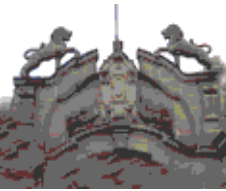


Addis Ababa
University
(Since 1950)



ADDIS ABABA UNIVERSITY
SCHOOL OF GRADUATE STUDIES
SCHOOL OF INFORMATION SCIENCE

APPLICATION OF CASE BASED RECOMMENDER SYSTEM IN
INVESTMENT SECTOR AND INVESTMENT ACTIVITY SELECTION TO
NEW INVESTORS: IN THE CASE OF ETHIOPIA.

**A thesis submitted to the school of Graduate Studies of Addis Ababa University in
Partial fulfillment for the Degree of Master of Science in Information Science**

BY

YIBELTAL CHANIE

June, 2013

ADDIS ABABA UNIVERSITY
COLLEGE OF GRADUATE STUDIES
SCHOOL OF INFORMATION SCIENCE

APPLICATION OF CASE BASED RECOMMENDER SYSTEM IN
INVESTMENT SECTOR AND INVESTMENT ACTIVITY SELECTION TO
NEW INVESTORS: IN THE CASE OF ETHIOPIA.

BY

YIBELTAL CHANIE

Names and Signature of Members of the Examining Board

Chair person, Examining Board

Signature

Advisor

Signature

Examiner

Signature

ACKNOWLEDGEMENT

First and foremost, I would like to give the almighty God who provided me everything to finish this thesis.

I gratefully thanks to my advisor Dr. Gashaw Kebede, for his valuable commitment, patience reading for each and every section of the thesis, encouragement, guidance and critical comments from the initial to the final level of this research that enable me to finish the thesis work.

My grateful thanks go to Dr. Million Meshesha who provides unreserved precious advices, comments important directions at the research proposal development.

I gratefully thank Ethiopian investment agency domain experts of each investment sector and also Mr. GirumTadesse (investment promotion senior expert), Yonas Latamo (promotion team coordinator), Seid Mohamed(director, licensing and registration directorate) who devoted their golden time for the interviews and consultation sessions throughout the research and for providing me valuable facilities and resources.

I would like to thank Ms. Haregwayne Mirotaw (information service expert), for providing relevant data cases, providing other necessary information and encouraging me by providing necessary material for my research work.

It is a pleasure to express my gratitude to all my family members for their encouragements and supports throughout my study. I especially thank my brother Mamaru and Getnet, Yimles for his spiritual and secular supports, advices, encouragements and asking me about the progress of my thesis work.

Finally, I would also like to thank sincerely all my friends those who helped me with their valuable support during the entire process of this thesis.

TABLE OF CONTENTS

CONTENTS	PAGES
ACKNOWLEDGEMENT-----	I
TABLE OF CONTENTS-----	II
LIST OF TABLES-----	VI
LIST OF FIGURES-----	VII
LIST OF ACRONYMS-----	VIII
ABSTRACT-----	IX
CHAPTER ONE	
1.0. INTRODUCTION-----	1
1.1. Background-----	1
1.2. Problem of statements-----	4
1.3. Objective of the Study-----	7
1.3.1. General Objective-----	7
1.3.2. Specific Objectives-----	7
1.4. Methodology-----	8
1.4.1. Literature Review-----	8
1.4.2. Data collection and preparation method-----	8
1.4.3. Case Representation-----	10
1.4.4. Implementation Tool-----	11
1.4.5. Testing and evaluation-----	12
1.5. Scope and limitation of the Study-----	13
1.6. Significance of the Study-----	13
1.7. Organization of the thesis-----	14
CHAPTER TWO	
2.0. LITERATURE REVIEW-----	15
2.1.Recommendation system-----	15
2.2. Architecture for Recommender System-----	16
2.3.Types of Recommender System -----	17

2.3.1. Knowledge based recommender systems-----	17
2.3.1.1.Knowledge based system (case based system) -----	17
2.3.1.1.1. Architecture of knowledge based system-----	18
2.4.The Knowledge engineering process-----	19
2.4.1. Knowledge acquisition-----	19
2.4.2. Knowledge representation-----	21
2.4.3. Knowledge validation-----	21
2.4.4. Knowledge based reasoning techniques-----	22
2.4.5. Case based Reasoning-----	22
2.5. Case based reasoning cycle-----	23
2.6. CBR Techniques -----	27
2.7. Advantage of case based reasoning-----	30
2.8.Disadvantage of case based reasoning-----	30
2.9.Integrating Rule-based and Case-based Reasoning-----	30
2.10. CBR System Performance Evaluation Methods -----	31
2.11. Knowledge based system development tools-----	32
2.11.1. Investment in Ethiopia-----	34
2.11.2. Investment related infrastructure in Ethiopia-----	35
2.11.3. Areas of opportunity to invest in Ethiopia-----	35
2.12. Related research work-----	37

CHAPTER THREE

3.0. KNOWLEDGE ACQUISITION AND MODELING-----	40
3.1. Knowledge Acquisition -----	40
3.1.1. Knowledge Acquisition from domain experts-----	40
3.1.2. Knowledge acquisition from investors-----	42
3.1.3. Knowledge Acquisition from Relevant Document -----	42
3.2. Attribute selection using data mining tools-----	43
3.2.1. Data Collection-----	43
3.2.2. Data Preparing and Cleaning-----	44

3.2.3. Attribute (Features) selection -----	45
3.3. Conceptual Knowledge modeling-----	46
3.3.1. Investment sector and investment activity selection-----	48
3.3.1.1.Risky investment sectors-----	49
3.3.1.2.Factors that Influences investment sector and investment activity selection decisions-----	51
3.3.1.2.1. Age -----	52
3.3.1.2.2. Gender-----	55
3.3.1.2.3. Capital -----	57
3.3.1.2.4. Location of investment-----	59
3.3.1.2.5. Number of employee-----	63
3.3.1.2.6. Types of investor-----	64
3.3.1.2.7. Form of ownership-----	65
3.3.1.3.Structures of Investment Companies-----	67
3.3.1.4.Investment Areas -----	67
3.3.1.4.1.Investment sectors and their respective investment activity-----	68
3.4.Knowledge Representation-----	73
3.5.Investment sector and investment activity selection Case Structure -----	74

CHAPTER FOUR

4.0. DESIGN AND IMPLEMENTATION OF THE PROTOTYPE -----	77
4.1. Introduction -----	77
4.2. Designing the Architecture of CBRISAIAS-----	77
4.3. Case-based Reasoning System for investment sector selection-----	79
4.3.1. Building the Case Base -----	82
4.3.2. Case Representation -----	83
4.3.3. Description of CBRISAIAS Case Attributes -----	83
4.3.4. Managing Connectors -----	87
4.3.5. Managing Tasks and Methods-----	88
4.3.5.1. Managing Tasks / CBR application-----	88

4.3.5.2.	Case Similarity, Matching and Ranking -----	91
4.3.5.3.	Managing Methods -----	92
4.3.5.4.	Deploy the case base recommender system -----	93
4.4.	Explanation Facilities -----	94

CHAPTER FIVE

5.0.	TESTING AND PERFORMANCE EVALUATION OF THE PROTOTYPE -----	96
5.1.	Introduction-----	96
5.2.	Case similarity testing-----	96
5.3.	Evaluations of the retrieval and reuse process by using statistical analysis-----	97
5.3.1.	Evaluation of the retrieval process-----	98
5.3.2.	Evaluation of the reuse process-----	100
5.3.3.	Comparison of the Performance of CBRISAIAS with Previous CBR Systems -----	101
5.4.	Testing the learning mechanism-----	102
5.5.	User Acceptance Testing -----	104
5.6.	Discussion on user acceptance and system performance using recall and precision, case similarity and reuse process -----	108

Chapter six

6.	CONCLUSION AND RECOMMENDATIONS-----	111
6.0.	Conclusion -----	111
6.1.	Recommendation-----	113
	Reference -----	114
	Appendix -----	117

LIST OF TABLES

Table 2.1: Potential Areas for Farming-----	36
Table 3.1: Shows the Case Structure for investment sector selection-----	75
Table 4.1: Descriptions and Weight of the Selected Attributes -----	85
Table 5.1: Sample of queries that are used in this experiment with their values-----	96
Table 5.2: Query similarity with their corresponding cases from the case base-----	97
Table 5.3: Relevant cases assigned by domain experts for the sample test case-----	99
Table 5.4: Recall and precision results for the sample test case-----	100
Table 5.5: Accuracy value of the reuse process -----	101
Table 5.6: A comparison of CBRISAIAS system with the previous CBR systems-----	101
Table 5.7: Sample of new investor case-----	102
Table 5.8: The CBRISAIAS system performance evaluation by the domain experts-----	106
Table 5.9: CBRISAIAS performance evaluation by the investor's -----	107
Table 5.10: Domain experts and investors feedback on closed ended questions-----	108

LIST OF FIGURES

Figure 2.1: Architecture of recommender system-----	16
Figure 2.2: Architecture of knowledge based system-----	18
Figure 2.3: The Major Components of the CBR System-----	23
Figure 2.4: Case based reasoning cycle-----	24
Figure 2.5: Software Architecture of jCOLIBRI-----	33
Figure 3.1: Hierarchical structure of investment sector and investment activity selection-----	48
Figure 3.2: Age factors on investment activity selection-----	55
Figure 3.3: Gender factor on investment activity selection-----	57
Figure 3.4: capital factors on investment activity selection-----	59
Figure 3.5: Location of investment-----	61
Figure 3.6: Type of employee-----	64
Figure 3.7: The hierarchy when form of ownership is partnership-----	66
Figure 3.8: The hierarchy when form of ownership is joint-----	66
Figure 3.9: The hierarchy when form of ownership is individual-----	67
Figure 4.1: CBRISAIAS Architecture -----	79
Figure 4.2: The Main Window of jCOLIBRI -----	80
Figure 4.3: Creating new CBR Application-----	81
Figure 4.4: Types of jCOLIBRI extensions -----	81
Figure 4.5: CBR applications-----	82
Figure 4.6: Defining Case Structures and similarity-----	84
Figure 4.7: jCOLIBRI Connector Schema-----	87
Figure 4.8: Managing Connector Configuration -----	88
Figure 4.9: Configure the CBR Application-----	89
Figure 4.10: Revision and retain tasks -----	91
Figure 4.11: Tasks and Methods Configuration -----	93
Figure 4.12: Window for Case Entry into the Case Base -----	94
Figure 4.13: The explanation facility-----	95
Figure 5.1: Revise process for the newly solved cases -----	103
Figure 5.2: Retaining process of the newly solved case for the future use-----	104

LIST OF ACRONYMS

AI	-----	Artificial Intelligence
KBS	-----	Knowledge-based system
CBRISAIAS	-----	Case Based Reasoning for Investment Sector and Investment Activity Selection
CBR	-----	Case Based Reasoning
PSM	-----	Problem Solving Method
GUI	-----	Graphical User interface
GAIA	-----	Group of Artificial Intelligence Applications
CSV	-----	Comma Separated Values
Plc	-----	Private limited company
US	-----	United States
UK	-----	United Kingdom
UNCTAD	-----	United Nations Conference on Trade and Development
ICT	-----	Information and Communication Technology
PROLOG	-----	PROgramming in LOGic

ABSTRACT

Investment is a commitment of funds, directly or indirectly, to one or more assets with the expectation to enhance future wealth. Each and every person can be specifically differentiated on various parameters in investment selection decision. Prior to purchasing any investment product or service, it is important that investors fully think about their unique needs and overall financial situation in order to determine whether the investment or service is right for them.

However, investment advice in Ethiopia has different problems. Among these problems, there are no sufficient and knowledgeable experts to give advice to investors on investment sector and investment activity selection. And also there is no consistency in advising system from expert to expert. As a result, majority of investors drop out of the investment projects or are not successful. The aim of this research is to develop case based recommender system for investment sector and investment activity selection that assists investment experts and investors to make timely decisions. To develop case based recommender system for investment sector and investment activity selection, important knowledge was acquired through interview and document analysis. Twelve domain experts and four investors were interviewed to elicit the required knowledge about investment sector and investment activity. The acquired knowledge was modeled using hierarchical structure.

The acquired knowledge was represented using feature value case base representation and implemented using jCOLIBRI programming tool. The main data source (case base) used to develop Case based recommender system for investment sector and investment activity selection (CBRISAIAS) is previous investor cases from EIA. Nearest neighbor retrieval algorithm is used to measure the similarity of new case (query) with cases in the case base. As a result, if there is a similarity between the new case and the existing case the system assigns the solution (recommended investment sector and investment activity) of previous case as a solution to new case. To determine the applicability of the prototype system in the domain area, the system has been evaluated by the domain experts and investors through visual interaction based on the criteria of easiness to use, time efficiency, applicability in the domain area and providing correct recommendation.

According to the evaluation through user acceptance 82% system performance is obtained from domain experts and 84% system performance is obtained from investors. And also the performance of the prototype is measured using recall, precision and accuracy measures, where the system achieves 85% recall, 64% precision and 87% accuracy.

Further study should be conducted that by adding other important attributes that have an influence in investment sector and investment activity selection. And also further study conducted by using hybrid system of rule based and case based recommender system to enhance the performance of the system, because the hybrid system eliminates the limitation of case based and rule based recommender system. Finally further research can be done by developing the case based recommender system using different local languages for the purpose of investors can easily communicate with the system by using their own languages.

1. Introduction

1.1. Background

Recommender systems enable people to share their opinions and benefit from each other's experience. They can be defined as "any system that produces individualized recommendations as output or has the effect of guiding the user in a personalized way to interesting or useful objects in a large space of possible options" [7]. Recommender systems can be described as an information filtering technology that is used to present information on items to the user that are in line with the users tastes. Recommender systems are used to support users in their decision-making in daily life situations in terms of pre-selecting information that might be of interest to them, where they confronted situations without sufficient experience in the available alternatives. Recommender systems can be described as an information filtering technology that is used to present information on items to the user that are in line with the users tastes. These systems involve predictive models, heuristic search, data collection, user interaction and model maintenance [7].

Recommender systems learn about user preferences over time and automatically find things of similar interest, thus reducing the burden of creating explicit queries. They dynamically track users as their interests change. However, such systems require an initial learning phase where behavior information is built up to form user profile. During this initial learning phase performance is often poor due to the lack of user information; this is known as the cold-start problem [7].

Recommendation process is a sequential process. It is not a simple sequential Prediction problem rather a sequential decision problem. At each point the recommender system makes a decision about the item/items. This decision should take into account the sequential process involved and the optimization criteria suitable for the recommender system. They suggest items to the user who can accept one of the recommendations. At the next stage a new list of recommended items is calculated and presented to the user. This sequential nature of the recommendation process, where at each stage a new list is calculated based on the users past ratings [9].

Recommender systems may be based on collaborative filtering (by user ratings), content-based filtering (by keywords), knowledge based recommender system that uses knowledge about users and products to pursue a knowledge based approach to generating a recommendation by reasoning about what products meet the user's requirements and hybrid filtering (by both collaborative and content-based filtering) [16,

17]. Case based recommender system is a part of knowledge based recommender system that exploits case based reasoning to generate personalized recommendations for exploiting the knowledge contained in past recommendation cases. These systems assume that the quality of a new recommendation depends on the quality of the recorded recommendation cases [18]. A case based recommender system maintains a set of cases of previously solved recommendation problems and their solutions. According to [19], case base is the product catalogue which is the solutions of the recommendation problem, and the problem is the user's query that is essentially a partial description of the desired product. A case could be thought of as a record in a database; collection of features and their values.

