****

**Application of Data Mining with Knowledge Based System for Diagnosis and Treatment of Cattle Diseases: The case of International Livestock Research Institute (ILRI) Animal Health Center Addis Ababa**

**By**

**Zerihun Fantahun Wele**

**July 2021**

**Application of Data Mining with Knowledge Based System for Diagnosis and Treatment of Cattle Diseases: The case of international livestock research institute (ILRI) animal health center**

**By**

**Zerihun Fantahun Wele**

**to**

**The Faculty of Informatics**

**of**

**St. Mary’s University**

**In Partial Fulfillment of the Requirements**

**for the Degree of Master of Science**

**in**

**Computer Science**

**July, 2021**

**ACCEPTANCE**

**Application of Data Mining with Knowledge Based System for Diagnosis and Treatment of Cattle Diseases: The case of international livestock research institute (ILRI) animal health center**

**By**

**Zerihun Fantahun Wele**

**Accepted by the Faculty of Informatics, St. Mary’s University, in partial fulfillment of the requirements for the degree of Master of Science in Computer Science**

**Thesis Examination Committee:**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Internal Examiner**

**Dr. Alemebante Mulu**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**External Examiner**

**Dr. Minale Ashagrie**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Dean, Faculty of Informatics**

**Dr. Alemebante Mulu**

**July 2020**

# DECLARATION

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

Zerihun Fantahun Wele

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Signature

Addis Ababa

Ethiopia

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

Dr. Temtim Assefa,

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Signature

Addis Ababa

Ethiopia

July 2021

# ACKNOWLEDGEMENT

God and many people have helped me to accomplish the research work on time and in successful manner. I am grateful to all who supported me to complete this research paper. I especially thank my advisor Dr Temtim Assef for his willingness to help me and providing constructive advices and comments from the initial to the finalization of this research work.

I would like to express my thanks to international livestock research institute (ILRI) animal health center experts: Dr. Tamrat Degafe and Dr. Yirgalem Shimelis who devoted their golden time for a number of days for the interview and system evaluation session for this research work.

I would like to thank my family who supported and encouraged me throughout the time of my study and the research work.

Finally, my thanks go to all the people who have supported me to complete the research work directly or indirectly.

# Table of content

Contents

[DECLARATION i](#_Toc79719008)

[ACKNOWLEDGEMENT ii](#_Toc79719009)

[Table of content iii](#_Toc79719010)

[List of Figures vii](#_Toc79719011)

[List of Tables viii](#_Toc79719012)

[ACRONYMS ix](#_Toc79719013)

[ABSTRACT xi](#_Toc79719014)

[CHAPTER ONE 1](#_Toc79719015)

[INTRODUCTION 1](#_Toc79719016)

[1.1 Background of the study 1](#_Toc79719017)

[1.2. Statement of the problem 2](#_Toc79719018)

[1.3. Objective of the study 4](#_Toc79719019)

[1.3.1. General objective 4](#_Toc79719020)

[1.3.2. Specific objective 4](#_Toc79719021)

[1.4 Scope and limitation of the research 4](#_Toc79719022)

[1.5. Significance of the research 5](#_Toc79719023)

[1.6. Research methodology 6](#_Toc79719024)

[1.6.1. Research design 6](#_Toc79719025)

[1.6.2. Literature review 6](#_Toc79719026)

[1.6.3. Interviewing Domain Experts 7](#_Toc79719027)

[1.6.4. Knowledge Acquisition 7](#_Toc79719028)

[6.4. Knowledge Representation 7](#_Toc79719029)

[1.6.5. Implementation Evaluation Methods 7](#_Toc79719030)

[1.7. Organization of the Study 8](#_Toc79719031)

[CHAPTER TWO 10](#_Toc79719032)

[LITERATURE REVIEW 10](#_Toc79719033)

[2.1 Introduction 10](#_Toc79719034)

[2.2 Overview of Cattle disease 10](#_Toc79719035)

[2.3 Overview of Data Mining 10](#_Toc79719036)

[2.3.1. What is Data Mining? 11](#_Toc79719037)

[2.3.2. Why Data Mining? 11](#_Toc79719038)

[2.4 Data Mining Tasks 12](#_Toc79719039)

[2.4.1 Classification 12](#_Toc79719040)

[2.4.2. Decision Tree Induction 13](#_Toc79719041)

[2.4.3 Rule based classification 15](#_Toc79719042)

[2.4.4 Naïve Bayes Classifier 16](#_Toc79719043)

[2.4.5. Clustering 17](#_Toc79719044)

[2.4.6. Regression 17](#_Toc79719045)

[2.5 Data Mining Process Model 17](#_Toc79719046)

[2.5.1 The KDD Process 18](#_Toc79719047)

[2.5.2. Sample Explore Modify Model Assess Process 19](#_Toc79719048)

[2.5.3. Cross Industry Standard Process for Data Mining Process 20](#_Toc79719049)

[2.4.4. Hybrid Process 20](#_Toc79719050)

[2.6 Evaluation of the Classification Model 22](#_Toc79719051)

[2.6.1 Prediction Accuracy 22](#_Toc79719052)

[2.6.2 True Positive Rate and False Positive Rate 22](#_Toc79719053)

[2.6.3 Precision and Recall 23](#_Toc79719054)

[2.7 Attribute Selection Measures 23](#_Toc79719055)

[2.7.1. Information Gain 23](#_Toc79719056)

[2.7.2. Gain Ratio 24](#_Toc79719057)

[2.7.3. Gini Index 25](#_Toc79719058)

[2.8. Implementation Tools 26](#_Toc79719059)

[2.9 Knowledge Based System 27](#_Toc79719060)

[2.9.1 Advantages of Knowledge Based System 27](#_Toc79719061)

[2.9.2 Types of Knowledge 28](#_Toc79719062)

[2.9.3 Architecture of Knowledge Based System 29](#_Toc79719063)

[2.10 Knowledge representation 31](#_Toc79719064)

[2.10.1 Semantic Networks 31](#_Toc79719065)

[2.10.2. Rule Based Knowledge Representation 32](#_Toc79719066)

[2.10.3. Logic Knowledge Representation 33](#_Toc79719067)

[2.11 Related Works 33](#_Toc79719068)

[CHAPTER THREE 38](#_Toc79719069)

[METHODS AND APPROACH 38](#_Toc79719070)

[3.1. Introduction 38](#_Toc79719071)

[3.2. Research Design 38](#_Toc79719072)

[3.2.1. Understanding of the Problem Domain 39](#_Toc79719073)

[3.2.2. Understanding of the Data 40](#_Toc79719074)

[3.2.3. Preparation of the Data 41](#_Toc79719075)

[CHAPTER FOUR 47](#_Toc79719076)

[KNOWLEDGE MODELING, EXPERIMENTATION, AND ANALYSIS 47](#_Toc79719077)

[4.1 Knowledge Modeling 47](#_Toc79719078)

[4.1 Selecting Modeling Technique 47](#_Toc79719079)

[4.2 Experimental Setup 48](#_Toc79719080)

[4.3.1 Methods of Training and Testing Option 48](#_Toc79719081)

[4.3,2 Developing Classifier Model Using Naïve Bayes 48](#_Toc79719082)

[4.3.3 Developing Classifier Model Using J48 Decision Tree 50](#_Toc79719083)

[4.3.4 Developing Classifier Model Using JRip Rule Based 52](#_Toc79719084)

[4.3.5 Performance Comparison of Naïve Bayes, J48 and JRip Models 53](#_Toc79719085)

[4.4 Classifying Model Performance Evaluation Metrics 54](#_Toc79719086)

[4.5 Rule Extraction from JRip Classification Algorithm 55](#_Toc79719087)

[CHAPTER FIVE 57](#_Toc79719088)

[KNOWLEDGE ACQUISITION AND REPRESENTATION 57](#_Toc79719089)

[5.1. Introduction 57](#_Toc79719090)

[5.2 Knowledge acquired using documents analysis 57](#_Toc79719091)

[5.3. Common Cattle Diseases 57](#_Toc79719092)

[5.3.1. Foot-and-Mouse disease (FMD) 58](#_Toc79719093)

[3.2.2. Lumpy Skin disease (LSD) 58](#_Toc79719094)

[5.3.3. Blackleg 59](#_Toc79719095)

[5.3.4. Anthrax 60](#_Toc79719096)

[5.3.6. Pneumonia 64](#_Toc79719097)

[5.3.8. Enteritis /Diarrhea 67](#_Toc79719098)

[5.3.9. Calf Diphtheria 68](#_Toc79719099)

[5.3.10. Amphistomiasis 69](#_Toc79719100)

[5.4 Data collection Using Interview 70](#_Toc79719101)

[5.5 Knowledge Representation 71](#_Toc79719102)

[CHAPTER SIX 73](#_Toc79719103)

[IMPLEMENTATION AND DISCUSSION 73](#_Toc79719104)

[6.1 Introduction 73](#_Toc79719105)

[6.2 Architecture of the System 73](#_Toc79719106)

[6.3 Combination of Data Mining and Knowledge Based System 78](#_Toc79719107)

[6.4 Structure of JRip Rule and PROLOG Rule 78](#_Toc79719108)

[6.5 Evaluation of the Prototype 79](#_Toc79719109)

[6.5.1 User Acceptance Testing 79](#_Toc79719110)

[6.6. Discussion 82](#_Toc79719111)

[CHAPTER SEVEN 86](#_Toc79719112)

[CONCLUSION AND RECOMMENDATION 86](#_Toc79719113)

[7.1. Conclusion 86](#_Toc79719114)

[7.2. Recommendation 87](#_Toc79719115)

[References 89](#_Toc79719116)

[Appendix I 94](#_Toc79719117)

[Appendix II 95](#_Toc79719118)

[Appendix III 96](#_Toc79719119)

[Appendix IV 98](#_Toc79719120)

# List of Figures

[Figure 2. 1 Data Mining Algorithm Implementation Classification Model [19] 13](#_Toc79155372)

[Figure 2. 2 Decision tree Structure [28] 14](#_Toc79155373)

[Figure 2. 3 The Five stage of KDD process [23] 18](#_Toc79155374)

[Figure 2. 4 The general architecture of Knowledge Based System [45] 29](#_Toc79155375)

[Figure 3. 1 Data Mining System framework [43] 39](#_Toc79155397)

[Figure 3. 2 Sample data after handling missing values 42](#_Toc79155398)

[Figure 3. 3 Illustrates sample WEKA ARFF files show as follows: 43](#_Toc79155399)

[Figure 3. 4 Imbalance dataset before SMOTE 44](#_Toc79155400)

[Figure 3. 5 Balanced Dataset after SMOTE 45](#_Toc79155401)

[Figure 4. 1 Illustrates sample WEKA result 49](#_Toc79155405)

[Figure 4. 2 Illustrates sample WEKA result 49](#_Toc79155406)

[Figure 4. 3 TP rate of classifiers 54](#_Toc79155407)

[Figure 5. 1 Diagnostic Procedure for cattle Disease 71](#_Toc79155414)

[Figure 6. 1 System Architecture of the Knowledgebase system 74](#_Toc79155424)

[Figure 6. 2 Welcoming windows of KBS cattle Disease Diagnosis user interface 76](#_Toc79155425)

[Figure 6. 3 Sample windows between the user and the system to identify cattle disease 77](#_Toc79155426)

[Figure 6. 4 Sample window of the system’s treat and prevent for the identified disease 78](#_Toc79155427)

# List of Tables

[Table 2. 1 Confusion Matrix system PROLOG [44]. 22](#_Toc79155711)

[Table 3. 1 Number of Attributes and Records 40](#_Toc79155712)

[Table 3. 2 Data Transformation by the cattle temperature range [4] 42](#_Toc79155713)

[Table 3. 3 Removing unnecessary attributes of the table 46](#_Toc79155714)

[Table 4. 1 Detailed Accuracy by Class for Naïve Bayes classification algorithm: 50](#_Toc79155715)

[Table 4. 2 Detailed Accuracy by Class for J48 classification algorithm: 51](#_Toc79155716)

[Table 4. 3 Detailed Accuracy by Class for JRip classification algorithm: 53](#_Toc79155717)

[Table 4. 4 Comparison of the results of the models 53](#_Toc79155718)

[Table 4. 5 Rules extracted by JRip Classification algorithm 55](#_Toc79155719)

[Table 5. 1 Mastitis basic microorganism 63](#_Toc79155195)

[Table 5. 2 Sample selected for interview 70](#_Toc79155196)

[Table 6. 1 Cattle disease diagnosis and treatment Evaluation performance result 81](#_Toc79155199)

[Table 6. 2 Comparison of different researchers in this study 83](#_Toc79155200)

# ACRONYMS

AI Artificial Intelligence

ARFF Attribute Relation File Format

C/I Contraindication

CART Classification & Regression Trees

CBR Case Based Reasoning

CommonKADS Common Knowledge Acquisition and Documentation Structuring

CRISP-DM CRoss Industry Standard Process for Data Mining

CSV Comma Separated Value

D/F Dosage Forms

D/I Drug Interaction

FMD Foot Mouse Disease

GUI Graphical User Interface

ID3 Iterative Dichotomiser 3

IIS Internet Information Server

ILRI international livestock research institute

IM Intramuscularly

IU International Unit

IV Intravenously

JDE Java Integrated Development Environment

JDK Java Development Kit

JPL Java to Prolog Library

KBS Knowledge Based System

KDD Knowledge Discovery Database

KDP knowledge discovery processes

KR Knowledge Representation

LSD Lumpy skin Disease

MSF Medicine Sans Frontieres

IPPER Repeated Incremental Pruning to Produce Error Reduction

SMOTE Synthetic Minority Oversampling Technique

S/E Side Effect

SEMMA Sample, Explore, Modify, Model Asses

SWI-PROLOG PROGraming in Logi

T/P True positive

VMD Veterinarian Medicine Doctors

WEKA Waikato Environment for Knowledge Analysis

# ABSTRACT

Ethiopia is one among the nations that possesses the largest livestock population in the African continent with an estimated 56 Million of cattle, 58 Million of sheep and goats and 10 Million of equines, 1 Million of camels and 57 Million of chicken. Ethiopia has great potential for increasing livestock production, both for local use and export. However, development has been constrained by numerous reasons.

In this study, the possibility of integrating data mining result with knowledge based system is realized and explored. The integration process begun by taking samples of ILRC dataset. The dataset is preprocessed and made suitable for mining steps. Due to several limitations in acquiring knowledge for knowledge base from domain experts in the area of diagnosis and treatment of cattle disease, integrated (manual and automated) knowledge acquisition techniques were used to acquire knowledge. Data mining has proven to induce hidden knowledge from large collections of datasets. Hence, data mining classifier, JRip is employed for knowledge acquisition step since it has performed best among the selected classifiers with an accuracy of 97.68%.

To identify the best prediction model for diagnosis and treatment of cattle disease, 6 experiments for three classification algorithms, namely J48 pruned, Naïve Bayes and JRip under ten-fold Cross- Validation test option and percentage split test option were conducted. Finally, by conducting objective and subjective interestingness measure, the researcher decided to use rules that are generated by JRip classification algorithm model for further use in the development of knowledge base system because it registered better performance than J48 and Naïve Bayes with 97.68%, 96.65% and 95.42% evaluation result in 10-fold cross validation respectively.

The prototype Knowledge Based System, which provides advice for Animal Health Workers about diagnosis and treatment of cattle disease was developed using SWI-Prolog 7.7.13 with NetBeans 8.2. The proposed Knowledge Based System has Knowledge Base, Inference Engines, Explanation Facility and User Interface. Then 70 test cases were prepared to evaluate the performance of the proposed system. Finally, system performance evaluation, testing and user acceptance testing were conducted. User acceptance testing is performed based on seven criteria of evaluation. Selected domain experts are trained and used the system to evaluate how much the KBS meets their requirements. The system on average scored 84.85% based on user acceptance evaluation.

**Keywords:** - Cattle disease, Data mining, Knowledge based system, Rule based, Integrated with the Knowledge Base System,

# CHAPTER ONE

# INTRODUCTION

## Background of the study

Ethiopia is one of the African countries with the most livestock, with an estimated 56 million cattle, 58 million sheep and goats, 10 million equines, 1 million camels, and 57 million chickens. Ethiopia's has great potential for increasing livestock production, both domestically and internationally. Livestock production is frequently poor due to technological, economic, and institutional restrictions. Milk, meat, and eggs are produced in 5.6 billion liters, 1.1 million tons, and 419 million eggs every year [48]. Statistics on total production, on the other hand, can vary greatly depending on the source. The production value estimates, which are also subject to change, do not include organic fertilizers (68 million tons) or animal draft power (617 million days per year). Since many of the animals rescored come from harsh environments, their typical lifespan is 18 to 22 years. Even though they may have greater health problems and a shorter lifetime than conventional cattle, these animals are not immune to disease [2].

Exports of live cattle and small ruminants account for a substantial portion of the livestock industry's revenues. According to estimates, live animal exports is about 10 percent earnings from the 69 percent of Ethiopia's total export. [48]

The entire supply of animal source foods in the country, including net trade, equates into an annual per capita intake of 9 kg meat, 56.2 liters of milk, and around 4 eggs. Cattle products, such as beef and cow milk, account for about 80% of total meat and milk consumption. Market transactions are concentrated in metropolitan regions, whereas self-consumption reigns supreme in rural areas [48].

A knowledge-based system (KBS) is a form of artificial intelligence (AI) that aims to capture the knowledge of human experts to support decision-making. Examples of knowledge-based systems include expert systems, which are so called because of their reliance on human expertise. The typical architecture of a knowledge-based system, which informs its problem-solving method, includes a knowledge base and an inference engine. The knowledge base contains a collection of information in a given field medical diagnosis, for example. The inference engine deduces insights from the information housed in the knowledge base. Knowledge-based systems also include an interface through which users query the system and interact with it. A knowledge-based system may vary with respect to its problem-solving method or approach. Some systems encode expert knowledge as rules and are therefore referred to as rule-based systems. Another approach, case-based reasoning, substitutes cases for rules. Cases are essentially solutions to existing problems that a case-based system will attempt to apply to a new problem.

An application system is a computer system that competes with a human expert's decision-making capacity. They are best suited for circumstances where a specialist is unavailable. Knowledge must be collected from domain experts and documented resources in order to create an expert system. Knowledge may be obtained from a variety of sources, including interviews with domain experts, document analysis, observation, and others.

The majority of developing-country governments have significant challenges in providing animal healthcare services. One of the difficulties is a lack of highly skilled human resources, as well as financial barriers to operating an animal clinic. Several studies have indicated that cattle diseases have a detrimental influence on exports to overseas markets [13]. The Ethiopian Drug Administration and Control Authority [40] estimates that the direct loss from infectious illness mortality is 8-10% of cattle. According to the findings of this study, cattle disease is one of the country's most significant issues. Many diseases have reduced output and created a substantial barrier to the worldwide market.

Studies indicate shortage of skilled manpower in healthcare, which contributes to poor diagnosis and treatment of cattle diseases especially in Ethiopia [9], mainly shortage of manpower and laboratory facilities, thus automatic processing and exploring technology is important in order to solve such problem using KBS. Thus, this research explores the applicability of using knowledge based system technology in the domain of cattle disease diagnosis and treatment by developing an integrated knowledge base system.

## Statement of the problem

When it comes to providing animal healthcare services, the majority of developing-country governments confront significant challenges. One of the difficulties is a lack of highly skilled human resources, as well as financial barriers to operating an animal clinic. Cattle disease, according to one research, have a detrimental impact on foreign market export [13]. The Ethiopian Drug Administration and Control Authority [40] estimates that the direct loss from infectious illness mortality is 8-10% of cattle. Cow illness is one of the country's most significant concerns, according to the study's conclusions. Many diseases have reduced production and created major barriers to entrance into the global market.

Ethiopia had a comparatively low number of veterinarians until the preceding decade. The veterinarian-to-animal ratio was around one-to-500,000 [8]. Despite the government's efforts to increase the number of veterinarians by establishing veterinary schools, the veterinarian-to-animal ratio remains inadequate. The government has recently deployed animal health aides to treat animals in rural areas. These animal health aides have just a basic grasp of animal diseases. Animal health aides, who work in Ethiopia's rural communities, are the primary means of treating cattle in the country. They will have no opportunity to learn about cattle diseases once they graduate.

The information they obtain through the agricultural development office is static and does not respond to the user's specific needs since the information is general. For example, if a new disease occurred in the country, a detailed symptom of the disease would not be known. During this situation, people would look for veterinarians [9]. However, veterinarians are rarely able to devote adequate time to assisting all requests. Moreover, in many instances, the response by veterinarians is not also on time. The use of a knowledge-based system could be a possible solution for treating on time. Studies indicate shortage of skilled manpower in healthcare, which contributes to poor diagnosis and treatment of cattle diseases especially in Ethiopia [9], mainly shortage of manpower and laboratory facilities, thus automatic processing and exploring technology is important in order to solve such problem using KBS. Thus, this research explores the applicability of using knowledge based system technology in the domain of cattle disease diagnosis and treatment by developing an integrated knowledge base system

Above all, the differential diagnosis of diseases takes a long time. The veterinary service cycle is delay in veterinarian clinics. Because some diseases have two or more similar symptoms. This is difficult, especially for inexperienced veterinarians. Knowledge sharing among experienced veterinarians and new ones is difficult due to the lack of well-experienced experts in all veterinary clinics in the developing country.

In this research, Data Mining results integrated with Knowledge based System for Cattle Diseases Diagnosis and Treatment is present to reduce the problems, which will use to facilitate fast disease diagnosis and on-time treatment by the veterinarians and educated producers of cattle. Because several diseases show very similar conditions that need a differential diagnosis and take a long, time to detect. In this regard, these studies answers to the following research questions: -

* What are the main attributes that can predict the type of cattle disease?
* Which classification algorithm is best to develop the prediction model for cattle disease diagnosis?
* How develop a knowledge-based system using data mining results?
* How evaluate the performance of knowledge based system?

## 1.3. **Objective of the study**

### 1.3.1. General objective

The general objective of this study is to integrate data mining result for the development of a knowledge base system for cattle diseases and treatment.

