment4	selected	0.25	2	10 fold cross validation
Experment5	selected	0.25	5	10 fold cross validation
Experment6	selected	0.5	2	10 fold cross validation
Experment7	all	0.25	2	Percentage split with 66%
Experment8	all	0.25	5	Percentage split with 66%
Experment9	all	0.5	2	Percentage split with 66%
Experment10	selected	0.25	2	Percentage split with 66%
Experment11	selected	0.25	5	Percentage split with 66%
Experment12	selected	0.5	2	Percentage split with 66%

Based on the above experiment parameters, the J48 decision tree experiment result is discussed in table 11.

Table 11: J48 classifier experimental result

Performance Measurement	**Experiment**											
	All attributes						**Selected Attributes**					
	Exp1	Exp2	Exp3	Exp7	Exp8	**Exp9**	Exp4	Exp5	Exp6	Exp10	Exp11	Exp12
Accuracy (%)	71.4	71.1	71.7	72.0	71.8	**72.3**	59.5	59.4	59.5	58.7	58.7	58.8
# of leaves	1890	1458	2642	1890	1458	**2642**	126	124	146	126	124	146
Size of tree	2543	1963	3538	2543	1963	**3538**	187	184	217	187	184	217
Time (sec)	0.97	0.67	0.92	0.73	0.70	**0.60**	0.37	0.34	0.34	0.32	0.32	0.32
TP Rate	0.714	0.711	0.717	0.720	0.718	**0.723**	0.595	0.594	0.595	0.587	0.587	0.588
FP Rate	0.071	0.072	0.071	0.070	0.715	**0.069**	0.101	0.101	0.101	0.103	0.103	0.103
Precision	0.714	0.711	0.716	0.718	0.718	**0.720**	0.586	0.585	0.586	0.583	0.583	0.583
Recall	0.714	0.711	0.717	0.720	0.714	**0.723**	0.595	0.594	0.595	0.587	0.587	0.588
F-Measure	0.713	0.710	0.715	0.717	0.646	**0.720**	0.586	0.586	0.586	0.583	0.583	0.583
ROC Area	0.911	0.910	0.914	0.913	0.912	**0.917**	0.842	0.842	0.843	0.842	0.842	0.842
CCI	75123	74818	75388	25891	25791	**25999**	62550	62543	62261	21115	21114	21118
ICI	30081	30387	29816	10051	10151	**9943**	42654	42661	42643	14826	14827	14823

Keys: Exp: experiment; sec: second; TP: True positive; FP: False Positive; ROC: Receiver Operating Characteristic; CCI: Correctly Classified Instances; ICI: Incorrectly Classified Instances

Literatures [89] showed that predictive accuracy reflects how good the training results with respect to the test data. The higher the predictive accuracy is the better algorithm. Taking into consideration, in table 11 the highest accuracy (72.3%) is observed in experiment9 (percentage split (66%) with values of confidence factor 5 and minNumobj 2 applied in all attributes). Experiment7 has the second and experiment8 comes in the 3rd place. Therefore, the researcher selected experiment9 to compare with other algorithms.

Scenario 2: PART Rule with all attributes and selected attributes

PART is also a tree algorithm which uses separate-and-conquer i.e. it builds a partial C4.5 decision tree in each iteration and makes the best leaf into a rule. Like J48 algorithm using confidence factor and minNumobj parameters, the PART algorithm experiment has done both all attributes and selected attributes through applying 10 fold cross validation and 66% percentage split test option. Table 9 experiment parameters was applied in this algorithm. The values of the parameters is also same with J48 algorithm. Table 12 showed the performance measurement of the PART classifier experiment in both all and selected attributes.

Table 12: PART experimental result

Performance Measurement	**Experiment**											
	All attributes						**Selected Attributes**					
	Exp1	Exp2	Exp3	Exp7	Exp8	**Exp9**	Exp4	Exp5	Exp6	Exp10	Exp11	Exp12
Accuracy (%)	71.3	70.8	71.4	72.2	71.5	**72.2**	59.4	59.4	59.5	58.8	58.7	58.8
# of rules	1253	896	1529	1253	896	**1529**	110	107	120	110	107	120
Time (sec)	32.98	40.13	41.2	32.85	38.8	**72.26**	1.81	1.75	2.11	2.12	1.84	2.27
TP Rate	0.713	0.708	0.714	0.722	0.715	**0.722**	0.394	0.594	0.595	0.588	0.587	0.588
FP Rate	0.072	0.073	0.072	0.062	0.071	**0.069**	0.107	0.101	0.101	0.103	0.103	0.103
Precision	0.712	0.707	0.713	0.718	0.712	**0.718**	0.480	0.585	0.586	0.583	0.583	0.583
Recall	0.713	0.708	0.714	0.722	0.715	**0.722**	0.394	0.594	0.595	0.588	0.587	0.588
F-Measure	0.711	0.706	0.712	0.718	0.712	**0.718**	0.433	0.585	0.586	0.583	0.582	0.583
ROC Area	0.912	0.909	0.915	0.916	0.913	**0.916**	0.841	0.842	0.843	0.841	0.839	0.841
CCI	74983	74456	75104	25935	25714	**25943**	62526	62521	62566	21120	21109	21123
ICI	30222	30749	30101	10007	10228	**9999**	42679	42683	42638	14822	14832	14819

Table 12 depicted that the highest accuracy (72.2) is observed in experiment 9 (percentage split (66%) with all attributes and value of confidence factor is 0.5 and minNumobj 2 followed by experiment7 i.e. default values percentage split (66%). Hence, the researcher selected experiment 9 to compare with other algorithms.

Scenario 3: JRip Rule with all attributes and selected attributes

JRip is a propositional rule learner, Repeated Incremental Pruning to Produce Error Reduction (RIPPER) and a very common and effective technique found in decision tree algorithms [90].

Using default mode of the JRip classifier, the comparison has done through applying 10 fold cross validation (CV) and 66% percentage split test option.

Table 13: JRip algorithm experiments result

Performance Measurement	All Attribute		Selected Attribute	
	Exp1 (10 fold CV)	**Exp2 (66% split)**	Exp3(10 fold CV)	Exp4 (66% split)
Accuracy	68.6	**69.0**	57.6	57.5
Number of rules	105	**105**	30	26
Time (sec)	133.94	**113.15**	13.55	9.98
TP rate	0.686	**0.690**	0.576	0.575
FP Rate	0.079	**0.077**	0.106	0.106
Precision	0.698	**0.706**	0.596	0.617
Recall	0.686	**0.690**	0.576	0.575
F-Measure	0.678	**0.688**	0.569	0.582
ROC curve	0.870	**0.873**	0.807	0.815
CCI	72118	**15817**	60568	13173
ICI	33087	**7094**	44637	9737

As presented in Table 13 JRip classifier which was implemented on percentage split (66%) experiment achieved the highest accuracy (69.0%) while 10 fold cross validation experiment is the closest with overall accuracy of 68.6. Therefore, the researcher selected JRip classifier with percentage split (66%) to compare it with the other classification algorithms.

Scenario 4: REPTree Rule with all attributes and selected attributes

REPTree is fast decision tree learner. Builds a decision/regression tree using information gain/variance and prunes it using reduced-error pruning (with back fitting). Only sorts values for numeric attributes once. Missing values are dealt with by splitting the corresponding instances into pieces (i.e. as in C4.5). Using default mode of the REPTree classifier, the comparison has done through applying 10 fold cross validation and 66% percentage split test option.

Table 14: REPTree algorithm experiment result

Performance Measurement	All Attribute		Selected Attribute	
	Exp1 (10 fold CV)	**Exp2 (66% split)**	Exp3(10 fold CV)	Exp4 (66% split)
Accuracy	71.0	**71.5**	59.3	58.8
Number of the size tree	3256	**3256**	242	242

Time (sec)	1.15	**1.17**	0.44	0.43
TP rate	0.710	**0.715**	0.593	0.588
FP Rate	0.072	**0.071**	0.102	0.103
Precision	0.707	**0.710**	0.584	0.583
Recall	0.710	**0.715**	0.593	0.588
F-Measure	0.707	**0.710**	0.584	0.582
ROC curve	0.915	**0.750**	0.842	0.842
CCI	74704	**25692**	62415	21119
ICI	30501	**10250**	42790	14822

Based on the experiments result described in table 14, the second experiment which is 66% percentage split with 71.5 percent accuracy level performs better as compared to another experiments. Thus, experiment 2 is selected to compare with other scenarios.

Scenario 5: Naïve Bayes with all attributes and selected attributes

Naïve Bayesian Classification is unsupervised classification technique which is based on the theory of Bayes. It can predict class membership probabilities, such as the probability that a given tuple belongs to a particular class [9]. This algorithm 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. Using default mode of the naïve Bayes classifier, the comparison has done through applying 10 fold cross validation and 66% percentage split test option.

Table 15: Naive Bayes algorithm experiment result

Performance Measurement	All Attribute		Selected Attribute	
	Exp1 (10 fold CV)	**Exp2 (66% split)**	Exp3(10 fold CV)	Exp4 (66% split)
Accuracy	61.4	**62.6**	56.7	57.1
Time (sec)	0.02	**0.02**	0.01	0.02
TP rate	0.614	**0.626**	0.567	0.571
FP Rate	0.096	**0.093**	0.108	0.107
Precision	0.604	**0.616**	0.556	0.561
Recall	0.614	**0.626**	0.567	0.571
F-Measure	0.605	**0.617**	0.554	0.560
ROC curve	0.859	**0.860**	0.819	0.815

CCI	64596	**22481**	59665	20525
ICI	40609	**13460**	45540	15416

In table 15, naïve bayes percentage split (66%) scored highest accuracy (62.6%) with default values of parameter using all attributes followed by 10 fold cross validation with same parameter. As a result, the researcher selected percentage split 66% to compare with other algorithms.

4.4. Comparing Models

For this research; two decision trees, two decision rules and one bayes algorithms were selected and experimented. Once experiment has done and best result is selected from each scenarios, the researcher compared these algorithms in order to select best models that predicts choice of contraceptive methods.

Accordingly selected models which are J48, PART, JRip, REPTree, and naïve bayes were compared with performance measurement of accuracy, TP rate, FP rate, time taken, precision, recall, F-measure and ROC curve. Table 16 showed the comparison between models selected from five scenarios.

Table 16: Comparison between models which is selected from each scenario

Performance Measurement	**J48**	PART	JRip	REPTree	Naïve Bayes
Accuracy	**72.3**	72.2	69.0	71.5	62.6
Time (sec)	**0.60**	72.26	113.15	1.17	0.02
TP rate	**0.723**	0.722	0.690	0.715	0.626
FP Rate	**0.069**	0.069	0.077	0.071	0.093
Precision	**0.720**	0.718	0.706	0.710	0.616
Recall	**0.723**	0.722	0.690	0.715	0.626
F-Measure	**0.720**	0.718	0.688	0.710	0.617
ROC curve	**0.917**	0.916	0.873	0.750	0.860

As depicted in table 16, among best models selected from each scenario, J48 has highest accuracy (72.3%) and PART is also the closest accuracy (72.2%) where as Naïve bayes scored the lowest accuracy (62.6%). Looking the true positive rate (sensitivity) of selected algorithm, J48 is also performed followed by PART. J48 algorithm is equal or better in others performance measurement parameters. The ROC curve, created by plotting the true positive rate and false

positive rate at various threshold settings, also showed that J48 algorithm showed the highest (0.917) since the higher the ROC is the better the model predicts.

J48 decision tree algorithm classified best compared to the rest of the classifiers in the domain of predicting choice of contraceptive methods even though there is 26.7% error rates. Therefore, the researcher conclude that J48 decision tree is the most appropriate in this particular domain problem. As a result, J48 algorithm selected as final model and rules are generated from this models which was used as an input for knowledge based system. The confusion matrix of this J48 classifier is attached in appendix II.

4.5. Generated Rules from Selected model

Pruned J48 tree model produced different rules. In order to evaluate the importance of the discovered knowledge/rules, whether they are acceptable or not and whether they match with what is already known in practice, the researcher has concluded to discuss with domain experts of MSIE. The key determinant variables for the uptake of contraceptive methods are different from method to method. The other objective on this research was selecting best rules that determine the choice contraceptive method. Thus, the researcher discussed key variables that determine the contraceptive method and hidden patterns that govern the choice of method. The following rules are selected that domain experts are agreed.

Bilateral Tuba-ligation (BTL) Rules

Rule1: (Age =Adult) AND (Number of Children =High family size) AND (Last Method Used=has never used before): BTL (2219.92)

Rule 2: (Age group =adult) AND (Last Method Used =has never used before) AND (Number of Children= Medium size family) AND (Education=comp Primary) AND (Residence= Rural) AND (Religion =Christian) AND (Region=Amhara) AND Marital Status=married): BTL (442.87/56.8)

Rule 3: (Age group = adult) AND (Number of Children = High family size) AND (Religion = Christian) and (Marital status=married) AND (Last method used=has never used before): BTL (5598.06)

Rule 4: (Age=Adult) AND (Number of Children=High family size) AND (Religion=Christian) AND (Marital Status =Married) AND (Last Method Used =Not currently using has used before)

AND (Residence=Urban) AND (Region=Addis Ababa) AND (Education=none/Some primary): BTL (821.79/49.64)

Rule 5: (Age group =Adult) AND (number of children=high family size) AND (Religion=Christian) AND (Marital Status = Married) AND (Last method used = (Not Currently Using has used before) AND (Residence=Urban) AND (Region=Addis Ababa) AND (Education=some primary): BTL (2040.77/206.92)

Rule 6: (Age group=Adult) AND (Number of children=high family size) AND (Marital status = Married) AND Last method used = Not currently using but has used before) AND (Residence = Rural) AND (Education=none/some primary): BTL (3202.52/17.41)

Rule 7: (Age group=Adult) AND (Number of children=high family size) AND (Religion=Christian) AND (Marital status = Married) AND Last method used = has used before but has used before) AND (Residence = Rural) AND (Education=none/some primary) AND Region (Oromia): BTL (588.21.52/9.1)

Rule 8: (Age group=Adult) AND (Number of children=high family size) AND (Religion=Christian) AND (Marital status = Married) AND Last method used = has used before but has used before) AND (Residence = Rural) AND (Education=none/some primary) AND (Region=Amhara): BTL (1098.42/36.72)

Rule 9: (Age group=Adult) AND (number of children =high family size) AND (Religion=Christian) AND (Marital Status= Married) AND (Last Method Used=Not currently using but has used before) AND (Residence = Rural) AND (Education =Comp Primary) AND (Region=Oromia): BTL(196.67/3.63)

Rule 10: (Age group=Adult) AND (number of children =high family size) AND (Religion=Muslim) AND (Last Method Used=Has never used before): BTL (289.56)

IUCD Rules

Rule 1: (Age group = early youth) AND (Marital status=married) AND (Number of children=medium family size) AND (Residence=Urban) AND (Education=none/Some primary): IUCD (1449.37/249.07)

Rule 2: (Age group = early youth) AND (Marital status=married) AND (Number of children=medium family size) AND (Residence=Urban) AND (Education= higher/Tertiary) AND (Region=SNNPR) AND (Religion=Christian): IUCD (403.92/57.56)

Rule 3: (Age group = early youth) AND (number of children=Medium family size) AND (Residence=Urban) AND (Religion=Christian) AND (Region=Oromia) AND (Last method used = has never used before): (135.4/8.5)

Rule 4: (Age group=early youth) AND (Marital status=married) AND (number of children=High family size) AND (Residence=Urban) AND (Education=Tertiary/higher) AND (Region=SNNPR) AND (Religion=Other) AND (Last contraceptive method=not currently using but has used before): IUCD (27.91/4.02)

Rule 5: (Age group = early youth) AND (number of children=Medium family size) AND (Residence=Urban) AND (Religion=Christian) AND (Region=Oromia) AND (Last method used = not currently using but has used before): IUCD (185.51/31.74)

Rule 6: (Age group = early youth) AND (marital status=married) AND (number of children=Medium family size) AND (residence= urban) AND (Education=comp secondary) AND (religion=Christian) AND (region= SNNPR): IUCD (328.02/29.44)

Rule 7: (Age group = early youth) AND (marital status=married) AND (number of children=Medium family size) AND (residence= rural) AND (region = Oromia) AND (Education=comp primary): IUCD (277.59/56.94)

Rule 8: (Age group =early youth) AND (Marital Status=Married) AND (Number of Children= Medium size family) AND (Residence =Urban) AND (Education=comp Primary) AND (religion=Christian) AND (region=Addis Ababa) AND (last Method used = has never used before): IUCD (132.28/35.24)

Rule 9: (Age group=youth) AND (Number of children=medium family size) AND (Marital status=married) AND (Religion=Christian) AND (residence=Rural) AND (Region=Oromia) AND (Education=Tertiary/higher) last method used=has never used before): (96.29/2.24)

Rule 10: (Age group=youth) AND (Number of children=medium family size) AND (Marital status=married) AND (Religion=Christian) AND (residence=urban): IUCD (6474.11/1693.83)

Rule 11: (Age group=youth) AND (Number of children=medium family size) AND (Marital status=married) AND (Religion=Christian) AND (residence=rural) AND (Region=Oromia)

AND (Education=Tertiary/higher) last method used=not currently using but has used before): IUCD (205.74/19.13)

Rule 12: (Age group=youth) AND (Number of children=high medium size) AND (Marital status=married) AND (last method used=has never used before) AND (region=Amhara) AND (Education=Tertiary/higher): IUCD (49.45/1.68)

Rule 13: (Age group= adolescent) AND (marital status=single) AND (number of children= no children) AND (Religion=Christian) AND (Region=Oromia) AND (Residence=Urban) AND (Education=none/some primary) AND (Last method used=has never used before): IUCD (29.94/3.07)

Rule 14: (Age group=adult) AND (Region=Oromia) AND (Residence=urban) AND (Education= tertiary/higher) AND (number of children=medium family size) AND (Religion=Christian): IUCD (14.44/1.01)

Rule 15: Age group=adult) AND (region=Addis Ababa) AND (Education=comp secondary) AND (number of children= high family size) AND (Marital status=married): IUCD (130.99/2.6)

Rule 16: (age group=adult) AND (number of children=low family size) AND (residence= rural) AND (Region=Oromia) AND (Religion=Christian) AND (marital status=married) AND (last contraceptive method=not current using has used before): IUCD (737.63/391.28)

Implant Rules

Rue 1: (Age group= adolescent) AND (marital status= single) and (number of children=no children) AND (Region=Addis Ababa) AND (Education=None/some primary): Implant (339.3/71.86)

Rule 2: (age group=adolescent) AND (marital status=single) AND (no of children=low family size) AND (Residence=rural) AND (Region=SNNPR) AND (education=comp primary): Implant (56.3/6.5)