Ethiopian Investment Agency (EIA) is a government agency established to promote, encourage and facilitate private investments in general and foreign investment in particular in Ethiopia. The favorable climate for foreign investment created by the overall economic conditions of the country is further augmented by specific incentives and administrative procedures to encourage investment [13].

Investment is an expenditure of capital by private individuals to establish a new business or to expand or upgrade a business that already exists. Legislation often seeks to provide incentives to promote private capital investment, especially by promoting participation of foreigners in the national economy. In Ethiopia, where investment has boomed in recent years, causing deleterious effects on the environment and natural resource base of the country, it is crucial that EIA be integrated with the current legal framework for investment [13].

According to [13], before going to Ethiopia to begin the investment procedure, it is advisable to consider the following:

1. Familiarize yourself with the Investment Code and the latest Investment Guide which can be downloaded from the website of the Ethiopian Investment Agency at www.ethioinvest.org
2. Decide on the Sectors of investment interest.
3. Calculate the investment capital required.
4. Have an idea of labor requirement.
5. Have an idea of Environmental situation.

The specific investment opportunities in Ethiopia promoted by the state are in agriculture (food crops, coffee, milk, horticulture, apiculture, floriculture); industry (leather goods, soft drinks, breweries, pasta,

edible oil, animal feed, flour mill); chemical industry (printing ink, adhesives, essential oils, pvc pressure pipes, soaps); paper and printing industry (cement paper bag, printing press, handmade paper, corrugated paper board); woodwork (fuel briquette, paint brush, straw board, ply wood, pencil); non-metallic mineral products (brick, marble, chalk, cement); textile industry (canvas, socks, surgical bandages, cotton); social services (education, health); hotel and tourism; and, mining and energy.

The legislative frame work in Ethiopia offers businessmen a number of opportunities for organizing business activities in the country. In Ethiopia, a business or an investment may be carried out in any one of the following forms: by an individual operating as a sole proprietorship, by two or more persons in a partnership agreement, and by a foreign company registered or incorporated abroad [13].

Obviously, investment constitutes the backbone of an economy. The Ethiopian investment sector plays a vital role in the industrial development of the country. According to citation of [1] Industrial development was earlier believed to have occurred because of large enterprises. However, starting in the late 1970s and early 1980s, investment has been perceived as the key agent for industrialization. It is recognized that investment provides not only employment opportunities to an increasing number of people in the country, but it is also an effective means of fighting poverty and income inequality. At the same time, investment serves as a training ground for emerging entrepreneurs. It is within this context that investment development became focal attention for governmental as well as nongovernmental organizations [1].

According to [26] Investment advice refers to any recommendations how to invest their deferred income for retirement regarding an investor's characteristics. Many professionals, including financial planners, bankers can provide investment advice that is specific to their financial situation and short and long term financial goals. Investment advising system requires that certain steps be taken by advisors, such as conducting a suitability analysis before advice can be given. This analysis considers individuals' total financial capital, gender, age, type of investor, form of ownership and their interest to ensure recommendations are appropriate for their specific situation [26].

Getting investment advice is a critical issue for investors since knowing the right investment area is a key factor to consider to new investors [21]. Investors tend to lose money by choosing wrong investment sector and investment activity. In the context of investing, the wise words of the vision emphasize that

success depends on the selection of your investment sector and investment activity that should fits with your personal and socio-economic characteristics [21].

In Ethiopian most investors select investment sectors and investment activity based on their interest without considering the amount of capitals, age, gender, location of investment such as availability of infrastructures in that specific location, availability of customer in that location, the availability of raw materials to run the investment and the availability of high employment potential to run the investment [4]. The goals of this research is to develop a case based recommender systems for investment Sector and investment activity selection to new investors by using previous investor cases as a case base.

1.2. Problem of statements

In Ethiopian investment advising system there are a number of problems faced by domain experts. The problems faced by domain experts in advising system are lack of appropriate, relevant and understandable information to give advice and guidance to their clients. Lack of guidance and criteria to give advising system for investors in the selection of investment sector and investment activity are another problems faced by domain experts [15].

In addition, there is no integration or collaboration between different domain experts that are found in different investment sectors to develop an organized guidance or advisory system to new investors because collection of ideas from different investors is important to develop well defined and organized guidelines to investors. For instance, one expert may have awareness about health related investments but have no more idea about investment in agriculture, engineering, technology etc these shows experts advice only the one most familiar with you [15].

According to one investor in horticulture and flowers farming, the main problems in Ethiopian investment agency's advice system was the advising services given in EIA are not fast and have no consistency advising styles in domain experts, due to this it takes more time to get investment advices to invest.

Another investor in Dairy farm & milk processing comments that there are not enough and experienced experts in Ethiopian investment agency office that are able to give advice on investment sector and investment activity selection more suitable to each investor's. It is also indicated that the advice given

by the domain experts vary from investor to investor even when investors have the same capital or characteristics.

Difficulty of getting investment advice is a critical issue for investors since knowing the right investment area is a key factor and also knowing the right company to invest is another factor to consider to new investors [21]. Investors tend to lose money by choosing the wrong investment sectors and investment activity without the consideration of their characteristics or location of investment or amount of capital needed to start investment. In the context of investing, the wise words of the vision emphasize that success depends on selecting the best investment activity that fits with your personal and socio-economic characteristics [21]. Even though all investors are trying to make money, each one comes from a diverse background and has different characteristics and capabilities, each individual follows specific investing vehicles and methods that are suitable for certain types of investors.

According to Ethiopian investors' promotion senior expert Ato Girum Tadesse, investors have a few factors to consider when looking for the right place to park their money such as amount of capital, location of investment, form of ownership and type of investor. An investment sector and investment activity selection decision also depends on a person's age, Gender and personal circumstances and investors interested areas of investment to invest. The senior expert further comments that because of lack of experienced domain experts to give investment advising system, investors are confused about where to invest, what fund to use and which investment Sector and investment activity is best to me to be successful.

According to [6] quantitative research carried out in the US identifies a similar range of factors, including amount of capital, wealth, age, marital status, gender, location of investment and level of education affecting the success of investors in investment activities. Additional evidence from the US and the UK supports the notion that individuals lack the knowledge and understanding to make pension investment sector and investment activity choices. Other Studies have shown that investors' characteristics are associated with their preferences of investment areas. For example, the study by [10] shows educated people prefer to invest in gold, Treasury bill, government bond etc. that need technical and complex investment instruments while female tend to have preference to invest safer instruments like time deposit and gold rather than risky instruments requiring technical knowledge like stock exchange.

Generally Investors to invest in different investment activity, it is important to know what works, how it works and why it works by considering amount of capital, location of investment , risk taker capability etc. For example, some of the linkages that work well in coffee with its relatively high margins (i.e. the weather condition of locations of investment sector) do not work well in other agricultural sectors with lower margins, such as beef. And also some of the micro finance linkages that work well in rural areas are does not work well in urban areas. The “why “is of critical importance in determining what it is possible to replicate, and where, what can be standardized, and what must be localized [12].

As shown above, therefore, one of the major problems faced in Ethiopian investment is the difficulty of getting fast, reliable, and consistent expert advice on areas of investments that are suitable to each investor’s characteristics and capabilities. The other problem, related to the first, is the inadequacy of the number of experienced experts and consulting firms who can give advice on investment issues in the country. So to fill the above stated problems in Ethiopian investment agency, this research aims to design a case based recommender system that helps investors get fast and consistent advisory service to select investment sectors and investment activity. As a result, investors can identify the areas of investment activity that have the highest potential of success and that match their personal characteristics.

To this end, this study attempts to explore and answer the following research questions:

- ✚ What are the major attributes or factors that have more influence in investment sector and investment activity selection in Ethiopia?
- ✚ What are the common investment sectors and investment activity in Ethiopia?
- ✚ What looks like the availability of domain experts in EIA to advice investors on investment sector and investment activity selection?
- ✚ How to acquire, model, represent, and implement the required knowledge for the proposed system?
- ✚ Is a case based recommender system applicable in investment sector and investment activity selection in Ethiopia?
- ✚ Is a case based recommender system providing acceptable performance in investment sector and investment activity selection in Ethiopia?

1.3. Objective of the Study

To contribute towards addressing the problems described in section 1.2 above, the study has the following general and specific objectives:

1.3.1. General Objective

The general objective of the study is to develop Case Based Recommender System that can give recommendation on the selection of investment sectors and investment activity in Ethiopia to foreign and domestic investors.

1.3.2. Specific Objectives

With the intention of a CBR system to provide recommendations and guidance to investment Sector and investment activity selection in Ethiopia the main specific objectives are:

- To understand the concept of case based recommender system and how it is designed by reviewing different literatures.
- To identify and collect the previous cases, facts, insights and rules that new investors need to know in order to select investment sectors and investment activity.
- To identify the main attributes that influence in the selection of investment Sector and investment activity to new investors.
- To identify the types of investment sectors and investment activity in Ethiopia
- To collect different expert ideas together and develop recommender systems to new investors.
- To identify suitable models, representation techniques and implementing tools for the proposed case based recommender system.
- To develop a prototype case based recommendation system to investors on the selection of investment Sector and investment activity that best matches with their characteristics.
- To determine the performance of the proposed recommender system using different evaluation techniques.
- To forward recommendations for further work.

1.4. Methodology

To realize the main goal of this study, different methods and techniques are used.

1.4.1. Literature Review

For proper understanding of the problem related to investment sector and investment activity selection, to collect necessary information related to recommender system and successful completion of this study relevant literatures such as books, journals, magazines, conference papers, manuals, and resources from internet are reviewed for achieving the research objective.

1.4.2. Data collection and preparation method

For the purpose of this study, both primary and secondary data collection methods are employed to collect the required domain knowledge. As primary sources, investment experts from EIA and investors from 12 different investment sectors have been interviewed. In addition, relevant literature from all possible sources including journal articles, investment related websites, manuals especially on Ethiopian investment, and guidelines are reviewed. Domain experts are selected by using purposive sampling techniques. Domain experts are selected based on their educational qualifications related to the domain area and their immediate position in the investment promotion. Investors are selected by using random sampling from different investment sectors.

Semi-structured interview techniques have been employed to acquire the required tacit knowledge from the selected domain expert. It allows the interviewer to change the order of the questions and add new question based on the participant response. Therefore, this interview focuses on the concept, procedures, guidelines and experience which domain expert used while advising in investment (see interview questions in Appendix D).

The primary source of dataset used to undertake this research was previous investor's case from EIA office. The dataset contains a total of 35235 investor cases record in a year from 2008-2012. The dataset of each year stored in different Microsoft excel file.

The researcher used purposive data selecting methods to select sample datasets for this research. Because there are large amount of data and the data includes all investors in different stages namely in pre-implementation stage (not start any activity to investment), implementation stage (at initial stage to implement the investment project) and operational stage (successful). Investors in Pre-implementation stage means investors does not start any activity to start the investment even accepting lands for implementation. Investors in implementation stage means investors start some activities like land acquisition but do not start production, manufacturing, export, import of goods and services. And investors in operational stage are investors starting of producing or manufacturing or export or import goods and services or other activity.

Operational investors are successful by starting production or manufacturing or export or import of goods and service. As a result the main objective is to select success investors in each investment project, the researcher selected investors who are at operational stage (who are successful) for this research as a case base. These successful investors were selected from each investment activity from the data found from EIA. There are totally 11500 operational investors from the total of 35235 investors. But since the data set having different problems such as missing value of majority attributes, redundancy, etc the researcher selects One thousand three hundred forty four (1344) operational cases of investors from total of 11500 operational s investors. Finally these 1344 cases represented as case base in a CSV format that are used as previously solved cases. From 1344 case used for this research 372 are from Agriculture; hunting and forestry sector, 61 are from construction sector, 61 from educational sector, 2 from fishing sector,38 from health and social work sector, 95 from hotel and restaurant sector, 385 from manufacturing sector,3 from mining and quarrying sector, 2 from Other community, social and personal service activities sector , 261 from Real estate, renting and business activities, 30 from Transport, storage and communication sector and 31 from Wholesale, retail trade and repair service sector. The number of samples from each investment sector depends on the total number of cases in the case base so the researcher selects the case from each sector proportionally.

1.4.3. Case Representation

After the required knowledge is acquired, the next task is knowledge (case) representation. Basically, case based reasoning (CBR) is a problem solving methodology that addresses a new problem by first retrieving past cases i.e. already solved similar case, and then reusing that case for solving the current problem.

Case representation in CBR makes use of familiar knowledge representation formalisms from AI to represent the experience contained in the cases for reasoning purposes. The case representation task is concerned with the selection of relevant attributes, the definition of indexes and Structuring the knowledge in a specific case implementation.

For this research, the acquired cases are represented using one of the different case representation methods that are appropriate for the research. Among the different case representation methods, feature-value case representation method is used. Other case representation methods like relational database case representation, predicate based representation and soft computing case representation methods have their own advantages and disadvantages. But, for this study feature-value case representation method is appropriate. Feature-value case representation method is a process of representing a case as a vector of attribute-value pairs, similar to the propositional representations used in Machine Learning (ML), that support nearest neighbor matching and instance-based learning. The reason for representing the cases using feature-value representation is that this approach supports nearest neighbor retrieval algorithm and it represents cases in an easy way [24, 25].

Cases are selected and retrieved in a ranked order based on their similarity for the given new case query. There are different case retrieval algorithms that different CBR systems use. For this research nearest neighbor retrieval algorithm is used to measure the similarity of new case(query) with cases in the case base.

1.4.4. Implementation Tool

There are many programming tools used to develop recommender system. Among these, there are two important CBR implementation frameworks available for free for teaching and academic research. These are myCBR which is developed by German Research Center for Artificial Intelligence and jCOLIBRI which is developed by Group of Artificial Intelligence Applications (GAIA). Also there is declarative programming language in artificial intelligence research called prolog that used for knowledge base system development. For this research jCOLIBRI has been selected.

The reason of the selection of this programming tool is the features and abilities of the tool in case based reasoning. In addition JCOLIBRI Case base recommender system has the following capabilities [55].

- jCOLIBRI supports the full CBR cycles (Retrieval, Reuse, Revise and Retain as discussed in literature review of section 4.5.1.1).
- jCOLIBRI is suitable for developing large scale applications
- jCOLIBRI is extensible, reusable.
- jCOLIBRI is more user friendly than others and it works well in external database such as plain texts.
- In jCOLIBRI Knowledge representation logic is based on local and real past cases rather than generalized rules/guidelines, so it is not complex and time consuming.
- In jCOLIBRI the usage of the model could reduce the time in similar planning applications and overcome uncertainty by avoiding to start from the beginning every time (enables fast introduction). But if the system is developed by coding like prolog programming it is complex, time consuming and always start from the beginning to develop a system.
- jCOLIBRI have also the main function of learning ability . That means in jCOLIBRI when new cases are coming in the domain area, the system have the capability to learn the new cases and update this new case in the existing case for the purpose of using in the future as a recommendation.

jCOLIBRI stores all the configuration data using different XML configuration files. When the application is executed, the framework core reads these files to know how to configure the CBR system. You can write or modify this configuration files by hand, however it can be a very tedious task. XML

intends to be a standard interchange language of data between computers, not to be managed directly by humans. [55].

So in this research the database will contain all the cases in the form of records grouped by attributes with **.txt** format. These cases are used as input for the system which is collected from EIA for decision making process.

1.4.5. Testing and evaluation

The developed prototype case based recommender system is tested and evaluated to ensure the performance of the system whether meeting the objective or not. The evaluation processes focus on system's user acceptance of the prototype, the retrieval performance of the system, the reuse performance of the system, and case similarity of the system. User acceptance measurements are concerned with issues how well the system addresses the needs of the user, whereas performance measurement determine if the system perform the required task successfully. User acceptance system evaluators use visual interaction methods together with questionnaire. System evaluators interact with the system by using appropriate cases. Based on that, they evaluate the performance of the system by using close ended questions.