### 1.3.2. Specific objective

To achieve the general objectives of the study, the following specific objectives are formulated:

* To acquire knowledge from domain experts, document analysis, and data mining.
* To conduct preprocessing of the International Livestock Animal health center cattle disease dataset.
* To select appropriate data mining algorithm to build the predictive model for cattle disease diagnosis.
* To create a predictive model based on cattle disease dataset.
* To construct the Knowledge base system using a rule based knowledge representation approach.
* To evaluate the performance and user acceptance of the knowledge based system.

## Scope and limitation of the research

As mentioned in the introduction section, there are different types of cattle diseases. Due to time limitation, only the top ten cattle diseases are considered. This study is, thus bound to propose a prototype of a Knowledge Based System for the diagnosis and treatment of ten common cattle diseases, namely blackleg, anthrax, Foot Mouse Disease (FMD), Lumpy Skin Disease (LSD), pneumonia, mastitis, enteritis, Amphistomiasis, Calf Disptheria and Internal Parasite are frequently occurring, according to ILR Animal Health Center dataset. Knowledge acquisition, modeling, representation, and development of the knowledge-based system for the diagnosis and treatment of selected diseases are the major goals of this research.

Hence, this study does not include the diagnosis and treatment of all diseases of cattle. The reason is that many of the viral and bacterial diseases need sophisticated laboratory equipment for the identification of them. It is, thus, limited to the differential diagnosis and treatment of diseases of cattle based on symptoms.

Generally, the study tries to implement a prototype Knowledge-Based System with a Java graphical user interface that plays the differential diagnosis of a limited number of diseases. The dataset was small in number because of this the researcher faced a class imbalance problem and forced to use the Synthetic Minority Oversampling Technique (SMOTE) to make the imbalance dataset balanced and get better prediction results.

## 1.5. Significance of the research

The importance of this study are animal health professionals especially who are working in rural towns. It will help the animal health assistants as decision support system while they diagnose cattle diseases in areas where highly qualified veterinarian professionals are unavailable. Today many investors are investing on animal farming. So, they can use this system to maintain the health condition of their cattle’s.

The benefit of this study is to provide the medical diagnosis and treatment of diseases of cattle. The veterinary who work in agricultural sectors can easily access the collected knowledge in the Knowledge Base System to support the cattle producers timely and quickly. Domain Experts and middle-level professionals can improve their knowledge and experience using the knowledge base system to diagnose and treat diseases.

KBS enables human experts to share their private knowledge among users in the agricultural fields to improve the quality and productivity of cattle production. Developing a Knowledge Base System would play a vital and critical role in documenting guidelines and knowledge. Experience of well-educated and veterinaries make it available for those animal health workers, which in turn serves as means of knowledge transfer such that the people of Ethiopia could get early and timely diagnosis and treatment where ever they are which also help the country in providing a quality health care service for the people who are living in remote areas. Developing and implementing a Knowledge Base System that provides advice to animal health workers. Cattle owners would play a great role because KBS provides the high-quality performance, which solves difficult problems in a domain as good as, or better than human experts do and can possess vast quantities of domain-specific knowledge to the minute details which can, in turn, serve as means of knowledge and experience transfer. After all, the ability of intelligent systems to capture and redistribute expertise has significant implications for the overall development of the country.

## 1.6. Research methodology

### 1.6.1. Research design

The research design, mixed research method design. It uses data mining and knowledge base development methodologies and type of information gathered have been elaborated. The techniques carried out in this study were knowledge acquisition, modeling, and knowledge representation and implementation method using appropriate knowledge programming tools. Before starting development of prototype knowledge based system, the researcher has investigated the problem area. This helps the researcher to get details understanding about the problems and limitations that challenge the domain experts in controlling and managing Cattle diseases. Since the research topic needs several experiments and domain experts’ opinion, the researcher used The research design a mixed research method design. It uses data mining and knowledge base development methodologies. The researcher also collected the dataset for data mining purpose from ILR Animal health center dataset.

### 1.6.2. Literature review

The researcher reviewed different researches, articles, and journal papers that related to cattle disease and treatments, data mining, Knowledge base systems, knowledge acquisition of knowledge-based systems, and integration of data mining results with knowledge base systems to get a conceptual understanding of the problem on the hand. In addition to this, the researcher also reviewed related works to identify the gap and formulate the problem and research questions of the study.

### 1.6.3. Interviewing Domain Experts

Structure and unstructured interviews employed to elicit tacit knowledge from domain experts. Since one of the specific objectives of this research is to acquire knowledge about type of cattle diseases treatment strategy from primary sources using interview and secondary source document analysis. For this research, selected five health experts purposively for interview about for Diagnosis and Treatment of Cattle Diseases and treatment depending on their professions, educational qualification level, and willingness.

### 1.6.4. Knowledge Acquisition

#### 1.6.4.1. Manual Knowledge Acquisition

The researcher used both interviews and document analysis to acquire knowledge. The researcher conducted the domain expert’s interview with Veterinaries who works on Cattle Diseases and treatment.

#### 1.6.4.2. Automated Knowledge Acquisition

The researcher used **Cr**oss **I**ndustry **S**tandard **P**rocess for **D**ata **M**ining (**CRISP-DM**) model to acquire knowledge from ILRI Animal health center cattle disease using the (Waikato Environment for Knowledge Analysis (WEKA) data mining tool. This model consists of six phases intended as a cyclical phase, namely Business understanding, data understanding, data preparation, Modeling, evaluation and Deployment.

### 6.4. Knowledge Representation

Since the knowledge that the researcher acquired from the data mining classification technique is in the form of rules the knowledge that the researcher acquires from document analysis and domain experts interview about diagnosis and treatment of cattle disease. Procedures, which are easy to convert to rules, the researcher forced to use rule based knowledge representation method that is the most predominant knowledge representation methods to develop the Knowledge base

### 1.6.5. Implementation Evaluation Methods

In order to mine the hidden knowledge from the pre-processed dataset and compare the performance of classifiers, the researcher used the WEKA 3.8.11 data-mining tool. To represent rules in the knowledge base and construct the prototype of cattle disease advising Knowledge Based System, the researcher uses SWI-Prolog. Java NetBeans IDE 8.2 with JDK-8u20 [10] was employ to integrate WEKA results with the Knowledge based system and develop the GUI of the proposed system. In order to evaluate the results and accuracy of the developed model of the study the researcher used Precision, Recall, F-measure and True Positive rate. The researcher also evaluates the KBS using system performance testing by preparing test cases and users' acceptance testing questionnaire which helps the researcher to make sure that whether the potential users would like to use the proposed system frequently and whether the proposed systems meets user requirements.

## 1.7. Organization of the Study

This study comprises seven chapters. Chapter two discusses conceptual and related works reviews that are relevant for this study. In this chapter, the researcher discussions about data mining process model, data mining tasks, classification algorithm. In addition to this, the researcher discusses the Knowledge-Based System such as methods of knowledge acquisition, knowledge representation, Knowledge Base System architecture, Knowledge Base development tools that are relevant for this study.

Chapter three discuss about methods and approach on this mainly focuses on how the research conducted including understanding the problem, data understanding and selection, preprocessed the data, data transformation and attribute selection.

Chapter four discusses about knowledge modeling, experimental and analysis. Here, the researcher presents selecting modeling techniques, experimental setup and explain which classification algorithm used and how to evaluate the model were discussed in detail in this chapter

Chapter five presents the knowledge acquisition and Representation. In this chapter, the researcher discussions about common cattle disease, symptoms of the disease, transmission, and prevention methods, clinical diagnosis and treatment of cattle disease. The focus here is on manual (domain expert interview and documents analysis) and automated knowledge acquisition techniques through data mining. In addition to this the researcher discussed about conceptual modeling for cattle disease diagnosis and treatment.

Chapter six discusses about implementation, evaluation and discussion of the proposed systems. In this chapter, the researcher discusses the components of the proposed system such as the knowledge base, the inference engine, the user interface and the knowledge update facility. Besides to this, the researcher discusses how he evaluates the proposed system using test cases and user acceptance testing mechanisms.

Finally, chapter seven conclusion and recommendation, which is the last chapter of the study, discusses the conclusion and forwards a recommendation for further investigation.

# CHAPTER TWO

# LITERATURE REVIEW

## 2.1 Introduction

The literature review of this study try to observe about of cattle disease, overview of data mining, Data mining process model, Data mining tasks,Attribute selectionmeasures, knowledge Based system, Knowledge based representation, Evaluation of knowledge classification model and different research which are related with the research topic are going to be discussed to support the research project and finally the literature review is going to be summarized.

## 2.2 Overview of Cattle disease

According to Aklilu [31], in Ethiopia, there are two types of cattle breeds: indigenous and exotic breeds. Indigenous cattle population is a dominant species in the country, which accounts for 99%. Borena, Horro, Fogera, Karayu, Arsi and Nuer are the most widely used indigenous breeds in the country [31].

The Fogera and Horro cattle are important indigenous animals with huge potential for dairy and meat production. The Fogera breed types are reared around Lake Tana in Amhara Regional State and Horro are reared in Eastern Welega in the west of Oromiya Regional State [38]. The Borena, renowned beef breed is found in the south and east of the country in the Southern Nation,

Nationalities and Peoples' Regional State (SNNPRS) and in the Somali Regional State. The Nuer breed in the southwest is considered to have tolerance to high tsetse challenge. Similarly, Arsi cattle are mainly found in the central high-lands of Ethiopia especially in Arsi, Shewa and Bale administrative region [38].

On the other hand, Holstein-Friesian, Jersey and Simmental are some of exotic cattle breeds. These exotic breeds, which are aimed to improve milk production in the country, are imported from other countries [14, 16].

## 2.3 Overview of Data Mining

Information is available at any time anywhere because of the rapid growth of World Wide Web and electronic information services. The invented machines are faster in producing, manipulating and disseminating information. In information age, an appropriate usage and organization of the information helps to be powerful and to achieve goals. Hence, information processing mechanisms such as automatic data summarization, information extraction and discovering hidden knowledge are very important [14].

### 2.3.1. What is Data Mining?

Data mining is the process of extracting or mining knowledge from large data sets. But, knowledge mining from data can describe the definition of data mining even if it is long. data mining have similar or a bit different meaning with different terms, such as knowledge mining from data, knowledge extraction, data/pattern analysis, data archaeology, and data dredging [14].

According to Olson[15], Data mining also considered as an exploratory data analysis. Generally, Data mining uses advanced data analysis tools to find out previously unknown (hidden), valid patterns and relationships among data in large data sets. It is the core field for different disciplines such as database, machine learning and pattern recognition.

It is a common practice to refer to the idea of searching applicable patterns in data using different names such as data mining, knowledge extraction, information discovery, information harvesting, data archaeology, and data pattern processing. Among these terms KDD and data mining are used widely [23].

Knowledge discovery was coined at KDD to emphasize the fact that knowledge is the end product of a data-driven discovery and that it has been popularized in the artificial intelligence and machine learning fields. According to Fayyad [25], KDD and data mining are two different terms. KDD refers to the overall process of discovering useful knowledge from data and data mining refers to a particular step in the process. Furthermore, data mining is considered as the application of specific algorithms for extracting patterns from data.

### 2.3.2. Why Data Mining?

Nowadays, massive amount of data is produced and collected incrementally. The possibility of gathering and storing huge amount of data by different organizations is becoming true because of using fast and less expensive computers. When organizational data bases keep growing in number and size due to the availability of powerful and affordable database systems the need for new techniques and tools became very important. These tools are used for helping humans to automatically identify patterns, transform the processed data into meaning full information in order to draw concrete conclusions. In addition, it helps in extraction of hidden knowledge from huge amount of digital data [24].

In the private sector industries such as banking, insurance, medicine, telecommunication and retailing use data mining to reduce costs, enhance research, and increase sales. Different organizations worldwide are used data mining techniques for applying and locating higher value customers and to reconfigure their product offerings to increase sales. In the public sector, data mining applications initially were used as a means for detecting fraud and waste of materials, but it grown for different purposes such as measuring and improving program performance [24].

First discuss the different data mining methods – predictive (classification), descriptive (clustering) methods and algorithms within each method.

## 2.4 Data Mining Tasks

There are two major goals in Data Mining: prediction and description. Prediction is often referred to as supervised learning, while descriptive includes unsupervised learning and visualization aspects of Data Mining [31]. Most data mining techniques are based on supervised learning, where a model is constructed explicitly or implicitly by generalizing from a sufficient number of training examples [20]. The underlying assumption of the supervised approach is that the trained model is applicable to future cases. Maimon and Rokach [24] stated that, Discovery methods are those that automatically identify patterns in the data. The discovery method branch consists of prediction methods versus description methods. Descriptive methods are oriented to data interpretation, which focuses on understanding (by visualization for example) the way the underlying data relates to its parts. Prediction-oriented methods aim to automatically build a behavioral model, which obtains new and unseen samples and is able to predict values of one or more variables related to the sample. It also develops patterns, which form the discovered knowledge in a way which is understandable and easy to operate upon. Some prediction-oriented methods can also help provide understanding of the data. In this study, predictive models are used for acquiring knowledge.

### 2.4.1 Classification

Classification is the task of assigning class labels to the data according to a model learned from the training data where the classes are known. Classification is one of the most common tasks in supervised learning, but it has not received much attention in temporal data mining [29]. The classification task is characterized by a well-defined definition of the class labels, and a training set consisting of reclassified examples. The task is to develop a classification model of some kind that can be applied to unclassified data in order to classify it. Decision trees, nearest neighbor are well-suited techniques for the classification task.

The developed model is based on the analysis of a set of training data whose class label is known and the derived model may be represented in various forms such as IF-THEN rules, decision trees, mathematical formulae, semantic network etc. [30]. Each technique employs a learning algorithm to identify a model that best fits the relationship between the attributes set and the class label of the input data. The following classification model illustrates the data mining algorithm implementation shows in Figure 2.2 as follows:

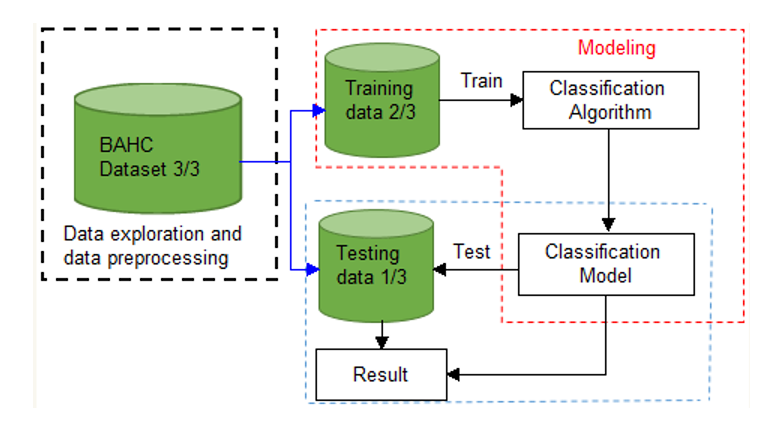


Figure 2. 1 Data Mining Algorithm Implementation Classification Model [19]

Figure 2.1 Data Mining Algorithm Implementation Classification Model [19]

### 2.4.2. Decision Tree Induction

Decision tree is one of the most used data mining techniques because its model is easy to understand for users. In decision tree technique, the root of the decision tree is a simple question or condition that has multiple answers. Each answer then leads to a set of questions or conditions that help us determine the data so that we can make the final decision based on it. The algorithms that are used for constructing decision trees usually work top-down by choosing a variable at each step that is the next best variable to use in splitting the set of items [34].

In decision tree construction, selection of splitting attributes is necessary in order to avoid irrelevant attributes by examining the effect of each attribute for the distinct class and its likelihood for improving the overall decision performance of the tree, since the feature with minimum impact on dependent variable may distort the tree's performance and the classification accuracy. One of the most attractive aspects of decision trees lies in their interpretability especially with respect to the construction of decision rules which is constructed from a decision tree simply by traversing any given path from the root node to any leaf [27]. Therefore, to make a decision tree model more readable, a path to each leaf can be transformed into an IF-THEN rule. In Figure 2.3, illustrates the root node and leaf node as follows [28].

Root Node

Internal Node

Internal Node

Leaf Node

Leaf Node

Leaf Node

Leaf Node

Figure 2. 2 Decision tree Structure [28]

Each leaf is assigned to one class representing the most appropriate target value. Alternatively, the leaf may hold a probability vector indicating the probability of the target attribute having a certain value. Instances are classified by navigating them from the root of the tree down to a leaf, according to the outcome of the tests along the path.

The many benefits in data mining that decision trees offer include the following [38,40]:

* Decision trees require very little data preparation whereas other techniques often require data normalization, the creation of dummy variables and removal of blank values.
* Uses a white box model i.e. the explanation for the condition can be explained easily by Boolean logic because there are mostly two outputs. For example, yes or no.
* Self-explanatory and easy to follow when compacted
* Able to handle a variety of input data: nominal, numeric and textual
* Able to process datasets that may have errors or missing values
* High predictive performance for a relatively small computational effort
* Available in many data mining packages over a variety of platforms ➢ Useful for various tasks, such as classification, regression, clustering and feature selection.

Some of the weaknesses of DT are [28,40]:

* Some DT can only deal with binary valued target classes, others are able to assign records to an arbitrary number of classes, but errors are prone when the number of training examples per class gets small. This can happen rather quickly in a tree with many levels and many branches per node.
* The process of growing a DT is computationally expensive. At each node, each candidate splitting field is examined before its best split can be found.
* Decision tree is less appropriate for estimation tasks where the goal is to predict the value of continuous such as income, blood pressure, or interest rate.
* Decision tree is also problematic for time-series data values a lot of effort is put into presenting the data in such a way that tends and sequential patterns are made visible.

### 2.4.3 Rule based classification

A rule is represented by the IF-THEN form, where the IF part is called the condition and the THEN part is called the action [27]. The basic unit and format of knowledge in rule-based reasoning is the rule. The IF-THEN rules are quite natural for humans and are easily understood by both programmers and domain experts. However, accurate description of the domain expert's knowledge in simple rules is often difficult.

Rule :( condition) => X

Where,

* Condition is a conjunction of attributes like (A1=v1) and (A2=v2) and …and

(An=vn) and

* X is a class label.

For example: (fever=high) ^ (diarrhea>= Yes) 🡪Enteritis

The advantage of IF-THEN rule is the rules are order independent i.e. regardless of the order of rules executed, the same classification of the classes is possible to reach [29]. PART and JRIP are algorithms are an example of rule based classifiers.

PART: It is a separate-and-conquer rule learner. The algorithm producing sets of rules called decision lists which are ordered set of rules. A new data is compared to each rule in the list in turn, and the item is assigned the category of the first matching rule (a default is applied if no rule successfully matches). PART builds a partial C4.5 decision tree in its each iteration and makes the best leaf into a rule. The algorithm is a combination of C4.5 and Repeated Incremental Pruning to Produce Error Reduction (RIPPER) rule learning [29].

JRip: It implements a propositional rule learner. JRip proposed a Repeated Incremental Pruning to Produce Error Reduction (RIPPER). It is an inference and rules--based learner (RIPPER) that can be used to classify elements with propositional rules. The RIPPERR algorithm is a direct method used to extracts the rules directly from the data [29]. JRip (Weka's implementation of the RIPPER rule learner) is a fast algorithm for learning "IF THEN" rules. Like decision trees rule learning algorithms are popular because the knowledge representation is very easy to interpret.

### 2.4.4 Naïve Bayes Classifier

The Naïve Bayes algorithm is a simple probabilistic classifier that calculates a set of probabilities by counting the frequency and combinations of values in a given dataset. The algorithm uses Bayes theorem and assumes all attributes to be independent given the value of the class variable. This conditional independence assumption rarely holds true in real world applications, hence the characterization as Naïve Bayes yet the algorithm tends to perform well and learn rapidly in various supervised classification problems [35]. Steps to calculate Naïve Bayes formula as follows:

Step 1: Convert the dataset into a frequency table

Step 2: Create Likelihood table by finding the probabilities

Step 3: Use Naïve Bayesian formula to calculate the posterior probability for each class.

The class with the highest posterior probability is the outcome of prediction.

Posterior probability P (C|X) = 𝐿𝑖𝑘𝑒𝑘𝑖ℎ𝑜𝑜𝑑 (𝑥|𝑐) ∗𝐶𝑙𝑎𝑠𝑠 𝑝𝑟𝑖𝑜𝑟 𝑃𝑟𝑜𝑏𝑎𝑏𝑖𝑙𝑖𝑡y (𝑐)

Predictor Prior Probability P(x)……. (2.1)

### 2.4.5. Clustering

Clustering is identifying similar groups from unstructured data. Clustering is the task of grouping a set of objects in a such a way that object in same group are more similar to each other than to those in other groups. Once the clusters are decided, the objects are labelled their corresponding clusters, and common features of the objects in cluster are summarized to form a class description [16]. The difference with the classification task is that clusters were unknown at the time the algorithm starts. In other words, there are no predefined classes it relies on [20]. The records are grouped together on the basis of self-similarity. The meaning of the results of clustering are to be determined by the user. Clustering is often done as a prelude to some other data mining task. Clustering technique is used for the clustering task [35].

### 2.4.6. Regression

Regression, sometimes also called estimation, is a kind of statistical estimation technique which is used to map each data object to a real value provided prediction value [22]. It deals with continuously valued outcomes and comes up with a value for some unknown continuous variable. The estimation approach has the great advantage that the individual records can be rank ordered according to the estimate. Uses of regression include prediction, modelling of causal relationships, and testing hypotheses about relationships between variables. Well suited techniques for regression tasks are (linear) regression models and none linear regration [35].

## 2.5 Data Mining Process Model

One of the greatest strength of data mining is reflected in its wide range of methodologies and techniques that can be applied to a host of problem sets [22]. Data mining tools perform data analysis and uncover important data patterns, contributing greatly to different business strategies including medical researchers. The widening gap between data and information calls for a systematic development of data mining tools that would turn data toms into golden nuggets of knowledge. Thus, patterns and knowledge from data mining is using for a sound judgement and proactive decision making in different organization including health care sectors. Broadly used methodologies in data mining are Knowledge Discovery in Database (KDD), Cross Industry Standard Process for Data mining (CRISP-DM), Sample Explore Modify Model Assess (SEMMA) and hybrid process [22].