Rule 3: (age group=adolescent) AND (marital status=married) AND (education=none/some primary) AND (number of children=low family size): Implant (171.92/18.62)

Rule 4: (age group=adult) AND (number of children=high family size) AND (religion=Christian) AND (residence=rural) AND (region=SNNPR): Implant (114.38/26.54)

Rule 5: (age group=youth) AND (number of children=medium family size) AND (marital status=single) AND (Religion=Christian) AND (Education=tertiary/higher): Implant (514.83/161.23)

Rule 6: (age group=early youth) AND (marital status=single) AND (number of children= no children) AND (Education=none/some primary) AND (region=Addis Ababa) AND (religion=Christian) AND (last method used=has never used before): Implant (311.71/135.46)

Rule 7: (age group=early youth) AND (marital status=single) AND (number of children= low family size) AND (Education=none/some primary) AND (region=Addis Ababa) AND (religion=Christian): Implant (379.17/111.73)

Rule 8: (age group=early youth) AND (marital status=single) AND (number of children= no children) AND (Education=tertiary/higher) AND (region=SNNPR) AND (religion=Christian): Implant (160.65/48.75)

Rule 9: (age group=early youth) AND (marital status=single) AND (number of children=medium family size) AND (Education=tertiary/higher) AND (region=Oromia) AND (religion=Christian) AND Last method used=has never used before): Implant (50.34/3.9)

Rule 10: (age group=early youth) AND (marital status=married) AND (number of children= low family size) AND (residence=rural) AND (region=Tigray): Implant (48.59/6.62)

Rule 11: (age group=early youth) AND (marital status=married) AND (number of children= low family size) AND (residence=urban) AND (education= comp primary) AND (region=Addis Ababa) AND (Religion=Christian): Implant (122.1/21.95)

Rule 12: (age group=early youth) AND (marital status=married) AND (number of children= low family size) AND (residence=urban) AND (education=none/some primary) AND (region=Addis Ababa): Implant (886.75/368.65)

Rule 13: (age group=early youth) AND (marital status=married) AND (number of children= high family size) AND (residence=rural) AND (region=Oromia) AND (religion=Muslim): Implant (47.73/6.89)

Rule 14: (age group=early youth) AND (marital status=married) AND (number of children= medium family size) AND (residence=rural) AND (region=SNNPR) AND (Religion=Christian) AND (last method used=not currently using but has used before: Implant (166.49/31.09)

Rule 15: (age group=early youth) AND (marital status=married) AND (number of children= medium family size) AND (residence=rural) AND (religion=Tigray) AND (region=Tigray): Implant (100.66/11.7)

Rule 16: (age group=early youth) AND (marital status=married) AND (number of children= low family size) AND (residence=urban) AND (region=Addis Ababa) AND (education=comp primary) AND (Religion=Christian): Implant (277.49/99.01)

Rule 17: (age group=early youth) AND (marital status=married) AND (number of children= medium family size) AND (Education=none/some primary) AND (region=Addis Ababa) AND (Religion=Christian): Implant (193.89/56.81)

Rule 18: (age group=early youth) AND (marital status=married) AND (number of children=medium family size) AND (region=Dire Dawa): Implant (36.82/3.25)

Rule 19: (age group=early youth) AND (marital status=married) AND (number of children= medium family size) AND (residence=rural) AND (region=SNNPR) AND (religion=Christian): Implant (71.24/2.99)

Rule 20: (age group=early youth) AND (marital status=married) AND (number of children= medium family size) AND (residence=rural) AND (region=Amhara) AND (religion=Christian): Implant (56.98/11.1)

Injection-Depo Rules

Rule 1: (age group=youth) AND (number of children=no children) AND (marital status=single) AND (education=comp primary): Injection-depo (446.62/164.5)

Rule 2: (age group=youth) AND (number of children=no children) AND (marital status=single) AND (region=Amhara) AND (residence=urban): Injection-depo (181.56/50.9)

Rule 3: (age group=youth) AND (number of children=low family size) AND (residence=rural) AND (Region=Oromia) AND (education>=comp secondary): injection-depo (160.22/48.41)

Rule 4: (age group=youth) AND (number of children=low family size) AND (residence=urban) AND (Region=Amhara): injection-depo (312.6/64.28)

Rule 5: (age group=youth) AND (number of children=high family size) AND (marital status=married) AND (religion=Christian) AND (Education=comp primary) AND (residence=urban): injection-depo (81.48/10.63)

Rule 6: (age group=youth) AND (number of children=medium family size) AND (marital status=married) AND (religion=Muslim) AND (education=none/some primary) AND (residence=rural) AND (Region=Oromia): injection-depo (49.84/13.44)

Rule 7: (age group=early youth) AND (marital status=single) AND (number of children=no children) AND (education=comp secondary) AND (Residence=rural) AND (Region=Oromia) AND (Religion=Christian): Injection-depo (173.72/36.56)

Rule 8: (age group=early youth) AND (marital status=single) AND (number of children=no children) AND (education=comp primary) AND (Residence=rural) AND (last method used=not currently using but has used before): Injection-depo (269.72/63.01)

Rule 9: (age group=early youth) AND (marital status=single) AND (number of children=no children) AND (education=none/some primary) AND (Region=Oromia): Injection-depo (202.42/43.81)

Rule 10: (age group=early youth) AND (marital status=single) AND (number of children=low family size) AND (education<=comp primary): Injection-depo (525.75/12.87)

Rule 11: (age group=early youth) AND (marital status=single) AND (number of children=medium family size) AND (education=comp secondary) AND (Region=Oromia): Injection-depo (200.43/68.47)

Rule 12: (age group=early youth) AND (marital status=single) AND (number of children=medium family size) AND (education= comp primary): injection-depo (747.17/164.74)

Rule 13: (age group=early youth) AND (marital status = single) AND (number children=medium family size) AND (education=none/some primary) AND (Residence=urban) AND (last method method=not currently using has used before) AND (Religion=Christian) AND (Region=Addis Ababa): injection-depo (178.99/54.83)

Rule 14: (age group=early youth) AND (marital status=married) AND (number of children=low family size) AND (education=comp primary) AND (Residence=urban) AND (Region=Amhara) AND (last method used=has never used before): Injection-depo (48.83/5.93)

Rule 15: (age group=early youth) AND (marital status=married) AND (number of children=high family size) AND (Residence=rural) AND (Religion=Muslim) AND (Region=Oromia) AND (education=comp primary): injection-depo (88.13/21.83)

Rule 16: (age group=early youth) AND (number of children=high family size) AND (Residence=Urban) AND (region=Oromia) AND (Religion=Other) AND (last method used=has never used before): Injection-depo (17.55)

Pills Rules

Rule 1: (Age group=early youth) AND (marital status=married) AND (Number of children=high family size) AND (Residence=urban) AND (Region=Oromia) AND (Religion=Muslim): Pills (80.7/2.24)

Rule 2: (Age group=early youth) AND (marital status=married) AND (Number of children=low family size) AND (Residence=urban) AND (Region=Oromia) AND (education=Tertiary/higher): Pills (75.93/21.61)

Rule 3: (Age group=early youth) AND (marital status=married) AND (Number of children=low family size) AND (Residence=urban) AND (education=Tertiary/higher): Pills (834.9/62.34)

Rule 4: (Age group=early youth) AND (marital status=married) AND (Number of children=low family size) AND (Residence=urban) AND (Religion=Christian) AND (education=comp secondary) AND (last method used=not currently using has used before): Pills (308.75/74.36)

Rule 5: (Age group=early youth) AND (marital status=married) AND (Number of children=no children) AND (region= (SNNPR OR Addis Ababa OR Oromia)) AND (education=Tertiary/higher): Pills (460.24/61.89)

Rule 6: (Age group=early youth) AND (marital status=married) AND (Number of children=no children) AND (Region=Oromia) AND (education=comp secondary) AND (religion=Christian) AND (last method used=not currently using but has used before): Pills (101.3/32.9)

Rule 7: (Age group=early youth) AND (marital status=married) AND (Number of children=no children) AND (Religion=Muslim) AND (education=comp secondary): Pills (95.71/6.2)

Rule 8: (Age group=early youth) AND (marital status=single) AND (Number of children=no children) AND (Residence=urban) AND (Region=Oromia) AND (education=tertiary/urban) AND (last method used=not currently using has used before): Pills (298.02/54.58)

Rule 9: (Age group=youth) AND (marital status=married) AND (Number of children=low family size) AND (Residence=urban) AND (Religion=Christian OR Muslim) AND (education=tertiary/higher) AND (region=Oromia): Pills (226.13/35.0)

Rule 10: (Age group=youth) AND (marital status=married) AND (Number of children=no children) AND (Residence=urban) AND (education=tertiary/higher): Pills (916.63/30.4)

Rule 11: (Age group=youth) AND (marital status=married) AND (Number of children=no children) AND (Residence=rural) AND (education=tertiary/higher) AND (region=Oromia): Pills (170.34/26.49)

Rule 12: (Age group=youth) AND (marital status=married) AND (Number of children=no children) AND (Residence=urban) AND (education=comp secondary): Pills (672.61/77.1)

Rule 13: (Age group=adolescent) AND (marital status=single) AND (Number of children=no children) AND (Residence=urban) AND (education=tertiary/higher): Pills (114.73/8.1)

Rule 14: (Age group=adolescent) AND (marital status=single) AND (Number of children=low family size) AND (Residence=urban) AND (education=tertiary/higher): Pills (268.1/4.55)

Rule 15: (Age group=adult) AND (Number of children=high family size) AND (religion=Muslim) AND (education=tertiary/higher): Pills (149.44/14.65)

4.5.1. Discussion on classification model's generated rules

From generated rules, it is observed that client's education, religion, number of children, residence, region, marital status, last method used and client age are key determinant factors in the uptake of any method. Though all these variables determine the choice of contraceptive method, but determinant variable vary from method to the method.

Generated rules showed that clients' age greater than 30 and her education level is comprehensive primary or low and has 3 or more children as well as has no past contraceptive history is high likely to use bilateral tubaligation (BTL). Whereas, youth or below age group and has 2 or below children has never used BTL.

IUCD can be received by any age group of clients and has any number of children and also has an experience on contraceptive or not. Religion is a key factor on IUCD uptake; Christian women are high likelihood receive it. In other words, Muslim women are less likely to use IUCD method.

Clients age between 15-24, and family size is 0-4 and her religion is Christian, there is high likely to use implant. On the contrary, Adult women and their family size is high as well as Muslims are less likely to use implant method.

A client age between 20 to 24, her education is comprehensive primary or low, marital status as well as the number of children is 2 and below, then there is high likely to use Injection-depo.

Tertiary/higher educated or Muslim clients as well as high family size women are high likely users of pills method.

4.6. Knowledge Acquisition from documents (explicit knowledge)

Literatures [78] [19] shows that choice of contraceptive methods is not only determined by socio demographic factors but also heath condition of the client, (medical eligibility of the client) which plays vital role during the uptake of contraceptive methods and lifestyle of the client. As described in chapter three, to construct the knowledge based system, data mining rules and other documents knowledge was required. The socio demographic factor was identified and best rules were generated using the knowledge discovery process. In this subtitle, the researcher explained key health condition determinant factors for the choice of contraceptive methods. This Medical eligibility criteria (MEC) was referred from World Health Organization (WHO) 5th edition of medical eligibility criteria [78]. Table 17 showed medical eligibility criteria for the choice of contraceptive methods.

Table 17: Medical eligibility criteria for contraceptive use

MEC	MEC values	IUCD	Implant	Injectable	Pills	Remark
Hypertension	140-159/90-99	Yes	Yes	No	No	
	>=160/>=100	Yes	Yes	No	No	
Smoking	Age<35	Yes	Yes	Yes	Yes	
	Age>=35	Yes	Yes	No	No	
Headache	Non migraines	Yes	Yes	Yes	Yes	
	Migraines with aura	Yes	Yes	No	No	
Medication	Certain anti convulsants	Yes	Yes	No	No	
	Rifampicin	Yes	Yes	No	No	
	Antiretroviral therapy	No	Yes	Yes	Yes	
HIV	Stage 3 or 4	No	Yes	Yes	Yes	
STI	Gonorrhoea Chlamydia	No	Yes	Yes	Yes	
	Other STIs & vaginitis	Yes	Yes	Yes	Yes	
	Increased risk of STIs	No	Yes	Yes	Yes	

PID	Current	No	Yes	Yes	Yes	
Sepsis	Puerperal and post abortion	No	Yes	Yes	Yes	
Vaginal Bleeding	Unexplained	No	No	Yes	Yes	
Cervical cancer	Pre-treatment	No	Yes	Yes	Yes	
Breast Cancer	Current	Yes	No	No	No	
Liver Diseases	Liver tumor	Yes	No	No	No	
	Hepatitis acute	Yes	Yes	No	No	
Venous Thromboembolism	History	Yes	Yes	No	No	
	Acute	Yes	No	No	No	
	Major Surgery	Yes	Yes	No	No	
Cardiovascular Disease	Stroke	Yes	Yes	No	No	
	Ishaemic heart disease	Yes	Yes	No	No	
Post-Partum and breastfeeding	<48 hours	Yes	Yes	No	No	
	48hr-4weeks	No	Yes	No	No	
	4week-6 week	Yes	Yes	No	No	
	6 week-6 month	Yes	Yes	No	No	

Keys: MEC: Medical eligibility criteria; BMI: Body mass Index; HIV: Human Immino Virus; STI: Sexual transmitted diseases; PID: Pelvic Inflammatory Disease;

In addition, table 18 showed that when the medical eligibility criteria value is 'Yes", the client can receive the specific contraceptive method. If "No", a given contraceptive method couldn't be received to the client in any circumstances.

4.7. Knowledge Acquisition from domain experts (Tacit Knowledge)

As described in chapter three, the conceptual framework of the study, knowledge base rules are generated from data mining results, documents and domain experts.

Knowledge from data mining and other documents are acquired. The next step is acquiring knowledge from domain experts. The researcher has conducted interview with domain experts on other determinant factor on the choice of contraceptive method. Five domain experts were selected for this purpose based on the education level, experience on the domain area and place of work as well as willingness to participate in this research. Table 18 showed that the education, work experience and location of selected domain experts. The reason only five is that staffs have other commitment, others are out of Addis Ababa, and others has low experience.

Table 18: Selected domain experts' profile

# of domain experts	Education level	Working Experience	Location/place of work	Willingness to participate
1	MD, Gynecologist	15 years	MSIE head office	Yes
1	MD	17 years	MSIE head office	Yes
1	MD, MPH	13 years	MSIE head office	Yes
1	HO, MPH	11 years	MSIE head office	Yes
1	RN Nurse	25 years	MSIE head office	Yes

Key: MD: Medical Doctor; MPH; Master of Public Health; RN: Registered Nurse

Thus, domain experts believed that life style of the client is also the determinant factor in the uptake of contraceptive method choice. The method choice is also depend on the fertility plan, sexual partner information and how a client perceive side effect of the method. Table 19 showed that how lifestyle of a woman can determine the choice of contraceptive methods.

Table 19: Life style of a client for the choice of contraceptive methods

Lifestyle criteria	Values	IUCD	Implant	Injection-depo	Pills	BTL
Fertility plan	Spacing	Yes	Yes	Yes	Yes	No
	limiting	Yes	Yes	No	No	Yes
Sexual partner	Single	Yes	Yes	Yes	Yes	Yes
	Multiple	No	No	No	No	No
Perceived side-effect	Not at all	Yes	No	No	No	Yes
	To some extent	Yes	Yes	No	No	Yes
	Accepted	Yes	Yes	Yes	Yes	Yes

Table 19 told us that if a client has multiple partner, she might be exposed to other disease like STI, HIV etc. So, contraceptive method is not advised to such kind of client. In addition, domain experts agreed that fertility intention is one of the most key variable for choice of contraceptive method.

4.8. Knowledge Modeling & Representation

Once the medical eligibility criteria and life style of a client determinants are identified, modeling them is the next key activity. Since data mining result is already modeled by the selected algorithm (discovered knowledge is modeled by decision tree), only domain expert and other document results are modeled.

The medical eligibility criteria are independent of each other i.e. one criteria result is not depending on the other criteria e.g. STI is not depend on headache so that the knowledge modeling is done individually to show the process. Like data mining, the discovered knowledge is modeled using decision tree. The following diagram shows sample knowledge modeling of medical eligibility criteria.

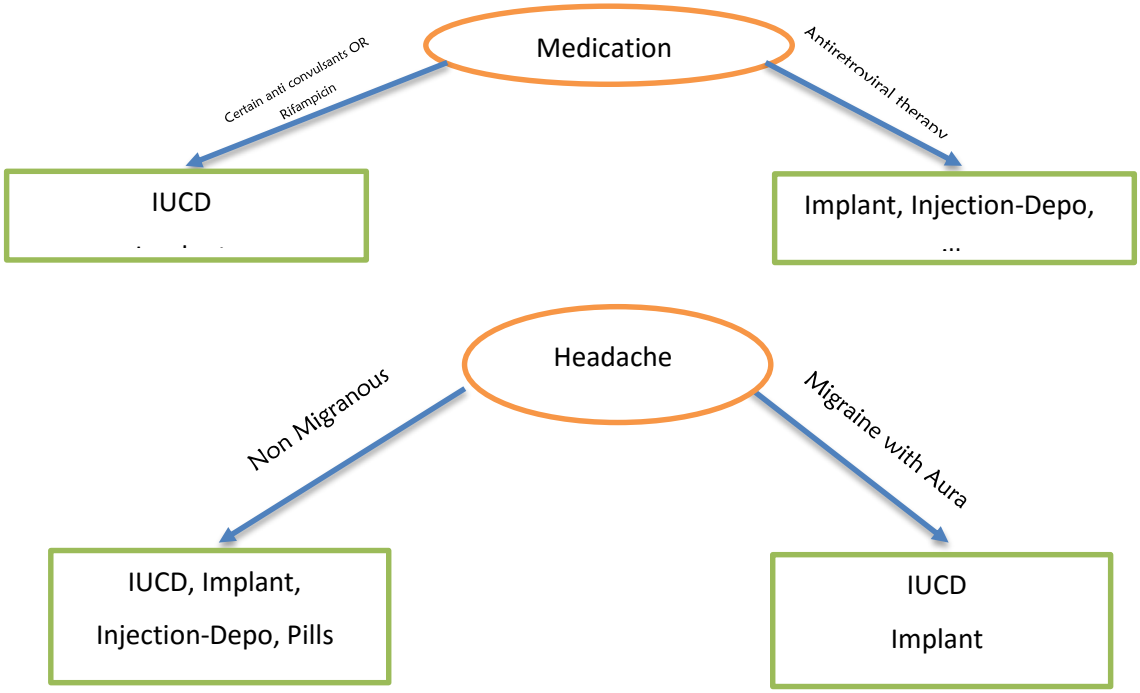


Figure 9: Sample Knowledge Modeling of Medical Eligibility Criteria

Figure 9 is explained as that if a client has certain anti convulsants OR Rifampicin then the client is not allowed to receive a contraceptive method of injection-depo or pills but she can receive either implant or IUCD. On other hand if a client is with antiretroviral therapy, IUCD is not allowed to a client but can be for other method. Similarly, a non-migranous client can receive any

method of contraception whereas migraine with aura client can receive only either IUCD or implant.