In addition to this, the standard measures of relevance (performance of the system) in the information retrieval (precision and recall) have been used to evaluate the performance of the prototype. As retrieval task of the CBR aims to retrieve relevant cases from the case base, precision and recall are useful measures of retrieval performance in CBR [23]. Recall is defined as the ratio of the number of relevant cases returned to the total number of relevant cases for the new case in case base [23, 24]. Whereas precision is the ratio of the number of relevant cases returned to the total number of cases for a give new case [23, 24]. For the implementation of this test, different threshold value is used to measure the similarity between the existing case and the new case. In this research the reuse process and the learning testing of the case base recommender system prototype are also evaluated. The retrieval process (i.e. recall and precision) and the reuse process of the prototype uses 52 sample cases that make up the case base as training and testing data. Both retrieval and reuse process uses leave-one-out cross validation testing proportion i.e. the evaluation is done for all cases by making one of the cases as a testing data and the rest of data as a training data(case base).the researcher conducted 52 experiments for both

retrieval and reuse evaluation of the system. The performance of the reuse task is measured by using accuracy. The researcher also made an experiment on case similarity testing to know how new cases are matched with the cases from the case base.

1.5. Scope and limitation of the Study

The study would focus on developing prototype case based recommender system for investment Sector and investment activity selection based on different characteristics of investors. The recommender system will not include other activities of the investment agency such as giving and renewal of investment license, calculation of taxation, etc. Moreover, the aim of the research is to develop a prototype case based recommender system since it is impossible to develop a full-fledged system within the given time and resources available for the research. Developing a full-fledged system demands one to construct maps of different levels for easy exploration, which in turn requires a long period of time.

The main limitation in this developed system is as previous investor cases are collected from EIA, the case structure for the prototype is made up using attributes that are recorded in the file of previous investors. Due to these attributes like level of education, marital status, level of risk taker which are considered in investment sector and investment activity selection are not included in the case structure for the prototype. Other limitation of the prototype is, the explanation facility does not give response or feedback based on investor's questions. It gives only a direct explanation when the system recommends a solution. For instance, if one investor not clear about form of ownership, there is no a possibility to ask explanation.

1.6. Significance of the Study

From this study, primary Ethiopian investment office investment experts and investors, specifically the inexperienced investment expert professionals are the immediate beneficiaries to enhance their day to day activities. The prototype has great significance to teach primary investment experts and investors in order to have well understanding about investment sector and investment activity selection mechanism. As a result, those investment experts can use the system for recommending investment sectors and investment activity to new investors based on their personal or economical characteristics. The developed prototype system uses by investors itself when highly qualified investment expert professionals are unavailable. The developed prototype case based recommender system is used to give advising services for both foreign and local investors. The case based recommender system is developed using the

knowledge of investment domain experts, investors and documentary sources which is used as organizational memory. Therefore, it gives better recommendation services where highly qualified investment expert professionals are not found.

1.7. Organization of the thesis

The study is organized into six chapters.

Chapter one is the introduction part, which contains the background of the study, problem statement, objectives, scope and limitations of the study, the significance of the study and methodology to carry out the research.

Chapter Two discusses about Review of literature on the knowledge based systems, about its background, architecture, development phases, and knowledge based system overview and investment opportunity, and investment selection decision are presented.

Chapter three discusses the knowledge acquisition, representation and conceptual modeling procedures.

Chapter four deals with implementation of Case based recommender system.

Chapter five presents the results found in the evaluation and testing process of the prototype case based recommender system.

Finally, **chapter six** focuses on the conclusion and recommendation based on the finding of the research and recommendations are proposed for future research.

Chapter Two

2.0. Literature review

2.1. Recommendation system

Recommender system is a part of a web-based application that uses data about users and their behavior to provide them with items which are the most relevant to them. The recommended items are typically things that are answerable to user taste, like books, music, news, etc. With the increase of the Internet usage and the available huge data, recommendations became a part of life [9]. No matter what the domain is, a huge amount of information is online and it becomes a difficult task to select items that are necessary. Recommender systems try to overcome this challenge and aim to map people with the correct items.

More formally, recommender systems can be defined as systems which generate personal suggestions as an output or guide users individually to reach relevant and useful items among a lot of possible options. So it can be said that the recommender system is a mapping between users and items involving a value of interest [18].

According to [31], the two basic entities which appear in any Recommender System are the user / customer and the item / product. A user is a person who utilizes the Recommender System providing his opinion about various items and receives recommendations about new items from the system.

Recommendation systems have arisen to provide convenient suggestions to the users. These systems can be used for different purposes in several domains from offering papers to researchers to helping consumers in e-commerce. There are recommendation systems in different domains such as films, television programs, video, music, books, news, images, and web pages [20]. It can be said that, recommendation systems basically aim to overcome the difficulty of finding proper information. Among the most famous ones, Amazon recommends books in book domain; Last.fm helps users to find the songs that they want to listen; and MovieLens tries to guide users to reach the movies they might like.

Their principal objective of recommender systems is that of complexity reduction for the human being, sifting through very large sets of information and selecting those pieces that are relevant for the active

user [10]. Moreover, recommender systems apply personalization techniques, considering that different users have different preferences and different information needs, so the goal of Recommender Systems is to generate suggestions about new items or to predict the utility of a specific item for a particular user. In both cases the process is based on the input provided, which is related to the preferences of that user. For instance, supposing the domain of book recommendations, historians are supposedly more interested in medieval prose [20].

2.2. Architecture for Recommender System

As shown fig 2.1, the basic components of a recommender system make interaction to give recommendation for the user. First, a user profile learning module (explicitly or implicitly) captures the preferences from the user. Once the system “knows” about the user’s tastes and interests, it performs a recommendation algorithm that compares and/or combines user profiles and item descriptions. Then item with a maximum gain to the user is recommended [10].

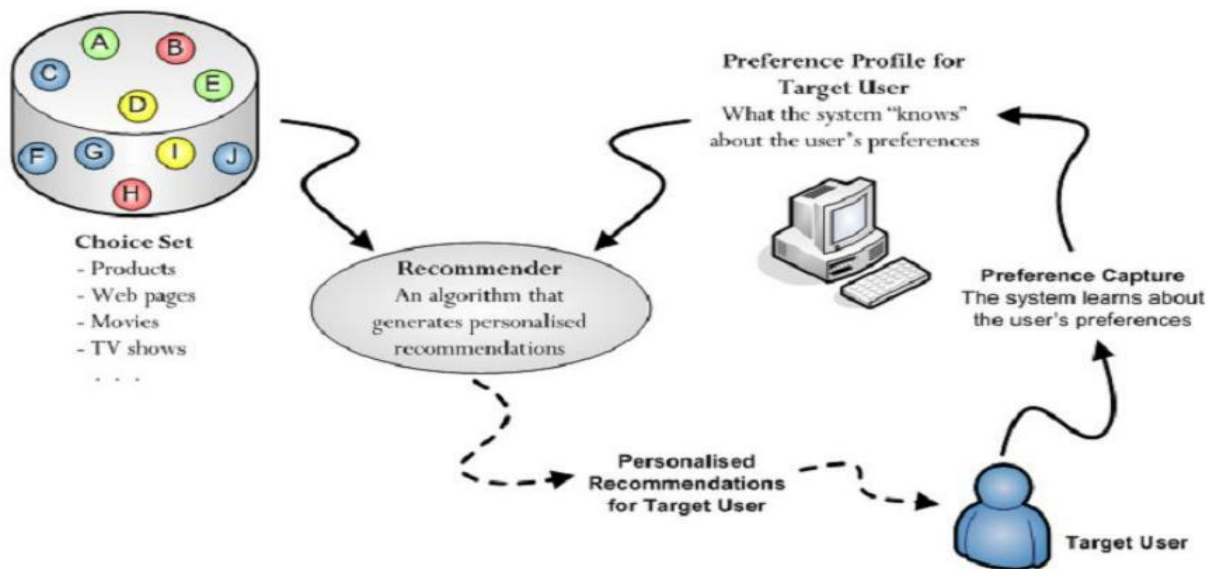


Figure 2.1: Architecture of recommender system (source: [10])

2.3. Types of Recommender System

Recommender systems can be classified into the following categories, based on how recommendations are made [9]:

- **Content-based recommender systems**, in which the user is recommended items similar to those the user preferred in the past.
- **Knowledge based recommender system** in which it depends either on explicit domain knowledge about the items or knowledge about the users to derive relevant recommendation.
- **Collaborative filtering systems**, in which the user is recommended items that people with similar tastes and preferences liked in the past.
- **Hybrid recommender system**: Due to the shortcomings proper of each of these strategies alone, combinations of content based and collaborative filtering have been investigated in the so called **hybrid recommender systems**.

Since this research concerned on knowledge based recommender system, the details of knowledge based recommender system are discussed as below:

2.3.1. Knowledge based recommender systems

Since our research is on knowledge based or case based recommender system the detail of all about knowledge based (case based) recommender system is discussed below.

2.3.1.1. Knowledge based system (case based reasoning system)

Knowledge Based System (KBS) is one of the major family members of the AI group. KBS can act as an expert on demand without wasting time, anytime and anywhere. KBS can save money by leveraging human expert, allowing users to function at higher level and promote consistency of work. One may consider the KBS as productive tool that have knowledge of more than one expert for long period of time. In fact a KBS is a computer based system which uses and generates knowledge from domain expert [32].

Human experts use their knowledge in particular field of expertise to solve day today activities. In the same way, knowledge based system handles problems; the computer needs an internal model of the world using the stored knowledge. All information is stored in such a way that it is readily accessible. To design knowledge based system, the expert knowledge was represented in a way that it supported for reasoning mechanism in computer languages. Representing knowledge into the expert system could offer potential advantages over human expertise. Because, knowledge based system can use the acquired knowledge permanently, consistently, easy to transfer and document expert knowledge [33].

2.3.1.1.1. Architecture of knowledge based system

Figure 2.2 below shows the building blocks of knowledge based system architecture adopted from [34].

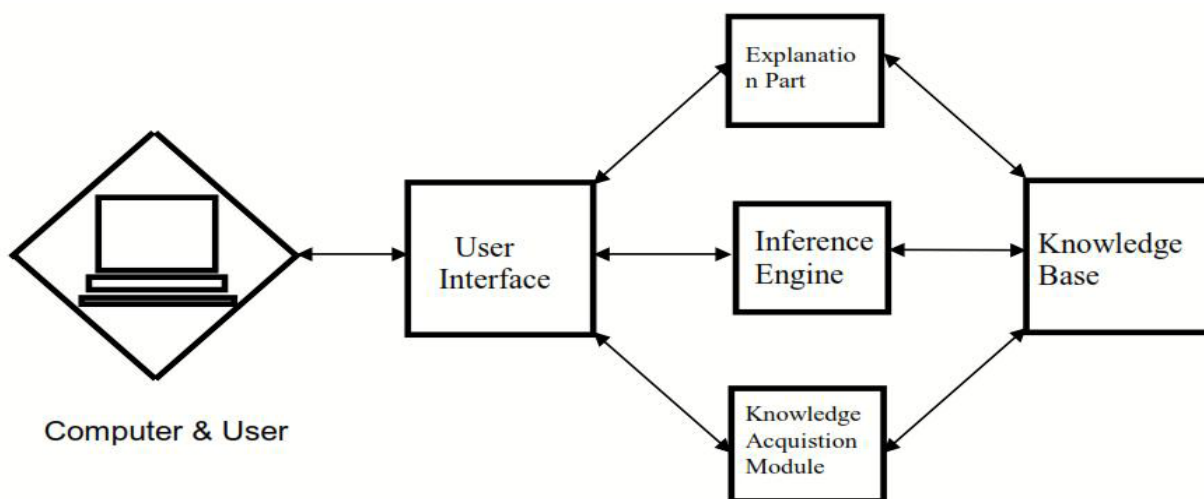


Figure 2.2: Architecture of knowledge based system

The architecture of knowledge based system consists of different components such as Knowledge Base, Knowledge acquisition module, inference engine, user interface and explanation module. The Knowledge Base contains all relevant knowledge acquired from domain experts. Knowledge also acquired from the user during their interaction with the system.

The knowledge acquisition module helps in the collection process of knowledge from the set of human experts as shown in Figure above. The inference engines formulate questions and assert the answers provided by the user in a natural language form. It provides a mechanism for conveying

recommendations to the end user. The explanation module provides a brief description to the user why the system arrived at a certain conclusion [35].

2.4. The Knowledge engineering process

The development of knowledge based system is the integration of many components. These are:

2.4.1. Knowledge acquisition

Knowledge acquisition is the process of acquiring relevant knowledge from human experts, books, documents, sensors, or computer files. The knowledge can be specific to the problem domain or to the problem-solving procedures, it can be general knowledge (e.g., knowledge about business) or it can be meta knowledge (knowledge about knowledge). Knowledge acquisition is the bottleneck in knowledge based system development today. Because, the trustworthiness and the performance of the knowledge based system mainly depends upon the acquired knowledge [36].

The knowledge acquisition process incorporates different methods such as interviews, questionnaires, record reviews and observation to acquire factual and explicit knowledge [37]. The performance of the expert systems depends upon the reliability, validity and accuracy of the elicited knowledge. The process of knowledge elicitation is affected by different contributing factors such as communication between the expert and ability of knowledge engineer. Therefore, effective elicitation techniques facilitate to acquire relevant knowledge form domain experts. The commonly used knowledge acquisition techniques are discusses as follows [37].

A. Interview

An interview technique is the process of interacting with domain expert on how they perform their tasks based on their expertise. Knowledge acquired through direct elicitation methods are procedural knowledge. Based on its structure, interview can be classified into structure, semi structure and unstructured interview [38].

- I. **Structured Interviews-** A structured interview method is questioning the domain expert directly. It is goal-oriented process. It forces organized communication between the knowledge engineer and the domain expert. The structure reduces the interpretation problems inherent in unstructured

interviews and allows the knowledge engineer to prevent the bias caused by the subjectivity of the domain expert [38].

- II. **A semi-structured interview** is an interview which has a guide that usually includes both closed-ended and open-ended questions. It is more flexible than structured one. In these kinds of interview the interviewer has a chance to change the order of questions and expand the dimension of questions based on the participants' responses [38].
- III. **Unstructured Interviews-** sessions are conducted informally, usually as a starting point. Unstructured interview techniques provide complete or well-organized descriptions of cognitive processes. There are many reasons that enforced to applying unstructured interview. Domain experts usually find it very difficult to express some of the most important elements of their knowledge. Through structured interview it is difficult to acquire the required knowledge. With good training and personal experience knowledge engineers can use unstructured interview to acquire relevant knowledge from domain expert [38].

Therefore, efficient and effective interview techniques largely depend on the ability of knowledge engineer to articulate their implicit knowledge. Because every interview is different in very specific ways and it is difficult to provide comprehensive guidelines for the entire interview process. Therefore, interpersonal communication and analytic skills of knowledge engineer is very important. On the other hand eliciting knowledge using indirect methods requires human intervention such as observation, document analysis, etc. [39].

B. Document Analysis

The final form of knowledge acquisition method is concerned with a detailed analysis of the existing document. This technique is used to collect relevant knowledge from the existed documents of different format. These documents include professional literature, brochures, manuals, guidelines, employee handbooks, reports, glossaries, course texts, and other relevant materials [40].

Knowledge elicitation methods can be classified into different types. Direct and indirect is the commonly known methods of knowledge elicitation. The way of classification depends upon how knowledge engineer directly obtains information from the domain expert [41].