### 2.5.1 The KDD Process

The knowledge discovery process (KDP), also called knowledge discovery in data base is defined as the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data [27].

As described in figure 2.1 below, Knowledge discovery in database (KDD) has five stages, such as selection, preprocessing, transformation, Data Mining and Interpretation or Evaluation.

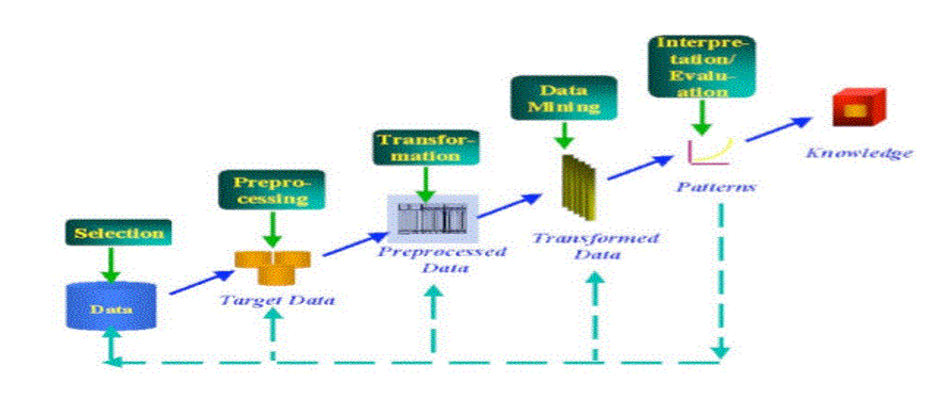


Figure 2. 3 The Five stage of KDD process [23]

**Selection**: This stage is concerned with creating a target data set or focusing on a subset of variables or data samples, on which discovery is to be performed by Understanding the data and the business area. Because, Algorithms alone will not solve the problem without having clear statement of the objective and understanding.

**Pre-processing**: This phase is concerned in removing noise or outliers if any, collecting the necessary information to model or account for noise, deciding on strategies for handling missing data fields, and accounting for time sequence information and known changes. On top of these tasks, deciding on DBMS issues, such as data types, schema, and mapping of missing and unknown values are parts of data cleaning and pre-processing.

**Transformation:** The transformation of data using dimension reduction or transformation methods is done at this stage. Usually there are cases where there are large numbers of attributes in the database for a particular case. With the reduction of dimension there will be an increase in the efficiency of the data-mining step with respect to the accuracy and time utilization.

**Data Mining**: This phase is the major stage in data KDD because it is all about searching for patterns of interest in a particular representational form or a collection of such representations. These representations include classification rules or trees, regression, clustering, sequence

modeling, dependency, and line analysis. Therefore, selecting the right algorithm for the right area is very important.

Evaluation: In this stage the mined data is presented to the end user in a Human viewable format. This involves data visualization, where the user interprets and understands the discovered knowledge obtained by the algorithms.

Using the Discovered Knowledge: Incorporating this knowledge into a performance system, taking actions based on the knowledge, or simply documenting it and reporting it to interested parties, as well as checking for and resolving for conflicts with previously acquired knowledge are tasks in this phase. Knowledge discovery in database (KDD), as a process consists of an iterative sequence of steps as discussed above. It is also clear that data mining is only one step in the entire process, though an essential one, it uncovers hidden patterns for evaluation.

### 2.5.2. Sample Explore Modify Model Assess Process

It is developed by the SAS institute. The acronym SEMMA stands for Sample Explore, modify, Model and Assess and refers to the process of conducting a data mining project [16]. The SAS institute considers a cycle with five stages for the process lists as following:

Sample: - this stage consists on sampling the data by extracting a portion of a large dataset big enough to contain the significant information, yet small enough to manipulate quickly. This stage is pointed out at 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 the modeling the data by allowing the software to search automatically for a combination of data that reliably predicts some desired outcomes.

Assess: - this stage consists on assessing the data by evaluating the usefulness and reliability of the findings from the data mining process and estimate how well it performs.

Although the SEMMA process is independent from data mining chosen tolls, it is linked to SAS Enterprise miner software and pretend to guide the user on the implementation of data mining applications. SEMMA offers an easy to understand process, allowing an organized and adequate development and maintenance of data mining project [16]

### 2.5.3. Cross Industry Standard Process for Data Mining Process

Data mining needs a standard approach which will help to translate business problems into data mining tasks, suggest appropriate data transformations and data mining techniques, and provide means for evaluating the effectiveness of the results and documenting the experience.

The Cross Industry Standard Process for Data Mining (CRISP-DM) project addressed parts of these problems by defining a process model which provides a framework for carrying out data. The CRISP-DM methodology provides essential support for those seeking to understand and practice data mining. The required process for success in data mining has been invented independently by many practitioners. By standardizing terminology, CRISP-DM has made it easy for practitioners to communicate about specific data mining projects and about the process in general. CRISP-DM also improves the training of new data miners by providing a detailed and standardized answer to the question “How should data mining be performed?” [16,21,33].

### 2.4.4. Hybrid Process

Hybrid DM process model combine aspects of both academic and industrial areas [33]. It was developed based on the CRISP-DM model by adopting it to academic research. The main differences and extensions include: Hybrid model provide more general; it introduces a data mining step instead of the modeling step. In addition, it introduces several new explicit feedback mechanisms and modification of the last step, since in the hybrid model, the knowledge discovered for a particular domain may be applied in other domains. Hybrid data mining model includes the following six steps [32].

Understanding of the problem domain: This step helps to work closely with domain experts to define the problem and determine the project goals by selecting key people and learning about current solutions to the problem. It also involves learning domain-specific terminology. Finally, project goals are translated into DM goals, and the initial selection of DM tools to be used later in the process is performed.

Understanding of the data: This step is used for collecting sample data and deciding which data is important. Data are checked for completeness, redundancy, missing values, plausibility of attribute values. Finally, the step includes verification of the usefulness of the data with respect to the DM goals.

Preparation of the data: This phase uses for preparing necessary data for subsequent operations. 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. The cleaned data are fed to next operations like reducing dimensionality, discretization and data generalization. The end results are data that meet the specific input requirements for the DM tools selected in first step.

Data mining: Data mining is the process of extracting or mining knowledge from large data sets. But, knowledge mining from data can describe the definition of data mining even if it is long. data mining have similar or a bit different meaning with different terms, such as knowledge mining from data, knowledge extraction, data/pattern analysis, data archaeology, and data dredging

Evaluation of the discovered knowledge: In this research ifferent classification models will be developed and evaluated using training and testing dataset. The experimental output of the classification models is analyzed and evaluated the performances accuracy using confusion matrix.

Use of Knowledge: After evaluating the discovered knowledge, the last step is using this knowledge for the industrial purposes. In this step the knowledge discovered is incorporated in to performance system and take this action based on knowledge or simply document it and report it to the interested parties and also check and resolve conflicts with previously acquired knowledge if any.

# 2.6 Evaluation of the Classification Model

In order to evaluate the performance of the classifier Prediction Accuracy, True Positive, False Positive, Precision, Recall and F-Measure are commonly used. Confusion matrix helps to see a breakdown of a classifier ‘s performance by showing how frequently instances of a class let us say class X are classified as class X or misclassified as some other class, say class Y [44].

Table 2.1 Confusion Matrix system PROLOG [44].

Table 2. 1 Confusion Matrix system PROLOG [44].

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | Predicted class | | Total instance |
| **+** | **-** |  |
| Actual Class | **+** | True positive | False Positive | Positive |
| **-** | False Negative | True Negative | Negative |

### 2.6.1 Prediction Accuracy

Prediction accuracy measures the proportion of instances that are correctly classified by the classifier.

Predictive Accuracy =

### 2.6.2 True Positive Rate and False Positive Rate

In contrary to Predictive Accuracy TP rate and FP rate values do not depend on the relative size of positive and negative classes [44].

True Positive rate (TP) is the proportion of positive or correctly classified instances as positive or correct instances.

True positive rate =

The false positive (FP)rate is measures the portion of negative instances that are erroneously classified as positive.

False positive tare =

### 2.6.3 Precision and Recall

Precision: is measuring the proportion of instances that are classified as positive that are really positive.

Precision =………………………..(2.5)

**Recall:** is what percent of positive/ negative tuples the classifier labeled as positive or negative for both true and false classes (Blackleg, Anthrax, FMD, LSD, Mastitis, Enteritis, pneumonia, internal parasite, Calf Diphatheria and Amphistomiasis).

**F – Measure:** is calculated as the harmonic mean of recall and precision.

F-measure =

## 2.7 Attribute Selection Measures

An attribute selection measure for developing decision tree is a heuristic for selecting the splitting criterion that best separates a given data partition of class-labeled training instances into individual classes. The attribute selection measure provides a ranking for each attribute describing the given training instances. The attribute that has the best score for the measure is chosen as the splitting attribute for the given instance. The tree node created for partition, let ‘s say D, is labeled with the splitting criterion, branches are grown for each outcome of the criterion, and the instances are partitioned accordingly [36]. This section describes three popular attribute selection measures, namely information gain, gain ratio, and Gini index.

### 2.7.1. Information Gain

Information gain for attribute selection measure is based on the work of Claude Shannon on information theory, which studied the value or information content of messages. Iterative Dichotomiser 3 (ID3) uses information gain for attribute selection measure. The notion used is as follows: - Let D, the data partition, be a training set of class labeled instances. Suppose the class label attribute has m distinct values defining m distinct classes, Ci (for i=1, …, m). The attribute with the highest information gain is selected as the splitting attribute. This attribute minimizes the information needed to classify the instances in the resulting partitions and reflects the least impurity in these partitions. Entropy (impurity) is used to measure the information content of the attributes. High entropy means the attribute is from a uniform distribution whereas low entropy means the attribute is from a varied distribution. Entropy is defined as follows. Let pi be the probability that an arbitrary instance, in D belongs to class Ci, estimated by |Ci, D|/|D|. Expected information (entropy) needed to classify an instance in D is given shows in the following equation 2.1:

**Entropy (E(D)) =∑mi=1 pi𝑙𝑜𝑔(𝑝𝑖) 𝑚 𝑖=1 …………. (2.8)**

Entropy(E(D)) - is the average amount of information needed to identify the class label of an instance in D. The smaller information required, the greater the cleanliness. At this point, the information we have is based solely on the proportions of instances of each class. A log function to the base 2 is used, because the information is encoded or measured in bits.

Suppose attribute A can be used to split D into v partitions or subsets, {D1, D2…, Dv}, where Dj contains those instances in D that have outcome aj of A. Information needed (after using A to split D) to classify D:

.........(2.9)

The smaller the expected information required, the greater the purity of the partitions. Information gained by branching on attribute A is given by

Gain(A)=E(D)-InfoA(D)……. (2.10)

Information gain increases with the average purity of the subsets. The attribute that has the highest information gain among the attributes is selected as the splitting attribute.

### 2.7.2. Gain Ratio

The information gain measure is biased toward tests with many outcomes. That is, it prefers to select attributes having a large number of values. This may result in the selection of an attribute that is non-optimal for prediction. C4.5, a successor of ID3, uses an extension to information gain known as gain ratio, which attempts to overcome this bias. It applies a kind of normalization to information gain using a split information‖ value defined analogously with Info(D) as:

…… (2.11)

This value represents the potential information generated by splitting the training dataset, D, into v partitions, corresponding to the v outcomes of a test on attribute A. Note that, for each outcome, it considers the number of tuples having that outcome with respect to the total number of tuples in D. It differs from information gain, which measures the information with respect to classification that is acquired based on the same partitioning [36]. The gain ratio is defined as:

GainRatio(A)=……..(2.5) …………..(2.12)

The attribute with the maximum gain ratio is selected as the splitting attribute. Note, however, that as the split information approaches 0, the ratio becomes unstable. A constraint is added to avoid this, whereby the information gain of the test selected must be large at least as great as the average gain over all tests examined.

### 2.7.3. Gini Index

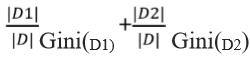
The Gini index is used in Classification & Regression Trees CART. Using the notation described above, the Gini index measures the impurity of D, a data partition or set of training tuples [7], as described in eq. 2.6 as follows:

Gini(D)=1 … …... (2.13)

Where pi is the probability that a tuple in D belongs to class Ci and is estimated by |Ci,D|/|D|. The sum is computed over m classes. To determine the best binary split on A, we examine all the possible subsets that can be formed using known values of A and need to enumerate all the possible splitting points for each attribute. If A is discrete valued attribute having v distinct values, then there are 2v-v possible subsets. When considering a binary split, we compute a weighted sum of the impurity of each resulting partition. If dataset D is split on A into two subsets D1 and D2, the Gini index gini(D) is defined as [36]:

GiniA (D)= Gini(D1) Gini(D2) ……... (2.14)

First, we calculate Gini index for all subsets of an attribute, then the subset that gives the minimum Gini index for that attribute is selected. The point, giving the minimum Gini index for a given

 …………. (2.15)

(continuous valued) attribute is taken as the split-point of that attribute. The reduction in impurity that would be incurred by a binary split on attribute A is

Gini(D) – GiniA(D) = …………… (2.16)

The attribute that maximizes the reduction in impurity (or has the minimum Gini index) is selected as the splitting attribute.

To summarize, the three measures for attribute selection are used mostly. Information gain is biased towards multi valued attributes. Whereas Gain ratio tends to prefer unbalanced splits in which one partition is much smaller than the others. Gini index is biased to multi valued attributes and has difficulty when the number of classes is large. The algorithm used each attribute of the data to make decisions by splitting the data into smaller subsets. All the possible tests are considered during decision making based on the information gain value of each attribute.

## 2.8. Implementation Tools

In order to mine hidden knowledge from the pre-processed dataset and compare the performance of classifiers, WEKA 6.8.1 is used. WEKA is chosen since it is proven powerful for data mining and used by many researchers for mining task and the researcher is familiar with the tool.

It contains tools for data preprocessing, clustering, regression, classification, association rules and visualization. WEKA is written in the Java language and contains a GUI for interacting with data files and producing visual results.

In addition, in order to develop an application, which maps the knowledge acquired from the data mining classifiers with knowledge based system Java NetBeans IDE 8.2 with JDK -8u20 is employed. NetBeans offers easy and efficient project management, has better support for latest Java technologies, and can be installed on all operating systems supporting Java. To represent rules in knowledge base and constructing the Rule based diagnosis disease and advising Knowledge based

## 2.9 Knowledge Based System

Knowledge based systems are sophisticated interactive computer programs which use high quality, specialized knowledge in some narrow problem domain to solve complex problems in that domain. It is a software system that contains a significant amount of knowledge in an explicit and declarative form. Knowledge Based Systems (KBS) have been referred to with a variety of names such as expert systems, intelligent assistants, epistemological systems and design and analysis systems. The two terms most popular in common usage, often used synonymously, are KBS and expert systems [37]. Knowledge based system emulates the behavior of human expert within a well-defined and narrow domain of knowledge [39]. 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.

The area of knowledge based systems (KBS) development has matured over the past two decades. It was started with first generation expert systems with a single flat knowledge base and a general reasoning engine, typically built in a rapid prototyping fashion. This has now been replaced by methodological approaches that have many similarities with general software engineering practice. Knowledge based system development is best seen as software engineering for a particular class of application problems. These applications problems typically require some form of reasoning to produce the required results [39]. Knowledge Based System (KBS) is one of the major family members of the AI group [41]. With availability of advanced computing facilities and other resources, attention is now turning to more and more demanding tasks, which might require intelligence. The society and industry are becoming knowledge oriented and rely on different experts‟ decision making ability. Indeed, KBS can act as an expert on demand without wasting time, anytime and anywhere.

### 2.9.1 Advantages of Knowledge Based System

Knowledge based systems is more useful in many situations than the traditional computer based information systems. Highlighted the following advantages of knowledge based system [42]:

* Time saving: the amount of time spent on doing the work manually is reduced.
* Quality improvement: the quality of decision made increases because there are fewer errors than if the decision performed by manually.
* Practical knowledge made available: knowledge based systems can assist experts in decision making even if they have that knowledge at hand; this improves the accuracy and timelines of the decision made.
* Infallible and 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.
* All day, every day: human experts have fixed working hours or are only available for a limited time throughout the day. They will also experience fatigue because of working long hours which might have a deleterious effect but Knowledge based system can work 24hr/7days/week.
* Updating knowledge. Knowledge based system can be updated easily by editing the rule base; but human expert takes to retrain.

### 2.9.2 Types of Knowledge

Frequent attempts have been made to give a systematic description of knowledge. Some attempts have been based on cognitive theories, whereas others have been formulated to serve as a basis for instructional design theories. Still another approach is to characterize knowledge from an epistemological point of view. Epistemological approaches are task dependent [43]. Within business and knowledge management, two types of knowledge are usually defined, namely explicit and tacit knowledge [44].

Tacit knowledge is the kind of knowledge that is difficult to transfer to another person by means of writing it down or verbalizing it. For example, that Addis Ababa is in the Ethiopia is a piece of explicit knowledge that can be written down, transmitted, and understood by a recipient. However, the ability to speak a language, drive car, play a musical instrument or design and use complex equipment requires all sorts of knowledge that is not always known explicitly, even by expert practitioners, and which is difficult or impossible to explicitly transfer to other users. Explicit knowledge is knowledge that can be readily articulated, codified, accessed and verbalized. It can be easily transmitted to others. Most forms of explicit knowledge can be stored in certain media. The information contained in encyclopedias and textbooks are good examples of explicit knowledge.

### 2.9.3 Architecture of Knowledge Based System

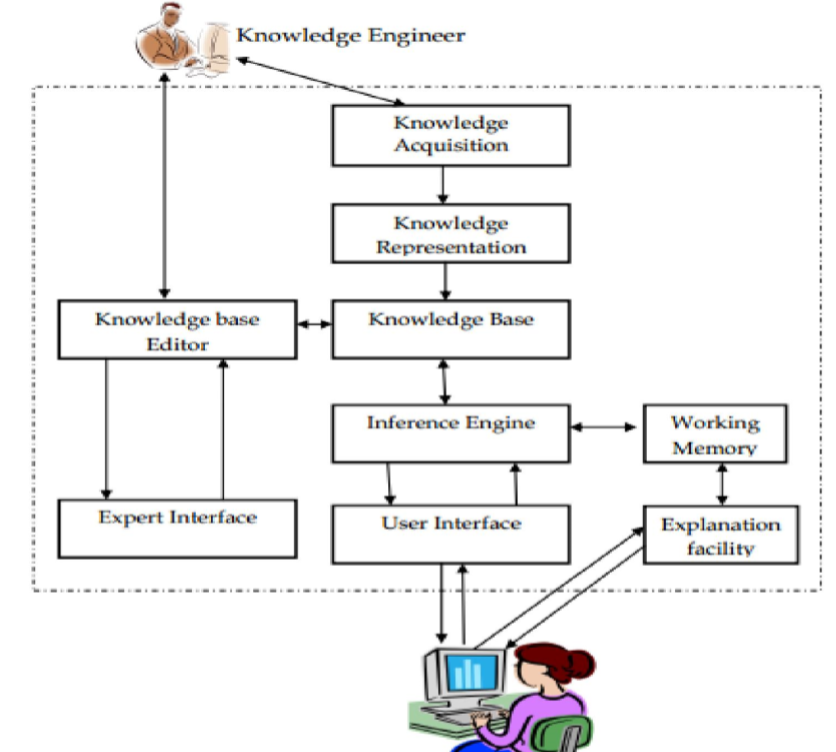


Figure 2. 4 The general architecture of Knowledge Based System [45]

Knowledge base: contains the knowledge necessary for understanding, formulating and for solving problems. It is a warehouse of the domain specific knowledge captured from the human expert via the knowledge acquisition module. To represent the knowledge production rules, frames, logic, semantic net, etc. is used [36]. The knowledge base stores all relevant information, data, rules, cases, and relationships used by the expert system. A knowledge base can combine the knowledge of multiple human experts [39].

Inference Engine: is a brain of expert systems. It uses the control structure (rule interpreter) and provides a methodology for reasoning. It acts as an interpreter which analyzes and processes the rules. It is used to perform the task of matching antecedents from the responses given by the users and firing rules. The major task of inference engine is to trace its way through a forest of rules to arrive at a conclusion [39].

The purpose of the inference engine is to seek information and relationships from the knowledge base and to provide answers, predictions, and suggestions in the way a human expert would. The inference engine must find the right facts, interpretations, and rules and assemble them correctly [39]. Two types of inference methods are commonly used, namely backward chaining and forward chaining. Backward chaining is the process of starting with conclusions and working backward to the supporting facts. Forward chaining starts with the facts and works forward to the conclusions [39].

User Interface: - is the interaction point between the user and the system. The user interface can be graphical user interface (GUI) or command line interface (CLI). In the course of integration of data mining with knowledge based system, Graphical User Interface for the integrator and Command Line Interface for the knowledge based system.

Explanation facility: - it provides information to user for the questions asked by the system. This facility is helpful to have clarity while answering to the questions asked by the system. In steps identifying disease as blackleg, anthrax, foot mouse disease, pneumonia, lumpy skin disease, mastitis, enteritis, and internal parasite the system displays questions. The user can get explanation of the question asked if he/she has no clarity with it.

Knowledge Acquisition: - It is the accumulation, transfer and transformation of problem-solving expertise from experts and/or documented knowledge sources to a computer program for constructing or expanding the knowledge base. It is a subsystem which helps experts to build knowledge bases [36].

The knowledge acquisition process incorporates typical fact finding methods like interviews, questionnaires, record reviews and observation to acquire facts and explicit knowledge. The knowledge acquisition process is usually comprised of three principal stages [7]:

Knowledge elicitation is the interaction between the expert and the knowledge engineer/program to elicit the expert knowledge in some systematic way. The knowledge thus obtained is usually stored in some form of human friendly intermediate representation. The intermediate representation of the knowledge is then compiled into an executable form (e.g. Production rules) that the inference engine can process.

This is a time intensive process, and automated knowledge elicitation and machine learning techniques areincreasingly common modern alternatives stages of knowledge acquisition [7].