Like medical eligibility criteria, the following decision tree tells us the life style of a client is modeled using table 20 result which is acquired from domain experts.

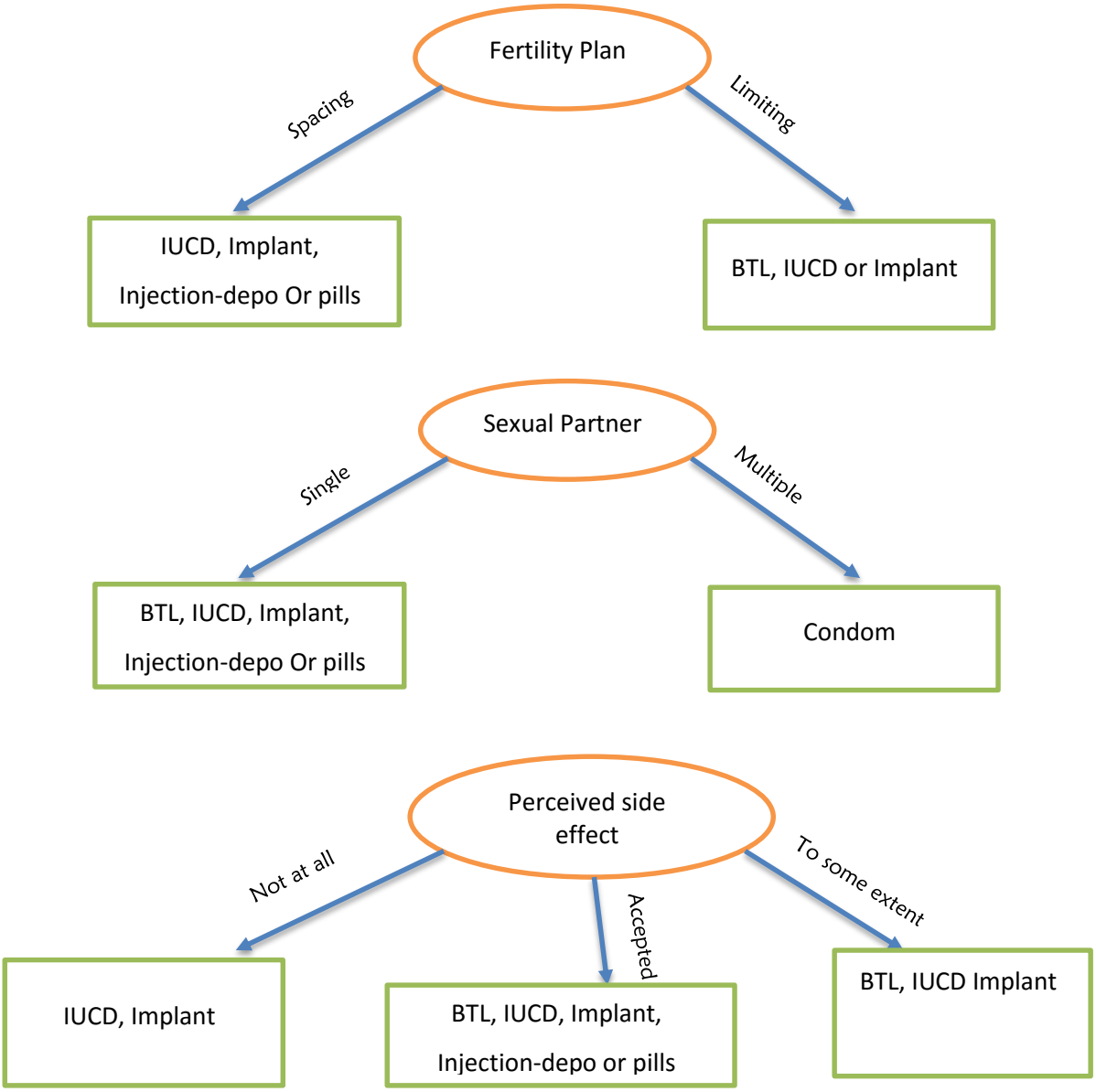


Figure 10: Decision tree of Life style characteristics for the choice of contraceptive methods

Figure 10 depicted that a client who want to spacing and needs for another child, BTL will never provide to her. If a client has multiple partner, she will never receive any of the methods rather will use condom.

The following knowledge modeling shows the integration of the acquired knowledge from domain experts and docucunts together. The reason to combine together is to describe the relationship between the MEC and lifestyle attributes.

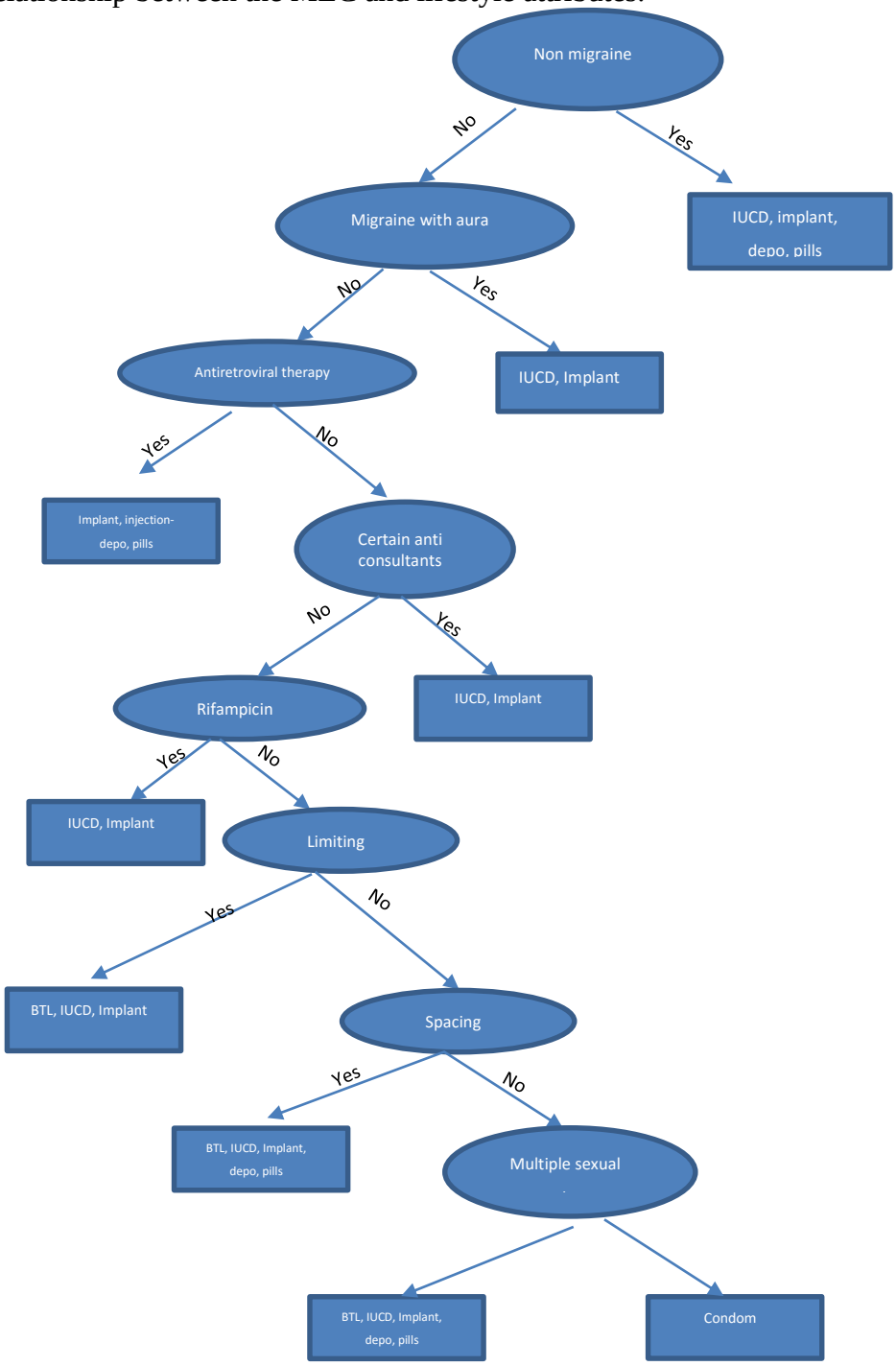


Figure 11: Decision tree: Integration of MEC and lifestyle attributes

Once knowledge is modeled using decision tree, the next step would be generating rules from decision rules. Using table 18 result, the researcher developed the following rules for each family planning method.

IUCD MEC Rules

Rule 1: (Hypertension = 140-159/90-99) OR (Hypertension = ">=160/>=100") OR (Smoking age = ">=35") OR (HIV stage 3 or 4=No) OR (Headache = Migraine with aura) OR (Medication = certain anti convulsants) OR (Medication = Rifampicin) OR (Post-partum = "<48hrs") OR (Post-partum = "4weeks- 6 weeks") OR (Post-partum = "6weeks – 6 months") OR (Liver Diseases =Liver Tumor) OR (Liver Diseases = Hepatitis acute) OR (Venus thromboembolism =history) OR (Venus thromboembolism = acute) OR (Venus thromboembolism = major surgery) OR (Cardiovascular disease = stroke) OR (Cardiovascular disease = Ishaemic heart disease) ➔ IUCD is eligible to the client

Rule 2: (Medication = antiretroviral therapy) OR (HIV stage 3 or 4=Yes) OR (STI = Gonorrhoea chlamydia) OR (STI=Increased Risk of STIs) OR (Current PID =Yes) OR (Spesis =Puerperal and post abortion) OR (Post-Partum = "48 hrs-4 weeks") OR (Unexplained Vaginal bleeding =Yes) OR (Cervical cancer pre-treatment =Yes) ➔ IUCD is not eligible to the client.

Implant MEC

Rule 1: (Hypertension = 140-159/90-99) OR (Hypertension = ">=160/>=100") OR (Smoking age = ">=35") OR (Headache = Migraine with aura) OR (Medication = certain anti convulsants) OR (Medication = Rifampicin) OR (Medication = antiretroviral therapy) AND (STI = Gonorrhea chlamydia) AND (STI = increased risk) AND (Current PID=No) OR (Sepsis= puerperal and post-abortion) AND (Post-partum = "<48hrs") OR (Post-partum = "48 hrs - 4weeks") OR (Post-partum = "4weeks – 6 weeks") OR (Post-partum = "6weeks – 6months") OR (cervical cancer pretreatment=No) OR (liver disease =Hepatitis acute) OR (Venus thromboembolism = history) OR (Venus thromboembolism = major surgery) OR (Cardiovascular disease = stroke) OR (Cardiovascular disease =Ishaemic heart disease) ➔ Implant is eligible to the client

Rule 2: (Unexplained Vaginal bleeding =Yes) OR (breast cancer =Yes) OR (liver disease = liver tumor) OR (venous thromboembolism= acute) ➔ Implant is not eligible to the client.

Injection-depo OR Pills Rules

The medical eligibility criteria for injection-depo and pills are the same so that rules are organized together.

Rule 1: (medication = antiretroviral therapy) OR (STI = gonorrhea chlamydia) OR (STI =Increased risk of STI) OR (sepsis = puerperal and post abortion) OR (vaginal bleeding=Yes/No) OR (current breast cancer= No) OR (cervical cancer pre-treatment=Yes/No) → Injection-depo OR Pills is eligible

Rules 2: (Hypertension = "140-159/90-99") OR (Hypertension = ">=160/>=100") OR (Smoking age = ">=35") OR (Headache = Migraine with aura) OR (Medication = certain anti convulsants) OR (Medication = Rifampicin) OR (Post-partum = "<48hrs") OR (Post-partum = 48 hrs - 4weeks) OR (Post-partum = 4weeks – 6 weeks) OR (Post-partum = 6weeks – 6months) OR (current breast cancer= Yes) OR (liver disease=liver tumor) OR (liver disease = hepatitis acute) OR (Venus thromboembolism = history) OR (Venus thromboembolism = acute) OR (Venus thromboembolism = major surgery) OR (Cardiovascular disease = stroke) OR (Cardiovascular disease= Ishaemic heart disease)→injection-depo OR Pills is not eligible

Using the decision tree described in figure 11, the researcher generated the rules which are acquired from domain experts..

BTL Rules

Rule 1: (Fertility plan = limiting) OR (Sexual partner = Single)→ BTL is eligible

Rule 2: (Fertility plan = spacing) OR (Sexual partner = multiple) →BTL is not eligible

IUCD Rules

Rule 1: (Fertility plan = Spacing/limiting) OR (Sexual partner = Single) OR (perceived side effect = Not at all) OR (perceived side effect =to some extent) OR (perceived side effect =Accepted)→ Client is eligible for IUCD

Rule 2: (Sexual partner = multiple)→ Client is not eligible for IUCD

Implant Rules

Rule 1: (Fertility plan =Spacing) OR (Sexual partner = Singles) OR (perceived side effect= accepted /to some extent)→ Client is eligible for Implant

Rule 2: (Sexual partner =multiple) OR (perceived side effect =Not at all) → Client is not eligible for Implant

Injection-Depo and Pills Rules

Rule 1: (Fertility plan = Spacing) OR (Sexual partner = Single) OR (perceived side effect = Accepted)→ Client is eligible for Injection-Depo OR Pills

Rule 2: (Fertility plan = limiting) OR (Sexual partner = multiple) OR (perceived side effect = Not at all) OR (perceived side effect= to some extent) → Client is not eligible for Injection-Depo OR Pills

Rules generated from domain experts tells us that knowing the fertility plan, sexual partner and perceiving side effect of contraceptive methods has factor in choice of contraceptive methods for a woman.

Chapter Five

KBS Design and Implementation of the Knowledge Based System

In chapter four, knowledge from data mining, documents and domain experts were acquired, modeled and represented. Once knowledge is acquired, the researcher discussed on how to integrate all these sources and form the knowledge base. Knowledge codification and implementation has done in this stage.

5.1. System Design

In this section, the researcher build the architecture of the data mining and other sources of knowledge integration in a way that how they can formulate the knowledge based system. Figure 12 showed the conceptual framework how data mining and other sources of knowledge are integrated with knowledge based system.

5.1.1. The Knowledge Base

To implement contraceptive method match knowledge based system (CMMKBS), the acquired knowledge is represented using production rules as 'if–then-rules'. This 'if-then-rules',

If first premise,

Second premise,

Then conclusion.

If the premises are true, then the actions are executed. When premises are examined by the inference engine, conclusion is executed if and only if the information supplied by the user satisfies the conditions part of the rules. Conditions are expressions involving attributes of socio-demographic, medical eligibility criteria and lifestyle of a client collectively together rules.

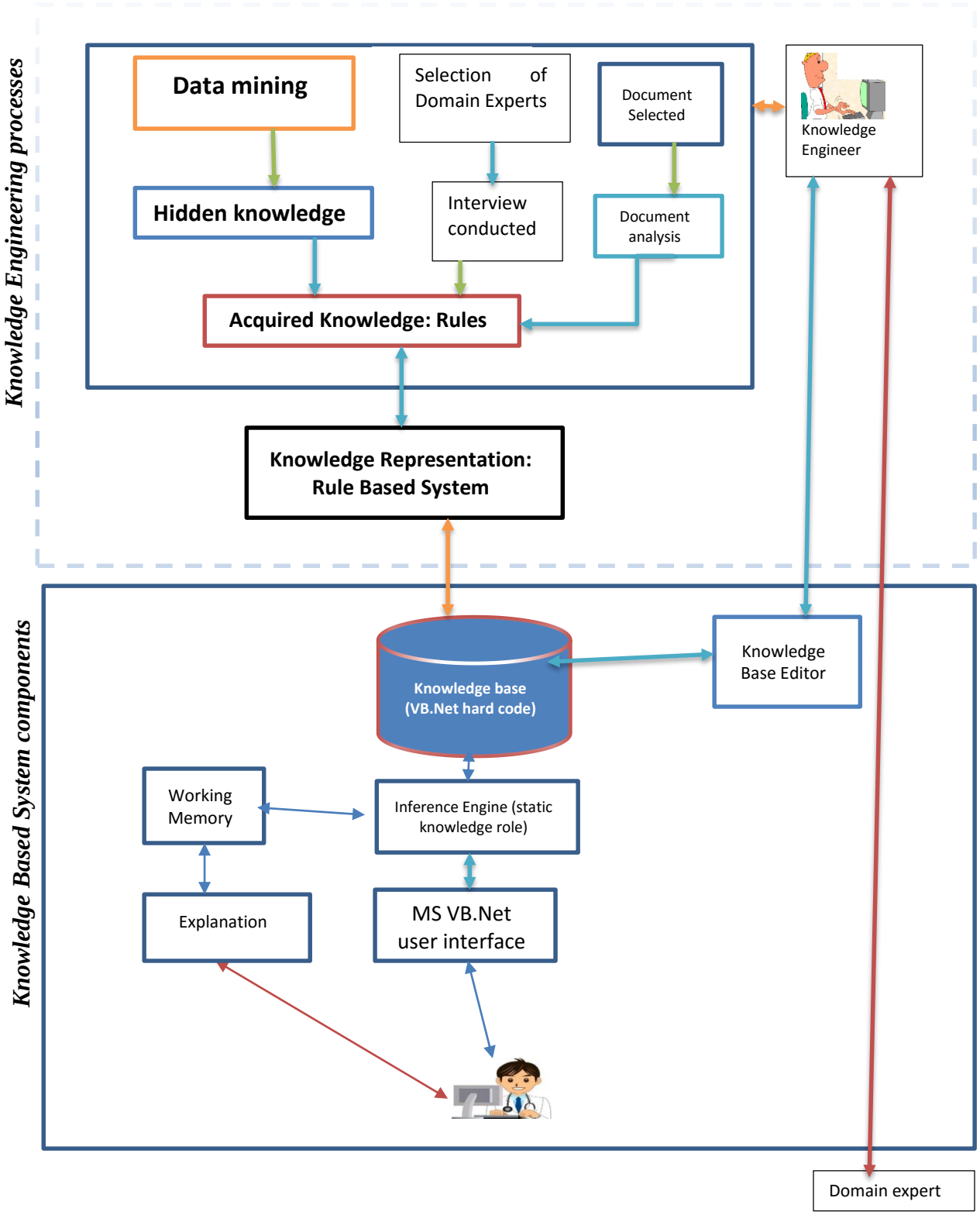


Figure 12: Architecture of the Choice of Contraceptive Method proposed system

These rules are the bases of knowledge base construction. Sample example of this if-then rules applied in Microsoft visual basic .net application looks as follows:

If Age group = early youth AND

If marital status=married AND

If Number of children=medium family size AND

If Residence=Urban AND

If Education=none/Some primary

Then IUCD would be best match to the client.