A direct method involves directly questioning a domain expert on how they do their job. In order to implement direct methods successfully, the domain expert has to reasonably articulate and willing to share his/her knowledge. However, in case of indirect methods the required knowledge is not requested directly. Instead, the result of the knowledge elicitation session must be analyzed in order to extract the required knowledge. Indirect methods are thought to be more suitable when knowledge is not easily expressed by the domain expert [41].

2.4.2. Knowledge representation

Acquired knowledge is structured so that it was ready for use in the process of knowledge representation. This activity involves preparation of a knowledge map and encoding of knowledge in the knowledge base.

2.4.3. Knowledge validation

Knowledge validation (or verification) involves validating and verifying the content of knowledge (e.g., by using test cases and confusion matrix) and user acceptance. The testing result of knowledge based system was validated by domain expert.

2.4.4. Inference engine

This activity involves the design of software to enable the computer to make inferences based on the stored knowledge for the specific domain problem. In other word inference engine is a programs that reason over extensive knowledge bases.

2.4.5. Explanation

This step involves the design and programming of an explanation facility.

Explanation module is program that answered how the knowledge based system arrived at certain conclusion. Explanation module addresses the issues of system user interactivity.

2.5. Knowledge based reasoning techniques

There are a number of knowledge based reasoning methods. The well-known reasoning approaches are ontology based reasoning, semantic network, neural network, fuzzy logic, case based reasoning and rule based reasoning. For the purpose of this research work case based and rule based reasoning approach are discusses as follows.

2.5.1. Case based Reasoning

Case-based reasoning (CBR) means “adapting old solutions to meet new demands, using old cases to account for new situations, using old cases to evaluate new solutions, or reasoning from precedents to interpret a new situation” [42]. CBR is more comfortable to make better decision in dynamically changing environment. People learn from their success and wrong activities to handle similar situations in the right manner and not to repeat their mistake of the past. CBR approach is more compatible to reuse previously solved problems and learning from experiences for future decision [24]. Similarly, CBR is an approach to incremental learning. Once a problem has been solved, CBR approaches use the solution to solve for future problems [43].

A case-based reasoned will be presented with a problem, either by a user or by a program or system [44]. The case-based reasoner then searches its memory of past cases (called the case base) and attempts to find a case that has the same problem specification as the case under analysis. If the reasoner cannot find an identical case in its case base, it will attempt to find a case or multiple cases that most closely match the current case.

In situations where a previous identical case is retrieved, assuming that its solution was successful, it can be offered as a solution to the current problem. In the more likely situation that the case retrieved is not identical to the current case, an adaptation phase occurs. During adaptation, differences between the current and retrieved cases are first identified and then the solution associated with the case retrieved is modified, taking these differences into account. The solution returned in response to the current problem specification may then be tried with the appropriate domain setting.

The structure of a CBR system is usually devised in a manner that reflects separate stages. However, at the highest level of abstraction, a CBR system can be viewed as a black box that incorporates the reasoning mechanism and the following external facets [44]:

- The input specification or problem situation
- The output that defines a suggested solution to the problem
- The memory of past cases, the case base, that are referenced by the reasoning mechanism

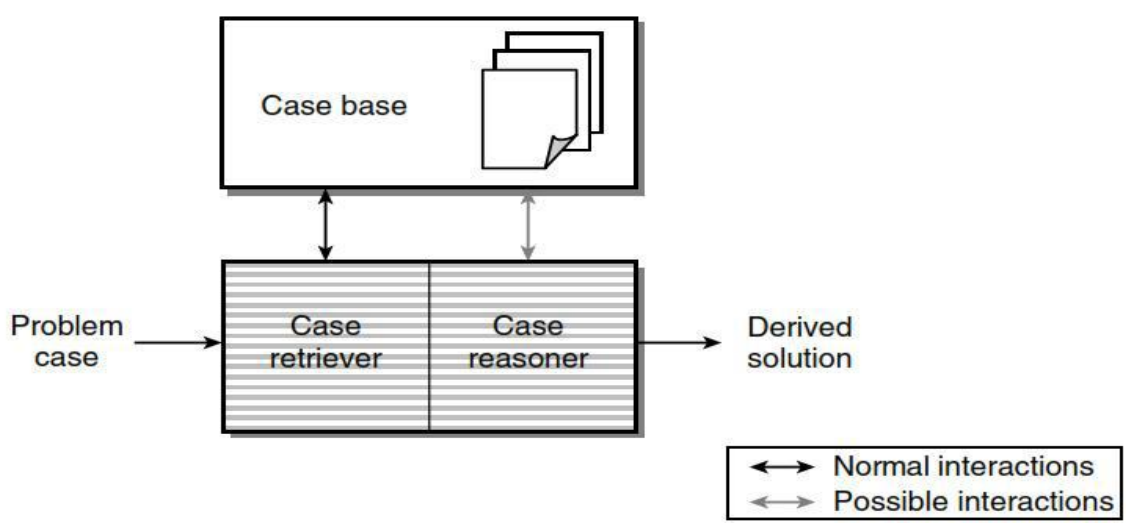


Figure 2.3: The Major Components of the CBR System [44].

In most CBR systems, the CBR mechanism, alternatively referred to as the problem solver or reasoner, has an internal structure divided into two major parts: the case retriever and the case reasoner [44]. The case retriever's task is to find the appropriate cases in the case base, while the case reasoner uses the cases retrieved to find a solution to the problem description given. This reasoning process generally involves both determining the differences between the cases retrieved and the current case, and modifying the solution to reflect these differences appropriately. The reasoning process may or may not involve retrieving additional cases or portions of cases from the case base.

2.5.1.1. Case based reasoning cycle

Case based reasoning life cycle incorporate four major components that make the reasoning mechanism successful. These are retrieval, reuse, revise and retain. Retrieval is the task that involves retrieving a

case from the collection of previously solved cases. The retrieved case is combined with the new case for later reuse into a solved case. Revise is a process that tests the success of a solution by applying into a real world environment, if repair is failed. When useful experience is retained the case is updated by a new learned case [43].

Cased based reasoning process generally involves both determining the differences between the retrieved cases and the current query case. It also involves modifying the retrieved solution to appropriately reflect these differences [44].

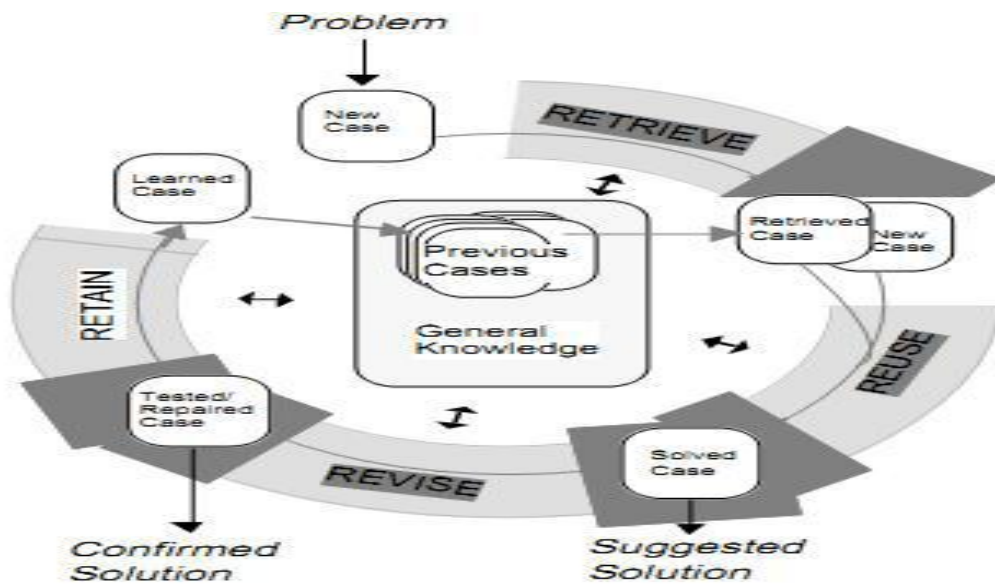


Figure 2.4: Case based reasoning cycle [44]

2.5.1.1.1. Retrieval

In CBR, retrieval is remembering previous cases stored in the case base to solve new problems at hand. The first step which is the most important step in CBR cycle is retrieval of previous cases that can be used to solve the target situation (new problem). Given a target problem, retrieve cases from memory those are relevant to solving it. A case consists of a problem description, its solution, and typically annotations about how the solution was derived [43]. Since retrieval is the first step in CBR, it affects the whole CBR system because others cycles are based on it. The selection of corresponding useful cases is then left to the CBR system which retrieves cases to be used for solving the problem by employing so called similarity measures. To retrieve relevant cases to the target problem, appropriate

similarity measurement should be used [43]. The first step is the retrieval of one or several cases considered to be useful to support solving the current problem.

To realize this retrieval task, CBR systems employ special similarity measures that allow the computation of the similarity between two problem descriptions. Because the interpretation of this similarity strongly depends on the particular domain, similarity measures are part of the general knowledge of the system. Different researchers describe the main tasks during retrieval of cases. For example [43] grouped case retrieval subtasks into three:

- **Identify features.** Involves indexing the problem with the most descriptive feature in order to match it with indexed matched cases. In other words, it identifies its descriptive properties and takes out the properties which don't describe the problem strongly.
- **Initially match.** Finding previous cases that match with the problem at hand and it retrieves a set of plausible candidates. That means it involves searching and similarity assessment to produce similar cases.
- **Select.** Selecting the best-matched case from the set of similar cases. It is based on the similarity assessment result that best matched case or set of cases is selected as output of the retrieval process.

The other prominent author in CBR [44] subdivides the retrieving process into:

- **Recall previous cases.** The main goal of this step is to retrieve "good" cases that have the potential to make relevant predictions about the new case.
- **Select the best subset.** The objective of this step is to select the best cases from the result of the first step. Sometimes it is appropriate to choose one best case; sometimes a small set is needed.

In CBR there are different case retrieval algorithms but the two most frequently used are nearest neighbor and induction case retrieval algorithms [44]. These algorithms can be used alone or in combination with each other.

I. Nearest Neighbor Retrieval Algorithm

Nearest-neighbor retrieval technique is to measure similarity between the source case and the case which we are searching [43]. The nearest neighbor algorithm measures the similarity of stored cases with a new input case, based on matching a weighted sum of features [44]. When a new case doesn't exactly match with old cases then this algorithm will return nearest match from CBR library. It is suitable when there are attributes that have numeric (continuous) value [42]. But the retrieval time of this algorithm increases linearly as the case in the case base increases.

The algorithm for nearest neighbor is as follows [24]:

-
- ✓ *For each feature in the input case*
 - ✓ *Find the corresponding feature in the stored case base*
 - ✓ *Compare the two values to each other and compute the degree of match*
 - ✓ *Multiply by a coefficient representing the importance of the feature to the match*
 - ✓ *Add the results to drive an average match score*
 - ✓ *This number represents the degree of match of the old case to the input.*
-

2.5.1.1.2. Reuse

After selecting one or several similar cases, the reuse step tries to apply the contained solution information to solve the new problem. Often a direct reuse of a retrieved solution is impossible due to differences between the current and the old problem situation. Then the retrieved solutions have to be modified in order to fit the new situation. How this adaptation is performed strongly depends on the particular application scenario [47].

In general, adaptation methods require additional general knowledge about the application domain. Because this leads to additional knowledge acquisition effort, many CBR systems used today do not perform case adaptation automatically, but leave this task to the user. Then, of course, the quality of the retrieval step influences the problem-solving capabilities of the entire CBR system primarily. Even if automatic adaptation is provided, the qualities of the retrieval result will strongly influence the efficiency of the system due to its impact on the required adaptation effort. After adapting the retrieved case automatically or manually to fit the current situation, a solved case is obtained containing a suggested solution for the current problem.

2.5.1.1.3. Revise

Depending on the employed adaptation procedure, the correctness of the suggested solution often cannot be guaranteed immediately. Then it becomes necessary to revise the solved case. How such a revision is performed, strongly depends on the particular application scenario. For example, it might be possible to apply the suggested solution in the real-world to see whether it works or not. However, often a direct

application of an uncertain solution is impossible due to the corresponding risks (e.g. medical diagnosis systems). Then the revision has to be performed manually by a human domain expert or by alternative methods such as computer simulation. Usually, the focus of the revise phase lays on the detection of errors or inconsistencies in the suggested solution and the initiation of further problem-solving attempts [43].

2.5.1.1.4. Retain

If the solved case has passed the revise step successfully, a tested/repaired case will be available representing a new experience that might be used to solve similar problems in the future. The task of the CBR cycle's last step is to retain this new case knowledge for future usage. Therefore, the new case may be added to the case base. In most cases, a general storage of all generated cases is not always useful. In order to enable better control of the retain process, various approaches for selecting cases to be retained have been developed [44]. These approaches often imply a reorganization of the entire case base when adding a new case, for example, by removing other cases.

Generally, the capability to acquire new case knowledge during a CBR system's lifetime principally adds these systems to the class of learning systems. Conversely, many CBR systems developed so far do not exploit this concept of the CBR cycle at all. This holds true especially for the commercially employed systems. Further, the original idea of the CBR cycle focuses on learning case knowledge.

2.6. CBR Techniques

CBR has different techniques [43]. Among these: Case Representation, Indexing, Storage, Retrieval and Adaptation are the commonly used techniques in any CBR research. Now, these techniques will be discussed briefly in the following subsections.

2.6.1. Case Representation

Case is a specific piece of knowledge representing an experience [43]. It contains the information which is content of case and situation where that information or experience can be used. Different type of data can be stored in a case. CBR community has a lack of consensus what information should be stored in a case. But, case should represent two things. The first one is functionality of information and the easiest

way in which information is obtained in the case. Whenever we want to choose representation format of a case, we have to keep in mind following factors [44].

- Language or shell in which we have to implement the CBR system. Selection of shell may reduce number of formats for case representation.
- Indexing and search mechanism. Format of case should be selected according to search mechanism. Case format should be able to interact with mechanism easily.
- Type or structure associated with content.

Cases also have to be structured for efficient case retrieval. There are two types of structures. Structure types are Common Structure and Hierarchical Structure. A memory model for the chosen form of case representation depends on following factors:

- The representation used in the case base.
- Number of features that are used to match cases during search.
- Number and complexity of cases.

Case is a combination of two components. Components are a description of a problem and its solution. Problem description consists of a set of attributes and values. Based on the description attributes values solutions are predicted.

2.6.2. Indexing

Index is a computational data structure can be held in memory and also can search quickly [43]. For example, databases use index to speed up retrieval of data. Information in a case can be two types:

1. Indexed information use for retrieval.
2. Unneeded information that may provide information to user but not used in retrieval.

Let us take the example of medical systems, where patient's age, sex, height and weight can be used as index features. That information is helpful for future retrieval. The patient's photograph can be included as an unneeded feature which can't be used in the retrieval. Picture should be helpful for doctor for remind patient. Index should have following features [43,44]:

- Be predictive.
- Should show the purpose for which case will be used.
- It should be easy to recognize it in the future.

- It must address the future use of the case base.

Indices can be selected in both ways: manually and automatically. If you choose index manually then you have to decide a purpose of case. Automated indexing has many methods including:

- In different-based indexing, index has to differentiate a case from another case.
- In the inductive learning method, features are identified which latterly uses as an index.
- Appropriate set of indices produce by similarity and explanation-based method for those cases which have same information.

2.6.3.Storage

One important aspect of the efficient CBR system is case storage [43]. It represents a logical view what is stored in case. For efficient retrieval, case base should be organized in a manageable way. These methods referred as a case-memory model. Two best models are Schank's and Kolodner's dynamic-memory model. 2nd is category-exemplar model of Porter and Bareiss.