## 2.10 Knowledge representation

Knowledge Representation (KR) is the area of artificial intelligence concerned with how knowledge can be represented symbolically and manipulated in an automated way by reasoning program. Knowledge is a progression from data to information, from information to knowledge and from knowledge to wisdom while representation is a combination of syntax, semantics and reasoning. Hence, KR is an area of research in AI which aims at representing knowledge in symbols to facilitate inference from those knowledge elements, creating new elements of knowledge [40].

As a branch of symbolic Artificial Intelligence, knowledge representation and reasoning aims at designing computer systems that reason about a machine-interpretable representation of the world, similar to human reasoning. In Artificial Intelligence, knowledge representation studies the formalization of knowledge and it’s processing within machines [41]. The object of knowledge representation is to express knowledge in computer-tractable form, such that it can be used to help agents perform well [47]. There are different knowledge representation techniques which are discussed in the following sections.

### 2.10.1 Semantic Networks

A semantic network is a graph whose nodes represent concepts and whose arcs represent relations between these concepts. They provide a structural representation of statements about a domain of interest. Semantic networks provide a means to abstract from natural language, representing the knowledge that is captured in text form more suitable for computation [41].

Semantic networks are closely related to another form of knowledge representation called frames systems. In fact, frame systems and semantic networks can be identical in their expressiveness, but use different representation metaphors. While the semantic network metaphor is that of a graph with concept nodes linked by relation arcs, the frame metaphor draws concepts as boxes, i.e. frames, and relations as slots inside frames that can be filled by other frames. Thus, in the frame metaphor the graph turns into nested boxes [41].

The semantic network form of knowledge representation is especially suitable for capturing the taxonomic structure of categories of domain objects and for expressing general statements about the domain of interest. Inheritance and other relations between such categories can be represented in and derived from subsumption hierarchies. On the other hand, the representation of concrete individuals or even data values, like numbers or strings, does not fit well the idea of semantic networks [41].

### 2.10.2. Rule Based Knowledge Representation

A common way to represent knowledge is with logic. It is important to emphasize that logic is not the knowledge itself; it is simply a way of representing knowledge. However, logic can be viewed as a form of meta-level knowledge about how to represent and reason with knowledge. What logic enables us to do is represent the knowledge possessed by an agent using a finite set of logical expressions plus a process (namely, the inference rules of logic) for generating a (potentially unlimited) set of other logical expressions that are part of the agent’s knowledge [39].

Rules come in the form of IF-THEN-constructs and allow expressing various kinds of complex statements. Rules can be found in logic programming systems, like the language Prolog, in deductive databases or in business rules systems. The IF-part is also called the body of a rule, while the THEN-part is also called its head. Typically, rule-based knowledge representation systems operate on facts, which are often formalized as a special kind of rule with an empty body. They start from a given set of facts and then apply rules in order to derive new facts, thus “drawing conclusions” [40].

It is important to distinguish between facts and their representations. Facts are part of the world, whereas their representations must have been coded in some way that can be physically stored within an agent. We cannot put the world inside a computer (nor can we put it inside a human), so all reasoning mechanisms must operate on representations of facts, rather than on the facts themselves. Because sentences are physical configurations of parts of the agent, reasoning must be a process of constructing new physical configurations from old ones. Proper reasoning should ensure that the new configurations represent facts that actually follow from the facts that the old configurations represent [20].

### 2.10.3. Logic Knowledge Representation

Both forms, semantic networks as well as rules, have been formalized using logic to give them a precise semantics. Without such a precise formalization they are vague and ambiguous, and thus problematic for computational purposes. The most prominent and fundamental logical formalism classically used for knowledge representation is the “first-order predicate calculus”, or first-order logic for short. First order logic allows one to describe the domain of interest as consisting of objects, i.e. things that have individual identity, and to construct logical formulas around these objects formed by predicates, functions, variables and logical connectives [40]. Similar to semantic networks, most statements in natural language can be expressed in terms of logical sentences about objects of the domain of interest with an appropriate choice of predicate and function symbols. Concepts are mapped to unary, relations to binary predicates. In particular, more complex restrictions that range over larger fragments of a network graph can be formulated in logic, where the intuitive graphical notation lacks expressivity [40].

## 2.11 Related Works

Different researches have been undertaken throughout the world in the past decades. Some of them that are related with works done on diagnosis and treatment of cattle disease are discussed below.

Agonifo and et al [42] developed a fuzzy expert system for diagnosis of cattle disease in Nigeria. In their research, they have developed a three layered architecture, namely application layer, business layer and storage layer. The system has given great advantages for the veterinary doctors. It gives reliable, high speed and accurate observations in diagnosing of cattle diseases. It can also be used in the absence of clinicians to diagnose diseases.

The system uses the Bayesian Belief Network methodology described. The system is parameterized using data collected from veterinary experts. A fuzzy rule base contains a set of fuzzy rules in a single IF-THEN rule in the form of membership value. There is a procedure for the fuzzy algorithm to diagnose the disease.

First to display the list of all the symptoms from the database. After getting the database to select the particular symptom of concern to be diagnosed and to select the level of severity of each selected symptom. Each selected symptom, search the database for all the diseases that has the symptom and then the weight of the symptom and multiply it by the appropriate linguistic variable. If select the disease with the highest value after summation has been carried out in step four and display its corresponding disease. This is the diagnosed disease is displayed in the application. Finally, the researchers have recommended that the work can be extended using the combined features of neural network and fuzzy logic to improve disease diagnosis in cattle.

Birhane [50], has developed a prototype KBS for diagnosis and treatment of diseases of sheep and goats. Common Knowledge Acquisition and Documentation Structuring (CommonKADS) knowledge engineering methodology is used to model the requirements of the system. Production rule is used to develop the prototype system as knowledge representation technique.

For this research, the rule based knowledge representation method is chosen because it clearly demonstrates the domain knowledge. There are already defined set of symptoms, syndromes and basic issues that should be addressed to confirm the presence of sheep and goat diseases. As a result, rule based representation method is more appropriate to represent and demonstrate the real domain knowledge in diagnosing sheep and goat diseases.

The researcher evaluates the performance of the result is by gathering information from the domain expert. Therefore, for the comfort of analyzing the relative performance of the system based on user evaluation to evaluate the performance. The overall average performance accuracy of the system is about 84.8%. This implies that the modeled system performs well in making right decisions in the diagnosis and treatment of sheep and goat diseases [40].

Derejaw [9], has developed a Web-based expert system for diagnosing cattle diseases. On his research, Knowledge is obtained from domain experts (veterinarians). For this research, the rule based knowledge representation method is chosen because it clearly demonstrates the domain knowledge. There are already defined set of symptoms, syndromes and basic issues that should be addressed to confirm the presence of cattle diseases. As a result, rule based representation method is more appropriate to represent and demonstrate the real domain knowledge in diagnosing cattle diseases.

The researcher evaluates the performance of the result by gathering information from the domain expert. Therefore, for the comfort of analyzing the relative performance of the system based on user evaluation to evaluate the performance. The overall average performance accuracy of the system is about 87.2%. This implies that the system performs well in making right decisions in the diagnosis of cattle diseases. Generally, we can conclude that the respondents are satisfied with the following criteria: user interface design, accuracy, response time and significance of the system [9].

Abdulkerim [4] has conducted a research on Towards Integrating Data Mining with Knowledge Based System: The Case of Network Intrusion detection and the general objective of this study was to construct a prototype knowledge base system which can update its knowledge base using the hidden knowledge extracted from intrusion dataset by using data mining techniques. The researcher used Knowledge Discovery in Database (KDD) model for the data mining task, Rule based knowledge representation approach to represent knowledge, Prototyping approach to develop the knowledge based system, WEKA 6.8.1 to mine hidden knowledge and compare the performance of classifiers, Java NetBeans IDE 7.3 with JDK 6 to develop an integrator application, PROLOG to represent rules in knowledge base, Precision, Recall, Fmeasure and True Positive rate to evaluate accuracy of the models and test cases to evaluate the performance of the KBS.

The system was aiming at utilizing hidden knowledge extracted by employing induction algorithm of data mining, specifically JRip from sampled KDDcup ‘99 intrusion data set. The integrator application then links the model created by JRip classifier to knowledge based system so as to add knowledge automatically. In doing so, the integrator understands the syntax of JRip classifier and PROLOG and converts from rule representation in JRip to PROLOG understandable format. The performance of the system was evaluated by preparing test cases. Twenty test cases were prepared for system performance test and provided to domain experts. For user acceptance test users were trained and evaluated the system. Generally, the system has scored 80.5% overall performance [4]. Hence, Abdulkerim recommended to apply integration of machine learning with knowledge based system in other domain areas than intrusion detection, especially in areas where there is shortage of domain expert to acquire knowledge.

Ahmad et al [45], conducted the research on web based cattle disease expert system diagnosis with forward chaining method in Indonesia. The researcher has used forward chaining and certainty factor method. Production rules are one of the many forms of knowledge representation used in the development of expert systems. Representation of knowledge with the rules of production, basically a rule (rule) in the form of IF-THEN. There are two stages of the model that are often used to calculate the level of confidence (CF), of a rule, namely digging out the results of interviews with experts and Manual calculation of CF Value. Finally, the modeled system performs not well known in making right decisions in the diagnosis of cattle diseases.

Rong and Daoliang [47], have developed a web based expert system for diagnosis of cow disease. Their system has been composed of three-tier web application which uses Internet Information Server (IIS), Microsoft SQL server 2000, Windows XP as the operating system and Windows XP with Internet Explorer (IE) on the client side. Their proposed system has adopted three algorithms, i.e. Case Based Reasoning (CBR), subjective Bayesian theory and Dempster-Shafer (D-S) evidential theory. Their developed system consists of four subsystems. These are case management, diagnosis, prevention and cure, and medication management. To diagnose the milk cow disease, the system first looks for a disease that can explain all the symptoms. If there is no such disease, it looked for a disease that can explain all but one symptom, etc., until it finds a possible disease.

Tesfamariam [49], has developed diagnosis and treatment of system for Visceral Leishmaniosis (VL) using integration of data mining and knowledge base system. The researcher decides to use rules that are generated by the PART classification algorithm model for further use in the development of knowledge base of the knowledge based system. Finally, [48], result study archives promising result with 95% and 86% of the system performance and user acceptance respectively.

Tadesse [ 55], conducted the research on integrated datamining result with knowledge based system for diagnosis and treatment cattle diseases in the Case of Debre Birhan Basso Animal Health Center. For this research, Hybrid model knowledge discovery approach engineering methodology to extract data mining from Debre Birhan Animal Health Center. The researcher used mixed (quantitative and qualitative) research design methods.

The researcher decides to use rules that are generated by the JRip classification algorithm model for further use in the development of knowledge base of the knowledge based system. Finally, [54], The researcher evaluates the performance of the result is by gathering information from the domain expert. Therefore, for the comfort of analyzing the relative performance of the system based on user evaluation to evaluate the performance. The overall average performance accuracy of the system is about 98.7% and 85.8% of the system performance and user acceptance respectively.

# CHAPTER THREE

# METHODS AND APPROACH

# 3.1. Introduction

This chapter deals about methodology i.e. data mining techniques used through the research process which includes general approach of the research design, the data collection, analysis and interpretation of artifacts. Having the description of the type of the study follows a hybrid model knowledge discovery approach which starts from understanding of the problem to use discovered knowledge.

## 3.2. Research Design

The research design a mixed research method design. It uses data mining and knowledge base development methodologies. In order to apply the Data mining technology, one must follow the standard process steps. This research used CRISP data mining methodology. Figure 3.1 illustrates the overall research development framework:

Understanding the problem

ILRI Cattle

Dataset

Da Da

Handing Missing value

Data transformation

Format data

Preparation of the data

Attribute selection

Data

Preprocessing

Naïve Bayes

J48

JRip

Performance evaluation

Generate rule

Knowledge representation

Knowledge elicitation

Developing knowledge base with Swing Prolog

Domain experts

knowledge base system evaluation

Figure 3. 1 Data Mining System framework [43]

### 3.2.1. Understanding of the Problem Domain

The researcher understands the problem domain by conducting some structured and unstructured sample interview questions for purposely selected domain experts. This initial step involves working closely with domain experts to define the problem and determine the project goals, identifying key people, like an animal health officer, veterinaries, laboratory technician, pharmacist and learning about the current solution to the problem. Here is a significant number of cattle are infected by disease before diagnosis and treatment of the disease. Finally, the project goal is translated into data mining goals, and the initial selection of data mining tools to be used later in the process is performed [17]. The acquired knowledge is refined with the consultation of the domain experts. Moreover, secondary sources of knowledge are gathered from the Internet, animal disease diagnosis and treatment guidelines (especially, The Drug Administration and Control Authority of Ethiopia, Standard Veterinary Treatment Guidelines for veterinary clinics), research papers and articles by using document analysis technique.

### 3.2.2. Understanding of the Data

Nowadays, data stored in medical databases are growing in an increasingly rapid way. Due to this tendency data mining application in animal healthcare today is excessive, because animal healthcare organizations today are capable of generating and collecting a large amounts of data. This increase in volume of data requires an automatic way for these data to be extracted when needed. With the use of data mining techniques, it is possible to extract interesting and useful knowledge and this knowledge can be used by physicians to determine diagnoses, prognoses and treatments for patients in healthcare organizations which improve work efficiency and enhance the quality of the decision making process [36]. The dataset for this study have been collected from

International livestock research center which is found in the Addis Abeba, Ethiopia and the organization have collected these data from 2013 to 2018 years. During these years the organization has used different types of data base systems and the records that exist in these data bases contain different attributes in number and type. For the year 2013 and 2014 the organization has used the same data base system and these data records have 21 attributes, for the year 2015 the organization used another data base system and these records have 23 attributes, for the year 2016 another data base system was used and these records have 28 attributes, for the year 2017 and 2018 another data base system was used and these records contain 34 attributes. The original size of the whole data before preprocessing was 18.3 MB. During the data understanding phase the application process for cattle disease diagnosis and treatment at the international livestock research center is studied, including the interviews from the veterinaries and guidelines, in order to identify the types of data collected from the diseased animal registered and stored in the international livestock research center databases in manual format.

The following table in the next page summarizes the number of attributes and number of records that the researcher has collected from international livestock research center.

Table 3. 1 Number of Attributes and Records

|  |  |  |
| --- | --- | --- |
| Year | No of Attributes | No of Records |
| 2013 | 21 | 680 |
| 2014 | 21 | 880 |
| 2015 | 23 | 1126 |
| 2016 | 28 | 1230 |
| 2017 | 34 | 1200 |
| 2018 | 34 | 1239 |
| Total | | 6355 |

### 3.2.3. Preparation of the Data

This step covers all activities needed to construct the final dataset, which constitutes the data that can be fed into data mining tool(s) in the next step. To be useful for data mining purpose, the database need to undergo preprocessing, in the form of data Cleaning, data reduction, data transformation and Waikato Environment for Knowledge Analysis (WEKA) understandable format.

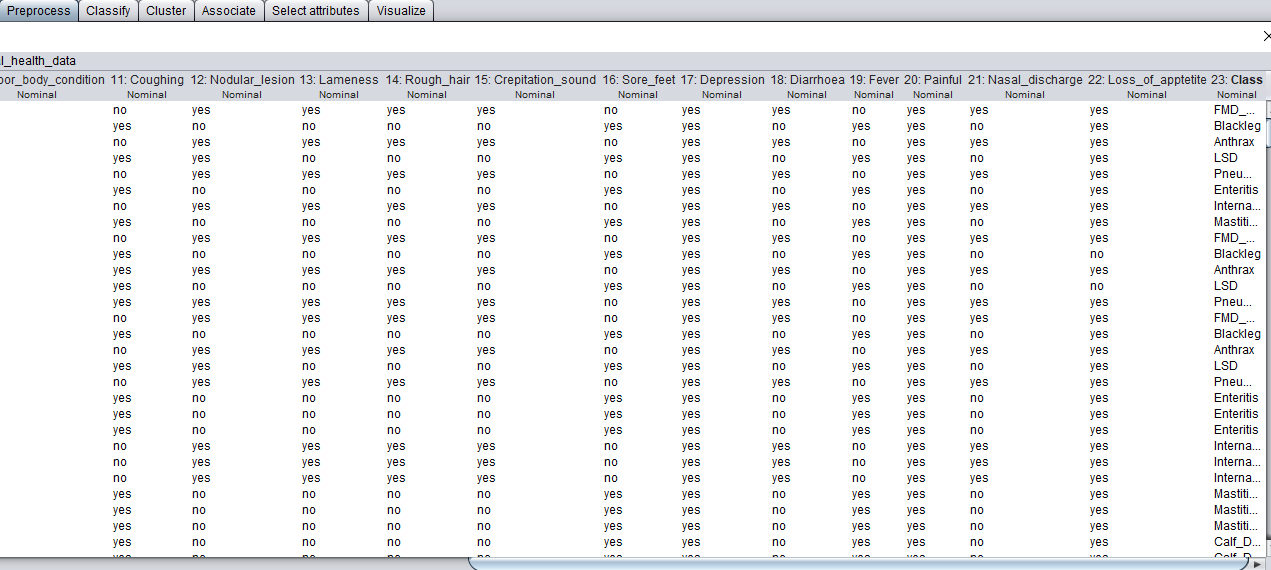
#### 3.2.3.1. Handling Missing Values

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 [18]. Anticipating that data will be 100% complete and error free is unrealistic when working with patient data which collected in complex health care systems. Data error identification is both an automated and a manual process, and required an iterative procedure that drew upon expertise from the clinical experts as well as statistical experts and the data warehouse engineer [19].

International livestock research center dataset has instances with total records of 6355 and 33 attributes including class label, among these records 123 (6.7%) are missing values in only temperature attribute. We use Matlab as a tool to handle missing values by using mean average techniques to fill the missing values.

This is one of the most frequently used methods. It consists of replacing the missing data for a given feature (attribute) by the mean of all known values of that attribute in the class where the instance with missing attribute belongs.

Figure 3. 2 Sample data after handling missing values



#### 3.2.3.2 Data Transformation

Data Transformation techniques can be used to reduce the number of values for a given continuous attribute by dividing the range of the attribute into intervals. [18].

In this study, the researcher performed generalization of the records by using categories for only temperature attribute which is grouped in low, normal and high. In the initial dataset, the body temperature is measured using degree centigrade. If temperature is more than 39, it is a high value, the range between 38 and 39 is normal and less than 38 is low. So, the researcher categorized them into three groups as indicated in Table 3.3.

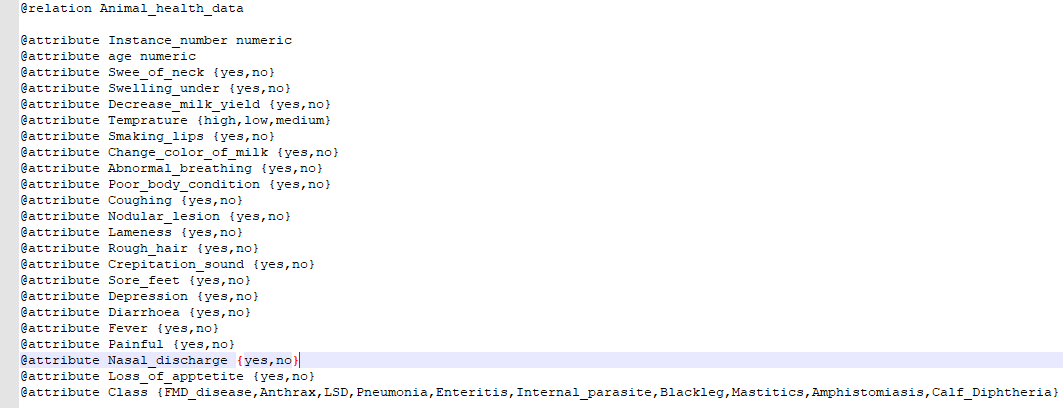
Table 3. 2 Data Transformation by the cattle temperature range [4]

|  |  |
| --- | --- |
| Value | Temperature in degree centigrade |
| Low | <38 |
| Normal | 38-39 |
| High | >39 |

#### 3.2.3.3 Format Data

Formatting transformations refers to primarily syntactic modifications made to the data that do not change its meaning, but might brokered by the modeling tool. After all, preprocessing the data is covered into Dot Comma Separated Value (. CSV) format and for the WEKA it’s again converted into dot attribute Relation File Format (. ARFF) file format.

Figure 3. 3 Illustrates sample WEKA ARFF files show as follows:



#### 3.2.3.3 Class Imbalance

The class imbalance problem is prevalent in many applications, including: fraud/intrusion detection, risk management, text classification, and medical diagnosis/monitoring. It typically occurs when, in a classification problem, there are many more instances of some classes than others. In such cases, standard classifiers tend to be overwhelmed by the large classes and ignore the small ones. Particularly, they tend to produce high predictive accuracy over the majority class, but poor predictive accuracy over the minority class [32].

A number of solutions to the class imbalance problem were proposed both at the data and algorithmic levels. At the data level, these solutions include many different forms of resampling such as over-sampling and under sampling. SMOTE (Synthetic Minority Oversampling Technique) is an oversampling approach which generates synthetic examples in a less application specific manner. The minority class is oversampled by taking each minority class sample and introducing synthetic examples along the line segments joining any/all of the k minority class nearest neighbors [32].

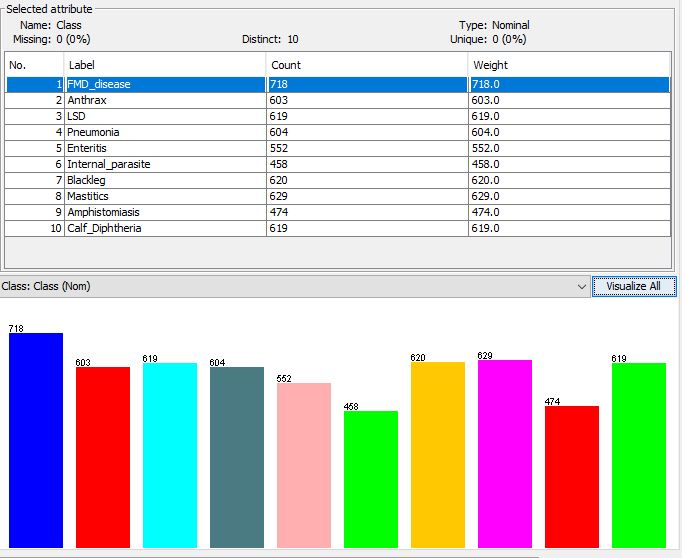


Figure 3. 4 Imbalance dataset before SMOTE

As we can observe in the figure below, the classes are imbalance and to avoid this, the researcher SMOTE the data set and discretize it before conducting the experiments and as a result the dataset increases from 5897 records to 6355 records.