This is mean to that a married woman and her age is between 20-24 who lives in urban place and has 3 to 4 children is high likely to prefer IUCD method.

As the result, the knowledge base of the prototype knowledge based system consists of rules generated by data mining model and based on medical eligibility criteria and lifestyle of the clients.

5.1.2. Inference Engine

The inference engine mimics the human experts' reasoning process. It achieves this by searching through knowledge base to find rules whose premises match what the user enter as input. This process continues until the inference mechanism is unable to match any premise with the inputs in the working memory. During the choice of contraceptive method processes, the inference engine starts from the medical eligibility criteria and lifestyle then socio demographic characteristics to prove the hypothesis i.e. contraceptive methods preference. If certain antecedents (socio demographic, life style and health condition) are evaluated as true, then it logically follows the consequents are proved and preferred contraceptive method to the client will be selected.

5.1.3. Working Memory

The working memory of the prototype knowledge based system holds user input that characterize preference of contraceptive methods. Functionally, this user input is used to match against the "if

(condition) part of the rules in knowledge base. It holds the information to the asked questions and the questions to be asked to find the best match contraceptive methods. In addition, the user input first matched with the cases represented in if-then form specifically the condition part (premises) of the rules. Then if the user input are proved to be true, the inference engine relates the same rule represented in the knowledge base. Then appropriate knowledge is searched and displayed for the users.

5.1.4. Service provider Interface

In this knowledge based system, the users interact with the system by which the system asks the user a series of questions on socio demographic data, medical eligibility criteria and life style condition of a client and the user responds either imputing the data or selecting from the list box on the basis of question. Finally, the system offers conclusions for the user request by communicating through the service provider interface.

5.1.5. Explanation

The explanation facility is added to give detail explanation about selected contraceptive method. This feature will help the service provider to counsel on the method to the level of details. This includes, the effectiveness of the method, benefit of the method and side effects of the method.

5.2. Rules Conversion

The objective of this study is to integrate results of data mining models, medical eligibility criteria and lifestyle of a client with knowledge base. To achieve this, Micro Soft Visual Basic .Net has been used through understanding J48 decision tree rules of data mining models, decision tree of medical eligibility criteria and life style of the client.

Data mining result is predicted by J48 decision tree and the result is represented as tree structure.

```
Age group = Early Youth
|  Marital Status = Married
|  |  Number of Children = Medium Size Family
|  |  |  Residence = Urban
|  |  |  |  Education = None/some prim: IUCD (1449.37/249.07)
|  |  |  |  Education = Tertiary/higher
|  |  |  |  |  Region = Tigray
|  |  |  |  |  |  religion = Christian: IUCD (34.54/13.64)
```

Figure 13: Sample tree structure of J48 model

The researcher formulates this tree structure as a form of condition and conclusion as like table 20.

Table 20: Sample rules' condition and conclusion extracted from decision tree

Rule	Condition	Conclusion
1	(age group=early youth **and** marital status married **and** number of children = medium size family **and** residence=urban **and** education = none/some primary) if and only if medication <>antiretroviral therapy OR HIV <> Yes OR PID <> Yes OR Sexual Partner <> Multiple	IUCD
2	(Age =Adult) **AND** (Number of Children =High family size) **AND** (Last Method Used=has never used before) if and only if Fertility plan <>spacing OR Sexual Partner <> Multiple	BTL
3	(age group=adolescent) **AND** (marital status=married) **AND** (education=none/some primary) **AND** (number of children=low family size) if and only if vaginal bleeding <> Yes OR Breast Cancer<>Yes OR Fertility plan <>limiting OR Sexual Partner <> Multiple	Implant
4	(age group=youth) **AND** (number of children=no children) **AND** (marital status=single) **AND** (education=comp primary) if and only if Hypertension<>"140-159/90-99" OR Headache <> migraine with aura OR Postpartum OR liver disease<>liver tumor OR Venus<>Acute OR perceived effect <> not at all OR fertility plan <> limiting or Sexual Partner <> Multiple	Injection-depo
5	(Age group=adult) **AND** (Number of children=high family size) **AND** (religion=Muslim) **AND** (education=tertiary/higher) if and only if Hypertension<>"140-159/90-99" OR Headache <> migraine with aura OR Postpartum OR liver disease<>liver tumor OR Venus<>Acute OR perceived effect <> not at all OR fertility plan <> limiting or Sexual Partner <> Multiple	Pills

Table 18 can read as that age group=early youth and marital status married and number of children = medium size family and residence=urban and education = none/some primary then clients preference would be IUCD if and only if the client medication is not antiretroviral therapy or HIV stage is not 3 or 4 or non PID and has multiple sexual partner. In other words, when conditions are fulfilled the conclusions (IUCD) would be draw.

This statement is not directly understand by Microsoft visual basic.Net development tool. Hence, this should be changed as if- then statement which can be understand by the VB.Net. So that, the above statement can be converted as:

```
        If ClientAge = "EARLY_YOUTH" And
         MaritalStatus = "Married" And
        NumebrOfChildren = "MEDIUM_FAMILY_SIZE" And
         Residence = "Urban" And
        Education= "None/Some Primary" Then
        PreferredMethod = "preferred Method to this client is: IUCD"
        End if
    End if
```

Figure 14: Sample rule converted by Microsoft VB.Net

Rules generated from the selected model which concludes the preferred method to the client, which are listed in chapter three, are converted to VB.Net in the same pattern.

Likewise, rules generated from medical eligibility criteria and life style are represented in the same pattern so the VB.net understands in the same manner.

5.3. Integration of Acquired Knowledge

The acquired knowledge from data mining, medical eligibility criteria and lifestyle of a woman needs to integrate together to form the knowledge based system.

Once knowledge is acquired and represented separately, the researcher discussed with domain experts to decide which source of knowledge will come first and continue to the next. Thus, medical eligibility criteria and lifestyle of a woman come first simultaneously followed by socio demographic variables. Among many attributes of MEC and lifestyle, a single attribute allow or deny contraceptive method even if the socio demographic attributes allows to use it to a woman. For example; a data mining result showed that a married woman and her age is between 20-24 who lives in urban place and has 3 to 4 children is high likely to use IUCD. But, if a client is STI of gonorrhea chlamydia or a client has multiple sexual partner then she couldn't receive IUCD.

The integration of these acquired knowledge structures looks:

If ***MEC*** *allow a method and*

If ***Lifestyle*** *allow a method*

Then ***Socio demographic*** *variables will execute*

Where socio demographic variables are executed

If first premises true

Then Contraceptive method will be displayed

Else second premises executed

Then contraceptive method will chosen

Else No method will be chosen

The condition will be executed in all condition and if there is no true premises then the KBS will execute no choice of contraceptive method.

The above example can be executed as follows

```
If MedicationGroup = "Antiretroviral therapy" Or
        HIVGroup = "Yes" Or
        STIGroup = "Gonorrhoea Chlamdia" Or
        STIGroup = "Increased risk of STIs" Or
        PIDGroup = "Yes" Or SepsisGroup = "Yes" Or
        PostPartumGroup = "48hours to <4weeks" Or
        VaginalBleedingGroup = "Yes" Or
        CervicalCancerGroup = "Yes" Or                    } Medical Eligibility Criteria
        SexualPartnerGroup = "Multiple" Then              } LifeStyle
        PreferredMethod = "Her health Condition and/or life style doesn't allow to use IUCD"
    Else
            If ClientAge = "EARLY_YOUTH" And
             MaritalStatus = "Married" And
            NumebrOfChildren = "MEDIUM_FAMILY_SIZE" And
             Residence = "Urban" And                      } Socio Demographic
            Education= "None/Some Primary" Then
            PreferredMethod = "preferred Method to this client is: IUCD"
            End if
        End if
```

Figure 15: Integration of acquired knowledge from different sources

The above example tells how medical eligibility criteria, lifestyle and socio demographic variables of a woman are integrated to form the contraceptive method choice knowledge based system.

5.4. Implementation of the KBS

Knowledge is generated from family planning data set, documents and expert knowledge so as knowledge base is constructed automatically as rules and facts which can be parsed via selected knowledge representation tool Microsoft VB.NET. The system, named as Contraceptive Method Match Knowledge Based System (CMMKBS), which supports the health provider in choosing contraceptive method to the client which fits to her lifestyle. In addition, it provides general information for the user about the selected contraceptive method.

When CMMKBS system displays the contraceptive method, it also tells to the chosen method effectiveness, side effects and advises to be taken by the client and this can be is implemented as module and has three. These are knowledge base module, Preference asker module and method explanation module.

Table 21: Module list of CMMKBS

No	Module Name	Description
1	Rule base	It contains a set of rules about contraceptive user's socio demographic, health condition and life style
2	Preference asker module	Is responsible for presenting questions to the health service provider in the form of variables' value
3	Module explanation module	It tells us the description about the method, advantage and disadvantage and side effect of the method.

The rule base module: the rule base is a collection of rules which are generated from data mining, documents and expert knowledge. After a discussion conducted with the domain experts, the researcher selected relevant 15-20 rules which are fit with real world practice of contraceptive methods from data mining model. In addition, for each method, two rules are generated from domain experts and two rules from documents. The rules first check the medical eligibility criteria and lifestyle of a client the will check the socio demographic data of a client. For example, the socio demographic data fulfill for the selection of IUCD but the medical eligibility prohibits to receive IUCD then the result will not be IUCD. See details of rules (codification of each method) of each contraceptive method in appendix III.

The preference asker module: This module creates an interaction with the health service provider. The module is designed in a way that key variable's response are listed out or receive input from the user. Figure 16 shows key preference asker module interface.

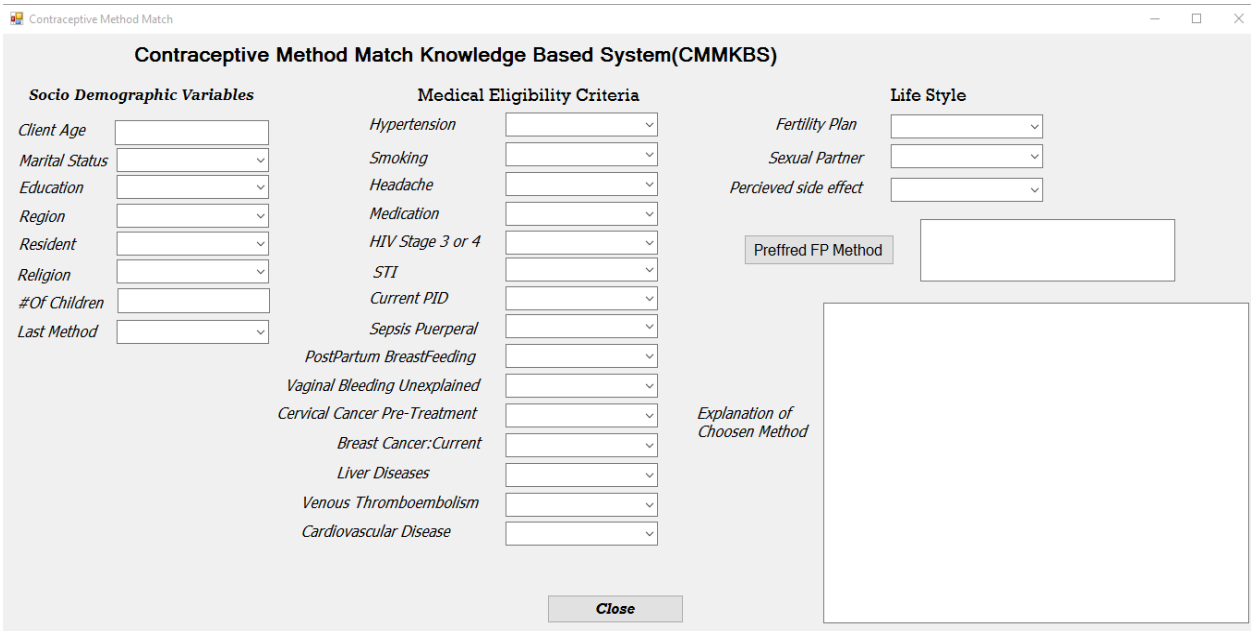


Figure 16: User interface of the CMMKBS

The explanation Module: Based on the variable response, it will match with rules from the rule base. If the preference asker response matches with the rule, then basic contraceptive method of that rule will be displayed. The explanation about the selected method would be presented. The main purpose of this module is present basic information of the method, advantage, disadvantage and side effects of preferred method. In figure 15, the method is displayed in the right bottom corner of the interface as "preferred method" and "explanation of chosen method" which is the function of this module.

Figure 17 and figure 18 showed below tells us that sample screen shoot of the system.

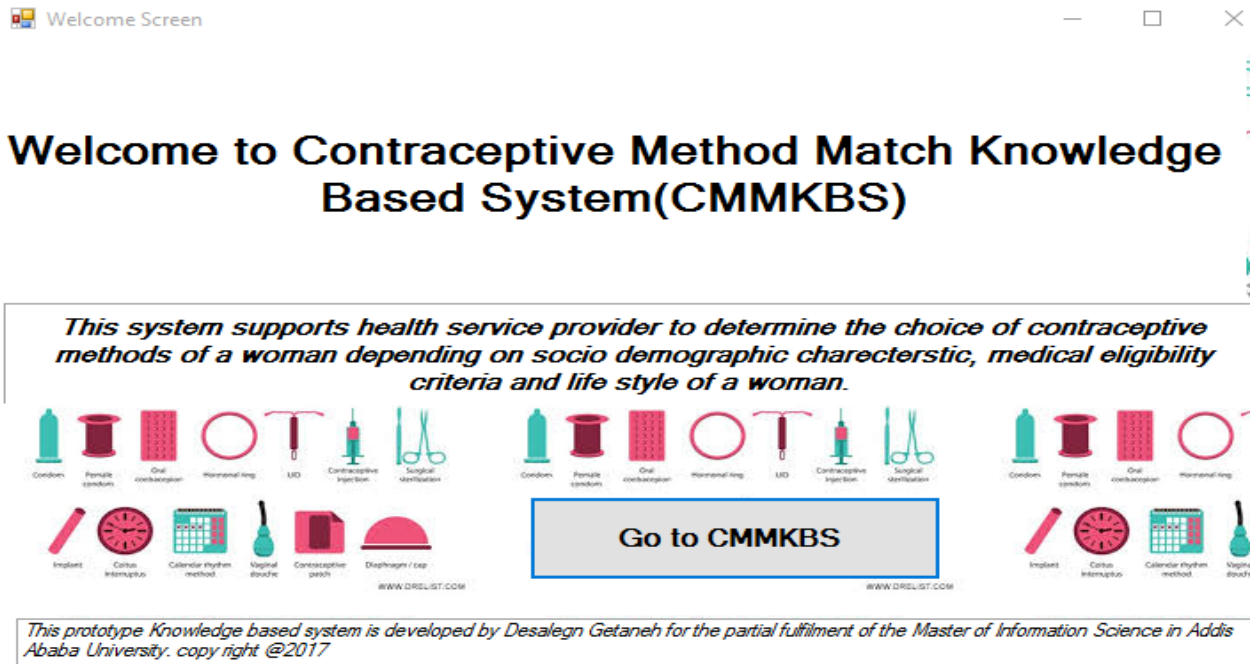


Figure 17: Welcome screen of the designed KBS system

Figure 17 describes that sample choice of contraceptive methods using the combination criteria of socio demographic, medical eligibility and life style of a woman.

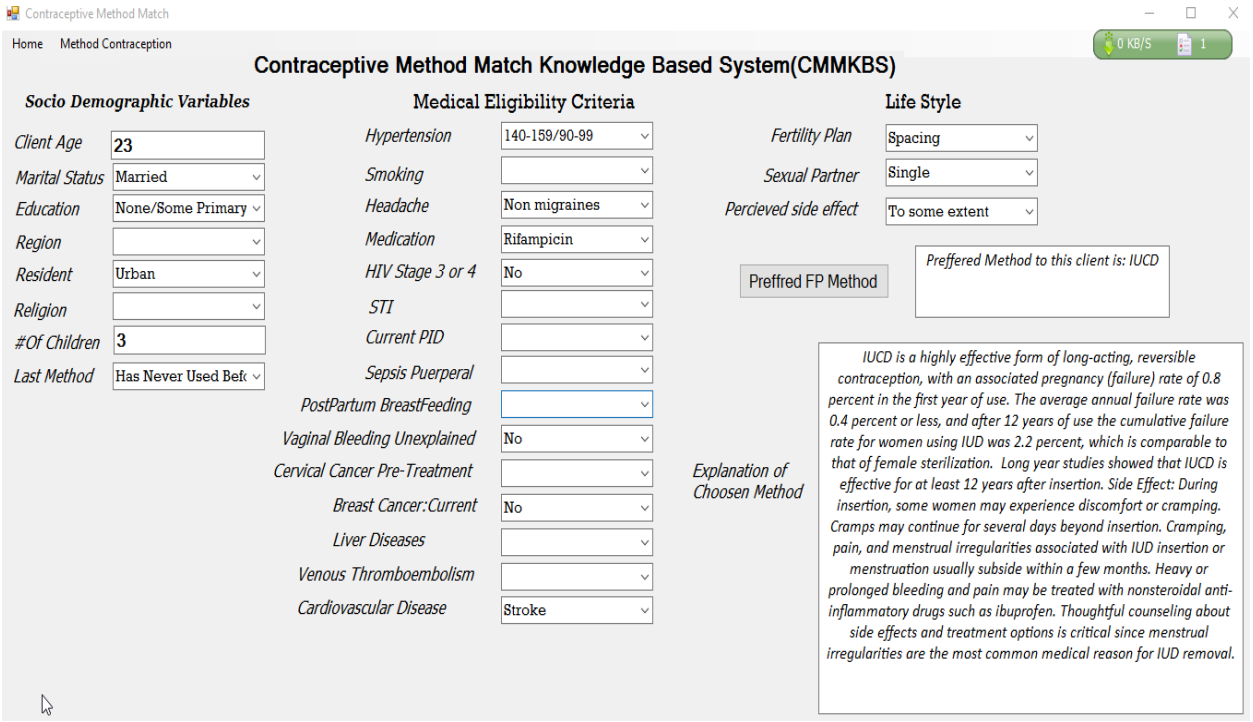


Figure 18: Sample contraceptive method match KBS result

5.5. Testing and Evaluation

Evaluation of the prototype knowledge based system is an important phase that helps to measure the performance of the developed system. In this study, CMMKBS is evaluated to recognize accuracy and efficiency of contraceptive method choice provided by the system. The evaluation conducted to confirm the user acceptance and capability of the system since this is a key point for this research. In addition, it also help to check whether the objective of this research work is achieved.