2.6.4. Retrieval

Case retrieval is a process of finding cases which are closest to current case [44]. For efficient case retrieval, there should be selection criteria which will judge a case. CBR major research area is retrieval. There are four retrieval techniques in which all of them are discussed in the CBR cycle section.

2.6.5. Case Adaptation

It is a technique to alter retrieved case for reproducing new solution for new problem [52]. It may be the most important step which adds intelligence. Case adaptation improves overall problem solving ability of CBR. These three techniques are used mostly in CBR systems [52].

1. Structural Adaptation

In structural adaptation, formulas and rules are directly applied to the stored solution in CBR library. When case applies to these rules and formulas, the CBR system adapts this case and match with new problems.

2. Derivational Adaptation

It is a technique to reuse the rules and formulas to produce a new solution to a current problem. Solutions which are retrieved must be stored as additional case in the CBR library so it reproduces new solution to the new case.

3. Null Adaptation

It uses no adaptation at all. It just applies whatever solution is retrieved to the current problem without adapting it. Null adaptation is useful for problems which involve complex reasoning.

2.7. Advantage of case based reasoning

A case based reasoning approach has tremendous advantages in the development of knowledge based system. The following are main the advantages of case-based reasoning [52].

- Ability to express specialized knowledge.
- Naturalness of representation
- Modularity
- Easy to knowledge acquisition
- Self-updatability.
- Handling unexpected or missing values.
- Inference efficiency.

2.8. Disadvantage of case based reasoning

Even though case based reasoning approaches have a numbers advantage. But, due to lack of sufficient cases, the construction and inference mechanism of a case-based system loss the required objective. Some of the limitation issues in case-based reasoning are [52]:

- Inability to express general knowledge
- Knowledge acquisition problems
- Inference efficiency problems and Provision of explanations

2.9. Integrating Rule-based and Case-based Reasoning

Cased based reasoning uses partial matching to draw a conclusion. If some of the given problem descriptions match with a given case, then the case is applicable to the proposed solution. It also tries to handle novel problems by referring previously solved cases. Rule based reasoning uses perfect matching

to apply a rule for a given problem. It doesn't handle missing information and unexpected data values [24].

Rules are suitable to represent general knowledge, whereas cases are suitable for representing specific situations. Rules in a rule based system have the abilities to represent experiential knowledge acquired from experts in a direct fashion. Cases are capable of representing specific historical knowledge. The problem here is that it is difficult to acquire complete and perfect knowledge in a complex domain. Cases are natural and easy to obtain. They can be collected from the historical record, repair logs or other sources [52].

Therefore, the integrated reasoning approach makes use of both existing knowledge and the past experiences. This integrated approach eliminates the drawbacks of each method and provides a better way to handle problems, which combine both inductive and deductive approaches [55].

2.10. CBR System Performance Evaluation Methods

Evaluation of knowledge base system includes both system performance (statistical analysis) and user acceptance [53]. The statistical analysis for CBR can be conducted for both retrieval and reuse process. The first task of CBR is to retrieve cases that are relevant to the new case [43]. As retrieval task of the CBR aims to retrieve cases relevant cases from the case base, precision and recall are useful measures of retrieval performance in CBR [22]. Recall is defined as the ratio of the number of relevant cases returned to the total number of relevant cases for the new case in case base [22]. Whereas precision is the ratio of the number of relevant cases returned to the total number of cases for a give new case ([22].

$$Recall = \frac{\text{Number of relevant cases retrieved}}{\text{Total number of relevant cases}}$$

$$Precision = \frac{\text{Number of relevant cases retrieved}}{\text{Total number of cases retrieved}}$$

Knowledge based system evaluation process involves to determine the suitability and desirability of the prototype [55]. Effective knowledge based system evaluation process incorporates both technical and non-technical aspects. The technical aspects include exploring of the code, examining the correctness of

reasoning techniques, checking the efficiency and performance of the system and debugging errors in the early age of a system development. The non-technical aspect includes system compatible with users' satisfaction, the easiness of the system, the quality of the user interface and the acceptability of the system in the real world environments. As a result of this user acceptance is conducted to assess the applicability of the system for the real life [55].

2.11. Knowledge based system development tools

There are different types of tools that can be used for developing a CBR system. Most of these tools are commercial and few of them are non-commercial. The following CBR tools are the main tools used for knowledge based system as indicated on the paper of [29].

I. jCOLIBRI

jCOLIBRI is a technological evolution of COLIBRI and it is an object-oriented framework in Java which is designed for building CBR systems. It is a java-based and uses JavaBeans technology for case representation and automatically generation of user interface. This framework is developed by the GAIA artificial intelligence group in Complutense University in Madrid. The framework is built in two hierarchical levels- upper and lower. The lower level consists of library of classes (Software modules) for full 4REs CBR cycle, also for definition of cases, attributes and connectors for access to outer databases. The upper level is "black box" – graphical interface, which allows non-complicated user CBR application generation based on lower level's modules.

jCOLIBRI supports full CBR cycle. At the retrieve stage the nearest N cases are retrieved. At reuse stage several methods for adaptation are available (direct proportion and also in ontology). At revise stage methods for revision of cases are realized, as well methods for new indices generation and methods for decision making (preference elicitation). At retain stage there are methods for query retaining to the case base for future use. iCOLIBRI allows retrieval from clustered and indexed case bases and submits program interfaces (connectors) to access text and XML files, as well standard and descriptive logic databases. These interfaces can be used for diagnostic systems database access. There are lots of CBR applications, developed on jCOLIBRI based: additional shells (abstract levels) for distributed CBR systems, statistical CBR systems, multi-agent supervisor systems for text file classification, and lots of CBR recommender systems.

For this study I used jCOLIBRI CBR tool to develop the recommender system among the different CBR tools.

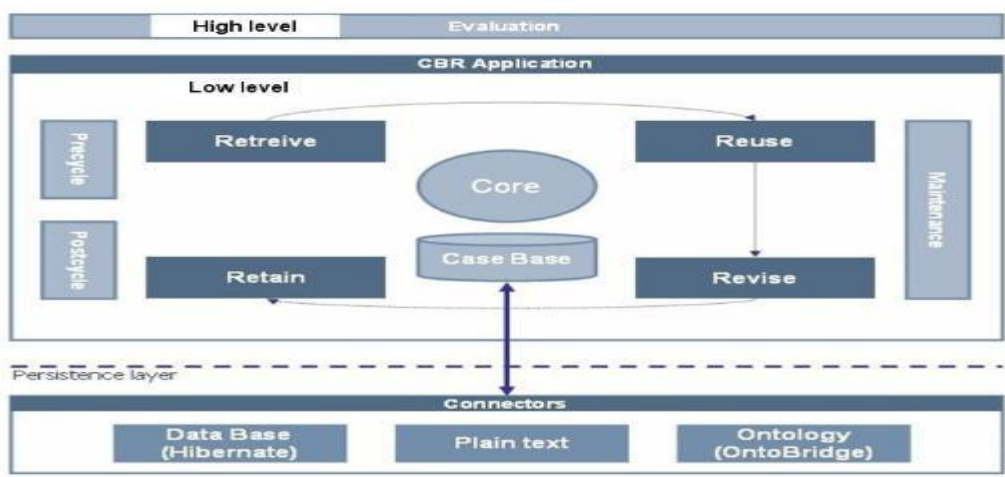


Figure 2.5: Software Architecture of jCOLIBRI [55]

II. myCBR

myCBR is one of the most popular CBR software platforms. It is a framework with certain capabilities and limitations. myCBR is developed by the German Research Center for Artificial Intelligence [55]. The platform has open source code written in Java and is accessible to all users. It can be easily modified by the users depending on the purpose. The purpose of myCBR is to minimize the efforts to create CBR applications.

The framework myCBR supports description of cases with various attributes: numeric, character, string, logical and class type. The templates of the cases are generated as classes or subclasses with a number of attributes, called slots. The CBR cases are objects of the class described by its attributes. Each attribute can participate in the class with its value and a weight that determines the significance of the attribute in relation to others. Attributes a weight of zero (0) is not considered when searching the case base database.

In myCBR the case and their attributes created manually or automatically. The automatic generation of attributes (slots) is done during the import procedure of the Comma Separated Value (CSV) file. Then to each column name of the CSV file is assigned an attribute with the same name to each row of the file the

new case (instance of the class) is created in the case base database. With regard to maintenance the CBR 4REs cycle phases, myCBR supports only Retrieve and Retain.

2.12. Investment in Ethiopia

Investment is an expenditure of capital by private individuals to establish a new business or to expand or upgrade a business that already exists. Legislation often seeks to provide incentives to promote private capital investment, especially by promoting participation of foreigners in the national economy.

Investment constitutes the backbone of an economy. The Ethiopian investment sector plays a vital role in the industrial development of the country. According to citation of [10] Industrial development was earlier believed to have occurred because of large enterprises. However, starting in the late 1970s and early 1980s, investment has become perceived as the key agent for industrialization. It is recognized that this sector provides not only employment opportunities to an increasing number of people in the country, but it is also an effective means of fighting poverty and income inequality. At the same time, investment serves as a training ground for emerging entrepreneurs. It is within this context that investment development became focal attention for governmental as well as nongovernmental organizations. This requires bringing the specific needs of the investors to the center of the policy-making process, and recognizing that investment are to be assisted not because they are small, but because of their capability to be efficient, innovative and able to compete in the local and international markets[10].

The Ethiopian Investment Agency (EIA) is a government agency established to promote, encourage and facilitate private investments in general and foreign investment in particular in Ethiopia. In Ethiopian investment agency there is lack of access to appropriate, relevant and understandable information and advice is one of the most important problems of small enterprises, in particular investments and micro enterprises [21]. This problem is encountered by EIA due to the fact that information system is not developed to enable proper collection; organization and dissemination in the country as a whole .There are also a few consultancy and advisory firms which are accessible to Ethiopian investment agency [21].

EIA operates as a one-stop-shop to enhance prompt services with the following functions:

- Providing pre and post investment services to investors;

- Collect, compile, analyze and disseminate information about investment opportunities in the country and advise upon request on the availability of partners for joint-ventures;
- Identify specific projects and invite interested investors to participate;
- Register and keep records of all technology transfer agreements relating to investments;
- Initiate, organize and participate in investment promotional activities such as exhibitions, conferences and seminars;
- Issue all legal permits including investment, work, residence and expatriate posts;
- Review, evaluate and forward policy recommendations to the concerned government body for approval;
- Perform such other functions that would enhance the attainment of its objectives.

2.12.1. Investment related infrastructure in Ethiopia

- | | |
|----------------------|------------------|
| ➤ Electricity Supply | ➤ Road transport |
| ➤ Telecommunications | ➤ Air Transport |
| ➤ Water supply | |

2.12.2. Areas of opportunity to invest in Ethiopia

Ethiopia's economy is still young with a vast untapped resources and a range of investment opportunities. The country has comparative advantages in textile and garments, agriculture, agro-processing, and leather and leather products. The areas, with most promising potential for investment in the country today are agriculture, agro-processing, textile and garment, leather and leather products, sugar, cement, chemical and pharmaceutical industry, tourism, mining and hydropower. The ongoing privatization program also offers enormous investment opportunities to private investors, particularly in the agricultural, manufacturing, and hotel and tourism sectors.

According to EIA the main areas of investment opportunities in the country consist of the following [21].

- | | |
|------------------|--------------------|
| 1. Agriculture: | 4. Mining |
| 2. Manufacturing | 5. Hydropower |
| 3. Tourism | 6. Social services |

2.12.3. Potential Areas for Agriculture and their suitable locations.

The estimated potential areas for the cultivation of the above mentioned agricultural products in all regional states of the country are presented in the following table.

No	Type of farming	Area (ha)	Region that the type of farming mostly grows
1	Rice	280,000	SNNP, Oromiya, Amhara, Benshangul Gumuz, and Somali
2	Maize	1,400,000	SNNP, Oromiya, Amhara, Benshangul Gumuz, Gambella and Somali
3	Horticulture	763,300	SNNP, Oromiya, Amhara and Dire Dawa
4	Coffee	426,000	SNNP, Oromiya, Amhara and Gambella
5	Tea	150,000	SNNP, Oromiya, Amhara and Gambella
6	Cotton	3,000,810	Tigray, SNNP, Oromiya, Amhara, Benshangul Gumuz, Gambella, Afar and Somali
7	Oil crops	1,601,323	Tigray, SNNP, Oromiya, Amhara, Benshangul Gumuz, Gambella, Afar and Somali
8	Pulse	3,274,469	Tigray, SNNP, Oromiya, Amhara, and Benshangul Gumuz
9	Rubber	200,000	SNNP and Gambella
10	Palm oil	450,000	SNNP, Oromiya and Gambella

Table 2.1: Potential Areas for Farming (Source: Ministry of Agriculture)

2.12.4. Institutional Framework

The Investment Proclamation of 2002, as amended in 2003, and the Regulations on Investment Incentives and Investment Areas Reserved for Domestic Investors of 2003, as amended in 2008, constitutes the main legal framework for both foreign and domestic investment in Ethiopia [21].

The EIA has restructured itself recently with a view to promoting more Foreign Direct Investment FDI and improving the services it renders to investors. According to [21] the major activities of the EIA and the one-stop-shop services it renders to investors are the following:

- promoting the country's investment opportunities and conditions to foreign and domestic investors;
- issuing investment permits, work permits, trade registration certificates and business licenses;

- registering technology transfer agreements and export oriented non-equity-based foreign enterprise collaborations with domestic investors;
- negotiating and, upon government approval, signing bilateral investment promotion and protection treaties with other countries;
- advising the Government on policy measures needed to create an attractive investment climate for investors; and
- Assisting investors in the acquisition of land, utilities, etc., and providing other pre and post-approval services to investors.

2.13. Related research work

Since there is no local research in our country on investment sector selection recommender system, the researcher review foreign researches on factors affecting investment decision and also review local researches related to case based recommender system. Some the related works conducted by foreign and local researchers in the knowledge based (case based) recommender system have been reviewed as follows.

There are a few works that attempted to predict investment choices by individuals. In [10] which used the 2003 data of Turkish Statistical Institution Budget Survey, factors affecting household's investment choices are predicted by using multinomial logit model. According to this work, household's investment choices are generally affected by age, educational background, gender, risk tolerance and income of household head. In another econometric work using the data of Turkish Statistical Institution 2004 and 2006 Household Budget Surveys by [10], demographic and socio-economic factors like age, house ownership, interest income, educational background and income are found related with investment choices.[10].

According to [27], factors that determine the investment decisions of individual investors. The primary goal of this research is to better understand the underlying factors that determine the investment decisions of individual investors in Bahrain. In this environment, understanding the basic factors that may underlie investors' differences in investment strategies may be imperative to a variety of stakeholders including the government, financial institutions and brokerage firms, business associations, and educational institutions. Understanding what determines the individual risk attitudes is a central aspect in this question. The author investigation on Five factors that

determine the investment decisions of individual investors namely Gender, age, income, educational level, and location of investment.

The [27] found that women take less investment risk and gender was the third most important determinant of investor style on investment decision (after age and income, educational level). Also the researcher found that young people are less risk averse than elder people in the same task context. So, young investors, unlike older investors, can adjust their current consumption downward and use some leisure time (work more) to compensate for losses in their portfolios. Younger investors also have more time to recover any lost value in an investment. The research finding also shows that investors having higher levels of education are more successful because university education provides them with knowledge and modern managerial skills, making them more conscious of the reality of the business world and thus in a position to use their learning capability to manage business. And also the author concludes that Selection of the place of investment location is one of the most vital factors for successful implementation of investment Project.