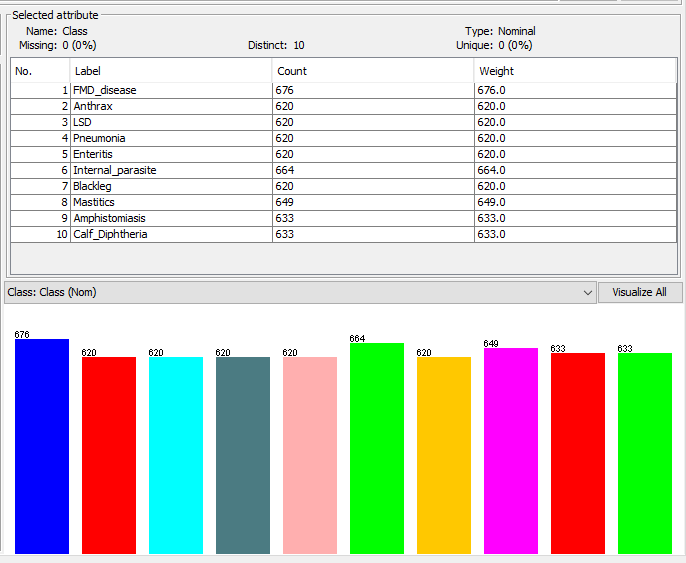


Figure 3. 5 Balanced Dataset after SMOTE

#### 3.2.3.4 Attribute Selection

Attribute selection searches through all possible combinations of attributes in the data and finds which a subset of attribute works best for prediction. Attribute selection also known as feature selection which select relevant attributes and remove redundant and/ or irrelevant attributes. In this research, roll number, date and kebele, removed because all are not important to determine the cattle disease. Species, sex and breed also removed because all are only identifiers of animals. We want to eliminate irrelevant attributes and evaluate attributes individually, so that we use the ranking method with Gain Ratio Attribute Evaluator were followed for selecting best attribute selection for data mining that come up 20 out of 33 attributes. Table 3.3 shows unnecessary attribute for diagnosis and treatment for cattle disease as follows:

Table 3. 3 Removing unnecessary attributes of the table

|  |  |
| --- | --- |
| Attribute name | Remark |
| Roll No, Date, Kebele, Species, Sex, Breed | Not important for the study |

# CHAPTER FOUR

# KNOWLEDGE MODELING, EXPERIMENTATION, AND ANALYSIS

## 4.1 Knowledge Modeling

Cattle disease diagnosis involves many steps beginning with collecting information about the cattle disease owner’s case history, symptoms, and performing a clinical examination by veterinaries. Although there are no laboratory tests in the Basso Animal Health Center to diagnose specifically cattle disease, the veterinary use various tests to make sure something else is not causing the 45 symptoms. The veterinary bases his or her diagnosis on the cattle owner case history of symptoms, including any functional problems caused by the symptoms and his or her observation. The veterinary then determines if the case history symptoms and degree of disability point to a diagnosis of a specific disease. Cattle disease diagnosis is one and basic part in the diagnosis of animal medical health conditions which needs careful examination and diagnosis. In the Clinical Diagnosis, a decision needs to be made whether the cattle disease is suspected or not. The input knowledge role consists of data about the diseased case such as, Fever, Cough, Painful, Diarrhea, Nasal discharge, Abortion, Loss of appetite, swelling of neck, Lameness, Crepitation sound, Change color of milk, Decrease milk yield, swelling udder, Shivering, Sore feet, Salivation, Rolling, Smacking lips, Depression, Nodular lesion, Abnormal breathing, Rough hair, Emaciation and Poor body condition.

In this study different experiments were conducted using various data mining methods to derive knowledge from preprocessed data for diagnosis of cattle disease from the preprocessed data. After preparation of the data, the next task is the mining process. As it has been stated in the previous sections, a total of 6355 data records were preprocessed to perform the experiment.

## Selecting Modeling Technique

Selecting appropriate model depends on data mining goals. Consequently, to attain the objectives of these research three classification techniques, namely the Naïve Bayes classification, decision tree classification and rule based classification has been selected for model building. The analysis was performed using WEKA environment. Among the different available classification algorithms in WEKA, Naïve Bayes, J48 and JRip algorithms are used for experimentation of this study. The reason why the researcher selected the above algorithms is because of the algorithms are easy to understand and interpret the results of the model. They have also advantages such as high tolerance to noise, and the ability to classify unseen patterns.

## Experimental Setup

In any data mining research before developing a model, we should generate a mechanism to test the model performance. For instance, in the supervised data mining task, such as classification, it is common to use classification accuracy measure, True Positive rate (TP), precision, recall and F-measure of the experts are used as to measure the performance of the developed data mining model.

### 4.3.1 Methods of Training and Testing Option

For the experimental setup, the cattle disease datasets were converted to ARFF (Attribute Relation File Format), because it is a suitable input file format for the WEKA system. In order to perform the experiment, the researcher used two methods to classify the dataset, the first method is k-fold (10-folds) cross validation and, the second method is percentage split.

Therefore, the dataset is randomly partitioned equally into ten parts. Consequently, 90% of the dataset is used for training and 10% for testing and the dataset are partitioned into percentages splits option (70%: 30%) that means 70% of the dataset is used for training and the remaining 30% for testing purpose. As the researcher explained on the experimental setup to build the model for each algorithm perform three experiments based on the two methods, namely the experiment performed based on K-fold (10 – fold) cross validation method called experiment I and the second experiment performed based on the percentage split method called experiment II.

### 4.3,2 Developing Classifier Model Using Naïve Bayes

**Experiment I:**

This experiment conducts under 10-fold cross-validation test option with default parameters of WEKA and the algorithm generates a model as Naïve Bayes and Correctly Classified Instances are 6170 which means 95.42 % and Incorrectly Classified Instances are 185 which means 5.15% from Total Number of 6355 of Instances and taking 0.62 seconds to build the model.

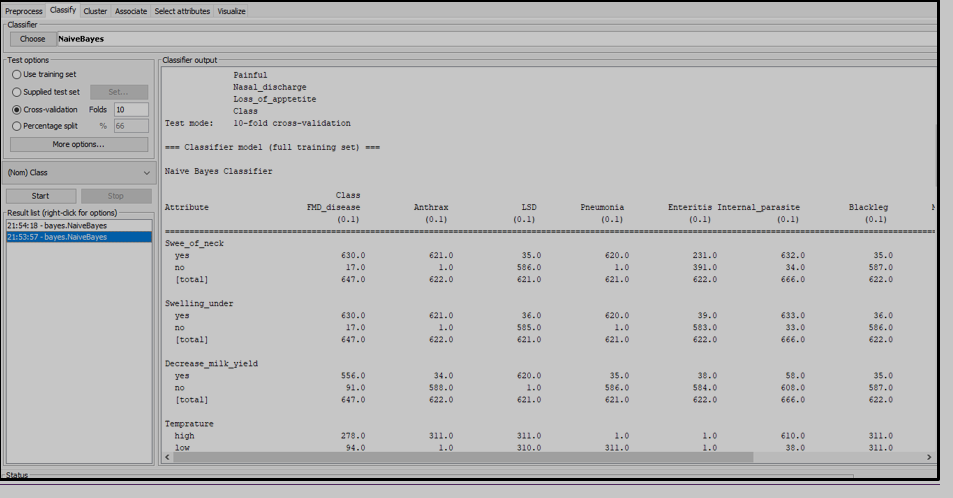


Figure 4. 1 Illustrates sample WEKA result

**Experiment II:**

Percentage split test option to train and test the classification model. Out of the 6355 total records 4449 (70%) of the instances were used as training dataset and the remaining, 1906 (30%) of the instances were used as a testing dataset. The Naïve Bayes learning algorithm scored an accuracy out of total 1906 number of testing instances 1850 (95.35 %) of them are classified correctly and the remaining 156 (6.65%) testing instances are misclassified or incorrectly classified.

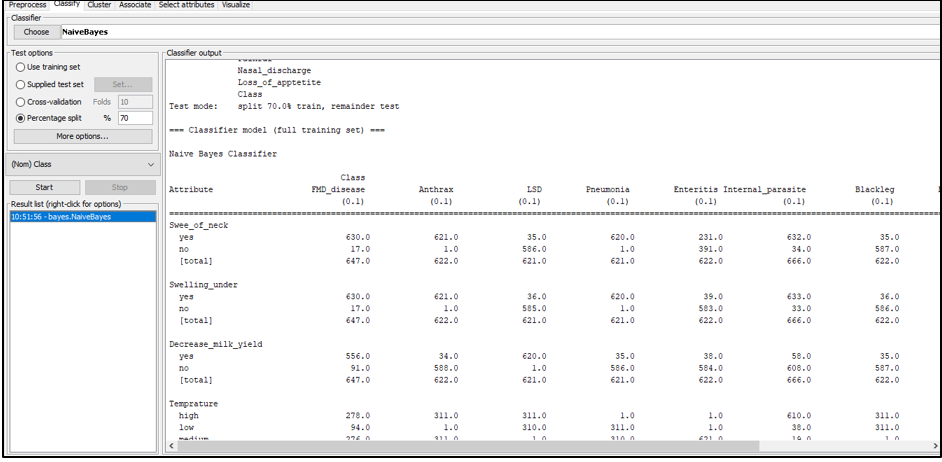


Figure 4. 2 Illustrates sample WEKA result

To conclude, the above two experiments namely experiment I and II performed in order to build the classification model using Naïve Bayes classification algorithm by applying k-fold cross validation and percentage split method in respectively on the experiments.

Experiment I and Experiment II shows that the Classification accuracy of the models based on the above two methods respectively. The first experiment was performed based on 10-fold cross validation method and classifies with 95.38%, accuracy rate, and the second experiment performed based on 70%:30% percentage split classifies with 95.45% accuracy rate.

In conclusion, when we compared the two experiments the first experiment performed based on K-fold cross validation dataset has a better accuracy performance than the second experiment performed by percentage split testing dataset.

Table 4. 1 Detailed Accuracy by Class for Naïve Bayes classification algorithm:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Detailed Accuracy by Class | | | | | |
|  | **TP Rate** | **Precision** | **Recall** | **F-Measure** | **Class** |
| 97.90% | 95.50% | 97.30% | 97.40% | FMD\_disease |
| 92.60% | 96.70% | 92.60% | 94.60% | Anthrax |
| 94.50% | 94.70% | 94.10% | 96.90% | LSD |
| 97.60% | 97.90% | 97% | 91.90% | Pneumonia |
| 96.50% | 99.80% | 96.90% | 98.30% | Enteritis |
| 93.40% | 95.10% | 94.23% | 96.60% | Internal\_parasite |
| 95.60% | 91.60% | 97.10% | 96.80% | Blackleg |
| 96.90% | 92% | 96.90% | 93.40% | Mastitics |
| 92.90% | 94.60% | 89.90% | 95.20% | Amphistomiasis |
| 96.90% | 96.70% | 96.90% | 93.40% | Calf\_Diphtheria |
| Weighted Average | 95.48% | 95.49% | 95.24% | 95.45% |  |

### 4.3.3 Developing Classifier Model Using J48 Decision Tree

**Experiment I:**

This experiment conducts under 10-fold cross-validation test option with default parameters of Weka and the algorithm generates a model as a decision tree with 32 number of Leaves and 61 size of the tree and correctly classified Instances are 6199 which means 96.65% and Incorrectly Classified Instances are 156 which means 3.35% from Total Number of Instances of 6355 and taking 0.13 seconds to build the model.

**Experiment II:**

This experiment conducted in percentage split test option to train and test the classification model. Out of the 6355 total records 4449(70%) of the instances were used as training dataset and the remaining, 1906(30%) of the instances were used as a testing dataset. The J48 learning algorithm scored an accuracy out of 1906 total number of testing instances 1845 (96.25%) of them are classified correctly and the remaining 166(3.75%) testing instances are incorrectly classified. The algorithm generates a model as a discussion tree with 32 number of leaves and 61 size of the tree and taking 0.16 seconds to build the model.

To conclude, the above two experiments namely experiment I and II performed in order to build the classification model using J48 classification Algorithm by applying k-fold cross validation and percentage split methods in respectively on the experiments. The first experiment was performed based on 10-fold cross validation method and classifies with 96.65 % accuracy rate, and the second experiment performed based on 70%:30% percentage split classifies with 96.25% accuracy rate.

So, when we compared the two experiments performed by different methods, the first experiment performed based on K-fold cross validation has a better accuracy performance than the second experiment performed by percentage split.

Table 4. 2 Detailed Accuracy by Class for J48 classification algorithm:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Detailed Accuracy by Class | | | | | |
|  | **TP Rate** | **Precision** | **Recall** | **F-Measure** | **Class** |
| 97.90% | 96.50% | 97.30% | 97.40% | FMD\_disease |
| 99.60% | 97.70% | 98.60% | 98.60% | Anthrax |
| 96.50% | 94.70% | 99.10% | 96.90% | LSD |
| 97.60% | 97.90% | 98% | 94.90% | Pneumonia |
| 96.90% | 94.80% | 96.90% | 97.30% | Enteritis |
| 98.40% | 95.10% | 98.00% | 98.60% | Internal\_parasite |
| 95.60% | 99.60% | 95.10% | 95.80% | Blackleg |
| 96.90% | 96% | 96.90% | 94.40% | Mastitics |
| 89.90% | 97.60% | 89.30% | 96.80% | Amphistomiasis |
| 96.90% | 96.70% | 96.90% | 95.40% | Calf\_Diphtheria |
| Weighted Average | 96.62% | 96.66% | 96.61% | 96.61% |  |

### 4.3.4 Developing Classifier Model Using JRip Rule Based

**Experiment I:**

In this experiment JRip rule induction algorithm is employed. Therefore, to generate IF-THEN rules from the experimental International Livestock research center dataset JRip algorithm with its default values of the parameter and 10-fold cross-validation test mode is employed. JRip correctly classified 6250 which means 97.48% instances and incorrectly classified 105 which means 2.32% instances. The algorithm generates 23 rules taking 0.37 seconds to build the model.

**Experiment II:**

For this experiment the K-fold cross validation method is changed into percentage split test option to train and test the classification model. Out of the 6355 total records 4449 (70%) of the instances were used as training dataset and the remaining, 1906 (30%) of the instances were used as a testing dataset. The JRip learning algorithm scored an accuracy out of 1906 total number of testing instances 1866 (97.25%) of them are classified correctly and the remaining 140 (2.55%) testing instances are misclassified or incorrectly classified. The algorithm generates 20 rules and taking 0.30 seconds to build the model.

To conclude, the above two experiments namely experiment I and II performed in order to build the classification model using JRip classification Algorithm by applying k-fold cross validation and percentage split methods in respectively on the experiments. The first experiment was performed based on 10-fold cross validation method and classifies with 97.78 accuracy rate, and the second experiment performed based on 70%:30% percentage split classifies with 97.25 accuracy rate. So, when we compared the two experiments performed by the two methods, the first experiment performed based on K-fold cross validation has a better accuracy performance than the second experiment performed by percentage split.

Table 4. 3 Detailed Accuracy by Class for JRip classification algorithm:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Detailed Accuracy by Class | | | | | |
|  | **TP Rate** | **Precision** | **Recall** | **F-Measure** | **Class** |
| 97.90% | 99.50% | 97.30% | 98.40% | FMD\_disease |
| 99.60% | 97.70% | 99.60% | 94.60% | Anthrax |
| 99.50% | 94.70% | 99.10% | 96.90% | LSD |
| 97.60% | 97.90% | 99% | 98.90% | Pneumonia |
| 96.90% | 99.80% | 96.90% | 98.30% | Enteritis |
| 98.40% | 95.10% | 98.00% | 98.60% | Internal\_parasite |
| 95.60% | 99.60% | 97.10% | 96.80% | Blackleg |
| 96.90% | 96% | 96.90% | 98.40% | Mastitics |
| 96.90% | 99.60% | 89.90% | 95.80% | Amphistomiasis |
| 96.90% | 96.70% | 96.90% | 98.40% | Calf\_Diphtheria |
| Weighted Average | 97.62% | 97.69% | 97.02% | 97.51% |  |

### 4.3.5 Performance Comparison of Naïve Bayes, J48 and JRip Models

Selecting a better classification technique for building a model, which performs best in handling the classification, is one of the aims of this research. For this reason, the three selected classification model with respective best performance accuracy is listed in below table 4.4:

Table 4. 4 Comparison of the results of the models

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Algorithms used | Correctly classified instances | | Incorrectly classified instances | | Time taken (second) |
| **10 Fold Cross Validation Test Option** | | | | | |
|  | No of correctly instance | Accuracy | No of incorrectly instance | Accuracy |  |
| Naïve Bayes | 6170 | 95.42% | 185 | 5.15% | 0.62 |
| J48 | 6199 | 96.65% | 156 | 3.35% | 0.41 |
| Jrip | 6250 | 97.48% | 105 | 2.32% | 0.37 |
| **Percentage Split Test Option** | | | | | |
| Naïve Bayes | 1850 | 95.35% | 156 | 6.65% | 0.52 |
| J48 | 1845 | 96.25% | 166 | 3.75% | 0.26 |
| JRip | 1866 | 97.25% | 140 | 2.55% | 0.30 |

In these experiments three algorithms are used, namely Naïve Bayes, J48 decision tree and JRip rule based. From each method totally six models are developed based on the two methods. As shown on the above comparison table (Table 4.4), all the results were almost all closely equal, but the difference lies in the execution period or the time taken to build the model.

The graphical representation of the algorithms with respect to classes for TP rate is indicated in Figure 4.1:

Figure 4. 3 TP rate of classifiers

## 4.4 Classifying Model Performance Evaluation Metrics

Evaluation of classifier datamining algorithms can be compared according to a number of measures. In comparing the performance of different classifier data mining algorithms to determine its classifications, some quantities that interpret the goodness of fit of a model, and error measurements must be considered.

The final comparative analysis of the models shows on the above table 4.4 JRip rule classifier algorithm with 10-fold cross validation test option are performed best classification accuracy of

97.48%. Whereas J48 and Naïve Bayes classifier algorithms with 10-fold cross validation test option performed classification accuracy of 96.65% and 95.42% of the result respectively.

In addition to those, the final comparative analysis of the models shows on the above table 4.4 the JRip classifier algorithm Percentage split test option in the testing dataset performed classification accuracy is 97.25% Whereas J48 decision tree and Naïve Bayes algorithm with percentage split test option performed best classification accuracy is 96.25% and 95.35% results respectively.

So, to sum up the model which was developed by the JRip rule classifier algorithm with 10-fold cross validation test option method was selected as the best classification model based on the accuracy evaluation methods used in this study. Hence, to evaluate the performance of the classifiers employed in this study True Positive rate, Precision, Recall and F-measure are used.

## 4.5 Rule Extraction from JRip Classification Algorithm

Having generated rules using JRip classifier, the next task is building or constructing the knowledge base. For this study, we devised an automatic construction of knowledge bases aligned with the data mining task. The overall task of the application is to extract rules from International Livestock Research Center dataset using JRip classifier and mapping JRip rules to Prolog rules.

Table 4. 5 Rules extracted by JRip Classification algorithm

|  |  |
| --- | --- |
| No | Rulues |
|  | (Decrease\_milk\_yield = yes) and (Temprature = high) and (Swee\_of\_neck = no) and (Loss\_of\_apptetite = no) => Class=LSD (310.0/0.0) |
|  | (Temprature = Normal) and (Fever = Yes) and (Decrease\_Milk\_yield = No) and (Nodular\_Lesion = No) and (Painful = No) and (Loss\_of\_Apptetite = Yes) and (Lameness = No) and (Depression = No) => Class=Pneumonia (7.0/1.0) |
|  | (Swee\_of\_neck = no) and (Temprature = high) and (Decrease\_milk\_yield = no) => Class=Blackleg (310.0/0.0) |
|  | (Temprature = medium) and (Smaking\_lips = yes) and (Decrease\_milk\_yield = no) and (Nodular\_lesion = yes) and (Coughing = yes) => Class=Anthrax (310.0/0.0) |
|  | (Temprature = High) and (Painful = No) and (Loss\_of\_Apptetite = Yes) and (Swelling\_Udder = No) and (Crepitation\_Sound = No) and (Fever = Yes) and (Depression = No) => Class=Calf\_Diphtheria (28.0/3.0) |
|  | (Fever = No) and (poor\_Body\_condition = No) and (Nodular\_Lesion = No) and (Temprature = Normal) and (Rough\_hair = No) and (Nasal\_Discharge = No) and (Loss\_of\_Apptetite = No) => Class=Enteritis (56.0/9.0) |
|  | (Temprature = High) and (Painful = No) and (Loss\_of\_Apptetite = No) and (Swelling\_Udder = No) and (Crepitation\_Sound = No) and (Fever = Yes) and (Depression = No) => Class=FMD\_disease (28.0/3.0) |
|  | (Temprature = High) and (Fever = Yes) (Decrease\_milk\_yield = no) and (poor\_body\_condition = Yes) and (Loss\_of\_Apptetite = Yes) => Class=Internal\_parasite (777.0/167.0) |
|  | (Depression = Yes) and (Nasal\_Discharge = Yes) and (Painful = Yes) => Class=Amphistomiasis (5.0/0.0) |
|  | (Temprature = low) and (Loss\_of\_Appetetite = Yes) and Swelling\_undder = Yes) => Class=Mastitics (775.0/155.0) |

# CHAPTER FIVE

# KNOWLEDGE ACQUISITION AND REPRESENTATION

## 5.1. Introduction

Knowledge acquisition is the process of acquiring relevant knowledge from domain experts and other sources of information such as books, Internet, databases, guidelines, manuals, journal articles, document. Knowledge acquisition is the process of eliciting, structuring and representing domain knowledge acquired from different sources [15]. Knowledge acquisition is the first step and time consuming task in the development of knowledge based system [17]. The researcher acquires knowledge using two types of knowledge acquisition methods which are documents analysis and interview.