Therefore, CMMKBS of this research work is evaluated both performance evaluation of the system and user acceptance test.

5.5.1. System performance Evaluation

System performance testing is done by preparing test cases. The test cases include samples of contraceptive method user data set, medical eligibility criteria and domain expert instances. The instances include 30 attributes with their respective values without the dependent class is given to the domain expert to confirm the test cases are provided the choice of contraceptive methods as BTL, IUCD, Implant, Injection-depo, pills. Considering the time it consumes to classify it manually, the researcher prepared only 30 instances for system performance testing. Based on the attributes and their respective value, domain experts (5 of the domain experts were participated) classify the instances. The same set of test instances are provided to CMMKBS and the outputs are compared with the domain experts 'contraceptive method recommendations. The result is discussed as follows using confusion matrix in table 24.

Table 22: Confusion matrix to test the system performance

	CMMKBS Choice of contraceptive Method						
Domain Expert Choice of contraceptive method		BTL	IUCD	Implant	Injection-depo	Pills	Total
	BTL	**2**	1	0	0	0	3
	IUCD	0	**5**	1	1	0	7
	Implant	0	0	**8**	0	0	8

	Injection-depo	0	0	0	**7**	0	7
	Pills	0	0	0	1	**4**	5
	Total	2	6	9	9	4	30

Table 21 depicts that out of 30 instances the domain expert grouped as 3 of them as BTL, 7 as IUCD, 8 as Implant, 7 of them as injection depo and 5 as pills. While the system classified as 2 of them as BTL, 6 of them as IUCD, 9 of them as Implant, 9 of them as injection-depo and 4 of them as pills. This shows that 2 of IUCD clients are incorrectly classified i.e. two instances are classified as false negative for IUCD and as false positive for injection and depo.

In general the system classified as 26 of them as correctly classified instances and 4 of them are incorrectly classified instances with an overall accuracy of 84%. Table 23 shows the performance measurement of the system.

Table 23: System performance measurement result

No	Performance Measurement	Result
	Accuracy	86.66
	TP rate	0.90
	FP Rate	0.13
	Precision	0.900
	Recall	0.866
	F-Measure	0.867
	ROC curve	0.78
	CCI	26
	ICI	4

In Table 23, we observed that domain experts' choice of contraceptive method is highly similar with CMMKBS choice of contraceptive method. Thus, it is encouraging to use the system for this purpose.

5.5.2. User Acceptance Test

The user acceptance test is key and more important in any system development. Thus, the researcher tested the CMMKBS how this system is performing in choice of contraceptive at the end of user perspective.

Five domain experts (due to time constraint the researcher couldn't include more domain experts) are participated in testing the system. The researcher provided explanation on how the system works. Then after domain experts test their own test cases to evaluate the system with a minimum of 10 cases. To minimize the subjectivity of the user test, the researcher adopt the following standards [91]. These are:

In order to test the user acceptance, the researcher adopts the questionnaire developed by [91]. Accordingly, the test is done using likert scale, as excellent, very good, good, fair and poor and scored as 5,4,3,2,1 respectively.
The questionnaire includes:

- Interaction with system and simplicity to use
- Efficiency in time
- System attractiveness
- The accuracy of the system in choosing contraceptive method that match with the client lifestyle, socio demographic and medical eligibility criteria.
- The explanation given to selected contraceptive use
- The applicability of the system in the domain area

The details of the user acceptance test result is compiled in table 24.

Table 24: User Acceptance test result

no	**Criteria**	poor	Fair	Good	Very good	Excellent	Average	Percent
1	Interaction with system and simplicity to use	0	0	1	2	2	4.2	84%
2	Efficiency in time	0	0	1	2	2	4.2	84%
3	System attractiveness	0	0	2	2	1	3.8	76%

4	The accuracy of the system in choosing contraceptive method that match with the client preference.	0	1	2	1	1	3.4	68%
5	The explanation given to selected contraceptive use	0	0	2	2	1	3.8	76%
6	Applicability of the system to the domain area	0	0	3	2	0	3.4	68%
	Total						3.8	76%

Table 23 depicts that user acceptance rate of the CMMKBS is 76 percent. Except response of question #4 & #6, others result is high score and encouraging.

In general, system performance test and user acceptance test of the system is acceptable and encouraging. The researcher believe that the system can be applicable to the domain area.

Chapter Six

Conclusion and Recommendation

6.1. Conclusion

Ethiopia is one of the most populous country in Africa with a fertility rate of 4.6. Eight percent of pregnancy is unwanted. This contributes for maternal mortality and child death. To make every child is wanted, family planning plays an essential role through delaying, spacing or limiting birth. But the CPR of Ethiopia is low and needs to work hard to address the demand of family planning. There are different reasons for the low uptake of family planning like demographic characteristics and poor counseling. Though there are different job aids to support the counseling, they are mainly focus on medical eligibility criteria not by socio demographic variables.

Empirical research design is employed to answer the research question and prototyping approach was used to develop the KBS. Moreover, as a research process model, knowledge engineering phases were followed.

Thus, in this study, data mining technique were employed to identify best classifier algorithm and key determinant variables for choice of contraceptive method. A hybrid data mining model is used for the knowledge discovery. Hence, a decision tree J48 algorithm is best classifier in family planning dataset to identify key determinants for the choice of contraceptive methods. According to the model, client age, number of children, education, residence, marital status, religion, region and contraceptive history determine the choice of contraceptive methods. Specific to contraceptive method; BTL is high depend on client age, number of living children, residence and education. IUCD is highly depend of religion of the client whereas implant depend on client age, religion, and region. Pills and injection-depo are also determined by education, religion, region, last method. Moreover, medical eligibility criteria and life style of a woman are also factor in choice of contraceptive methods.

To support the health service provider for the choice of contraceptive methods, knowledge based system called CMMKBS is designed through acquiring the life style, socio demographic and medical eligibility criteria of a woman. Rule based knowledge representation system is used to

represent the acquired knowledge. Finally, codification has done using Microsoft visual basic .net application to provide the choice of contraceptive methods.

Finally, System performance evaluation and user acceptance test has been conducted to evaluate the designed system performance. As a result, 86.6 % of accuracy was scored for system performance and 76% was scored for user acceptance test. So, the research conclude that integrating data mining rules, explicit and tacit knowledge is possible to develop knowledge based system can be implemented in the domain area.

6.2. Recommendation

Based on the finding, the researcher recommends the following points.

- Since religion has factor for the choice of contraceptive methods especially for IUCD, family planning program implementer can design intervention to address this gap.
- Socio demographic data like partner occupation and wealth index which are determinants for uptake of contraceptives were not included this thesis dataset. In the future study, experimenting the mining algorithm including this attribute might increase the performance of the prediction algorithm
- This CMMKBS has no learning component. So, researchers can work in this component in the future to make the system learn new facts.
- This study focus on supporting service provider in choice of contraceptive method and designed KBS system is desktop application. In the future, researchers can design web based KBS as well as support directly the client without the involvement of service provider to choose contraceptive method.

References

[1] Department of Economic & Social Affairs Population, "World Population Prospects: The 2015 Revision," World Health Organization, New York, 2015.

[2] O. W. N. Online, "one nation," One World Nation Online, 01 January 2016. [Online]. Available: http://www.nationsonline.org/oneworld/ethiopia.htm. [Accessed 18 July 2017].

[3] Cenral Statistics Agency, "Ethiopian Demographic and Health Servey," The DHS Program ICF,, Rockville, Maryland, USA, 2016.

[4] Inter-Agency Group, Maternal mortality in 1990-2015: Ethiopia, Nework: WHO, UNICEF, UNFPA, World Bank Group, and United Nations Population Division, 2015.

[5] Yirgu Gebrehiwot, Ann Moore, Tamara Fetters, Yohannes Dibaba Wado, Akinrinola Bankole, Susheela Singh, Hailemichael Gebreselassie and Yonas Getachew, "The Estimated Incidence of Induced Abortion In Ethiopia," Guttmacher Institute, Addis Ababa, 2014.

[6] Family Planning and Population, Division of RH, Health Benefits of Family planning, Geneva: World Health Organization, 1995.

[7] House of Peoples Represenatives, "Proclamation of the Constitution of the Federal Democratic Rebublic of EThiopia," *Federal Negarit Gazeta*, vol. 01, no. 01, pp. 1-38, 1995.

[8] Tedros Adhanom Ghebreyesus(PhD),, "NATIONAL GUIDELINE FOR FAMILY PLANNING SERVICES IN ETHIOPIA," Federal MInistry of Health, Addis Ababa, 2011.

[9] Jiawei Han, Micheline Kamber and Jian Pei, Data Mining: Concepts and Techniques Third Edition, Newyork: Morgan Kaufmann, 2012.

[10] Tutorialspoint, "Learning Data mining: Data Mining Tasks," Tutorial Point, 01 01 2017. [Online]. Available: http://www.tutorialspoint.com/data_mining/dm_tasks.htm. [Accessed 03 08 2017].

[11] M. Durairaj and V. Ranjani , "Data Mining Applications In Healthcare Sector," *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH*, vol. 2, no. 10, pp. ISSN 2277-8616 , 2013.

[12] V. V. Kulkarni and Megha M Muttin, "Knowledge based Systems, a support to the Preventive Maintenance," *International Journal of Innovations in Engineering and Technology (IJIET)*, vol. 4, no. 1, pp. 104-111, 2014.

[13] Hussein H. Owaied, Mahmoud Malek Abu and Hazim A. Farhan, "An application of Knowledge Based System," *IJCSNS International Journal of Computer Science and Network Security,* vol. 1, no. 3, pp. 208-213, 2010.

[14] Tipawan Silwattananusarn and KulthidaTuamsuk, "Data Mining and Its Applications for Knowledge Management: A Literature Review from 2007 to 2012," *International Journal of Data Mining & Knowledge Management Process (IJDKP),* vol. 2, no. 5, pp. 14-24, 2012.

[15] Central Statistical Agency, "Ethiopia 2017 Population and Housing Census Support," Central Statistical Agency;, Addis Ababa, 2016.

[16] Central Statistical Agency , "Population Projection of Ethiopia for All Regions At Wereda Level from 2014 – 2017," Central Statistical Agency ;, Addis Ababa, 2013.

[17] Farrokh-Eslamlou H, Aghlmand S, Khorasani-Zavareh D, Mohammad Alizadeh Charandabi S, Moghaddam Tabrizi F and Jahanfar S, "Structured versus routine family planning counselling for contraception (Protocol), Issue 7," Cochrane Database of Systematic Reviews, 2014.

[18] Central Statistical Agency, "Ethiopia Demographic and Health Survey," Central Statistical Agency;, Addis Ababa, 2005.

[19] Beniot Kasala, A decade of changes in contraceptive use in Ethiopia: in depth analysis, Addis Ababa: UNFPA, 2012.

[20] School of Public Health, "Performance Monitoring and Accountability 2020 (PMA2020): Ethiopia," Bill & Melinda Gates Institute for Population & Reproductive Health, Addis Ababa, 2015.

[21] Teshager Mersha, "Client Satisfaction Serveys," Marie Stopes International Ethiopia, Addis Ababa, 2015.

[22] Alexandra Alvergne and Eshetu Gurmu, "Contraceptive discontinuation and method choice in Ethiopia," Population Association of America, 2016.

[23] Abrham G/Giorgis, "MSc Thesis on Apllication of Data mining to predict the likelihood of contraceptive method use among women aged 15-49," Addis Ababa Univesity, Addis Ababa, 2012.

[24] K4Health, "Family Planning: A global handbook for providers," Knowledge for Health, 12 01 2017. [Online]. Available: https://www.fphandbook.org. [Accessed 06 07 2017].

[25] WHO, "Medical Eligibility Criteria Wheel," Worled Health Organization Press, 2015. [Online]. Available: http://apps.who.int/iris/bitstream/10665/173585/1/9789241549257_eng.pdf?ua=1. [Accessed 01 06 2017].

[26] R. Anand and U. D. Jeffrey, Mining of Massive datasets, California: Stanford University, 2010/11.

[27] Mümine KAYA KELEŞ, "AN OVERVIEW: THE IMPACT OF DATA MINING APPLICATIONS ON VARIOUS SECTORS," *TECHNICAL JOURNAL ISSN 1846-6168,* vol. 11, no. 3, pp. 128-132, 2017.

[28] M. J. Berry and G. S. linoff, Data Mining Tehcniques for marketing, sales and customer relashinship management 2nd editionn, Indiana: Wiley publishing, 2004.

[29] A. Azevedo and M. . F. Santos, "KDD, SEMMA AND CRISP-DM: A PARALLEL OVERVIEW," *presented at the Parallel overview ,* no. ISBN: 978-972-8924-63-8, 2008.

[30] Usama Fayyad, Gregory Piatetsky-Shapiro and Padhraic Smyth , "From data mining to knowledge discovery in Databases," *AI Megazine,* vol. 17 , no. 3, 1996.

[31] K. LUKASZ A. and P. MUSILEK, "A survey of Knowledge Discovery and Data," *The Knowledge Enginerring Review,* vol. 21:1, no. 10.1017/S0269888906000737 , pp. 1-24, 2006.

[32] Affliate*, "KDD, SEMMA AND CRISP-DM: A PARALLEL OVERVIEW," 2008.

[33] Ermias Abebe, *MSc Thesis of Predicting the status of HIV Patients during Drug Regimen Change:,* Gondar: University of Gondar, 2015.

[34] J. Jackson, "DATA MINING: A CONCEPTUAL OVERVIEW," *Communications of the Association for Information Systems,* vol. 8, pp. 267-296, 2002.

[35] Maryam Daneshmandi and Marzieh Ahmadzadeh , "A Hybrid Data Mining Model to Improve Customer improve modeling," *Maryam Daneshmandi et.al / Indian Journal of Computer Science and Engineering (IJCSE),* vol. 3, no. ISSN : 0976-5166, 2012.

[36] K. CIOS WITOLD, P, ROMAN and S AND KURGAN, A, The Knowledge Discovery Process, USA: Springer, Inc, 2007.

[37] D. mining, "Data mining Tasks," Tutorials Point, 01 06 2016. [Online]. Available: https://www.tutorialspoint.com/data_mining/dm_tasks.htm. [Accessed 08 05 2017].

[38] WideSkill, "Data Mining Task," WideSkill.com, 2015. [Online]. Available: http://www.wideskills.com/data-mining-tutorial/05-data-mining-tasks. [Accessed 16 05 2017].

[39] Vipin Kumar, Introduction to Data Mining, Minnesota: University of Minnesota, 2006.

[40] Mustapha Belouch, Salah El Hada and Mohamed Idhammad, "A Two-Stage Classifier Approach using RepTree algorithm for Network Intrusion Detection," *(IJACSA) International Journal of Advanced Computer Science and Applications,* vol. 8, no. 6, pp. 389-394, 2017.

[41] Vaishali S. Parsania, N. N. Jani and Navneet H Bhalodiya, "Applying Naïve bayes, BayesNet, PART, JRip and OneR Algorithms on Hypothyroid Database for Comparative Analysis," *NTERNATIONAL JOURNAL OF DARSHAN INSTITUTE ON ENGINEERING RESEARCH & EMERGING TECHNOLOGIES,* vol. 3, no. 1, pp. 61-64, 2014.

[42] S.Vijayarani and M.Divya, "An Efficient Algorithm for Generating Classification Rules," *International Journal of Computer Science And Technology,* vol. 2, no. 4, pp. 512-515, 2011.

[43] Wikipedia, "Knowledge Base," Wikipedia, 26 July 2017. [Online]. Available: https://en.wikipedia.org/wiki/Knowledge_base. [Accessed 10 August 2017].

[44] Technopedia, "Knowledge based system definition," Technopedia, 01 01 2017. [Online]. Available: https://www.techopedia.com/definition/7969/knowledge-based-system-kbs. [Accessed 31 06 2017].

[45] Krishnamoorthy C.S. and Rajeev S., "Artificial Intelligence and Expert Systems," CRC Press LLC , 1996.

[46] Ejigu Tefera, "MSc Thesis on developing knowledge based system for cereal crom diagnosis and treatment: the case of Kulumsa Agriculture Center," Addis Ababa University, Addis Ababa, 2012.

[47] Raman R.V. and K. Prasad V.K., "Applications of Knowledge Based Systems in Mining," in *Proceedings of the Twentieth International Symposium on the Application*, Johannesburg, 1987.

[48] Chee-Fai TAN, "A Prototype of Knowledge-Based System for Fault Diagnosis inAutomatic Wire Bonding Machine," *Turkish J. Eng. Env. Sci.,* vol. 32, pp. 235-244, 2008.

[49] Reid G. Smith, "Knowledge-Based Systems Concepts, techniques and examples," in *Canadian High Technology Show*, Lansdowne Park, Ottawa, 1985.

[50] A. K, Mohamed and Lious F, Knowledge Based approach and systesm, World acadamy of science and engineering and technology, 2005.

[51] David Poole, Alan Mackworth and Randy Goebel, Computational Intellegence: A logical approach, Newyork: Oxford University Press, 1998.

[52] C Chakraborty, "www.myreaders.info," www.myreaders.info, 01 06 2010. [Online]. Available: http://www.myreaders.info/07_Expert_Systems.pdf. [Accessed 28 08 2017].

[53] Brian R. Gaines and Mildred L. G. Shaw, Eliciting Knowledge and transferring it effectively to a knowledge based system, Alberta: Knowledge Science Institute: University of Calgary, 1992.

[54] Abdulkerim Mohammed, "MSc Thesis on Towards Integrating Data Mining with Knowledge Based System: The Case of Network Intrusion detection," Addis Ababa University, Addis Ababa, 2013.

[55] C.S. Krishnamoorthy and S. Rajeev, Artificial Intelligence and Expert Systems for Engineers, CRC Press LLC, 1996.

[56] Priti Srinivas Sajja and Rajendra Akerkar, "Advanced Knowledge Based Systems: Model, Application and Research," in *Advanced Knowledge Based Systems: Model, Application and Research*, India, TMRF e-Book, 2010, pp. 1-10.

[57] Kamran Parsaye and Mark Chignell, Expert Systems for Experts, Newyork: John Wiley & Sons, Inc., 1998.

[58] Yuliadi Erdani, Acqueisition of Human Exper Knowledge for rule based knowledge based systems Using Ternary Grid, Germany: Univeristy of Duisburg-Essen, 2005.

[59] De Kock, E, Chap 6: Expert Systems and knowledge acqueisition, Pretoria: University of Pretoria, 2003.