Biazen [28], has done research study on application of case based recommender system in field of study selection in the case of higher education in Ethiopia. The objective of the author is to develop a prototype case base recommender system that assists the students in their field of study selection process. The system provides recommendation to the students based on previously solved cases and new query given by the student. The author uses 105 cases which are collected from successful students as case base. These cases are used as an input for the system to provide recommendation. After accepting the input the system calculates similarity between existing case and new queries that are provided by the students and provides solution or recommendation by taking best cases to the new query. This recommendation enables students to make decision easily. In this study, the author used JCOLIBRI case base development tool to develop the prototype of case based recommender system because JCOLIBRI contains both user interface which enables students to enter their query and programming codes with the help of Java script language. After developing the prototype of the system, testing of the prototype for case base recommender system was done to evaluate the performance of the system. Based on user acceptance of prototype testing, the average performance of the system is 77.2% and 80.2% by the domain experts and students respectively.

Getachew [29], has done research study on application of case-based reasoning for anxiety disorder diagnosis. The main goal of this research is developing a prototype case-based reasoning system that can give decision support for anxiety disorder diagnosticians at a different level of expertise. Overcoming the limitations of a rule-based knowledge base system such as incremental learning and specific knowledge acquisition are the instigation of this research. For the implementation of the prototype, successfully solved cases are acquired from Amanuel Mental Specialized Hospital. In addition, the main parameters are identified in consultation with anxiety disorder experts. Then, the implementation of the prototype using jCOLIBRI case-based reasoning framework is realized. Finally, testing of the prototype case-based reasoning system is done to evaluate the performance of the system. The testing of the prototype is performed from two sides. The first one is testing in terms of precision and recall and registered 71% and 82% respectively. In addition to this, the average solution similarity using methods Leave One Out evaluator and Hold Out evaluation achieved performance of 73% and 75.5% respectively. The second one is the performance of the system is evaluated by the potential users" of the system and achieved 83.2% performance.

Halil Tunal, et al (2010), tried to discuss about course recommendation using data mining techniques called association rule. According to the author, students often need guidance in choosing adequate courses to complete their academic degrees. Course recommender systems have been suggested in the literature as a tool to help students make informed course selections. The main focus of the author is on the effectiveness of the incorporation of data mining in course recommendation. The system is based on the following collaborative filtering algorithms: user-based and item-based (discussed earlier). According to the author, the system can predict the usefulness of courses to a particular student based on other users' course ratings. To get accurate recommendations, one must evaluate as many courses as possible. Based on the evaluation results, the author suggests C4.5 as the best algorithm for course recommendation. The system cannot predict recommendations for students who have not taken any courses at the University.

Similarly, the proposed knowledge based system in this study is conducted to explore the applicability of case based system for investment sector selection. The main objective of the research study is to assign investors in different investment sector and investment activity based on personal characteristics, socio economic characteristics and form of ownership. Therefore, the proposed case based recommender systems can assist investment expert and investors during the investment advice process and to recommend the appropriate investment sectors.

CHAPTER THREE

3.0. KNOWLEDGE ACQUISITION AND MODELING

3.1. Knowledge Acquisition

In knowledge engineering, there are two most important steps during the development of knowledge-based systems that every knowledge engineer should consider. The first one is acquiring the required knowledge from domain experts, investors and relevant documents and the second one is representing the acquired knowledge with the appropriate knowledge representation method.

Knowledge acquisition is the process of acquiring relevant knowledge from domain experts and other sources of information such as books, databases, guidelines, manuals, journal articles, computer files, etc. Knowledge acquisition is the process of eliciting, structuring and representing (formalizing) domain knowledge acquired from different sources. Knowledge acquisition is the first step and time consuming task in the development of knowledge based system [46].

The process of knowledge acquisition of this research encompasses some basic activities such as gathering the needed knowledge, analyzing that knowledge, identifying important concepts (investment sectors and investment activities) and finally modeling them in using hierarchical structure.

3.1.1. Knowledge Acquisition from domain experts

Primary sources of knowledge are collected from domain experts in EIA. To gather the required knowledge semi-structured interview technique is used. Since one of the main focuses of this chapter is eliciting relevant tacit knowledge from the domain experts, twelve (12) domain experts from each investment sectors are selected using purposive sampling technique. As a result of this, investment officers from each investment sector have been interviewed to obtain the required knowledge on the domain area.

The domains of interview with expert covered issues such as how the expert interact with investors, what are the criteria consider to assign investors in different investment sector and investment

activity, which investment sectors are more risky and what are the possible investment sectors and investment activity recommended to the investors (see full interview questions in appendix D).

During the extensive discussion, the researcher tries to acquire the relevant tacit knowledge which is significant to generate the proposed case. In addition the domain experts are actively participated throughout the research work and they are consulted to confirm the correctness of the acquired knowledge. During face to face communication, the acquired knowledge from domain experts has been recorded manually by using pen and paper sheet.

Investment experts are devoted to providing investment promotion services in the institution. According to some domain experts, the investigation of investment application starts by collecting some relevant information such as Gender, location of investment, age ,amount of capital, type of investor i.e. whether domestic or foreign investor, form of ownership of the investor, interested investment areas, how many temporarily and permanent employees needed. But even if these investor's files are collected, mostly experts assign investors in different investment sector and investment activity based on the interest of investor.

All domain experts provided almost similar information about investment sector and investment activity selection advising system and they said since there is no investment guideline or criteria to assign or recommend investors in different investment activity, experts simply asks interested investment area of investors and then gives some explanation about the selected investment activity rather than use a guideline or criteria to recommend investors in suitable investment sector or investment activity. Since the main goal of this research is to identify the factors that affect in investment sector and investment activity selection, the researcher asking questions to domain experts about “what are the main attributes having effect on investment sector and investment activity selection ? ”. The researcher also asks the effect of some attributes that found from different secondary sources to get confirmation from domain experts whether it has effect or not. Based on the above questions domain expert said that considering some of the profile of investor and other attributes is important to determine the types of investment sector and investment activity to investors; because some investment sectors and investment activity have unique associations with geographical location, age factors, gender factor, form of ownership factor, type of investor factor, and amount of capital needed factor (see the detail in section 3.3.1.2).

Domain expert finally conclude that considering the above factors to assign investors in different investment sector and activity is a very important thing to decrease the failure of investors. Because investors mostly choose investment sectors based on the opinion of others or the interest of investment sector without considering the risk level of investment activity, without considering the time period of investment, without considering the location factor of investment activity etc. Hence, experienced and knowledgeable experts can identify the best investment sector and investment activity to each investor through considering different demographic characteristics of investors, socio-economic factors and other factors to make immediate decision.

3.1.2. Knowledge acquisition from investors

Primary sources of knowledge are also collected from investors that invest in different investment sectors. To gather the required knowledge semi-structured interview technique is used. Since one of the main focuses of interviewing with investors is to eliciting relevant tacit knowledge, four (4) investors were selected using random sampling technique.

The domains of interview with investors covered issues such as how to get advising systems from domain experts, what are the problems in advising systems of EIA and is there any experienced investment experts that gives a brief advice on how to invest and where to invest (see the interview question in appendix II). Based on the questions raised investors responded that there are not enough and experienced experts in Ethiopian investment agency office that are able give advice more suitable to each investor's. Also the advice system is different from expert to expert due to this reason investment experts mostly recommend investors on the familiar investment sectors and investment activity. Investors said that it must need well organized investment advising guidelines to investors. Difficulty of getting investment advice is a critical issue for investors since knowing the right investment area is a key factor and also knowing the right company to invest in is another factor to consider to new investors. Investors tend to lose money by choosing wrong investment sectors and investment activity.

3.1.3. Knowledge Acquisition from Relevant Document

Document analysis involves gathering knowledge from existing documentations. Hence, document analysis has been carried out to acquire explicit knowledge which is found in various secondary

sources of knowledge. In order to elicit knowledge for this research relevant documents which are related to investment sector and investment activity selection process have been review. The relevant documents used in this study are: Articles that are published in different journals, research papers, Manuals, Guideline and forms that are used in the process of investment sector and investment activity selection. Investment Websites related to investment sector and investment activity selection decisions especially Ethiopian investment agency website are also reviewed. As the result, relevant and technical knowledge were extracted and structured in a manner that suitable for knowledge modeling and finally knowledge representation.

The main data source (previous investor case base) used for developing CBRISAIAS system for this research is previous investor's cases found from EIA office. And also the list of investment sectors, list of investment activity; legal frameworks on the amount of capital needed for foreign investors, investment areas reserved for government, for domestic investor, for foreign investor have been collected from the manual of investment in Ethiopia.

The detail of this knowledge acquired from different sources that focus on investment sector and investment activity selection is discussed, structured and conceptually modeled in section 3.3.

3.2. Attribute selection using data mining tools

In this research, in order to solve the problem on the collected data set of investor cases, there are a series of activities that are undertaken. The major activities are discussed as follows.

3.2.1. Data Collection

Since the data to be used for this research has been collected from EIA, any field work to gather the data was not required. But, integration and preprocessing the dataset was held in order to suit it with the intended purpose. These datasets (previous investor's data) which were collected in year 2008, 2009, 2010, 2011 and 2012. Originally, the dataset consists of 19 attributes and 35,235 records, which include the relevant information concerning investors, and type of investment sector and investment activity.

3.2.2. Data Preparing and Cleaning

Data preprocessing helps to fill some missing values; to detect some outliers that may put at risk the result of data mining; to remove or correct some noisy data. In relation to this, data normalization, need to be performed. Moreover, to conduct the experimentation, the dataset must be prepared in the appropriate format. To do this the following activities are done.

The data originally were available in excel format. So, in order to use it in Weka it should be transformed to CSV or ARFF format. The researcher preprocesses the data for cleaning the data using Weka software. Then, attributes that have so many missing values can easily be detected and removed. Values that are not in Weka compatible form are modified. For instance, Weka doesn't handle values with space unless they are in single quotation. Therefore, since it is difficult to put all such values in single quotation, the researchers preferred to put them as one token by using underscores (_) to fill spaces.

The major problem of the original data set that needs data preprocessing are:

- Attributes have so many missing values
- In the original dataset there is an error on changing the attribute values from attribute to attribute. For instance some records have age value in gender attribute and gender value in age attribute.
- Another problem in the dataset is there is no common writing way of attribute values i.e. some attribute values are written in abbreviation format and some attribute values are written in stand form.
- The spellings of some attribute value are different in different places.
- The value of land size requirements are recorded in different unit of measurement like hectare, Sq.M, M2', based on these the researcher wasted more time to changes different units in to a common unit of measurement i.e. square meter (Sq.M) .

Therefore, the researcher takes more time to correct and normalize the above stated problem of the dataset step by step manually, using Ms excel and Weka tool.

Then, after going through the data cleaning, the data is saved as CSV file format in which the values are saved in comma delimited form in order to create an ARFF format file. Finally the data that were converted into ARFF file format has been used for the experimentation for attribute selection.

3.2.3. Attribute (Features) selection

Attribute selection can be used to investigate which (subsets of) attributes are the most important ones. In data mining task, one can get some attribute that has little or no impact on the overall investment sector and investment activity selection output. As mentioned above there are many attributes in the data that recorded in investor's dataset. As a result the researcher performs an attribute selection task using Weka attribute selection algorithm.

Attribute selection methods in Weka contain two parts: these are:

- **A search method:** such as ranking, best-first, forward selection, random, exhaustive, genetic algorithmic.
- **An evaluation method:** such as information gain, correlation-based, wrapper, chi-squared etc.

For this research, the researcher uses a search method of ranking and an evaluation method of information gain attribute selection. The reasons for selecting information gain evaluation method and ranking search methods are: information gain attribute evaluation evaluates the worth of an attribute by measuring the information gain with respect to the class. Information gain evaluator is used to select the best attribute at each node in the tree. Such a measure is referred to as an attribute selection measure or a measure of the goodness of split. The attribute with the highest information gain (or greatest entropy reduction) is chosen as the test attribute for the current node. This attribute minimizes the information needed to classify the samples in the resulting partitions and reflects the least randomness or "impurity" in these partitions.

The attribute with the highest information gain is considered as the most discriminating attribute of the set under consideration. So, an attribute that yields maximum information gain will be chosen for data set partitioning. Then, a node is created and labeled with the chosen attribute, branches are formed for each value of the attribute, and the samples are partitioned accordingly. The same criteria will then be applied to each split sample. The iterative divide and conquer process executes until no further split is required. Ranking search method is used to rank attributes by their

individual evaluation from highest information gain value to lowest information gain value. The highest information gain value (i.e. ranking first) is the most important attribute for investment sector and investment activity selection. Based on the experiment using information gain evaluation methods and ranking search methods the importance of attributes in ranking order are Gender, Age, Type of investor, form of ownership, interested investment activity, current investment activity, woreda, zone, investment sector, capital, region, land size requirement, and temporary employee, permanent employee .

After selecting attributes using attribute selection algorithm, the researcher were consulted with domain experts for the purpose of validating the selected attributes are important in investment sector and investment activity selection decision.

3.3. Conceptual Knowledge modeling

Once the required knowledge is acquired from pervious investor cases, investment experts and other relevant documents, the next step is modeling the knowledge. During the knowledge modeling phase, the acquired knowledge (elicited by various techniques) is represented in a knowledge model. A knowledge model is a structured representation of knowledge using symbols to represent pieces of knowledge and the relationships between them. Knowledge models include symbolic character based languages such as logic, diagrammatic representations such as networks and ladders, tabular representation such as matrices and structured text such as hypertext. The generation and modification cycle of a knowledge model is an essential part of the knowledge modeling phase. The model helps to ensure that all stakeholders in a proposed system understand the language and terminology being used and quickly conveys information for validation and modification where necessary.

During the knowledge acquisition stage, knowledge engineer collects both tacit and explicit knowledge. The knowledge engineer will try to understand both the tacit and the explicit part of the knowledge and then use simple visual diagrams to stimulate discussion amongst users and knowledge experts. Then knowledge engineer has to construct the conceptual model from what has been discussed during the knowledge acquisition stage. This communicates the knowledge to the knowledge engineer who transforms the model into workable computer programs or codes.

There are different conceptual modeling techniques and for this study hierarchical structure is used to model how investment sector and investment activity selection is performed. To make the acquired knowledge reasonable for knowledge representation it is modeled using hierarchical structures. The context of this hierarchical structure is used to demonstrate clearly the investment sector and investment activity selection which are implemented by using jCOLIBRI programming tool.

The model was built by the knowledge engineer after the core concepts are extracted from domain experts, investors and secondary source of data (document analysis). After acquiring the required knowledge, the knowledge engineer makes a discussion with domain experts to validate the knowledge that acquired from different source is correct for investment sector and investment activity selection. Mainly investment sector and investment activity selection is done in taking consideration of the attributes such as type of investor, form of ownership, Age, Gender and amount of capital needed to start investment. In this study, the conceptual modeling technique is used to show how investment sector and investment activity selection is held on.

Hierarchical structure was used to model the knowledge. The hierarchical structure as shown in figure 3.1 is derived from the knowledge acquired from in the consultations of experts, investors and secondary sources. These hierarchical structures are the base for the prototype knowledge based system development. The prototype knowledge base system is developed based on the model presented in these hierarchical structures. The prototype follows the procedures presented in the hierarchical structure to recommend investment sectors and investment activity. In the following hierarchical structure figure 3.1 the main factors of investment sector and activity selection to reach a better decision and the fundamental procedures during the investment sector and investment activity selection are structured.