## 5.2 Knowledge acquired using documents analysis

The researcher uses different documents which are Guideline for Diagnosis, Treatment and Prevention of cattle disease in Ethiopia which is prepared by Ethiopian Federal Ministry of Animal Health and an article Animal health safety which is prepared by Johan van Griensven and Ermias. The researcher also conducts interview with Veterinary Doctors that works in the research and treatment of Cattle disease in ILRI. The knowledge acquired from documents and interviews are similar but the researcher used to validate knowledge obtained from one source by another source. Accordingly, the researcher presents the knowledge that is acquired from documents and interview as follows:

## 5.3. Common Cattle Diseases

According to International Livestock Research Animal Health Center dataset common Cattle diseases are selected. Those are listed below: -

* Foot-mouse Disease (FMD)
* Lumpy Skin Disease (LSD)
* Blackleg
* Anthrax
* Internal Parasite
* Amphistomiasis
* Calf Diphtheria
* Enteritis
* Mastitis
* Pneumonia

### 5.3.1. Foot-and-Mouse disease (FMD)

Foot-and-Mouse disease (FMD) is a highly communicable viral infection of cattle, pigs, sheep, goats, buffalo, and artiodactyl wild life species characterized by fever, vesicle in the mouse, on the muzzle, gums, pharynx, teats, and interdigital cleft. It is caused by an Aphthovirus that is transmitted by contact and through milk. The disease spreads by direct contact or indirectly through infected water, manure, hay and pastures [3].

**Clinical symptoms**: -The clinical sign includes appetence, fever, shacking or rolling of the feet, reduced milk yield, high temperature, shivering followed by smacking the lips, and pregnant animals may abort [4].

**Diagnosis**: - Clinical signs are indicative and confirmed by Foot Mouse Disease (FMD) serology.

**Treatment and prevention**

* Drug treatment: - No specific treatment, however, supportive treatments against secondary bacterial infection are necessary.
* Control and prophylaxis: - vaccination, test and quarantine infected herds

### 3.2.2. Lumpy Skin disease (LSD)

Lumpy skin disease is an infectious, eruptive, occasionally fatal disease of cattle caused by a virus associated with the nettling poxvirus in the genus Capri poxvirus of the family Poxviridae [55]. Pneumonia is a common sequel in animals with lesions in the mouth and respiratory tract. It is a highly infectious viral disease of cattle and buffalo characterized by boxlike intracutaneous firm nodules, edema of the lips, superficial lymph nodes swelling, and lyphangitis. The disease is widely distributed in Ethiopia imposing severe economic loss due to damage of hide [4].

**Clinical symptoms**

Painful swelling, fever, nasal discharge, emaciation, and hypersalivation followed by characteristic eruptions on the skin and other parts of the body are characteristics; the nodules are circumscribed, round, slightly raised, firm, and painful and involve the entire Curtis and the mucosa of the gastrointestinal, respiratory and genital tract. The regional lymph nodes are swollen and edema develops in the udder, brisket, and legs. Secondary infection causes extensive suppuration and sloughing lesions [4].

Diagnosis: - The widespread nodular lesion of the skin and mucous membranes and biphasic fever are indicative [4].

**Treatment and prevention**

* Drug treatment: There is no effective treatment, but secondary bacterial infections are prevented by administration of broad-spectrum antibiotics.
* Prevention: - vaccination with sheep/ goat pox virus or LSD strain.

### 5.3.3. Blackleg

Blackleg is an acute, febrile disease of cattle and sheep caused by Clostridium chauvoei characterized by emphysematous swelling of the heavy muscle and severe toxemia. Blackleg is not transmitted directly from sick animals to healthy animals by mere contact. Lameness, loss of appetite, rapid breathing, depression and high fever are the signs observed in animals infected by this disease. The sick animal usually dies within 12 to 48 hours. In most cases the animal is found dead without being previously observed sick. Blackleg is mostly occurring in the northern part of Ethiopia [50]. Cattle within the age of six months to two years old are mainly affected by this disease, though at any age and condition is affected. Blackleg is common in Ethiopia during dry period of the year [4].

**Clinical symptoms**

Depression, anorexia, rumen stasis, high-fever (41-42 0 C) and painful. The marked lameness with muscle swelling of the upper part of the affected leg with capitation may follow. At necropsy affected tissues are filled with rancid serosanguineous fluid and gas pockets, which crepitate when squeezed and the muscle appear dry [4].

* Lameness.
* Loss of appetite.
* Rapid breathing.
* Fever.
* Unwillingness to move.

**Diagnosis**

The clinical signs and postmortem findings are indicative; the epidemiology and bacterial isolation are confirmatory [4].

**Treatment and Prevention**

**Non-drug treatment**

* Drainage and slashing of affected tissues to allow oxygen into the tissue, plus supportive treatment with parental fluids, analgesics, etc.

**Drug treatment**

* Procaine penicillin G, 22,000IU/kg, IM or SC q 24 h for 3 to 5 days or Benzathine penicillin or similar repository preparations, q 48-72 h.
* S/E, C/I and D/I: Hypersensitivity reactions to penicillin’s, simultaneous administration of chloramphenicol, tetracycline or phenylbutazone.
* D/F: 200,000 IU/ml to 400000 IU/ml
* W/P: meat 14 days and milk 3 days
* Local antibiotic treatment Aug. Oxyteracycline spray 5% of the site of the wound is helpful.

**Prevention**

* Vaccination with C. chauvoeibacterin.

### 5.3.4. Anthrax

In acute anthrax of cattle and sheep, there is an abrupt fever and a period of excitement followed by depression, stupor, respiratory or cardiac distress, staggering, convulsions, and death. Often, the course of disease is so rapid that illness is not observed and animals are found dead. Anthrax, which is caused by bacterium, occurs in areas where animals have previously died of anthrax. Even though anthrax has appropriate vaccination, in Ethiopia, still it occurs frequently [50]. One of the common signs in cattle with anthrax is a progression from a normal appearance to dead in a matter of hours. Weakness, fever, excitement followed by depression, difficulty in breathing, uncoordinated movements and convulsions are other signs of anthrax additionally, after death, the animal's body rapidly decomposes.

**Clinical symptoms**

The clinical sign in anthrax disease includes in ruminant species, acute illness is characterized by abrupt onset of fever, signs of abdominal pain, abortion, diarrhea, and milk production is decreasing. Chronic infection is rare in cattle and is manifested by localized edematous swelling on the ventral neck, Thorax and shoulders [4].

**Diagnosis**

Anthrax is diagnosed by examining blood (or other tissues) for the presence of the bacteria. Samples must be collected carefully to avoid contamination of the environment and to prevent human exposure to the bacteria [4].

**Treatment and prevention**

**Supportive therapy**

Hyperimmune serum plus antibiotics

* Penicillin 22,000 IU/kg, IM, q 12 h for 2 days, then daily for 3 days or Benzathine penicillin or other repository preparations, q 48-72 h; the initial dose should be administered IV. For C/I/, S/E, D/I, D/F, W/P, see page 34
* Oxyteracycline 6-11 mg/kg, IM, or IV, q 12-24 h. Initially, divided the daily dose into two doses.
* S/E, C/I, D/I: renal impairment, last 2-3 weeks of ingestion in pregnant animals and up to 4 weeks of age in neonates. Gastrointestinal symptoms are more severe with Oxyteracycline among the tetracycline’s; discoloration of the teeth when used during pregnancy and drug interactions with anti-acids, dairy products, calcium salts, iron salts, magnesium salts, zinc salts and warfarin.
* D/F: injection, 5, 10%
* W/P: meat 21 days, milk 7 days
* Amoxicillin 5-10 mg/kg q 24 h for 3-5 days; for C/I, S/E, D/F, D/I, and W/P are similar to penicillin

**Vaccination:**

Caution: animals that have died of anthrax should be burned in a closed incinerator. Animals should not be vaccinated within 2 months of anticipated slaughter; antibiotics should not be administered with one week of vaccination.

**5.3.5. Mastitis**

It is, a complex and costly disease of dairy cows, that results from the interaction of the cow and environment including milking machine and microorganism [51]. In Ethiopia, even though the disease of mastitis has been known locally, it has not been studied systematically, making information available on the prevalence of disease and associated economic loss inadequate [52]. Unfortunately, mastitis is not always easy to detect in its early stages, particularly when the redness and swelling of the udder is not obvious. If left untreated, severe clinical mastitis may cause the death of the cow [53].

Mastitis, an inflammation of the mammary gland, is almost always due to infection by bacteria of mycotic pathogens. Although over 135 microorganisms have been reported because the disease, Staphylococcus aureus, Staphylococcus agalactiae, Str. uberis, Str. dysgalactiae, other treptococci, Arcanobacterium pyogenes, Mycoplasma spp, Nocardia asteroides and Coliforms are the most agents in Mastitis disease [4].

**Clinical symptoms**

Clinical mastitis is manifested by inflammation of the udder and often accompanied by abnormal milk secretions. The signs depend on the organism involved. Systemic signs could also be observed. Subclinical mastitis is the most common [4].

The clinical sign is usually non-specific, but the following gives the clue for the type of agent involved.

Table 5. 1 Mastitis basic microorganism

|  |  |
| --- | --- |
| Microorganism | Clinical finding |
| Staphylococcus aureus | Sever swelling, purulent milk with clots |
| Mycoplasma species | Drop in milk production, infection of all quarters simultaneously |
| Arcanobacterium pyogenes | Profuse, foul-swelling, purulent discharge |
| Mycoplasma bovis | Rapid onset |

**Diagnosis**: - It is based on clinical signs, and identification of the pathogen. A test to detect subclinical mastitis include California Mastitis Test, or direct somatic cell count [4].

**Treatment and Prevention**

**Drug Treatment:**

Treatment depends on the type of infecting organism and the stage of mammary gland damage.

* Subclinical infections: Inframammary infusion, q 48 h for 3 times, applied separately into every quarter parachute or acute staphylococcal mastitis:
* Systematic treatment

Procaine penicillin G 22, 000 IU/kg, aqueous suspension, IM or SC q 24 h for 3 to 5 days or Benzathine penicillin G or a similar repository preparation, q 48-72 h. For S/E, C/I, D/F, D/I, W/P,

* Amoxicillin or Ampicillin 10 mg/kg q 24 h, IM
* W/ P: Meat 6 days; milk 96 hr.
* Long acting penicillin preparation before dying the cow C/I, S/E, D/I see page 34
* Oxyteracycline 10mg/kg, q 24 h, IV
* W/P: milk 96 hour; meat 14-28 days; For C/I, S/E, D/I, see page 34

**Inframammary infusion:**

* Benzathine Cloxacillin, 500mg for 3 days
* S/E: like Benzathine or procaine penicillin (see page 34)
* C/I: hypersensitivity and not recommended for food producing animals
* D/F: Intramammary suspension 500, 625 mg/dose

**Prevention**

* Disinfection of the teat before and after milking
* Dry cow therapy with long acting penicillin preparation should be applied before drying the cow.

### 5.3.6. Pneumonia

It is inflammation of the pulmonary parenchyma usually accompanied by inflammation of the bronchioles and often by pleurisy. It is the main respiratory problem of dairy cattle. The occurrence of this problem might be due to stress, workload and movement of animals during drought period that can favors the bacteria to multiply due to the immune status of the animals were suppressed.

Environmental risk factors include close confinement and poor ventilation of the house, high temperature that encourage replication of the diseases [35].

Important pathogens associated with pneumonia in cattle are Mannheimia haemolytica serotype A, Haemophilus sommus, P, multocida, Mycobacterium Bevis, Mycoplasma mycoides subspecies mycoides small colony type, Parainfluenza virus 3 bovine respiratory syncytial virus, bovine herpes virus 1 and bovine viral diarrhea [4].

**Clinical symptoms**

It is manifested clinically by an increase in respiratory rate, cough, salivation, impotence, abnormal breath sounds on auscultation and, in most bacterial pneumonias, by evidence of toxemia. Broncho-pneumonia is usually accompanied by a moist, painful cough; interstitial pneumonia is characterized by frequent, hacking coughs, often in paroxysms. The presence of nasal discharge depends on accompanying inflammation of the upper respiratory tract [4].

**Diagnosis**

The clinician has to decide whether there is pneumonia and if there is, then determine the nature and cause of the pneumonia [4].

**Treatment and prevention**

**Non-drug treatment**

* The treatment of pneumonia depends on the etiology
* Nursing: sick animals should be provided with shelter; animals would be given good quality long-stem pastures.

**Drug treatment**

* Procaine pencillin G, 22000 IU/kg; maintenance q 24 h , PO for 3 to 5 days or Benzathine pencillin G or similar respiratory preparation q 48-72 h. for S/E, C/I, D/I, D/F W/P, see page 34
* Ampicillins tetrahydrate 22 mg/kg IM, SC, q 24 h for 3-6 days. For S/E, C/I, D/I, D/F

W/P, see page 34

* Oxyteracycline hydrochloride 11 mg/kg, q 24 h or long acting formulation, 20 mg/kg, IM, q 48 h for 3-5 days. S/E, C/I, D/I, D/F W/P, see page 35
* Tylosin 44 mg/kg, IM, q 24 h for 3-5 days
* S/E: allergic reaction in all species and gastrointestinal disturbance
* C/I: animals with impaired liver function
* D/F: powder 10, 20, and 30%; injection, 50, 200, 150 and 220 mg/ml and tablet 200mg
* D/I: theophylline, warfarin and beta-adrenergic drugs
* W/P: adult, meat 7 days, milk 4 days; calves, meat 14 days

**5.3.7. Internal Parasite (Gastrointestinal)**

Internal parasites are infectious of the gastrointestinal tract with nematodes, cestodes and trematodes. The common stomach worms of cattle are Haemonchus placei, Ostertagia ostertagi and Trichostronglyus axeii. Ostertagia and Trichostronglyus infections are characterized by profuse, watery diarrhea that usually is persist. Signs of anemia, hyperproteinemia and edema, particularly the lower jaw and sometimes along the ventral abdomen manifests these infections together with haemonchus infections [4].

**Clinical Symptoms**: Ostertagia and Trichostronglyus infections are characterized by profuse, watery diarrhea that usually is persist. Signs of anemia, hyperproteinemia and edema, particularly the lower jaw and sometimes along the ventral abdomen manifests these infections together with haemonchus infections [4].

Diagnosis: animals with poor body condition have anemia and diarrhea is suggestive; confirmed by fecal examination. Clinical signs, grazing history and season may give a presumptive diagnosis of internal parasite infection in cattle. The diagnosis confirmed by finding worm eggs on a fecal exam [4].

**Treatment:**

Reduce exposure to parasite burdens by moving calves onto paddocks that have not been grazed for a long period. Avoid too much grazing pressure higher pasture covers reduce the number of larvae consumed as most larvae live close to the ground. Calves which are healthy and well fed will be less susceptible to gastrointestinal parasites than weaker calves. Grazing paddocks with adult cattle or sheep after calves will reduce the larvae load on the pastures.

Calves need to be drenched before worm burdens get too high. Can be administered as a pour on, oral or injectable. Oral combination drenches tend to be the most effective drench in young calves.

Talk to your vet re avoiding drench resistance - factors include using the correct dose, returning treated stock to contaminated pastures, not treating healthier animals, only drenching when necessary based on faecal eggo counts (FEC) and symptoms.

* Haemonchus placei: Tetramisole 15 mg/kg
* S/E: frothing, salivation, tremore, transient head shaking, licks of lip, urination, defecation, vomiting, ataxia, collapse and death due to respiratory failure
* C/I: within 14 days of treatment of organophosphorus compound or diethyl carbamate. Don’t exceed dose 4.5 gm per animal
* D/F: Bolus, 150,600,700,1000,1200,1500, and 2000 mg; powder or granule, 10,20, and 30% ; injection, 30 and 100 mg/ml
* Ostertagia ostertagi: Albendazol 7.5 mg/kg
* Trichostronglyus axeii: Ivermectin 200 mg/kg
* S/E: Ataxia, depression, tremors, mydriasis, listlessness, msculoskeletal pains, oedema of the faceor extremeties itching and popular rash
* C/I: calves less than 12 weeks of age and lactating animals
* D/F: Bolus, 1.75 gm and 5 gm; Suspension, 800mg/ml
* W/P: meat 28 days and don’t use in lactating animals
* Control: To be eﬀective, a parasite control program must reduce the numbers of worms in all classes of cattle and control the number of worms on the pasture. Control of internal parasites in cattle must kill all stages in the animal and help control the number of larvae and eggs on pastures. Adult cattle usually have more resistance to internal parasites than younger cattle, but deworming older cattle can help reduce pasture contamination.

### 5.3.8. Enteritis /Diarrhea

It is swelling or inflammation of the small intestines and can present as a variety of symptoms in the bovine. Enteritis also called diarrhea is a common in new born calves characterized by progressive dehydration and death, some times in a few as 12 hr. it is caused by multitude of bacterial, viral and non-infectious agents [4].

Clinical Finding:

**Subacute form**: diarrhea persists for several days and result in malnutrition and emaciation. It is common in diary calves [4].

Acute form: The major signs are diarrhea, dehydration, profound weakness, and death within one to several days of onset. The sign depends on the etiology [4].

**Diagnosis**: Slight to moderate distension of right abdomen; ﬂuid-rushing and splashing sounds on auscultation and ballottement Diarrhea and dehydration.

**Treatment and prevention:**

**Drug Treatment:**

* Sulfadimidine, initial dose: 140mg/kg, IV; maintenance dose: 70mg/kg, IV q 24 h for5-7 days.
* S/E: crystallization in urinary tract, cutaneous eruption, hypothyroidism and idiosyncratic toxicities.
* C/I: pregnant and lactating animals
* D/I/: thiopentone sodium and warfarin
* D/F: bolus, 5g; injection 330,333 and 160mg/ml; powder, 8,10,16,20,25, and 30%
* WP: meat 21 days; it should not be used in lactating cows
* Procaine pencillin, G, 22,000IU/kg, IM or SC q 24 h for 3-5 days
* Benzathine pencillin or similar respiratory preparations, q 28-72 h. S/E, C/I, D/I, D/F W/P,

### 5.3.9. Calf Diphtheria

Affects throat, ulceration and swelling of the affected parts.

Caused by Bacterium-Fuso-bacterium nicrophurum.

Symptoms: Temperature rises to 40-41 °C, painful coughing and difficulty in breathing, dribbling of saliva, and inability to swallow feed.

Diagnosis:

* The diagnosis of calf diphtheria is usually based on the clinical signs.
* For one-off cases rule out other problems such as BVD and foreign bodies by getting your vet to do a thorough oral
* examination
* Bacteriology can be also useful.
* A post-mortem can confirm the ulcerative nature of the disease, particularly in calves with the laryngeal for

Treatment:

* Early prompt treatment is important as early treatment is much more effective
* Separate the infected animals and isolate them
* Antibiotics and pain killers are effective in most cases
* The laryngeal form is much more resistant to treatment. Get veterinary advic

Prevention:

Fusobacterium necrophorum is a normal inhabitant of cattle intestines and the environment. Under unhygienic conditions, infection may be spread on feeding troughs and dirty milk buckets. Some of the contributory factors for occurrence of this disease include abrasions in the oral mucosa (such as those from erupting molar teeth), poor nutrition and the presence of other diseases present in young calves. If animals are closely confined, the spread of this infectious disease can be prevented by thoroughly cleaning and disinfecting of all calf feeders. Young calves must be examined daily to identify early stages of the disease.

### 5.3.10. Amphistomiasis

Amphistomiasis is a vector-borne, infectious blood disease in cattle caused by the rickesttsial parasites Anaplasma marginale and Anaplasma centrale. It is also known as yellow-bag or yellow-fever.

Symptoms

* Anemia
* Fever
* Weight loss
* Breathlessness
* Jaundice
* Uncoordinated movements
* Abortion
* Death

Treatment

Tetracycline is often used for clinical amphistomiasis. However it cannot be used in every country.  
  
General supportive care is also important for anemic animals. Blood transfusions are of limited benefit.

The incubation time for the disease to develop varies from two weeks to over three months, but averages three to four weeks. Adult cattle are more susceptible to infection than calves.  
  
The disease is generally mild in calves under a year of age, rarely fatal in cattle up to two years of age, sometimes fatal in animals up to three years of age, and often fatal in older cattle.  
  
Once an animal recovers from infection, either naturally or with normal therapy, it will usually remain a carrier of the disease for life. Carriers show no sign of the disease but act as sources of infection for other susceptible cattle.

Prevention

Typically, cases of amphistomiasis increase in late summer and fall as insect vectors increase. Therefore, control of vectors is key to preventing amphistomiasis. If necessary, herd treatment with oxytetracycline injection every 3 to 4 weeks during high risk times may be necessary will prevent clinical disease but animals can become carriers. Chlortetracycline also known as CTC can reduce the risk of amphistomiasis. A consistent intake of the correct amount of mineral is crucial to a amphistomiasis is prevention programmer. CTC is available in medicated feed, free choice salt-mineral mixes or medicated blocks. In some places, vaccines are available to increase resistance to amphistomias.

# 5.4 Data collection Using Interview

In this research study for knowledge acquisition process interviewing the domain experts through purposive sampling technique was used. This because, purposive sampling technique helps to acquire relevant knowledge from more experienced and educated experts rather than less experienced and educated experts. Totally five (5) veterinary experts were participated in the interview process for this research study, three from domain experts and two from veterinary experts from ILAHC. The selection criteria of domain experts for the study are based on the professions/expertise, educational qualification level, year of experience and their immediate position in the cattle disease diagnosis and treatment.

Table 5. 2 Sample selected for interview

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | Education level | Experience in year | Position | Sex |
| 1 | DVM | 11 | Expert of Clinical Veterinary Medicine | M |
| 2 | DVM(Specialized In surgeon) | 9 | Expert livestock resource Development | M |
| 3 | DVM | 15 | CVM and veterinary medicine Teacher | F |
| 4 | Professor | 18 | Research and Technology Transfer | M |
| 5 | DVM | 7 | Animal health assistant | F |

As the domain experts explained during discussion, there are two types of diagnosis that the veterinarian uses to treat the cattle. These are tentative diagnosis and differential diagnosis.