[60] DeJen Alemu Abtewe, Msc Thesis on APPLICATION OF KNOWLEDGE BASED SYSTEM FOR WOODY PLANT SPECIES IDENTIFICATION, Addis Ababa: Addis Ababa University, 2009.

[61] Mahalakshmi G.S and Geetha T.V, "Chapter 2: Representing Knowledge Effectively Using Indian logic," in *Advanced Knowledge Based Systems: Model, Applications and Research V1*, TMRF e-Book, India, 2010, pp. 22-28.

[62] LAWRENCE B. HOLDER, ZDRAVKO MARKOV and INGRID RUSSELL, "ADVANCES IN KNOWLEDGE ACQUISITION AND REPRESENTATION," World Scientific Publishing Company, Arlington, 2006.

[63] Namarta Kapoor and Nischay Bahl, "Comparative Study of Forward and Backward Chanining in Artificial Intellegence," *International Journal Of Engineering And Computer Science*, vol. 5, no. 4, pp. 16239-16242, 2016.

[64] Janet L. Kolodner, "An Introduction to Case-Based Reasoning," in *Artificial Intelligence Review* , vol. 6, College of Computing, Georgia Institute of Technology, Atlanta, GA 30332-0280, USA, Morgan-Kaufmarm Publishers, Inc, 1992, pp. 3-12.

[65] J. L. Crowley, "Intelligent Systems: Reasoning and Recognition," 2012.

[66] K. Morgan, "Knowledge Representation: an Introduction to Artificial Intellegence:," in *Prentice Hall*, London, UK, 2004/05.

[67] M. Mehdi Owrang O, "Database systems techniques and tools in automatic knowledge," in *Knowledge-Based Systems Vol 1*, Washington, United States of America:Acadamic Press, 2008, pp. Chap 8: pp 201-248.

[68] Enrico Coiera, Robert Baud, Luca Consoles, Jorge Cruz, John Durnick, Pierre Frutiger, Peter Hucklenbroich, Anthony Rickards and Klaus Spitzer, The Role of Knowledge Based Systems in Clinical Practices, 1993.

[69] Vincent Lemaire, Carine Hue and Olivier Bernier , "Correlation Analysis in classifiers for contraceptive use dataset," ResearchGate, France, 2010.

[70] Aalok Ranjan Chaurasia, "Contraceptive Use in India: A Data Mining Approach," *International Journal of Population Research*, vol. 2014, no. Article ID 821436, p. 11 pages, 2014.

[71] Ma. Dolores C. Tongco, "Purposive Sampling as a Tool for Informant Selection," *A Journal of plants, people, and applied Research Ethnobotany Research & Applications*, vol. 5, pp. 148-157, 2007.

[72] Richard Adanu et el, Medical Elgibility criteria for medical use 5th edition, Geneva: World heath Organization, 2015.

[73] Dr.habil.sc.ing and professor Janis Grundspenkis, "Fundamentals of Artificial Intellegence: Knowledge Representation and Networked Schemes," Faculty of Computer Science and Information Technology, Riga Technical University, Riga, 2006.

[74] P-H. Speel, A. Th. Schreiber, W. van Joolingen, G. van Heijst and G.J. Beijer, "Conceptual Modelling for Knowledge-Based Systems," Encyclopedia of Computer Science and Technology, Marce Dekker Inc., Newyork, 2001.

[75] Kiong Siew Wai, Abd. Latif B. Abdul Rahman, Mohd Fairuz Zaiyadi and Azwan Abd Aziz, "Expert System in Real World Applications," Generation5, 29 05 2005. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=2A4A0CD5626721ECCF1DD016D5BCAD3B?doi=10.1.1.117.9964&rep=rep1&type=pdf. [Accessed 01 09 2017].

[76] Data School, "Simple guide to confusion matrix in machine learining," Data School, 25 03 2014. [Online]. Available: http://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/. [Accessed 09 05 2017].

[77] Paula Miranda, Pedro Isaias and Manuel Crisóstomo, "Evaluation of Expert Systems: The Application of a Reference Model to the Usability Parameter," in *Universal Access in HCI, Part I*, Vols. Springer-Verlag Berlin Heidelberg 2011, Berlin Heidelberg, Springer-Verlag, 2011, pp. 100-109.

[78] L. Kurgan and Krzysztof J. Cios, "Discretization algorithm that uses class-attribute interdependence maximization," in *International conference on Artificial Intelegence*, 2001.

[79] Ian H. Witten and Eibe Frank, Data Mining: Practical Machine Learning Tools and Techniques,Second Edition, 500 Sansome Street, Suite 400, San Francisco, CA 94111: Morgan Kaufmann Publishers, 2005.

[80] Croce Danilo, Decision Tree Algorithm: Weka tutorial, Web Mining e Retrieval , 2010.

[81] Sam Drazin and Matt Montag, "Decision Tree Analysis Using Weka," University of Miami, Coral Gables, 2006.

[82] Anita Wasilewska, Writer, *Testing Classfier Accuracy*. [Performance]. SUNY Stony Brook University, 2011.

[83] Wikipedia, "Data Mining Algorithms In R/Classification/JRip," Wikipedia, 09 01 2016. [Online]. Available: https://en.wikibooks.org/wiki/Data_Mining_Algorithms_In_R/Classification/JRip. [Accessed 17 08 2017].

[84] Solomon GebreMariam, "MSc Thesis on A SELF-LEARNING KNOWLEDGE BASED SYSTEM FOR DIAGNOSIS AND TREATMENT OF DIABETES," Addis Ababa University, Addis Ababa, 2013.

[85] EngenderHealth, "Engender Health," Engender Health, 01 01 2016. [Online]. Available: https://www.engenderhealth.org/our-work/family-planning/long-acting-and-permanent-methods.php. [Accessed 20 08 2017].

[86] P. e. a. Chapman, "CRISP-DM 1.0 - Step-by-step data mining guide," 2000.

[87] Wikipedia, "Attribute Selection," 05 09 2017. [Online]. Available: https://en.wikipedia.org/wiki/Feature_selection. [Accessed 06 09 2017].

[88] MIS and Research Unit, "MSIE 2016 Annual Performance Report," MIS and Research Unit, Addis Ababa, 2017.

[89] S. Dinakaran and P. Ranjit Jeba Thangaiah, "Role of Attribute Selection in Classification Algorithms," *International Journal of Scientific & Engineering Research*, vol. 4, no. 6, pp. 67-71, 2013.

[90] Osiris Villacampa, "A PhD Dessertation on Feature Selection and Classification Methods for Decision Making: A Comparative Analysis," Nova Southeastern University, 2015.

Appendix I: Checklist for the elicitation of knowledge from the experts

1. What procedure you follow while you make a woman informed choice for contraceptive use?
2. List criteria's where a woman is eligible to use each contraceptive methods.
3. List criteria's where a woman is not eligible to use each contraceptive methods.
4. What is the relationship between these criteria's?

Appendix II: J48 Sample decision tree

=== Run information ===

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

Relation: Family Planning User-weka.filters.supervised.attribute.ClassOrder-R1-C0-weka.filters.AllFilter-weka.filters.supervised.instance.Resample-B0.0-S1-Z100.0-weka.filters.supervised.instance.ClassBalancer-num-intervals10-weka.filters.supervised.instance.ClassBalancer-num-intervals100

Instances: 105205

Attributes: 10

Region

Last Method Used

FTEU

Residence

Age group

Education

religion

Marital Status

Number of Children

Current Method Used

Test mode: split 66.0% train, remainder test

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

J48 pruned tree

Age group = Early Youth
| Marital Status = Married
| | Number of Children = Medium Size Family
| | | Residence = Urban
| | | | Education = None/some prim: IUCD (1449.37/249.07)
| | | | Education = Tertiary/higher
| | | | | Region = Tigray
| | | | | | religion = Chirstian: IUCD (34.54/13.64)
| | | | | | religion = Muslim
| | | | | | | FTEU = Yes: Pills (6.52/1.49)
| | | | | | | FTEU = No: Implant (2.24)
| | | | | | religion = Other: IUCD (1.49)
| | | | | Region = Oromia
| | | | | | religion = Chirstian: IUCD (488.4/171.91)
| | | | | | religion = Muslim
| | | | | | | Last Method Used = Has Never Used Before: IUCD (53.66/11.86)
| | | | | | | Last Method Used = Not currently using but has before: Pills (107.67/48.32)
| | | | | | | Last Method Used = Short term user: IUCD (14.77/5.82)

| | | | | | | Last Method Used = Long term User: IUCD (1.49)
| | | | | | religion = Other
| | | | | | | Last Method Used = Has Never Used Before: IUCD (6.49/2.01)
| | | | | | | Last Method Used = Not currently using but has before: Pills (18.07/2.99)
| | | | | | | Last Method Used = Short term user: Pills (0.0)
| | | | | | | Last Method Used = Long term User: Pills (4.02)
| | | | | Region = Addis Ababa
| | | | | | religion = Chirstian: IUCD (728.33/320.76)
| | | | | | religion = Muslim: IUCD (59.52/28.17)
| | | | | | religion = Other
| | | | | | | Last Method Used = Has Never Used Before: IUCD (5.97)
| | | | | | | Last Method Used = Not currently using but has before: Pills (25.64/1.49)
| | | | | | | Last Method Used = Short term user: Pills (0.0)
| | | | | | | Last Method Used = Long term User: Pills (0.0)
| | | | | Region = Amhara: IUCD (261.15/99.92)
| | | | | Region = SNNPR
| | | | | | religion = Chirstian: IUCD (403.92/57.56)
| | | | | | religion = Muslim: IUCD (51.39/11.08)
| | | | | | religion = Other
| | | | | | | Last Method Used = Has Never Used Before: Pills (8.53/1.49)
| | | | | | | Last Method Used = Not currently using but has before: IUCD (27.91/4.02)
| | | | | | | Last Method Used = Short term user: IUCD (0.0)
| | | | | | | Last Method Used = Long term User: IUCD (0.0)
| | | | | Region = Dire Dawa
| | | | | | Last Method Used = Has Never Used Before: IUCD (10.08/1.12)
| | | | | | Last Method Used = Not currently using but has before: Pills (52.41/24.24)
| | | | | | Last Method Used = Short term user: IUCD (2.99)
| | | | | | Last Method Used = Long term User: Pills (0.0)
| | | | Education = Comp primary
| | | | | religion = Chirstian
| | | | | | Region = Tigray: IUCD (21.17/6.25)
| | | | | | Region = Oromia
| | | | | | | Last Method Used = Has Never Used Before: IUCD (135.4/8.5)
| | | | | | | Last Method Used = Not currently using but has before: IUCD (185.51/31.74)
| | | | | | | Last Method Used = Short term user: IUCD (21.46/0.56)
| | | | | | | Last Method Used = Long term User: Pills (6.52/1.49)
| | | | | | Region = Addis Ababa
| | | | | | | Last Method Used = Has Never Used Before: IUCD (132.28/35.24)
| | | | | | | Last Method Used = Not currently using but has before: IUCD (264.8/103.56)
| | | | | | | Last Method Used = Short term user: Implant (12.31/4.48)
| | | | | | | Last Method Used = Long term User: Implant (3.17/1.49)
| | | | | | Region = Amhara: IUCD (94.64/12.53)
| | | | | | Region = SNNPR: IUCD (316.12/39.93)

| | | | | | Region = Dire Dawa: Pills (7.64/2.61)
| | | | | religion = Muslim
| | | | | | Region = Tigray
| | | | | | | FTEU = Yes: IUCD (4.48)
| | | | | | | FTEU = No: Implant (3.36)
| | | | | | Region = Oromia: IUCD (43.08/20.69)
| | | | | | Region = Addis Ababa: Pills (39.19/21.08)
| | | | | | Region = Amhara: IUCD (10.11/4.14)
| | | | | | Region = SNNPR: IUCD (29.67/7.27)
| | | | | | Region = Dire Dawa: Pills (1.57/0.56)
| | | | | religion = Other
| | | | | | Region = Tigray: IUCD (2.99)
| | | | | | Region = Oromia: Pills (13.53/4.48)
| | | | | | Region = Addis Ababa: Pills (6.0/2.99)
| | | | | | Region = Amhara: IUCD (0.0)
| | | | | | Region = SNNPR: IUCD (29.92/6.04)
| | | | | | Region = Dire Dawa: IUCD (0.0)
| | | | Education = Comp secondary
| | | | | religion = Chirstian
| | | | | | Region = Tigray: IUCD (64.33/25.51)
| | | | | | Region = Oromia: IUCD (319.36/53.62)
| | | | | | Region = Addis Ababa: IUCD (439.84/171.11)
| | | | | | Region = Amhara: IUCD (179.94/42.59)
| | | | | | Region = SNNPR: IUCD (328.02/29.44)
| | | | | | Region = Dire Dawa
| | | | | | | Last Method Used = Has Never Used Before: IUCD (4.1/1.12)
| | | | | | | Last Method Used = Not currently using but has before: Pills (5.32/1.3)
| | | | | | | Last Method Used = Short term user: Implant (0.56)
| | | | | | | Last Method Used = Long term User: Pills (0.0)
| | | | | religion = Muslim
| | | | | | Region = Tigray: IUCD (2.99)
| | | | | | Region = Oromia
| | | | | | | Last Method Used = Has Never Used Before: IUCD (13.47/3.02)
| | | | | | | Last Method Used = Not currently using but has before: Pills (27.59/12.5)
| | | | | | | Last Method Used = Short term user: Pills (5.52/1.49)
| | | | | | | Last Method Used = Long term User: Pills (1.01)
| | | | | | Region = Addis Ababa: Pills (60.68/16.42)
| | | | | | Region = Amhara
| | | | | | | Last Method Used = Has Never Used Before: IUCD (5.97)
| | | | | | | Last Method Used = Not currently using but has before: Pills (11.55/1.49)
| | | | | | | Last Method Used = Short term user: Pills (0.0)
| | | | | | | Last Method Used = Long term User: Pills (2.01)
| | | | | | Region = SNNPR: IUCD (26.87)

| | | | | | Region = Dire Dawa: Pills (4.02)

| | | | | religion = Other

{ Since the tree is long, I couldn't include here; there are missed tree}

=== Summary ===

Correctly Classified Instances	25999.0559	72.3363 %
Incorrectly Classified Instances	9942.8413	27.6637 %
Kappa statistic	0.6541	
Mean absolute error	0.1538	
Root mean squared error	0.2805	
Relative absolute error	48.0523 %	
Root relative squared error	70.1124 %	
Total Number of Instances	35941.8972	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.987	0.023	0.916	0.987	0.950	0.938	0.990	0.975	BTL
	0.785	0.094	0.675	0.785	0.726	0.654	0.920	0.712	IUCD
	0.643	0.063	0.717	0.643	0.678	0.606	0.906	0.735	Pills
	0.615	0.074	0.673	0.615	0.643	0.560	0.896	0.711	Injection - Depo
	0.580	0.091	0.613	0.580	0.596	0.499	0.865	0.632	Implant
Weighted Avg.	0.723	0.069	0.720	0.723	0.720	0.653	0.917	0.754	

=== Confusion Matrix ===

a	b	c	d	e	<-- classified as
7238.88	96.52	0	0	0	a = BTL
241.85	5626.76	374.72	386.66	541.92	b = IUCD
176.04	994.86	4576.97	602.55	764.51	c = Pills
144.96	586.33	690.99	4390.99	1324.12	d = Injection - Depo
97.91	1032.83	737.98	1148.09	4165.46	e = Implant