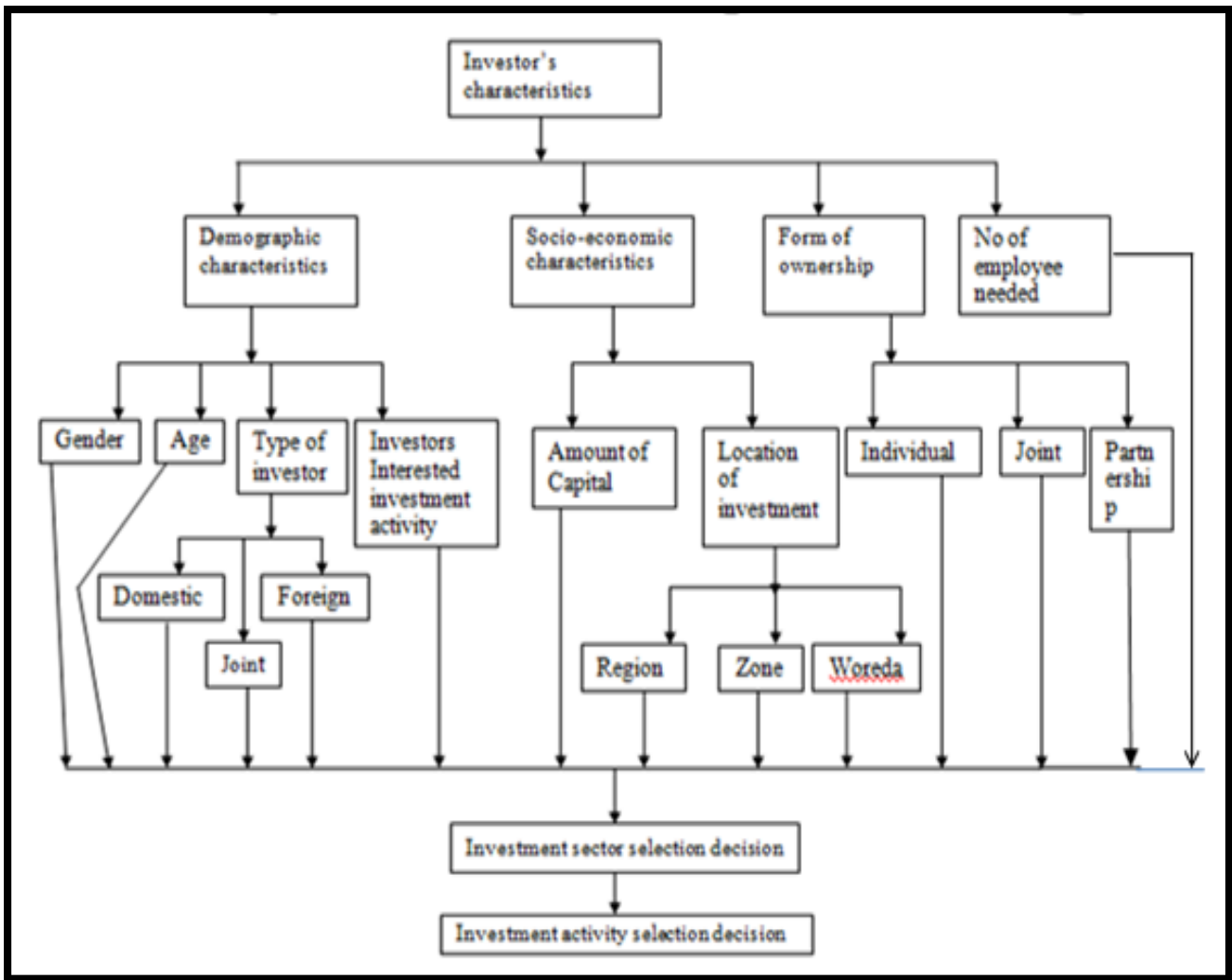


Figure 3.1: Hierarchical structure of investment sector and investment activity selection

3.3.1. Investment sector and investment activity selection

Investment is a commitment of funds, directly or indirectly, to one or more assets with the expectation to enhance future wealth. Direct investment may take in the forms of either physical assets or financial assets that are trade or non-trade in a financial market. Investors may hold non traded financial assets by investing their funds on bank products, such as saving accounts and time deposit. These types of investment are relatively less risky, can be sold more easily, and have a shorter investment period. However, investors may also choose to invest their funds in traded money market, investors with a long-term investment horizon may invest their money into capital

market instruments. These kinds of investment are riskier, but offer higher expected returns than that of money market instruments [49].

Mason [49], Provided important insights into how business angels evaluate investment opportunities. When they first come to selection of an investment sector or investment activity their first question is to consider how well it fits with their own personal characteristics, including age, gender nationality, location of investment, amount of capital, and level of risk tolerance in the sector. Level of risk tolerance mostly depends on the gender and age difference of investor.

According to [51] potential investors consider their goals, time horizon, financial stability, and risk tolerance when making investment selection. Risk tolerance could be highly correlated with an investors' likelihood of achieving his or her desired financial goals. If one considers the above situation, it would be in a better position to determine the types of investments best suited for your specific goals and objectives.

3.3.1.1. Risky investment sectors

According to different EIA investment promotion senior expert the major risky investment activities in the country are the following:

1. Industrial (manufacturing) investment sector

A major risk in this sector is the dependency of raw materials that import from other country. This is mainly due to raw materials that are either not found in Ethiopia (e.g. all metal and plastics must be imported), not easily accessible (e.g. due to poor roads), or are partly of poor quality. The heavy dependency of Ethiopian manufacturing on imported raw materials is a noticeable and continuing concern for operations and expansion. Another problem for the Ethiopian manufacturing industry is the cases of operating below capacity. Reasons for manufacturers working below full capacity are Shortage of raw materials, shortage of spare parts, shortage of foreign exchanges, problem with electricity and water, repeated breaking of machinery, outdated of technology, and lack of skilled manpower's.

2. Mining investment sector

Mining is one of the world's most dangerous occupations. Over the years, many serious accidents have occurred in various parts of the world, often with significant loss of life. Miners face many dangers on the job. The first is cave-ins or mine collapses; occur when the walls and ceilings of underground mineshafts have not been properly secured. And the second risk are since Mining involves the use of many toxic and dangerous chemicals. Chemicals are frequently used to transform the ores from their natural state into usable commodities. Accidents occur when the chemicals are not securely stored. Miners working with these chemicals need to have adequate ventilation to prevent the risk of inhaling dangerous fumes. And finally the risk of mining is there is the probability of absent of minerals in the targeted earth (area) after the process of mining or digging. Due to this it needs testing of mining materials for safety and effectiveness, but the testing materials are costly and not easily accessible.

3. Milk and dairy processing

Milk and dairy products have some unique characteristics that may affect the use of the emerging hedging mechanisms. Milk production is produced day in and day out. Therefore, milk is a flow product. By extension, dairy products such as commodity cheese, butter, and nonfat dry milk are also flow products. The major risks in this sector is for regulating fresh dairy product flow to the market or distribute to the customer in daily rather than for storing for later sales. Milk and dairy products are easily perishable or contaminated if stored for a week due to lack of customer, high temperature, decrease of price and lack of refrigerator. And also it is difficult to instantaneously change the volume of milk production in response to changes in current market milk price because if someone increases the volume of the milk and there are no customers in that day the milk will be contaminated or outdated even with refrigeration or pasteurization.

So unlike other storable commodities such as grains, corn, soybeans, or cotton, dry milk and commodity cheese as flow products may not be suitably stored for later to sell when price increases.

4. Agricultural sectors

Agribusiness investment in Ethiopia depends on vagaries of the environment and nature. Agricultural risk occurs because agribusiness enterprise is affected by many uncontrollable events that are often related to weather, drought, physical hazard to the factory site and technological failure of the firm. In addition to lacking adequate infrastructure for irrigation and transportation, the negative effects of climate change on crop production are expected to be pronounced in Ethiopia due to dependency on rainfed agriculture. This risk affects the efficient conversion of input to output.

For export-oriented agribusinesses, storage facilities, railroads and ports are crucial. In addition to these the perish ability of agricultural products requires special infrastructure such as cold storage and refrigerated transport. For instance flower exporting needs cold storage and refrigerated transport to export from Ethiopia to other countries in order to protect the dryness or perish-ability of flowers. But the refrigerated transport service is costly and difficult to easily access at anytime and anywhere.

5. Information and communication technology (ICT)

ICT is the technology that is not highly adopted in Ethiopia. The major risks in ICT investment activity were since there is no skilled manpower that operates or used the technology; the customers may not be motivated to use the technology. As a result the numbers of customers are not sufficient. On the other hand since most of the technology is import from other country it is costly and difficult to customers to buy and use this technology. Technology needs adoption to customers because customers are not familiar with the technology. The other risk of ICT is may be outdated or useless due to the creation of new technology.

3.3.1.2. Factors that Influences investment sector or investment activity selection decisions

Each and every person can be specifically differentiated on various parameters in investment selection decision. Prior to purchasing in any investment product or service, it is important that you fully think about your unique needs and overall financial situation in order to determine whether the

investment sector and investment activity is right for you. Financial products and services offer many different risks and benefits, and before you invest you should understand the features of each investment sector and investment activity. Some of the key issues that you should be aware of as is the gender and age difference in the investment's level of risk taker, the period of time you plan to investment , the accessibility of different infrastructures in the investment location, your liquidity needs and the fees and costs associated with the investment. You should also fully understand your capital needs and the overall risk that you can afford to bear with the investment [10].

To make effective investment decision, investor needs to select the right investment sector and investment activity among different alternatives at the right time. In order to choose investment sectors and investment activity, investor has to evaluate alternative investments and specify criteria to minimize those alternatives and rank the lifted ones [10].

Investors have to make a lot of decisions. For example, an investor has to decide which investments to select; how much money to put in each investment and the best time to be purchasing that investment and where the location of investments is should be. Moreover, some types of investments may not be appropriate or suitable or even accessible for all or some individual investors [3].

Since investors can differ in their investment decisions in many ways. So the major factors that influence investment sector and activity selection decisions are on the investor's demographic characteristics such as age, and gender, type of investor(nationality) , form of ownership and socio-economic factors such as capital, location of investment or environment. The detail of each factor is discussed below:

3.3.1.2.1. Age

The first factor you should consider to determine where to put your investment and how much to invest is your age. Your life stage and circumstances are likely to affect what types of investments you choose. As you change life stages and circumstances, you are likely to want to change the proportion of different investments. As a general rule of thumb, the younger and middle age investor (age below 50) you are and/or the more time you have to reach your financial goal, the more investment risk you can afford to take. If you are young, you might consider that you can take on more risk when making longer term investment. But if you are closer to retirement (like

ages above 50s) you are likely to be more careful about taking risk because you want to protect your savings and investments in order to have money to survive when you retire [3].

Another advantage of being young is, it is not very important to put in a lot of money for investment if you have very long term goal. On the other extreme, if you are older aged and thinking about retirement you are just starting to save for retirement, you should invest the maximum amount you can afford and you should also put your money in a relatively safe investment. As a result you can live comfortably when you retire [50]. According to [50] Older workers invest more conservatively than younger workers and invest in short time period investment activity than long time period investment activity and also conclude that risk aversion increases with age such that older people are more risk averse than younger people.

In investment sector and investment activity selection decision there is significant relationship between age and the time periods of investment made by the people because the time duration of investment can vary from a few hours to few months or even several years. The two main classes of investments on the basis of period of time are Short-term Investment and Long- Term Investment. Investments made for a period of one to three years are termed as short-term investments and that are invested for more than three years are termed as long-term investments [10].

Short term investment has its liquidity. Liquidity ratios measure ability of the firm to meet its short-term (less than a year) obligations and reveal short-term financial strength and weakness [10]. Liquidity is the ability to convert an investment into cash quickly. Some investments are less liquid and may not be easily converted into cash when you need it. So as you are older investor it may be important that you have access to your investments and be able to convert them into cash quickly in case of emergency or need for additional income. Older investors are preferable to invest in higher liquidity investment sectors because high liquidity investment has higher margin of safety and ability to meet its short-term obligations. Short term investments are usually considered to be less risky in comparison to long term investments based on this most older investors are invest in short term time horizon investment activities [50].

For instance from pervious investors data from EIA male and female investors under the age group Less than 50 have preferred long-term investments. On the contrary older investors (i.e. age above 50) typically are more interested in preservation of capital and have a shorter-term investment horizon.

But when investors are invest in partnership or in joint its gender and age attribute value is “Plc” instead of male, female or exact age value of investor as shown in fig 3.2 and 3.3. **Plc** is a business arrangement in which two or more domestic or international enterprises agree to pool their resources for the purpose of accomplishing a specific task. This task can be a new project or any other business activity. In a plc, each of the participants is responsible for profits, losses and costs associated with it. A “plc” partnership is may be a group of men’s and women’s investor, young and older investors and also having different skills and experience. Because of these plc investors has the capacity to take risks and participate in all investment sectors and investment activity. If a partnership (plc) business loses money, the losses are divided among the partners in the same way as profits would have been, which is less risky.

Therefore, based on the time period of investment, long term investment sectors are the following:

- Manufacturing (industry)
- Textile and leather production
- Construction
- Sport center (such as athletics, football etc)
- Mining

On the other hand, based on the time period of investment, short term investment sectors are the following:

- | | |
|---|-----------------------------------|
| • Real state | • agriculture |
| • Construction machinery rental service | • Education services(such as KG, |
| • Hotels, tourism, loge and restaurant | elementary, secondary, high and |
| • Tour operation and travel agent | preparatory school , colleges and |
| • health and social work | university, and training center) |

The figure below shows the category of age value of investors and its effect on investment sector and investment activity selection.

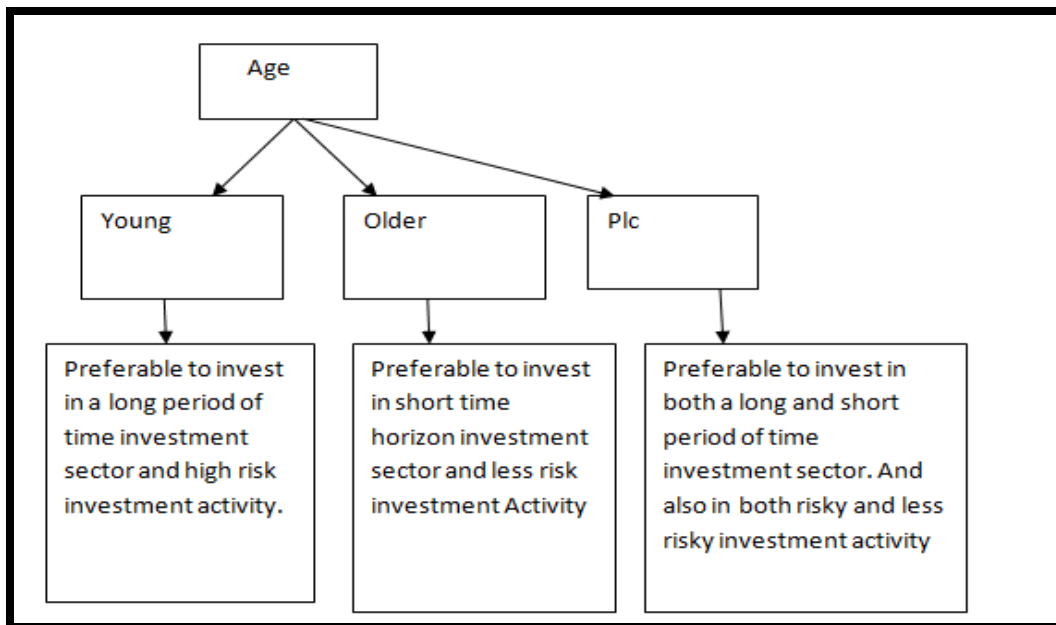


Figure 3.2: Age factors on investment activity selection

Figure 3.2 shows the age factors in different category of either young, old or plc. The reason for using the attribute value of age and gender as “plc” as shown in figure 3.2 and 3.3 is because when investors are invest in partnership it is a group of two or more investors, so it is difficult to assign the age and gender value of all investors. As a result of this, partnership or joint investors registered their age and gender value as plc. But when “form of ownership” of investors is “individual”, investors register the exact value of their age and gender. Generally when investor’s “form of ownership” is either “partnership” or “joint”, the attribute value of gender or age is always “plc” as shown in figure 3.7 and 3.8 instead of male or female or value of exact age.