1. A tentative diagnosis is a preliminary suspicion of patient status, which is usually made by physicians according to patient narrative right at admission. It largely depends on the experiences and professional knowledge of physicians.

Tentative diagnosis is used to treat cattle which are infected with the disease. Some diseases show specific clinical signs when the animal is infected. For example, Lumpy Skin Disease, which is caused by a virus can be diagnosed and treated quickly without comparing symptoms. This is because, the disease has easily identified symptom. These kinds of diseases can be treated by asking the case history from animal owners.

1. Differential diagnosis is a process wherein a doctor differentiates between two or more conditions that could be behind a person's symptoms. When making a diagnosis, a doctor may have a single theory as to the cause of a person's symptoms. They may then order tests to confirm their suspected diagnosis

Differential diagnosis: This is used to diagnose certain diseases which have two or more similar clinical signs. In this case, in order to identify a disease, the veterinarian should diagnose the difference between the infected animal by using clinical symptoms.

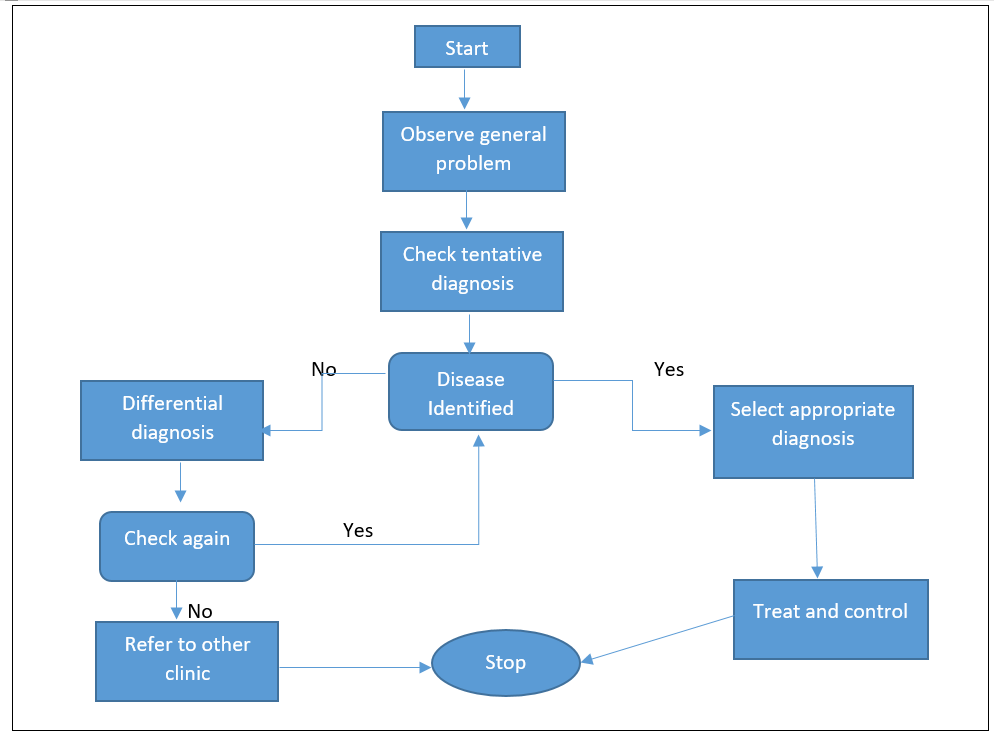


Figure 5. 1 Diagnostic Procedure for cattle Disease

# 5.5 Knowledge Representation

Since the knowledge that the researcher acquired from the data mining classification technique is in the form of rules the knowledge that the researcher acquires from document analysis and domain experts interview about diagnosis and treatment of cattle disease. Procedures, which are easy to convert to rules, the researcher forced to use rule based knowledge representation method that is the most predominant knowledge representation methods to develop the Knowledge base. The inference engine evaluates the portion of a statement and concludes whether a goal is satisfied or not. If the goal is not satisfied, the inference engine proceeds to the next rule until the conditions are satisfied [38]. The following rules are which are incorporated in the knowledge base.

**Rule1**: (Temperature = High) and (Painful = No) and (Loss\_of\_Apptetite = No) and (Swelling\_Udder = No) and (Crepitation\_Sound = No) and (Fever = Yes) and (Depression = No) => Class=FMD\_disease (28.0/3.0)

**Rule2:** (Decrease\_milk\_yield = yes) and (Temperature = high) and (Swee\_of\_neck = no) and (Loss\_of\_apptetite = no) => Class=LSD (310.0/0.0)

**Rule3:** (Loss\_of\_Apptetite = Yes) and (Fever = Yes) and (Nasal Discharge = No) and (Swee\_of\_neck = no) and (Decrease\_milk\_yield = no) and (Temperature = high) => Class=Blackleg (310.0/0.0)

**Rule4**: (Temperature = medium) and (Smaking\_lips = yes) and (Decrease\_milk\_yield = no) and (Nodular\_lesion = yes) and (Coughing = yes) => Class=Anthrax (310.0/0.0)

**Rule5:** (Temperature = low) and (Loss\_of\_Appetetite = Yes) and Swelling\_undder = Yes) => Class=Mastitics (775.0/155.0)

**Rule6:** (Temperature = Normal) and (Fever = Yes) and (Decrease\_Milk\_yield = No) and (Nodular\_Lesion = No) and (Painful = No) and (Loss\_of\_Apptetite = Yes) and (Lameness = No) and (Depression = No) => Class=Pneumonia (7.0/1.0)

**Rule7:** (Temperature = High) and (Fever = Yes) (Decrease\_milk\_yield = no) and (poor\_body\_condition = Yes) and (Loss\_of\_Apptetite = Yes) => Class=Internal\_parasite (777.0/167.0)

**Rule8:** (Fever = No) and (poor\_Body\_condition = No) and (Nodular\_Lesion = No) and (Temperature = Normal) and (Rough\_hair = No) and (Nasal\_Discharge = No) and (Loss\_of\_Apptetite = No) => Class=Enteritis (56.0/9.0)

**Rule9**: (Temperature = High) and (Painful = No) and (Loss\_of\_Apptetite = Yes) and (Swelling\_Udder = No) and (Crepitation\_Sound = No) and (Fever = Yes) and (Depression = No) => Class=Calf\_Diphtheria (28.0/3.0)

**Rule10** (Depression = Yes) and (Nasal\_Discharge = Yes) and (Painful = Yes) => Class=Amphistomiasis (Loss\_of\_apptetite = no) => Class=Amphistomiasis (929.0/297.0)

# CHAPTER SIX

# IMPLEMENTATION AND DISCUSSION

## Introduction

Findings of data mining are the base for the development of knowledge base system through the Integrator application. Facts we get from Data mining can be represented as a rule in the knowledge base. After knowledge acquisition is done using a JRip rule induction algorithm, which performs on data collected from ILR the facts extracted is represented in the knowledge base system. The main challenge here is how to use his knowledge extracted from data mining for knowledge based system.

## Architecture of the System

Classification algorithms are widely used in various medical applications. Data classification is a two phase process in which first step is the training phase where the classifier algorithm builds a classifier with the training set of tuples and the second phase is classification phase where the model is used for classification and its performance is analyzed with the testing set of tuples [62]. This system was designed with the progression of conceptual design that refined the systems’ architecture. Of course, the conceptual design was essential to stabilize the architecture of data mining result with knowledge based system for cattle diseases diagnosis and treatment.

Throughout design iterations, the design of data mining result with knowledge based system for cattle diseases, diagnosis and treatment was expended into system architecture to ensure that it supported the cattle disease and treatments. Figure 5.1 shows the overall system design and framework of integrating data mining induced hidden knowledge about cattle disease diagnosis and treatment based on International Livestock Research Center dataset with knowledge based system.



ILRI Database

Domain Expert

Knowledge acquisition

Data mining

Rule

Knowledge

Knowledge

Integrator

Knowledge representation

Knowledgebase system (inference engine, editor module, explanation facility, user interface,

Knowledgebase



Users

Figure 6. 1 System Architecture of the Knowledgebase system

**Data Mining**: - Data preprocessing subtask is applied for effective data selection to enhance the performance of the cattle disease diagnosis and treatment system. It contains the feature extraction component to extract the required features from ILRC dataset for the process of disease diagnosis and treatment. During the preprocessing stage of instances are classified as normal and diseased cattle, and also diseased cattle are further classified into Ten sections represented into FMD\_Disease, Anthrax, LSD, Pneumonia, Enteritis, Internal parasite, Blackleg, Mastitics, Amphistomiasis and Calf\_Diphtheria

**Integrator**: - The JRip model is integrated with knowledge based system automatically for designing intelligent diagnosis and treatment of cattle disease system.

**Rule Library**: - Is a container of rules about cattle disease which are generated by a JRip rule induction algorithm after mapped by the Integrator to Prolog understandable format.

**Knowledge based system**: - Let us discuss components of the knowledge Based System for Diagnosis and Treatment of cattle disease as follows: Common component of the system consists of the following subsystems: rule library, Inference engine, knowledge base and user interface.

**Knowledge Base**: - The researcher stores all knowledge that is collected from domain experts, document analysis and data mining in the knowledge based as set of rules using a rule-based knowledge representation method.

**Inference Engine**: - An inference engine is the brain of the Knowledge Based System, which directs the system how it can derive a conclusion by looking for possible solutions from the knowledge base and recommend the best possible solution. Since the objective of the proposed Knowledge Based System is for diagnosis and treatment of cattle disease and the Prolog’s built-in inference mechanism is backward chaining, the researcher prefers to use a backward inference mechanism which is a goal derived that tries to prove or disprove the goal.

For instance, if the user wants to write a prescription for the diseased cattle after he/she has diagnosis the diseased cattle, he/she should answer the questions that are asked by the system. To do so, the system needs facts and rules about the drugs which are appropriate for the specific diseased cattle.

**User Interface**: - KBS cattle disease diagnosis and treatment used a simple user interface to display the information. When KBS cattle disease diagnosis and treatment identified a possible type of cattle disease, the system would present treatments and prevention of the disease on the screen and the users can ask detail or more explanation if they did not satisfy with the provided suggestions in such away the users could observe it straight away. For example, when the type of disease was identified, the typical disease descriptions, treatments and prevention would appear on the computer screen and asked whether they need more explanation or not. The user could view different type of disease and treatments for the cattle disease. The first page of the user interface welcomes user’s home page as shown in the figure below figure 6.2:

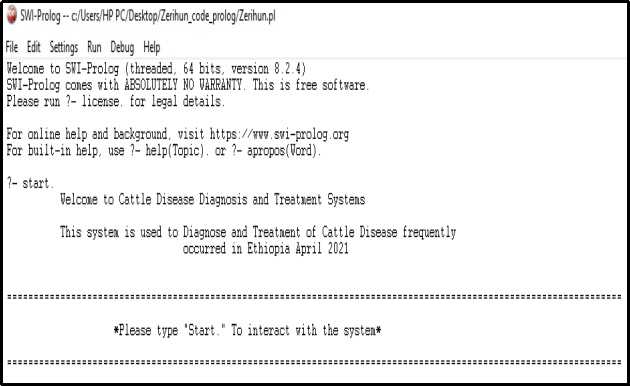
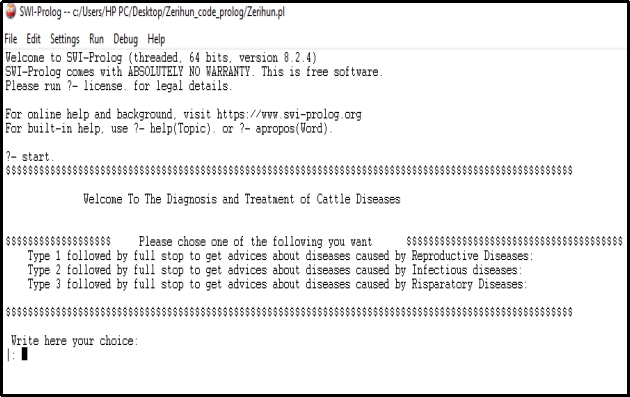


Figure 6. 2 Welcoming windows of KBS cattle Disease Diagnosis user interface

After the welcoming window of Cattle disease diagnosis and treatment system user interface is displayed, the user interface allows the users to interact with the system by starting the system as indicated in Figure 6.3 below. The user can start the system by typing “start” and putting full stop (.) at the end.



**Figure 6. 3 Sample windows between the user and the system to identify cattle disease**

After the disease is identified and described well its recommended treatments are provided as it is presented in the following sample window.

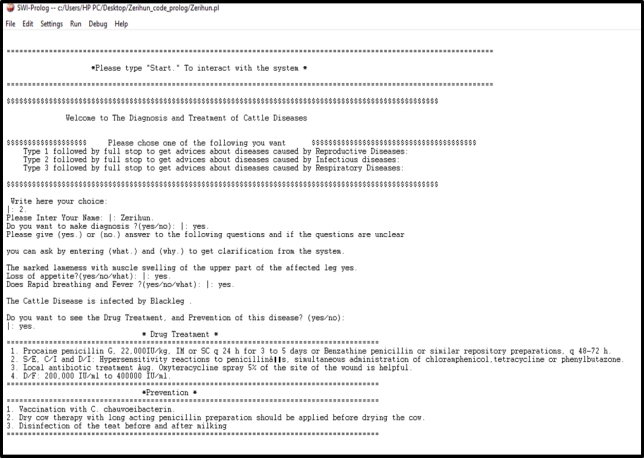


Figure 6. 4 Sample window of the system’s treat and prevent for the identified disease

## 6.3 Combination of Data Mining and Knowledge Based System

Accordingly, for diagnosis and treatment of cattle disease the model constructed by JRip rule based is used for cattle disease diagnosis. SWI prolog programming was used to integrate diagnosis and treatment of the cattle disease model created using a JRip rule with the Knowledge Based System automatically. For this study, we used Weka swi-prolog tools to construct a predictive model.

## Structure of JRip Rule and PROLOG Rule

JRip algorithm generates rules in the form of (condition)… (Conclusion) format. The algorithm generates 23 rules from the ILRC dataset. The format of some of the rules is indicated in the table 6.1. The condition part contains attribute, a comparison operator and value. Two or more conditions are joined by = and’. After the conditions the ‘=>’ meaning implies follows. The concluding part of the rule has the format class=’disease’, for example class= LSD, class= Pneumonia.

## Evaluation of the Prototype

System evaluation is the basic issue for design science research which is intended with artifact development. The developed knowledge base system, cattle disease diagnosis and treatment is tested and evaluated to check whether the objectives of the research are achieved or not. The other one is preparing rules and feed the proposed system with these rules and give the same rule for domain experts and compare the results of the proposed system and the domain experts such that we can make sure that the proposed system could replace the domain experts in his or her absence. In addition to this, evaluation can be done by conducting a user acceptance testing which will help the researcher to make sure whether the proposed system is user friendly or not.

### 6.5.1 User Acceptance Testing

In this way, four users were selected and had been given the chance to use and interact with the system. The user selected were two Veterinarian Medicine Doctors (VMD) and three animal health assistant professional from ILRC.

To evaluate the Cattle disease diagnosing system in this study, eight (8) Veterinarian Medicine Doctors (VMD) and three (3) animal health assistant professional from ILRC. Before starting the evaluation, the researcher explained the objective of the developed system and how the system interacts with the users. This explanation helps the evaluator get full understanding how they consult the system in getting advice.

Then after, the domain experts were allowed to interact with the system by running number of cases having similar parameter with the facts incorporated in the knowledge base. After the consultation of the system, to assess the user acceptance of the prototype knowledge based system, questionnaires were distributed. Using these questionnaires, domain experts‟ feedback about the developed system was gathered for further analysis.

The type of questionnaires distributed for feedback collection from the evaluators were closed ended and open ended questionnaires focusing on easiness, time efficiency, accuracy of Cattle disease diagnosis and treatment. The questionnaires also focused on the applicability of the system in diagnosing Cattle disease, problem solving ability and the significance of the system in the domain area.

The questions are divided into two parts; the first seven questions are close ended questions which helps system evaluators to check on the user interface design aspects, easiness of the system to use, attractiveness, correctness of the decision, adequacy of knowledge content, the problem solving ability and significance of hybrid system in Cattle disease diagnosis and treatment. On the other hand, the remaining eight questions are open ended questions which used to collect constructive feedback from the system evaluator’s based on the system’s conclusion.

The evaluators were allowed to rate the options as excellent, very good, good, fair, and poor for these closed ended questions. Therefore, for easiness of analyzing the relative performance of the prototype based on the user evaluation after the interaction with the system, the researcher assigned numeric value for each of the options given in words.

The values are given as Excellent = 5, Very good = 4, Good = 3, Fair = 2, and Poor = 1. Based on the given scale, system evaluators provide a value for each closed ended questions. Thus, this method helps the researcher to manually examine the user acceptance based on evaluator’s response.

The user acceptance of the system is measured manually as follows: The Table below indicates the feedbacks obtained from the domain experts (evaluators) on systems interaction as calculated based on the given scales.

C:\Users\HPPC~1\AppData\Local\Temp\ksohtml\wps348D.tmp.jpg

Where, AVS average score,

SV scale value,

TNR total number of respondent and nr is number of respondent.

To get the result of user acceptance an average performance is calculated out of 100%.

C:\Users\HPPC~1\AppData\Local\Temp\ksohtml\wps348E.tmp.jpg

Where NS is the number of scale and AVP is average performance. The following Table summarizes the results obtained from the respondents.

Table 6. 1 Cattle disease diagnosis and treatment Evaluation performance result

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **No** | **Evaluation Questions** | **Poor(1)** | **Fair(2)** | **Good(3)** | **Very good(4)** | **Excellent (5)** | **Average score** | **Average performance (%)** |
| 1 | Simplicity to use and interact with the prototype system | 0 | 0 | 2 | 4 | 4 | 4.2 | 84 |
| 2 | The attractiveness the prototype system | 0 | 0 | 1 | 4 | 5 | 4.4 | 88 |
| 3 | The system is more efficient in time? | 0 | 1 | 0 | 4 | 5 | 4.3 | 86 |
| 4 | How accurately does a system reach a decision in diagnosing Cattle disease | 0 | 0 | 3 | 4 | 3 | 4.1 | 82 |
| 5 | Does the system incorporate Sufficient and practical knowledge? | 0 | 0 | 2 | 5 | 3 | 4.1 | 82 |
| 6 | The ability of the prototype system in making right conclusions and recommendations | 0 | 0 | 2 | 4 | 4 | 4.2 | 84 |
| 7 | How do you rate the contribution of the system in the domain area? | 0 | 0 | 1 | 4 | 5 | 4.4 | 88 |
|  | Total average | | | | | | 4.24 | 84.85 |

Note AVG = average, Perf. = performance

As indicated in Table 5.1 above, 20% of the respondents rated ‘easiness of the prototype’ as good, for the same questions 40% of respondent respond as very good, and the remaining 40% responded excellent. In the same way, for question ‘attractiveness of the prototype’ 10% of the respondents rated as good, 40% of them as very good and the rest 50% of them respond as excellent. Similarly, for question ‘efficient in time’ 10% of the respondents rated the criterion as fair, 40% respondent evaluated as very good and the remaining 50% of them respond as excellent. At the same time for criteria the accuracy of the prototype to make correct decision’30% respondent rated as good, 40% of the respondent respond as very good and the remaining 30% evaluated as excellent. Likewise, for the criteria does the prototype incorporate adequate knowledge, 20% of the respondents rated it as good, 50% as very good and the rest 30% as excellent. Again 20% of the respondent’s rate as good, 40% respondents as very good and the reaming 40% respond as excellent for the criteria of ‘the ability of the system in making right conclusions and right recommendations’ and the rest. Finally, for the question related to ‘significance of the knowledge base system in the domain area’ 10% of the respondent rated as good, 40% of them evaluated as very good and the rest 50% of them respond as excellent.

To summarize table 5.1 above based on the responses of ten system evaluator, the average performance obtained is 4.24 on a scale of 5. This value is the result obtained from the values assigned for each close ended question. The result indicates that about 84.85% of users are satisfied by the performance of the knowledge based system. It means that the proposed knowledge based system gain about 84.85% of user acceptance.

## 6.6. Discussion

At the beginning, this study has four research questions to answer and let us discuss how these Questions have been answered by this study. The first research question of this study was ‘What are the main attributes that can predict the type of cattle disease?’ To answer this question,

information gain method and the domain expert’s interview were used and this study finds out that all instances of the rule generated discloses that the strong and significant attribute for predicting the cattle disease performance is a swell of neck.

The second question was “Which classification algorithm is best to develop the prediction model for cattle disease diagnosis? To answer this question, two experiments for three classification algorithms namely Naïve Bayes, J48 pruned, and JRip under 10-fold Cross-Validation test option/mode and percentage split were conducted and the experiments showed that JRip classification algorithm is the best classification algorithm to develop the prediction model that can predict the type of diagnosis that should be given to the diseased cattle is infected by disease because it registers better performance with 97.48% evaluation result and the researcher decided to use the results for further use in the development of knowledge base of KBS.

The third question was “How develop a knowledge-based system using data mining results?

To answer this question, a prototype Knowledge Based were developed using the knowledge that is acquired from domain Experts’ Interview and documents analysis with Prolog for treatment purpose and constructing disease diagnosis model, data mining technology is used to acquire hidden knowledge about the behavior of different disease from the given dataset. WEKA was integrated with SWI prolog programming was used to integrate diagnosis and treatment of the cattle disease model created using a JRip rule with the Knowledge Based System automatically.