Appendix III: Choice of Contraceptive Methods Microsoft VB.Net Code

```
Public Class ContraceptiveMethodMatch
  Dim ClientIndicator As String

  Private Sub DetermineFPMethod()
    'set Indicators

    Dim ClientAge As Integer
    Dim ClientAgeGroup As String
    If txtClientAge.Text <> "" Then
      If IsNumeric(txtClientAge.Text) Then
        ClientAge = CInt(txtClientAge.Text)
        If ClientAge >= 30 And ClientAge <= 49 Then
          ClientAgeGroup = "ADULT"
        ElseIf ClientAge >= 25 And ClientAge <= 29 Then
          ClientAgeGroup = "YOUTH"
        ElseIf ClientAge >= 20 And ClientAge <= 24 Then
          ClientAgeGroup = "EARLY_YOUTH"
        ElseIf ClientAge >= 15 And ClientAge <= 19 Then
          ClientAgeGroup = "ADOLESCENT"
        Else
          ClientAgeGroup = "OTHER"
        End If
      End If
    End If
    Dim NumebrofChildren As Integer
    Dim NumebrOfChildrenGroup As String
    If txtNumberOfChildren.Text <> "" Then
      If IsNumeric(txtNumberOfChildren.Text) Then
        NumebrofChildren = CInt(txtNumberOfChildren.Text)
        If NumebrofChildren = 0 Then
          NumebrOfChildrenGroup = "NO_CHILDREN"
        ElseIf NumebrofChildren >= 1 And NumebrofChildren <= 2 Then
          NumebrOfChildrenGroup = "LOW_FAMILY_SIZE"
        ElseIf NumebrofChildren >= 3 And NumebrofChildren <= 4 Then
          NumebrOfChildrenGroup = "MEDIUM_FAMILY_SIZE"
        ElseIf NumebrofChildren >= 5 Then
          NumebrOfChildrenGroup = "HIGH_FAMILY_SIZE"
        Else
          NumebrOfChildrenGroup = "OTHER"
        End If
      End If
    End If
    Dim LastMethodGroup As String = cboLastMethod.SelectedItem
    Dim Educationgroup As String = cboEducation.SelectedItem
    Dim MaritalStatusGroup As String = cboMaritalStatus.SelectedItem
    Dim RegionGroup As String = cboRegion.SelectedItem
    Dim ResidenceGroup As String = cboResidence.SelectedItem
    Dim ReligionGroup As String = cboReligion.SelectedItem
    Dim HypertensionGroup As String = cboHypertension.SelectedItem
    Dim SmokingGroup As String = cboSmoking.SelectedItem
    Dim HeadacheGroup As String = cboHeadache.SelectedItem
    Dim MedicationGroup As String = cboMedication.SelectedItem
    Dim HIVGroup As String = cboHIV.SelectedItem
    Dim STIGroup As String = cboSTI.SelectedItem
    Dim PIDGroup As String = cboPID.SelectedItem
    Dim SepsisGroup As String = cboSepsis.SelectedItem
    Dim VaginalBleedingGroup As String = cboVaginalBleeding.SelectedItem
```

```
        Dim CervicalCancerGroup As String = cboCervicalCancer.SelectedItem
        Dim BreastCancerGroup As String = cboBreastCancer.SelectedItem
        Dim LiverDiseasesGroup As String = cboLiverDiseases.SelectedItem
        Dim VenouSGroup As String = cboVenous.SelectedItem
        Dim CardioVascularGroup As String = cboCardiovascular.SelectedItem
        Dim PostPartumGroup As String = cboCardiovascular.SelectedItem
        Dim FertilityPlanGroup As String = cboFertilityPlan.SelectedItem
        Dim SexualPartnerGroup As String = cboSexualPartner.SelectedItem
        Dim PercievedSideEffeCtGroup As String = cboPercivedSideEffect.SelectedItem

        'Has Never Used Before
        'Not Currently Using but has used before
        'Short Term User
        'Long Term User
        ' None/Some Primary
        'Comp Primary
        'Comp Secondary
        'Tertiary/Higher
        'Single
        'Married
        'Divorced
        'Widowed
        'RULE 1
        Dim PreferredMethod As String
        'BTL
        If SexualPartnerGroup = "Multiple" Or FertilityPlanGroup = "Spacing" Then
            PreferredMethod = "Her health Condistion and/or life style doesn't allow BTL procedure"
        Else
            If ClientAgeGroup = "ADULT" And NumebrOfChildrenGroup = "HIGH_FAMILY_SIZE" And
LastMethodGroup = "Has Never Used Before" Then
                PreferredMethod = "Preffered Method to this client is: BTL"
            ElseIf ClientAgeGroup = "ADULT" And LastMethodGroup = "Has Never Used Before" And
NumebrOfChildrenGroup = "MEDIUM_FAMILY_SIZE" And Educationgroup = "Comp Primary" And ResidenceGroup
= "Rural" And ReligionGroup = "Christian" And RegionGroup = "Amhara" And MaritalStatusGroup = "Married" Then
                PreferredMethod = "Preffered Method to this client is: BTL"
            ElseIf ClientAgeGroup = "ADULT" And NumebrOfChildrenGroup = "HIGH_FAMILY_SIZE" And
ReligionGroup = "Chirstian" And MaritalStatusGroup = "Married" And LastMethodGroup = "Has Never Used Before"
Then
                PreferredMethod = "Preffered Method to this client is: BTL"
            ElseIf ClientAgeGroup = "ADULT" And NumebrOfChildrenGroup = "HIGH_FAMILY_SIZE" And
ReligionGroup = "Christian" And MaritalStatusGroup = "Married" And LastMethodGroup = "Not Currently Using but has
used before" And ResidenceGroup = "Urban" And RegionGroup = "Addis Ababa" And (Educationgroup = "None/Some
Primary" Or Educationgroup = "Comp Primary") Then
                PreferredMethod = "Preffered Method to this client is: BTL"
            ElseIf ClientAgeGroup = "ADULT" And NumebrOfChildrenGroup = "HIGH_FAMILY_SIZE" And
MaritalStatusGroup = "Married" And LastMethodGroup = "Not Currently Using but has used before" And
ResidenceGroup = "Rural" And Educationgroup = "None/Some Primary" Then
                PreferredMethod = "Preffered Method to this client is: BTL"
            ElseIf ClientAgeGroup = "ADULT" And NumebrOfChildrenGroup = "HIGH_FAMILY_SIZE" And
MaritalStatusGroup = "Married" And LastMethodGroup = "Not Currently Using but has used before" And
ResidenceGroup = "Rural" And Educationgroup = "None/Some Primary" And (RegionGroup = "Oromia" Or
RegionGroup = "Amhara") Then
                PreferredMethod = "Preffered Method to this client is: BTL"
            ElseIf ClientAgeGroup = "ADULT" And NumebrOfChildrenGroup = "HIGH_FAMILY_SIZE" And
MaritalStatusGroup = "Married" And ReligionGroup = "Christian" And LastMethodGroup = "Not Currently Using but has
used before" And ResidenceGroup = "Rural" And Educationgroup = "Comp Primary" And RegionGroup = "Oromia"
Then
                PreferredMethod = "Preffered Method to this client is: BTL"
```

```
        ElseIf ClientAgeGroup = "ADULT" And NumebrOfChildrenGroup = "HIGH_FAMILY_SIZE" And
ReligionGroup = "Muslim" And LastMethodGroup = "Has Never Used Before" Then
            PreferredMethod = "Preffered Method to this client is: BTL"
            txtChoosenMethod.Text = PreferredMethod
        End If
    End If
    'IUCD Clients Rule
    If MedicationGroup = "Antiretroviral therapy" Or HIVGroup = "Yes" Or STIGroup = "Gonorrhoea Chlamdia" Or
STIGroup = "Increased risk of STIs" Or PIDGroup = "Yes" Or SepsisGroup = "Yes" Or PostPartumGroup = "48hours to
<4weeks" Or VaginalBleedingGroup = "Yes" Or CervicalCancerGroup = "Yes" Or SexualPartnerGroup = "Multiple"
Then
        PreferredMethod = "Her health Condistion and/or life style doesn't allow to use IUCD"
    Else
        If ClientAgeGroup = "EARLY_YOUTH" And MaritalStatusGroup = "Married" And NumebrOfChildrenGroup =
"MEDIUM_FAMILY_SIZE" And ResidenceGroup = "Urban" And Educationgroup = "None/Some Primary" Then
            PreferredMethod = "Preffered Method to this client is: IUCD"
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And MaritalStatusGroup = "Married" And
NumebrOfChildrenGroup = "MEDIUM_FAMILY_SIZE" And ResidenceGroup = "Urban" And Educationgroup =
"Tertiary/Higher" And RegionGroup = "SNNPR" And ReligionGroup = "Christian" Then
            PreferredMethod = "Preffered Method to this client is: IUCD"
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And MaritalStatusGroup = "Married" And
(NumebrOfChildrenGroup = "MEDIUM_FAMILY_SIZE" Or NumebrOfChildrenGroup = "HIGH_FAMILY_SIZE")
And ResidenceGroup = "Urban" And Educationgroup = "Tertiary/Higher" And RegionGroup = "SNNPR" And
ReligionGroup = "Other" And LastMethodGroup = "Not Currently Using but has used before" Then
            PreferredMethod = "Preffered Method to this client is: IUCD"
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And NumebrOfChildrenGroup = "MEDIUM_FAMILY_SIZE" And
ResidenceGroup = "Urban" And RegionGroup = "Oromia" And ReligionGroup = "Christian" And LastMethodGroup =
"Not Currently Using but has used before" Then
            PreferredMethod = "Preffered Method to this client is: IUCD"
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And MaritalStatusGroup = "Married" And
NumebrOfChildrenGroup = "MEDIUM_FAMILY_SIZE" And ResidenceGroup = "Urban" And RegionGroup =
"SNNPR" And ReligionGroup = "Christian" And Educationgroup = "Comp Secondary" Then
            PreferredMethod = "Preffered Method to this client is: IUCD"
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And MaritalStatusGroup = "Married" And
NumebrOfChildrenGroup = "MEDIUM_FAMILY_SIZE" And ResidenceGroup = "Rural" And RegionGroup = "Oromia"
And Educationgroup = "Comp Secondary" Then
            PreferredMethod = "Preffered Method to this client is: IUCD"
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And MaritalStatusGroup = "Married" And
NumebrOfChildrenGroup = "MEDIUM_FAMILY_SIZE" And ResidenceGroup = "Urban" And RegionGroup = "Addis
Ababa" And ReligionGroup = "Christian" And LastMethodGroup = "Has Never Used Before" And Educationgroup =
"Comp Primary" Then
            PreferredMethod = "Preffered Method to this client is: IUCD"
        ElseIf (ClientAgeGroup = "EARLY_YOUTH" Or ClientAgeGroup = "YOUTH") And MaritalStatusGroup =
"Married" And NumebrOfChildrenGroup = "MEDIUM_FAMILY_SIZE" And ResidenceGroup = "Rural" And
RegionGroup = "Oromia" And ReligionGroup = "Christian" And (LastMethodGroup = "Has Never Used Before" Or
LastMethodGroup = "Not currently Using but has has used before") And Educationgroup = "Tertiary/Higher" Then
            PreferredMethod = "Preffered Method to this client is: IUCD"
        ElseIf ClientAgeGroup = "YOUTH" And MaritalStatusGroup = "Married" And NumebrOfChildrenGroup =
"MEDIUM_FAMILY_SIZE" And ResidenceGroup = "Urban" And ReligionGroup = "Christian" Then
            PreferredMethod = "Preffered Method to this client is: IUCD"
        ElseIf ClientAgeGroup = "YOUTH" And MaritalStatusGroup = "Married" And NumebrOfChildrenGroup =
"HIGH_FAMILY_SIZE" And RegionGroup = "Amhara" And Educationgroup = "Tertiary/Higher" And
LastMethodGroup = "Has Never Used Before" Then
            PreferredMethod = "Preffered Method to this client is: IUCD"
        ElseIf ClientAgeGroup = "ADOLESCENT" And MaritalStatusGroup = "Single" And NumebrOfChildrenGroup =
"NO-CHILDREN" And ReligionGroup = "Christian" And RegionGroup = "Oromia" And Educationgroup = "None/Some
Primary" And LastMethodGroup = "Has Never Used Before" Then
            PreferredMethod = "Preffered Method to this client is: IUCD"
```

```
        ElseIf ClientAgeGroup = "ADULT" And NumebrOfChildrenGroup = "MEDIUM_FAMILY_SIZE" And
ReligionGroup = "Christian" And RegionGroup = "Oromia" And ResidenceGroup = "Urban" And Educationgroup =
"Tertiary/Higher" Then
            PreferredMethod = "Preffered Method to this client is: IUCD"
        ElseIf ClientAgeGroup = "ADULT" And NumebrOfChildrenGroup = "HIGH_FAMILY_SIZE" And
MaritalStatusGroup = "Married" And RegionGroup = "Addis Ababa" And Educationgroup = "Comp Secondary" Then
            PreferredMethod = "Preffered Method to this client is: IUCD"
        ElseIf ClientAgeGroup = "ADULT" And NumebrOfChildrenGroup = "LOW_FAMILY_SIZE" And
MaritalStatusGroup = "Married" And RegionGroup = "Oromia" And LastMethodGroup = "Not Currently Using but has
used before" Then
            PreferredMethod = "Preffered Method to this client is: IUCD"
        End If
        txtChoosenMethod.Text = PreferredMethod
    End If
    'Implant Rules
    If VaginalBleedingGroup = "Yes" Or BreastCancerGroup = "Yes" Or LiverDiseasesGroup = "Liver Tumour" Or
VenouSGroup = "Acute" Or FertilityPlanGroup = "Limiting" Or SexualPartnerGroup = "Multiple" Or
PercievedSideEffeCtGroup = "Not at all" Then
        PreferredMethod = "Her health Condistion and/or life style doesn't allow to use Implant"
    Else
        If ClientAgeGroup = "ADOLESCENT" And NumebrOfChildrenGroup = "NO_CHILDREN" And
MaritalStatusGroup = "Single" And RegionGroup = "Addis Ababa" And Educationgroup = "None/Some Primary" Then
            PreferredMethod = "Preffered Method to this client is: IMPLANT"
        ElseIf ClientAgeGroup = "ADOLESCENT" And NumebrOfChildrenGroup = "LOW_FAMILY_SIZE" And
MaritalStatusGroup = "Single" And ResidenceGroup = "Rural" And RegionGroup = "SNNPR" And Educationgroup =
"Comp Primary" Then
            PreferredMethod = "Preffered Method to this client is: IMPLANT"
        ElseIf ClientAgeGroup = "ADOLESCENT" And NumebrOfChildrenGroup = "LOW_FAMILY_SIZE" And
MaritalStatusGroup = "Married" And Educationgroup = "None/Some Primary" Then
            PreferredMethod = "Preffered Method to this client is: IMPLANT"
        ElseIf ClientAgeGroup = "ADULT" And NumebrOfChildrenGroup = "HIGH_FAMILY_SIZE" And
ResidenceGroup = "Rural" And ReligionGroup = "Christian" And RegionGroup = "SNNPR" Then
            PreferredMethod = "Preffered Method to this client is: IMPLANT"
        ElseIf ClientAgeGroup = "YOUTH" And NumebrOfChildrenGroup = "MEDIUM_FAMILY_SIZE" And
MaritalStatusGroup = "Single" And ReligionGroup = "Christian" And Educationgroup = "Tertiary/Higher" Then
            PreferredMethod = "Preffered Method to this client is: IMPLANT"
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And NumebrOfChildrenGroup = "NO_CHILDREN" And
MaritalStatusGroup = "Single" And ReligionGroup = "Christian" And Educationgroup = "None/Some Primary" And
RegionGroup = "Addis Ababa" And LastMethodGroup = "Has Never Used Before" Then
            PreferredMethod = "Preffered Method to this client is: IMPLANT"
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And NumebrOfChildrenGroup = "LOW_FAMILY_SIZE" And
MaritalStatusGroup = "Single" And ReligionGroup = "Christian" And Educationgroup = "None/Some Primary" And
RegionGroup = "Addis Ababa" Then
            PreferredMethod = "Preffered Method to this client is: IMPLANT"
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And NumebrOfChildrenGroup = "NO_CHILDREN" And
MaritalStatusGroup = "Single" And ReligionGroup = "Christian" And Educationgroup = "Tertiary/Higher" And
RegionGroup = "SNNPR" Then
            PreferredMethod = "Preffered Method to this client is: IMPLANT"
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And NumebrOfChildrenGroup = "MEDIUM_FAMILY_SIZE" And
MaritalStatusGroup = "Single" And ReligionGroup = "Christian" And Educationgroup = "Tertiary/Higher" And
RegionGroup = "Oromia" And LastMethodGroup = "Has Never Used Before" Then
            PreferredMethod = "Preffered Method to this client is: IMPLANT"
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And NumebrOfChildrenGroup = "LOW_FAMILY_SIZE" And
MaritalStatusGroup = "Married" And ResidenceGroup = "Rural" And RegionGroup = "Tigray" Then
            PreferredMethod = "Preffered Method to this client is: IMPLANT"
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And NumebrOfChildrenGroup = "LOW_FAMILY_SIZE" And
MaritalStatusGroup = "Married" And ResidenceGroup = "Urban" And RegionGroup = "Addis Ababa" And
ReligionGroup = "Christian" And Educationgroup = "Comp Primary" Then
            PreferredMethod = "Preffered Method to this client is: IMPLANT"
```

```
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And NumebrOfChildrenGroup = "LOW_FAMILY_SIZE" And
MaritalStatusGroup = "Married" And ResidenceGroup = "Urban" And RegionGroup = "Addis Ababa" And
Educationgroup = "None/Some Primary" Then
            PreferredMethod = "Preffered Method to this client is: IMPLANT"
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And NumebrOfChildrenGroup = "HIGH_FAMILY_SIZE" And
MaritalStatusGroup = "Married" And ResidenceGroup = "Rural" And RegionGroup = "Oromia" And ReligionGroup =
"Muslim" Then
            PreferredMethod = "Preffered Method to this client is: IMPLANT"
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And NumebrOfChildrenGroup = "MEDIUM_FAMILY_SIZE" And
MaritalStatusGroup = "Married" And ResidenceGroup = "Rural" And RegionGroup = "SNNPR" And ReligionGroup =
"Christian" And LastMethodGroup = "Not Currently Using but has used before" Then
            PreferredMethod = "Preffered Method to this client is: IMPLANT"
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And NumebrOfChildrenGroup = "MEDIUM_FAMILY_SIZE" And
MaritalStatusGroup = "Married" And ResidenceGroup = "Rural" And RegionGroup = "Tigray" And ReligionGroup =
"Christian" Then
            PreferredMethod = "Preffered Method to this client is: IMPLANT"
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And NumebrOfChildrenGroup = "LOW_FAMILY_SIZE" And
MaritalStatusGroup = "Married" And ResidenceGroup = "Urban" And RegionGroup = "Addis Ababa" And
ReligionGroup = "Christian" And Educationgroup = "Comp Primary" Then
            PreferredMethod = "Preffered Method to this client is: IMPLANT"
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And NumebrOfChildrenGroup = "MEDIUM_FAMILY_SIZE" And
MaritalStatusGroup = "Married" And ReligionGroup = "Christian" And RegionGroup = "Addis Ababa" And
Educationgroup = "None/Some Primary" Then
            PreferredMethod = "Preffered Method to this client is: IMPLANT"
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And NumebrOfChildrenGroup = "MEDIUM_FAMILY_SIZE" And
MaritalStatusGroup = "Married" And RegionGroup = "Dire Dawa" Then
            PreferredMethod = "Preffered Method to this client is: IMPLANT"
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And NumebrOfChildrenGroup = "MEDIUM_FAMILY_SIZE" And
MaritalStatusGroup = "Married" And ResidenceGroup = "Rural" And (RegionGroup = "SNNPR" Or RegionGroup =
"Amhara") And RegionGroup = "Christian" Then
            PreferredMethod = "Preffered Method to this client is: IMPLANT"
        End If
        txtChoosenMethod.Text = PreferredMethod
    End If
    'Injection Depo Rules
    If HypertensionGroup = ">=160/>=100" Or HypertensionGroup = "140-159/90-99" Or SmokingGroup =
"Age>=35yrs" Or HeadacheGroup = "Migraines with aura" Or MedicationGroup = "Certain anti convulsants" Or
MedicationGroup = "Rifampicin" Or PostPartumGroup = "48hours to <4weeks" Or PostPartumGroup = "<48hours" Or
PostPartumGroup = "4weeks to <6weeks" Or PostPartumGroup = "6weeks to <6months" Or BreastCancerGroup = "Yes"
Or LiverDiseasesGroup = "Liver Tumour" Or LiverDiseasesGroup = "Hepatitis acute/flare" Or VenouSGroup = "History"
Or VenouSGroup = "Acute" Or VenouSGroup = "Major Surgory with prolonged immobilizaion" Or FertilityPlanGroup =
"Limiting" Or SexualPartnerGroup = "Multiple" Or PercievedSideEffeCtGroup = "Not at all" Or
PercievedSideEffeCtGroup = "To some extent" Then
        PreferredMethod = "Her health Condistion and/or life style doesn't allow to use Injection-Depo"
    Else
        If ClientAgeGroup = "YOUTH" And NumebrOfChildrenGroup = "NO_CHILDREN" And MaritalStatusGroup =
"Single" And Educationgroup = "Comp Primary" Then
            PreferredMethod = "Preffered Method to this client is: Injection-Depo"
        ElseIf ClientAgeGroup = "YOUTH" And NumebrOfChildrenGroup = "NO_CHILDREN" And
MaritalStatusGroup = "Single" And RegionGroup = "Amhara" And ResidenceGroup = "Rural" Then
            PreferredMethod = "Preffered Method to this client is: Injection-Depo"
        ElseIf ClientAgeGroup = "YOUTH" And NumebrOfChildrenGroup = "LOW_FAMILY_SIZE" And RegionGroup
= "Oromia" And ResidenceGroup = "Rural" And Educationgroup = "Comp Secondary" Then
            PreferredMethod = "Preffered Method to this client is: Injection-Depo"
        ElseIf ClientAgeGroup = "YOUTH" And NumebrOfChildrenGroup = "LOW_FAMILY_SIZE" And RegionGroup
= "Amhara" And ResidenceGroup = "Urban" Then
            PreferredMethod = "Preffered Method to this client is: Injection-Depo"
```

```
        ElseIf ClientAgeGroup = "YOUTH" And NumebrOfChildrenGroup = "HIGH_FAMILY_SIZE" And
MaritalStatusGroup = "Married" And ReligionGroup = "Christian" And Educationgroup = "Comp Primary" And
ResidenceGroup = "Urban" Then
            PreferredMethod = "Preffered Method to this client is: Injection-Depo"
        ElseIf ClientAgeGroup = "YOUTH" And NumebrOfChildrenGroup = "MEDIUM_FAMILY_SIZE" And
MaritalStatusGroup = "Married" And ReligionGroup = "Muslim" And Educationgroup = "None/Some Primary" And
ResidenceGroup = "Rural" And RegionGroup = "Oromia" Then
            PreferredMethod = "Preffered Method to this client is: Injection-Depo"
        ElseIf ClientAgeGroup = "YOUTH" And NumebrOfChildrenGroup = "NO_CHILDREN" And
MaritalStatusGroup = "Single" And ReligionGroup = "Christian" And Educationgroup = "Comp Secondary" And
ResidenceGroup = "Rural" And RegionGroup = "Oromia" Then
            PreferredMethod = "Preffered Method to this client is: Injection-Depo"
        ElseIf ClientAgeGroup = "YOUTH" And NumebrOfChildrenGroup = "NO_CHILDREN" And
MaritalStatusGroup = "Single" And Educationgroup = "Comp Primary" And ResidenceGroup = "Rural" And
LastMethodGroup = "Not Currently Using but has used before" Then
            PreferredMethod = "Preffered Method to this client is: Injection-Depo"
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And NumebrOfChildrenGroup = "NO_CHILDREN" And
MaritalStatusGroup = "Single" And Educationgroup = "None/Some Primary" And RegionGroup = "Oromia" Then
            PreferredMethod = "Preffered Method to this client is: Injection-Depo"
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And (NumebrOfChildrenGroup = "LOW_FAMILY_SIZE" Or
NumebrOfChildrenGroup = "MEDIUM_FAMILY_SIZE") And MaritalStatusGroup = "Single" And Educationgroup =
"Comp Primary" Then
            PreferredMethod = "Preffered Method to this client is: Injection-Depo"
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And NumebrOfChildrenGroup = "MEDIUM_FAMILY_SIZE" And
MaritalStatusGroup = "Single" And Educationgroup = "Comp Secondary" And RegionGroup = "Oromia" Then
            PreferredMethod = "Preffered Method to this client is: Injection-Depo"
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And NumebrOfChildrenGroup = "MEDIUM_FAMILY_SIZE" And
MaritalStatusGroup = "Single" And Educationgroup = "None/Some Primary" And ResidenceGroup = "Urban" And
LastMethodGroup = "Not Currently Using but has used before" And ReligionGroup = "Christian" And RegionGroup =
"Addis Ababa" Then
            PreferredMethod = "Preffered Method to this client is: Injection-Depo"
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And NumebrOfChildrenGroup = "LOW_FAMILY_SIZE" And
MaritalStatusGroup = "Married" And Educationgroup = "Comp Primary" And ResidenceGroup = "Urban" And
LastMethodGroup = "Has Never Used Before" And RegionGroup = "Amhara" Then
            PreferredMethod = "Preffered Method to this client is: Injection-Depo"
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And NumebrOfChildrenGroup = "HIGH_FAMILY_SIZE" And
MaritalStatusGroup = "Married" And Educationgroup = "Comp Primary" And ResidenceGroup = "Rural" And
ReligionGroup = "Muslim" And RegionGroup = "Oromia" Then
            PreferredMethod = "Preffered Method to this client is: Injection-Depo"
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And NumebrOfChildrenGroup = "HIGH_FAMILY_SIZE" And
ResidenceGroup = "Urban" And ReligionGroup = "Other" And RegionGroup = "Oromia" Then
            PreferredMethod = "Preffered Method to this client is: Injection-Depo"
        End If
        txtChoosenMethod.Text = PreferredMethod
    End If
    'Pills Rules
    If HypertensionGroup = ">=160/>=100" Or HypertensionGroup = "140-159/90-99" Or SmokingGroup =
"Age>=35yrs" Or HeadacheGroup = "Migraines with aura" Or MedicationGroup = "Certain anti convulsants" Or
MedicationGroup = "Rifampicin" Or PostPartumGroup = "48hours to <4weeks" Or PostPartumGroup = "<48hours" Or
PostPartumGroup = "4weeks to <6weeks" Or PostPartumGroup = "6weeks to <6months" Or BreastCancerGroup = "Yes"
Or LiverDiseasesGroup = "Liver Tumour" Or LiverDiseasesGroup = "Hepatitis acute/flare" Or VenouSGroup = "History"
Or VenouSGroup = "Acute" Or VenouSGroup = "Major Surgory with prolonged immobilizaion" Or FertilityPlanGroup =
"Limiting" Or SexualPartnerGroup = "Multiple" Or PercievedSideEffeCtGroup = "Not at all" Or
PercievedSideEffeCtGroup = "To some extent" Then
        PreferredMethod = "Her health Condistion and/or life style doesn't allow to use Oral Contraceptive Pills"
    Else
        If ClientAgeGroup = "EARLY_YOUTH" And MaritalStatusGroup = "Married" And NumebrOfChildrenGroup =
"HIGH_FAMILY_SIZE" And ResidenceGroup = "Urban" And ReligionGroup = "Muslim" And RegionGroup =
"Oromia" Then
```

```
            PreferredMethod = "Preffered Method to this client is: Oral Contraceptive Pills"
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And MaritalStatusGroup = "Married" And
NumebrOfChildrenGroup = "LOW_FAMILY_SIZE" And ResidenceGroup = "Urban" And Educationgroup =
"Tertiary/Higher" Then
            PreferredMethod = "Preffered Method to this client is: Oral Contraceptive Pills"
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And MaritalStatusGroup = "Married" And
NumebrOfChildrenGroup = "LOW_FAMILY_SIZE" And ResidenceGroup = "Urban" And Educationgroup = "Comp
Secondary" And LastMethodGroup = "Not Currently Using but has used before" Then
            PreferredMethod = "Preffered Method to this client is: Oral Contraceptive Pills"
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And MaritalStatusGroup = "Married" And
NumebrOfChildrenGroup = "NO_CHILDREN" And ReligionGroup = "Muslim" And Educationgroup = "Comp
Secondary" Then
            PreferredMethod = "Preffered Method to this client is: Oral Contraceptive Pills"
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And MaritalStatusGroup = "Married" And
NumebrOfChildrenGroup = "NO_CHILDREN" And Educationgroup = "Tertiary/Higher" And (RegionGroup = "SNNPR"
Or RegionGroup = "Addis Ababa") Then
            PreferredMethod = "Preffered Method to this client is: Oral Contraceptive Pills"
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And MaritalStatusGroup = "Married" And
NumebrOfChildrenGroup = "NO_CHILDREN" And Educationgroup = "Comp Secondary" And RegionGroup =
"Oromia" And ReligionGroup = "Christian" And LastMethodGroup = "Not Currently Using but has used before" Then
            PreferredMethod = "Preffered Method to this client is: Oral Contraceptive Pills"
        ElseIf ClientAgeGroup = "EARLY_YOUTH" And MaritalStatusGroup = "Single" And NumebrOfChildrenGroup
= "NO_CHILDREN" And ResidenceGroup = "Urban" And Educationgroup = "Tertiary/Higher" And RegionGroup =
"Oromia" And LastMethodGroup = "Not Currently Using but has used before" Then
            PreferredMethod = "Preffered Method to this client is: Oral Contraceptive Pills"
        ElseIf ClientAgeGroup = "YOUTH" And MaritalStatusGroup = "Married" And NumebrOfChildrenGroup =
"LOW_FAMILY_SIZE" And ResidenceGroup = "Urban" And Educationgroup = "Tertiary/Higher" And RegionGroup =
"Oromia" Then
            PreferredMethod = "Preffered Method to this client is: Oral Contraceptive Pills"
        ElseIf ClientAgeGroup = "YOUTH" And MaritalStatusGroup = "Married" And NumebrOfChildrenGroup =
"NO_CHILDREN" And ResidenceGroup = "Urban" And (Educationgroup = "Tertiary/Higher" Or Educationgroup =
"Comp Secondary") Then
            PreferredMethod = "Preffered Method to this client is: Oral Contraceptive Pills"
        ElseIf ClientAgeGroup = "YOUTH" And MaritalStatusGroup = "Married" And NumebrOfChildrenGroup =
"NO_CHILDREN" And ResidenceGroup = "rural" And Educationgroup = "Tertiary/Higher" And RegionGroup =
"Oromia" Then
            PreferredMethod = "Preffered Method to this client is: Oral Contraceptive Pills"
        ElseIf ClientAgeGroup = "ADOLESCENT" And MaritalStatusGroup = "Single" And NumebrOfChildrenGroup =
"NO_CHILDREN" And ResidenceGroup = "Urban" And Educationgroup = "Tertiary/Higher" Then
            PreferredMethod = "Preffered Method to this client is: Oral Contraceptive Pills"
        ElseIf ClientAgeGroup = "ADOLESCENT" And MaritalStatusGroup = "Single" And NumebrOfChildrenGroup =
"LOW_FAMILY_SIZE" And ResidenceGroup = "Urban" And Educationgroup = "Tertiary/Higher" Then
            PreferredMethod = "Preffered Method to this client is: Oral Contraceptive Pills"
        ElseIf ClientAgeGroup = "ADULT" And NumebrOfChildrenGroup = "HIGH_FAMILY_SIZE" And
ReligionGroup = "Muslim" And Educationgroup = "Tertiary/Higher" Then
            PreferredMethod = "Preffered Method to this client is: Oral Contraceptive Pills"
        End If
        txtChoosenMethod.Text = PreferredMethod
    End If
    If txtChoosenMethod.Text = "Preffered Method to this client is: IUCD" Then
        txtExplanation.Text = "IUCD is a highly effective form of long-acting, reversible contraception, with an associated
pregnancy (failure) rate of 0.8 percent in the first year of use. The average annual failure rate was 0.4 percent or less, and
after 12 years of use the cumulative failure rate for women using IUD was 2.2 percent, which is comparable to that of
female sterilization.  Long year studies showed that IUCD is effective for at least 12 years after insertion. Side Effect:
During insertion, some women may experience discomfort or cramping. Cramps may continue for several days beyond
insertion. Cramping, pain, and menstrual irregularities associated with IUD insertion or menstruation usually subside
within a few months. Heavy or prolonged bleeding and pain may be treated with nonsteroidal anti-inflammatory drugs
such as ibuprofen. Thoughtful counseling about side effects and treatment options is critical since menstrual irregularities
are the most common medical reason for IUD removal."
```

```
        ElseIf txtChoosenMethod.Text = "Preffered Method to this client is: IMPLANT" Then
            txtExplanation.Text = "The contraceptive implant is one or two small plastic rods which are about 4cm long and sit
under the skin of the inside of your upper arm and prevents pregnancy for 3 to 5 years (depending on the type).  The
contraceptive implant works by slowly releasing a hormone (progestin) in your blood and the hormone stops the body
releasing an egg each month. The hormone also makes it harder for the man's sperm to get into the womb. Inserting the
implant is a simple procedure that takes about five minutes. A trained person inserts the implant into the inside of your
upper arm. Your arm is numbed before insertion so it is generally not painful.  Once inserted the implant does not move
around your body. It stays where it is inserted. You can feel it under the skin. Dependent on when you get the implant
inserted, it can take up to seven days for the implant to start working. You can use a condom during this time. It stops
working when it is taken out, at which point your normal fertility returns."
        ElseIf txtChoosenMethod.Text = "Preffered Method to this client is: Injection-Depo" Then
            txtExplanation.Text = "The injection provides two to three months contraceptive protection depending on the type
you use. It contains the hormone progestin and it works by slowly releasing a hormone that stops the body producing an
egg each month. The hormone also makes it harder for the man's sperm to get into the womb. The injection is usually
given in the arm, thigh or buttock by a trained provider. It only takes a few minutes to have the injection.  It can take up to
seven days before the injection starts to work if you have not previously used a contraceptive. You can use a condom for
those 7 days."
        ElseIf txtChoosenMethod.Text = "Preffered Method to this client is: Oral Contraceptive Pills" Then
            txtExplanation.Text = "The pill works well at preventing a pregnancy. Its ability to stop a pregnancy largely
depends on a person using it properly. If one hundred women took the pills every time they were supposed to for a year
and carried on with their normal sex life then only one of those women would fall pregnant during that period. However, it
is hard for many women to remember to take something daily.  With typical use up to 8 women out of 100 would fall
pregnant. Risk of pregnancy is greatest when a woman starts a new pill pack 3 or more days late, or misses 3 or more pills
in a pack."
        ElseIf txtChoosenMethod.Text = "Preffered Method to this client is: BTL" Then
            txtExplanation.Text = "The Bilateral Tubaligation procedure is done and this would be given "
        Else
            txtExplanation.Text = " No method is choosen"
        End If
        If SexualPartnerGroup = "Multiple" Then
            txtChoosenMethod.Text = " Condom"
        End If
    End Sub

    Private Sub cmdPrefferedFPMethod_Click(sender As Object, e As EventArgs) Handles cmdPrefferedFPMethod.Click
        DetermineFPMethod()
    End Sub

    Private Sub PercivedSideEffect_SelectedIndexChanged(sender As Object, e As EventArgs) Handles
cboPercivedSideEffect.SelectedIndexChanged

    End Sub

        Private Sub CMMKBSToolStripMenuItem_Click(sender As Object, e As EventArgs) Handles
CMMKBSToolStripMenuItem.Click
        Me.Show()
    End Sub
End Class
```