The final question was “How evaluate the performance of knowledge based system?” To answer this question, to evaluate the Cattle disease diagnosing system in this study, eight (8) Veterinarian Medicine Doctors (VMD) and three (3) animal health assistant professional from ILRC. Then 70 test cases were prepared to evaluate the performance of the proposed system. Finally, system performance evaluation, testing and user acceptance testing were conducted. User acceptance testing is performed based on seven criteria of evaluation. Selected domain experts are trained and used the system to evaluate how much the KBS meets their requirements. The system on average scored 84.85% based on user acceptance evaluation. Here table 6.6 illustrates the researcher tries to summarize the related works in the following table for clear review:

Table 6. 2 Comparison of different researchers in this study

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Author | Title | Techniques | Evaluation Criteria’s | |
|  |  |  | User acceptance Testing | System performance Testing |
| Derejaw Lake | Developed on Web-based expert system for diagnosing cattle diseases | Interviews, document analysis | - | 87.2% |
| Abdulkerim | Integrating Data Mining with Knowledge Based System | Interview & Document analysis | - | 80.5% |
| Birhane Bekele | Developed a prototype KBS for diagnosis and treatment of diseases of sheep and goats | Interview & Document analysis | 85.8% | 90.22% |
| Tadesse Beyene | Integrating Data Mining Results with Knowledge Based System for Diagnosis and Treatment of Cattle Diseases: The Case of Debre Birhan Basso Animal Health Center | Interview, Document Analysis, Naïve Bayes ,J48 & JRip | 85.8% | 98.7% |
| Tesfamariam Mulugeta | Integrating Data Mining Results with the Knowledge Based System for Diagnosis and Treatment of Visceral Leishmaniosis | Interview, Document Analysis, PART,J48 & JRip | 86% | 95% |
| Zerihun Fantahun | Integrating Data Mining Results with Knowledge Based System for Cattle Disease Diagnosis and Treatment | Interview, Document Analysis, J48,JRip & Naïve Bayes | 84.85% | 97.48% |

# CHAPTER SEVEN

# CONCLUSION AND RECOMMENDATION

## 7.1. Conclusion

Ethiopia is one among the nations that possesses the largest livestock population in the African continent with an estimated 56 Million of cattle, 58 Million of sheep and goats and 10 Million of equines, 1 Million of camels and 57 Million of chicken [47]. Ethiopia has great potential for increasing livestock production, both for local use and export. However, development has been constrained by numerous reasons. Cattle disease is the main constraint.

In this study, the possibility of integrating data mining result with knowledge based system is realized and explored. The integration process begun by taking samples of ILRC dataset. The dataset is preprocessed and made suitable for mining steps. Due to several limitations in acquiring knowledge for knowledge base from domain experts in the area of diagnosis and treatment of cattle disease, integrated (manual and automated) knowledge acquisition techniques were used to acquire knowledge Data mining has proven to induce hidden knowledge from large collections of datasets. Hence, data mining classifier, JRip is employed for knowledge acquisition step since it has performed best among the selected classifiers with an accuracy of 97.68%.

To identify the best prediction model for diagnosis and treatment of cattle disease, 6 experiments for three classification algorithms, namely J48 pruned, Naïve Bayes and JRip under ten-fold Cross- Validation test option and percentage split test option were conducted. Finally, by conducting objective and subjective interestingness measure, the researcher decided to use rules that are generated by JRip classification algorithm model for further use in the development of knowledge base system because it registered better performance than J48 and Naïve Bayes with 97.68%, 96.65% and 95.42% evaluation result in 10-fold cross validation respectively.

The prototype Knowledge Based System, which provides advice for animal health workers about diagnosis and treatment of cattle disease was developed using SWI-Prolog 7.7.13 with NetBeans 8.2. The proposed Knowledge Based System has Knowledge Base, Inference Engines, and Explanation Facility and User Interface. Then 70 test cases were prepared to evaluate the performance of the proposed system. Finally, system performance evaluation, testing and user acceptance testing were conducted. User acceptance testing is performed based on seven criteria of evaluation. Selected domain experts are trained and used the system to evaluate how much the KBS meets their requirements. The system on average scored 84.85% based on user acceptance evaluation. However, further exploration and study has to be done to refine and yield a better Knowledge based system which can be deployed in real animal agriculture and to provide advice veterinaries so that they can take timely and appropriate actions for a certain cattle disease diagnosis and treatment.

Moreover, this study has covered the way for local researchers on using automatic knowledge acquisition techniques for the development of knowledge based system and motivates them to apply this approach than the conventional knowledge acquisition approach.

## 7.2. Recommendation

In this study promising result is achieved in integrating data mining with knowledge based system for diagnosis and treatment of cattle disease and providing advice to animal health workers. Some challenges have been encountered which hinder the system of scoring a better achievement.

The first one is in the course of integration, two interfaces, namely; graphical user Interfaces (for the Integrator) and command line interface for diagnosis and treatment of cattle disease KBS has been used. A challenge has encountered in bringing the Integrator and diagnosis and treatment KBS together under one interface. This is reflected on user acceptance test in which evaluator rated the simplicity to use and interact with the system below very good.

The other challenge encountered is about using knowledge which the KBS has already used previously before re-running the Integrator application following a change in the numbers of the dataset. In addition, the designed prototype KBS supports ten types of disease, namely FMD, Blackleg, Enteritis, LSD, Pneumonia, Internal Parasite, Anthrax, Mastitis, Amphistomiasis and Calf\_Diphtheria

Hence the researcher believes further researches have to be done to boost the benefits of Integration of data mining with the knowledge based system and the following are recommended for future study:

* Building hybrid knowledge based system which is capable of employing rule based reasoning and case based reasoning with integrated data mining techniques.
* Further combinations such as artificial intelligence, soft computing and other clustering algorithms can be used to improve the diagnosis accuracy.
* It is known that most of Ethiopian agricultural areas have no facilities like electricity and a computer access. So, further research must be conducted to take the integrated knowledge based system in to mobile application.

# References

[1] B. Akalu, "Review on Epidemiology of Bovine Tuberculosis in Ethiopia," Academic Journal of Animal Diseases, vol. 6, no. 3, pp. 56-66, 2017.

[2]"Physiology of Cattle," [Online]. Available: https://www.farmsanctuary.org/wp-content/uploads/2012/06/Animal-Care-Cattle.pdf. [Accessed 28 02 2021].

[3] S. Dr. Ashok "Common Cattle Diseases: Symptoms, Treatment and Prevention," vol. 453, no. 446, pp. 77-85, 2014.

[4] The Drug Administration and Control Authority of Ethiopia, Standard Veterinary Treatment Guidelines for veterinary clinics, 2006.

[5] A. Rajendra, P. Akerkar and S. Srinivas, "Knowledge-Based systems," Jones and Bartlett, Sudbury: MA 01776, 2010.

[6] p. Sajja and R. Akerkar, "Advanced Knowledge Based Systems," Model, Applications and Research, vol. 1, p. 1 – 11, 2010.

[7] A. Mohammed, "Towards Integrating Data Mining with Knowledge Based System: The Case of Network Intrusion detection," M.Sc. Thesis, Adis ababa University, 2013.

[8] M. Tefera, "Global Crisis and the Challenge of Veterinary Teaching in Ethiopia," Global Veternarian 5, vol. 6, pp. 294-30, 2010.

[9] L. Derejaw, "Web-based expert system for diagnosing cattle diseases,” M.Sc. Thesis in Computer Science, Addis Ababa University, Addis Ababa, Ethiopia, 2013.

[10] NetBeans, 06/04/2015. [Online]. Available: https://netbeans.org/features/.

[11] [Online]. Available: https://en.wikipedia.org/wiki/Cattle. [Accessed 15 01 2021].

[12] K. Mehmed and B. John, Data mining concepts, Models, Methods, and Algorithms, USA: Wiley -IEEE Press, 2003.

[14]. R. Osmar, Principles of Knowledge Discovery in Databases CMPUT690, 1999.

[15]. D. Olson, "Advanced data mining techniques", Springer, 2008.

[23]. U. Fayyad, "Knowledge Discovery and Data Mining: Towards a Unifying Framework", Gregory , 1996.

[25]. U. Fayyad, "From data mining to knowledge discovery in databases", 1996.

[24]. J. Seifert, "Data mining: An overview of National security issues," 2004, pp. 201-217.

[22]. K. Mehmed and B. John, Data mining concepts, Models, Methodes, and Algorithms, USA: Wiley -IEEE Press, 2003.

[27]. R. Raman and K. Prasad, "Applications of Knowledge Based Systems in Mining," APCOM 87. Proceedings of the Twentieth International Symposium on the Application of Computers and Mathematics in the Mineral Industries, Vols. 167- 180, pp. 1-11, 1987.

[26]. S. Institute, SAS Enterprise Minor SEMMA, 2016.

[28]. J. Jackson, "Data Mining: A Conceptual Overview," 2002.

[21]. W. Daimler, "CRISP-DM: Towards a Standard Process Model for Data Mining", 2001.

[33]. A. Azevedo, "KDD, SEMMA and CRISP-DM: a parallel overview," IADIS European conference on data mining, 2008, pp. 128-185, ISBN: 978-972-8924-63-8.

[32]. K. Cios, "The Knowledge Discovery Process In Data Mining," Springer US, 2007, pp. 9-24.

[31]. A. Meseret, "A Combined Reasoning System For Knowledge Based Network Intrusion Detection," Master thesis, Addis Ababa University, Addis Ababa,Ethiopia, 2016.

[20]. S. S. a. K. Kaur, "A Review on Diagnosis of Diabetes in Data Mining," International Journal of Science and Research, vol. 4, no. 6, pp. 2406-2408, 2015.

[24]. Maimon,O., Rokach,L., Discovery Handbook, Data Mining and Knowledge, Second Edition ed., Ramat Aviv, Israel: Springer New York, Dordrecht Heidelberg London, 2010.

[29]. S. V.Sathiya and N. Dr.Sai Satya, "Data Mining Tasks Performed By Temporal Sequential Pattern," International Journal of Research and Computational Technology, vol. 2, no. 3, pp. 1-6, 2012.

[30]. H. Jiawei, K. Micheline and P. Jian, Data Mining Concepts and Techniques, Third ed., New York: Morgan Kaufmann Publishers is an imprint of Elsevier, 2012.

[19]. K. Dorina, "Student Performance Prediction by Using Data Mining Classification Algorithms," International Journal of Computer Science and Management Research , vol. 1, no. 4, pp. 686-690, 2012.

[34]. L. Rokach, "Top-down induction of decision trees classifiers-a survey," IEEE Transactions on Systems, Man, and Cybernetics, 2005, p. 476–487.

[38]. R. Haryana, "Decision Tree: Data Mining Techniques Department of Computer Science," India, 2010.

[40]. L. Rokach, "Top-down induction of decision trees classifiers-a survey," IEEE Transactions on Systems, Man, and Cybernetics, 2005, p. 476–487.

[16]. P. Tina R. and S. Mrs. S. S., "Performance Analysis of Naive Bayes and J48 Classification Algoithm for Data Classification," International Journal Of Computer Science And Applications , vol. 6, no. 2, pp. 256-261, 2013.

[17] . E. W. T. Ngai, L. Xiu and D. C. K. Chau, "Application of data mining techniques in customer relationship management," A literature review and classification. Expert Systems with Applications, vol. 36, pp. 2592-2602 , 2009.

[35]. A. B. Michael J. and S. L. Gordon, Data Mining Techniques For Marketing, Sales, and Customer Relationship Management, Second ed., Indianapolis, Indiana: Wiley Publishing, Inc., 2004.

[36]. H. Jiawei, K. Micheline and P. Jian, Data Mining Concepts and Techniques, Third ed., New York: Morgan Kaufmann Publishers is an imprint of Elsevier, 2012.

[37]. R.V.Raman and K. Prasad. Applications of Knowledge Based Systems in Mining Engineering. Proceedings of the Twentieth International Symposium on the Application of Computers and Mathematics in the Mineral Industries, Volume1, Johannesburg, SAIMM. pp. 167 – 180,1987

[39]. Chee-Fai Tan. A Prototype of Knowledge-Based System for Fault Diagnosis in Automatic Wire Bonding Machine, Tubitak, Melaka-MALAYSIA,2008.

[41]. P. Sajja & R. Akarkar. Advanced Knowledge Based System: Models, Applications & Research, Vol 1, PP1-11, Jones & Bartlett publisher, Sudbury, MA, USA,2010

[42]. M.S.Abdullah, I.Benest and C.Kimble. Knowledge Modeling Techniques for Developing Knowledge Management Systems, 3rd European Conference on Knowledge Management, Dublin, Ireland, ISBN:0-9540488-6-5, pp. 15-25,2002

[43]. T. d. Jong, Types and Qualities of knowledge, Educational Psychologist, 1996.

[44]. Tripathi ,K P., A Review on Knowledge-based Expert System:Concept and Architecture," Artificial Intelligence Techniques - Novel Approaches & Practical Applications, 2011.

[45]. Abdulkerim Mohammed. “Towards Integrating Data Mining with Knowledge Based System: The Case of Network Intrusion detection”. M.Sc. Thesis, Addis Ababa University, June, 2013.

[46]. RAMANI, R. V. and PRASAD. K. V.K. “Applications of knowledge based systems in mining engineering.” APCOM 87. Proceedings of the Twentieth International Symposium on the Application of Computers and Mathematics in the Mineral Industries. Volume 1: Mining. Johannesburg, SAIMM, pp. 167- 180. 1987.

[47]. Abraham, Ajith. "Rule‐Based Expert Systems." Handbook of measuring system design. 2005

[48]. FAO “The future of livestock in Ethiopia. Opportunities and challenges in the face of uncertainty.” Rome. 48 pp. Licence: CC BY-NC-SA 3.0 IGO. 2019.

[49] M. Tesfamariam, "Integrating Data Mining Results with the Knowledge Based System for Diagnosis and Treatment of Visceral Leishmaniasis," MSC Thesis, Gondar University, Gondar, 2015.

[50] B. K. Birhane, "Developing Knowledge Based System for Diagnosis and Treatment of Diseases of Sheep and Goats," Msc Thesis, Debre birhan, 2015.

[51] T. Tadesse, M. Mulisa and F. Teka, "Bovine Mastitis: Prevalence and Isolation of Major Pathogens in Dairy Farms of Selected Sites in Addis Ababa, Ethiopia," Applied Journal of Hygiene, vol. 3, no. 3, pp. 31-37, 2014.

[52] S. Hunderra, Z. Ademe and A. Sintayehu, "Dairy Cattle Mastitis In and Around Sebeta, Ethiopia," Intern J Appl Res Vet Med, vol. 3, no. 4, pp. 332-338, 2005.

[53] [Online]. Available: http://kb.rspca.org.au/What-is-mastitis-in-dairy-cows\_221.html. [Accessed 12 3 2021].

[54] B Tadesse, “Integrating Data Mining Results with Knowledge Based System for Diagnosis and Treatment of Cattle Diseases: The Case of Debre Birhan Basso Animal Health Center,” MSC Thessis, Debre Birehan Universtiy, 2018

[55] A. Zelalem, D. Hailu and G. Getachew, "Assessment of Distribution and Associated Risk Factors of Lumpy Skin Disease in Selected Districts of West Wollega Zone, Western Ethiopia," Academic Journal of Animal Diseases, vol. 4, no. 3, pp. 130-140, 2015.

# Appendix I

**Interview questions**

The importance of this interview questions is to extract tacit knowledge from experts in disease of cattle diagnosis, which help for the development of Knowledge Based System for disease diagnosis and treatment. The interviewer writes the answerers replied using pen and notebook. I would like to thank for your collaboration and precious information.

1. Which types of cattle diseases are common in Ethiopia?

2. How a disease is identified when cattle are infected by a certain disease?

3. What are the overall symptoms that are shown when the diseases are occurrence?

4. What are the steps to diagnose a certain disease?

5. How can treat and control the cattle diseases?

# Appendix II

**User acceptance Testing evaluation criteria**

Dear Evaluator,

This evaluation form is prepared aiming at measuring to what extend does diagnosis and treatment of cattle disease KBS is useable and acceptable by end users in the area of Animal health center. Therefore, you are kindly requested to evaluate the system by labeling (√) symbol on the space provided for the corresponding attribute values for each criteria of evaluation.

I would like to appreciate your collaboration in providing the information.

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

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **No** | **Evaluation Questions** | **Poor(1)** | **Fair(2)** | **Good(3)** | **Very good(4)** | **Excellent (5)** | **Average score** | **Average performance (%)** |
| 1 | Simplicity to use and interact with the prototype system |  |  |  |  |  |  |  |
| 2 | The attractiveness the prototype system |  |  |  |  |  |  |  |
| 3 | The system is more efficient in time? |  |  |  |  |  |  |  |
| 4 | How accurately does a system reach a decision in diagnosing Cattle disease |  |  |  |  |  |  |  |
| 5 | Does the system incorporate Sufficient and practical knowledge? |  |  |  |  |  |  |  |
| 6 | The ability of the prototype system in making right conclusions and recommendations |  |  |  |  |  |  |  |
| 7 | How do you rate the contribution of the system in the domain area? |  |  |  |  |  |  |  |

# Appendix III

**Initial list of original attributes with their description**

|  |  |  |  |
| --- | --- | --- | --- |
| **Attributes Name** | **Description** | **Data Type** | **Domain Value** |
| Number | Counting, measure and label | Numeric | {Continuous value} |
| Date | Diseased cattle date of registration Date/Time | Numeric | {Yy/mm/DD format} |
| Kebele | Address of the cattle | Numeric | {Continuous value} |
| Sppices | Cattle species | Nominal | {Bovine} |
| Sex | Cattle gender | Nominal | {Male, Female} |
| Breed | Cattle offspring | Nominal | {Local, Cross} |
| Swee\_of\_neck | Cattle neck swells | Binary | {Yes, No} |
| Swelling\_under | A bag like organ containing the cattle glands | Binary | {Yes, No} |
| Decrease\_milk\_yield | Reduce the milk | Binary | {Yes, No} |
| Temprature | The rectal temperature of the cattle | Numeric | {Continuous} |
| Smaking\_lips | Involving a loud smacking of the lips | Binary | {Yes, No} |
| Change\_color\_of\_milk | Normal milk color change | Binary | {Yes, No} |
| Abnormal\_breathing | Fast breathing depending on age a | Binary | {Yes, No} |
| Poor\_body\_condition | Failing of state of physical fitness | Binary | {Yes, No} |
| Coughing | To force air through your throat with a short | Binary | {Yes, No} |
| Nodular\_lesion | The cattle body is inflamed | Binary | {Yes, No} |
| Lameness | Cattle impair freedom of movement | Binary | {Yes, No} |
| Rough\_hair | Standing the cattle hair | Binary | {Yes, No} |
| Crepitation\_sound | Cattle make a sound of crackling sound | Binary | {Yes, No} |
| Sore\_feet | Cattle foot pain | Binary | {Yes, No} |
| Depression | Decrease and loss of interest in pleasurable activities | Binary | {Yes, No} |
| Diarrhoea | Cattle intestine evacuation with More or less fluid stools | Binary | {Yes, No} |
| Fever | Cattle body temperature that is higher than normal | Binary | {Yes, No} |
| Painful | Causing pain to cattle body | Binary | {Yes, No} |
| Nasal\_discharge | Mucus flows out of cattle nose | Binary | {Yes, No} |
| Loss\_of\_apptetite | Cattle decrease feeding | Binary | {Yes, No} |
| Age | The age of the cattle | Nominal | {Adult, young, calf} |
| Class | Confirmative diagnosis of cattle disease | Nominal | {Blackleg, FMD, LSD, anthrax, mastitis,  Enteritis, Pneumonia,  Internal Parasite, Amphistomiasis, Calf Diphtheria} |

# Appendix IV

**An integrated knowledge based system for diagnosis of cattle disease diagnosis**

**Sample prolog code**

:-[Classifier]:-

Start:-

Write (‘Select the Value of Temperature: From The option’’),read (temp),

Write (‘Select the Value of Smacking lips: From The option’’),read (smacking),

Write (‘Select the Value of Sore Feet: From The option’’),read (sore),

Write (‘Select the Value of Fever: From The option’’),read (fever),

Write (‘Select the Value of Lameness: From The option’’),read (lame),

Write (‘Select the Value of Depression: From The option’’), read (depression),

Write (‘Select the Value of Painful: From The option’’), read (painful),

Write (‘Select the Value of Crepitation Sound: From The option’’), read (sound),

Write (‘Select the Value of Loss of Appetite: From The option’’),read (appetite),

Write (‘Select the Value of Diarrhea: From The option’’),read (diarrhea),

Write (‘Select the Value of Nodular lesion: From The option’’),read (lesion),

Write (‘Select the Value of Nasal discharge: From The option’’),read (discharge),

Write (‘Select the Value of Coughing: From The option’’),read (cough),

Write (‘Select the Value of Abnormal Breath: From The option’’),read (breath),

Write (‘Select the Value of Swell of Udder: From The option’’),read (udder),

Write (‘Select the Value of Decrease Milk Yield: From The option’’),read (milkyield),

Write (‘Select the Value of Change Color of Milk: From The option’’),read (colormilk),

Write (‘Select the Value of Swell of Neck: From The option’’),read (neck),

Write (‘Select the Value of Rough Hair: From The option’’),read (hair),

Write (‘Select the Value of Poor Body Condition: From The option’’),read (condition),

Jpl\_call(‘writetotext.Write To Text’, writeToSample,[ Temperature,Smacking\_lips,Sore\_Feet,Fever,Lameness,Depression,Painful,Crepetation\_sound,L oss\_of\_Appetite,Diarrhea,Nodular\_lesion,Nasal\_discharge,Coughing,Abnormal\_breath,Swell\_of \_udder,Decrease\_milk\_yield,Change\_color\_of\_milk,Swell\_of\_neck,Rough\_hair,Poor\_body\_con dition],\_),

wkpl\_read\_arff(‘Tadeset.arff’, Mainarff),

wkpl\_set\_classIndex(Mainarff,20),

wkpl\_classifier(‘weka.classifiers.rules.JRip’, Classifier),

wkpl\_build\_classifier(Mainarff, Classifier),

decrement (Number, Index),

Wkpl\_get\_instance(dataset, Index, Instoclassify),

Wkpl\_classify\_instance(Instoclassify, Classifier, Prediction),

Write(Prediction),nl,

Pred\_value(Prediction).

Pred\_value(Prediction):-

(Prediction==0.0) write (‘FMD’);

(Prediction==1.0) write (‘Blackleg’);

(Prediction==2.0) write (‘Enteritis’);

(Prediction==3.0) write (‘LSD’);

(Prediction==4.0) write (‘Pneumonia’);

(Prediction==5.0) write (‘Internal Parasite’);

(Prediction==6.0) write (‘Anthrax’);

(Prediction==7.0) write (‘Mastitis’),

(Prediction==8.0) write (‘Amphistomiasis’),

(Prediction==9.0) write (‘Calf Diphtheria’), nl.