Appendix IV: Evaluation Questionnaire

Dear Evaluator,

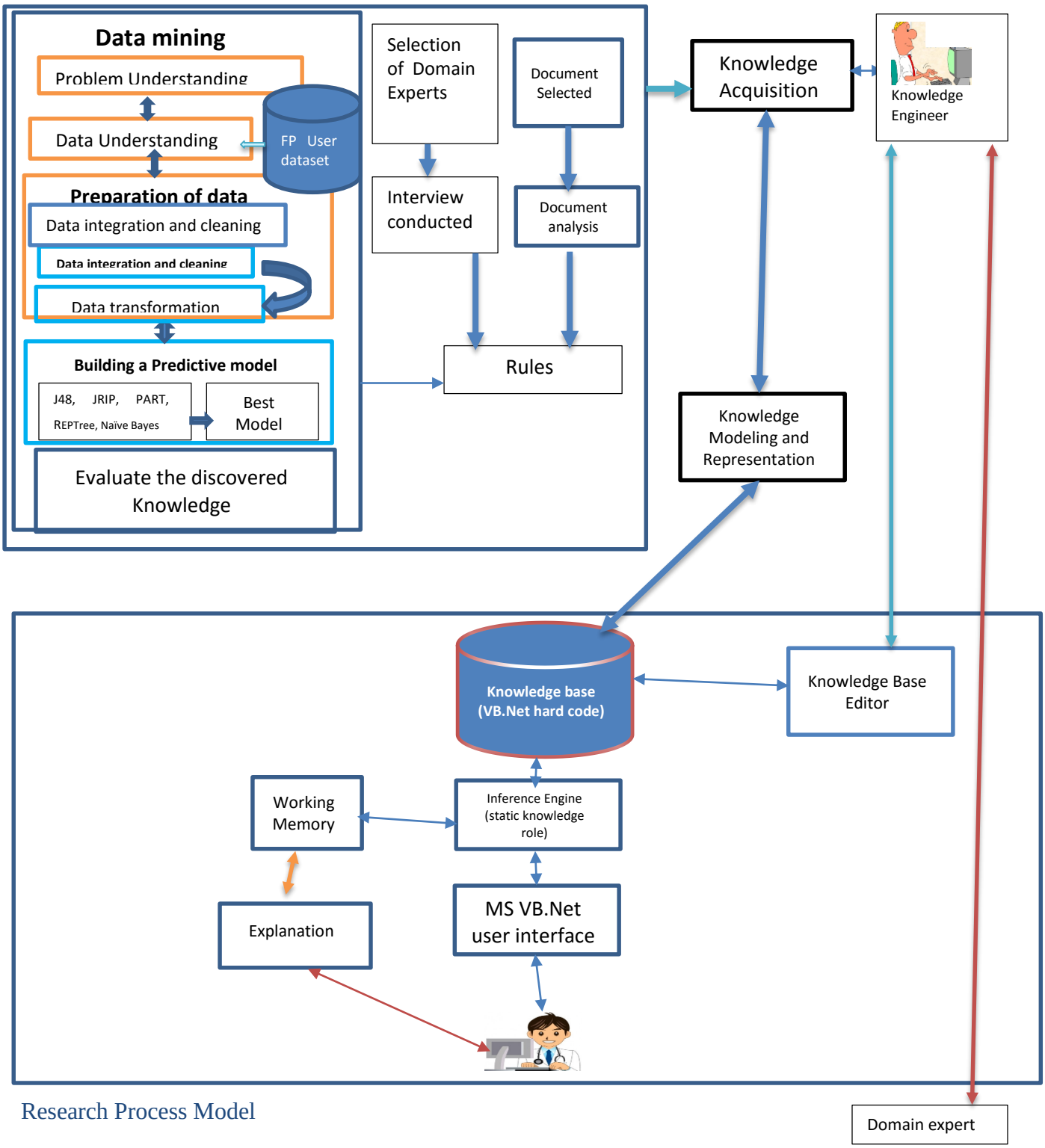
First, I would like to appreciate your collaboration in providing the information in advance. You can refuse at any time.

The aim of this evaluation is measuring to the extent of the CMMKBS system is acceptable by users (health service providers). Thus, you are kindly requested to evaluate the system by labeling (X) symbol on the space provided for each criteria.

no	**Criteria**	poor	Fair	Good	Very good	Excellent
1	Interaction with system and simplicity to use					
2	Efficiency in time					
3	System attractiveness					
4	The accuracy of the system in choosing contraceptive method that match with the client lifestyle.					
5	The explanation given to selected contraceptive use					
6	Applicability of the system to the domain are					

NB: The values for all attributes in the table are rated as: Excellent=5, Very good =4, Good=3, Fair= 2 and Poor =1.

Appendix V: Research Process Model



Research Process Model