3.3.1.2.2. Gender

The differences between men and women investors on investment decision are focusing on two factors that are relevant in the labor market: Risk taking, and reaction to competition. Women prefer jobs that are less risky and less competitive than men’s, and then this could explain part of the gender differences in the labor market or investment.

When it comes to investing, men are focused on much higher riskier investment sectors whereas women will never go for riskier investments sectors and they would rather spread their funds over buying several lower risk level investments. And also men are more often perceived to be

competent, confident, and independent while women are more often perceived to be concerned about the feeling of others [56].

Majority of male investors put their money in capital market, while majority of female investors invest their funds in banking industry. In terms of risk behavior, both male and female investors tend to be risk averters. However, the proportion of risk seeker is higher for male investors than for female investors. This indicates that male investors tend to be more risk tolerant than do female investors. In addition, women seem to be perceived as more conservative investors and are offered less risky investments by brokers [56].

Women with low risk tolerance are less likely to save in the short term as well as to save regularly, and unwilling to take a chance on losing any of their income by investing in risky assets. This was particularly important for women with no retirement saving plan as well as those with a defined contribution retirement plan. Women with low risk tolerance may be less likely to save, and when they do save, are less likely to choose assets that have greater growth over time, leaving them financially unprepared for retirement. But when investors are investing in partnership or in joint, it is a collection of both women's and men's investors. As a result investors are participating in both risky and less risky investors sectors and investment activity. If investment owner is in partnership or in joint form the gender and age is register as plc [10]. The figure 3.3 shows the category of gender value and level of risk taker in the age difference.

For instance previous investor's data that found from EIA shows that the majority of investment sectors that women investors participate are:

- Real state ,renting and business activity sector (such as construction machinery rental , and construction machinery leasing)
- Been keeping and honey production; animal diary and fattening
- Education sectors(such as kg, primary school, secondary school and high school)
- Hotel and restaurant sectors(such as apartment, hotels and tourism)
- Transport, storage and communication sector (such as tour operation , travel agent and car rental)
- Textiles production activity such as protective clothing, batiks, jerseys, bed linen, corporate wear, T-shirts, bridal wear, as well as baby wear; launder service, bread and pasta production and Processed foods sector such as syrups, juices, peanut butter and cakes, dried

vegetables). The following figure shows the category of investor's gender value. The reason for using plc is discussed in 3.3.1.2.1.

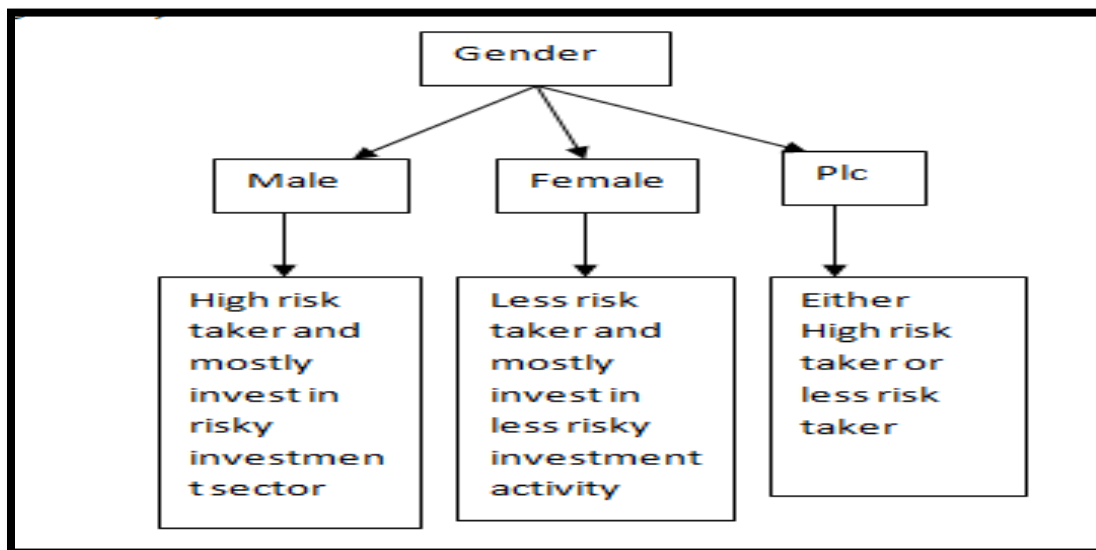


Figure 3.3: Gender factor on investment activity selection

3.3.1.2.3. Capital

A common fatal mistake for many failed businesses is having insufficient operating capital. Business owners underestimate how much money is needed and they are forced to close before they even have had a fair chance to succeed. A considerable number of people have unrealistic expectations when it comes to the funds needed to start a business. They often lack the necessary start-up funds and can't come up with adequate financing [49].

It is imperative to ascertain how much money your business will require; not only the costs of starting, but the costs of staying in business. It is important to take into consideration that many businesses take a year or two to get going. This means you will need enough capital to cover all costs until sales can eventually pay for these costs [49].

In financial markets, not all investors feel equally competent in making investment decisions. In general, investors with less annual capital feel less competent as an investor with higher annual capital. A person with higher capital feels more successful and more powerful in daily life. This feeling can carry over to the domain of financial decision-making [3].

The minimum capital required to start investment in Ethiopia by foreign investor is US\$ 200, 000 per project. However the minimum entry capital required by foreign investor investing in areas of engineering, architectural or other technical consultancy services, accounting and audit services, project studies or business and management consultancy services is US\$ 100,000 per project. While in joint venture the minimum capital contributed by the foreign partner(s) must not be less than US\$ 150,000. However, the amount capital required for joint investors in the investment areas of engineering, architectural or other technical consultancy services, accounting and audit services, project studies or business and management consultancy services or publishing must not be less than US\$ 50,000. But for domestic investors there is no restriction of amount of capitals needed for starting investment in any investment activity [21].

Investors having low amount of capital, invest more conservatively than investors having high amount of capitals. This suggests higher liquidity needs among workers with lower capitals [49]. So low capital investors are tend to be risk averters, while high capital investors tend to be risk seekers. This implies that income may affect investor's risk behavior. The investors that have high annual incomes are marginally preferred long term investments than short term investments [49]. The figure below shows the amount of capital required by investors to start investing:

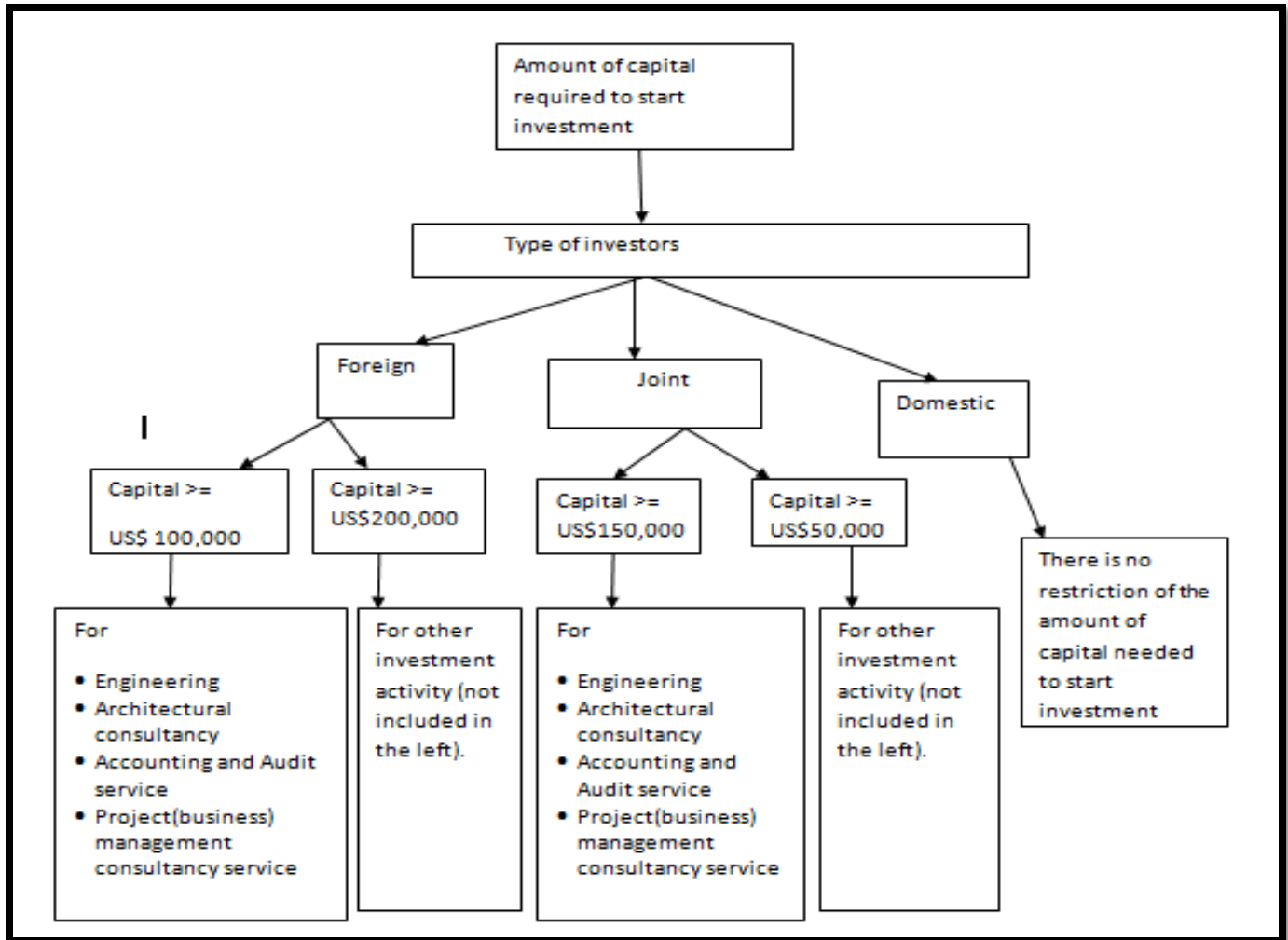


Figure 3.4: capital factors on investment activity selection

3.3.1.2.4. Investment Location

An investment project location is also critical for the success of a business. A good location may facilitate a struggling business to ultimately grow, whereas an investment project situated in a poor location will be at a disadvantage. Some factors to consider when establishing your investment location is where the targeted customers live, the traffic, accessibility, and parking, the physical distance from competitors, access of infrastructure and the condition and safety of a building [8].

Infrastructure and facilities (such as transportation infrastructure, communications, buildings, water and power supply) is a key driver in strengthening the national economy and enhancing Ethiopian

productivity. Infrastructure aids economic development of a state by creating access to regional and national markets. Infrastructure availability promotes both foreign and domestic investors because infrastructure growths are associated with greater accessibility and reduction in transportation costs and maximization of profit [21].

The level of infrastructure development in an economy influences the cost and efficiency of business operations. The limited availability and poor quality of roads and bridges in most Ethiopian regions and towns have been major problems to effective transportation of products from the rural areas to various markets, resulting in high post-harvest losses and rendering investments less profitable. So Poor infrastructure causes increase in transaction cost and limits access to both local and global markets which ultimately discourages both domestic and foreign investors in developing countries [36].

Throughout the literature, access to infrastructure has been shown to attract investors. Available land has also been cited as an important location factor [8]. Manufacturing firms concerned with land availability tend to avoid cities and densely populated locations. Specially manufacturing investment sectors needed efficient infrastructures such as electricity facility, road facility, communication network, and water facility and also needed to consider availability of customers, availability of employees and free from disasters [8]. So in Ethiopia the major industrial zones (locations) that are given for investment manufacturing sectors are Akaki, Kaliti, Nazareth, D/zeit, Dukem, Sebeta, Suilulta, Nifas silk, Adama , Mekele, Dire Dawa, Combolicha, Hawassa and Bahir Dar (from personal interview of Ato Girum Tadesse).

According to [36] Selection of the place of factory location is one of the most vital factors for successful implementation of investment Project. While selecting the location of an investment, the following considerations need be taken into account [36]:

- Transport of raw materials and finished goods is a very important factor and these should be taken into account. But more availability of transport facilities is not enough, the freight charges, conditions of roads and easy availability of service facilities are also important. Hence, to save in the cost of transport and quickness in delivery, a factory should be at such places where raw materials are available. Wider access to the market will create both new business opportunities and increased competition, leading to further increases in profitability. For

example, services in more remote areas of a region that were previously sheltered by distance may be exposed to competition with larger and more efficient entities that are centrally located.

- In addition, to locate a factory where other forms of communications are also available will give an added advantage such as railway postal and telegraphic services, etc.
- Manpower availability is another important factor for consideration.
- If the finished goods of the industry have local consumers' demand and whether it is likely to occupy the demand market in the near future, as well .

Locations of investment project include regions, zone, and woreda or city as shown in fig 3.5 below:

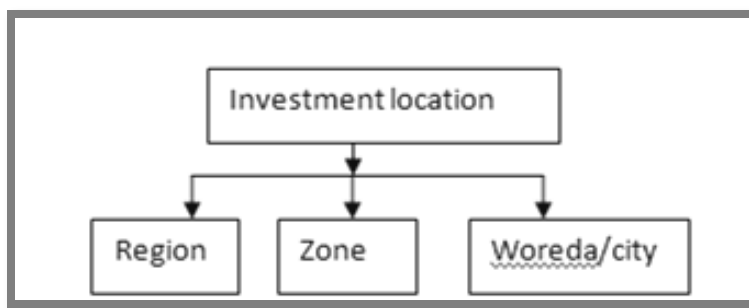


Figure 3.5: Location of investment

3.3.1.2.4.1. **Investment activity in Ethiopian different regions**

In Ethiopia there are nine regions (location of investment) such as Addis Ababa, Afar, Amhara, B.Gumze, Dire Dawa, Gambella, Harrire, Oromia, Somali, SNNPR, and Tigray that are open for both domestic and foreign investors in different investment sectors. For instance the three main coffee growing regions in Ethiopia are: Harar, SNNPR, and Oromia. The country has more genetic diversity among its coffee varieties other than any other countries. And also SNNPRS and Oromia is the most important Apiculture (honey and bees wax) producing region in Ethiopia.

Based on the previous investor's case, the characteristics and the major investment activities involve in each region are discussed below in detail:

I. Addis Ababa

Addis Ababa is a city administration and has 10 sub cities. Addis Ababa is one of the major industrial zone in the country and the majority of investment sectors in Addis Ababa are industry (i.e. manufacturing), hotel and tourism, commercial real estate, whole seal and trade, renting and

business activity, educational service, health and social services, tour operation and travel agent (see list of investment activity in each investment sector in 3.3.1.6).

II. Dire Dawa

Dire Dawa is the second city administration in Ethiopia. The major investment sectors in Dire Dawa are in the areas of agriculture (such as lives stock, animal fattening and coffee processing), construction, business center, hotel and tourism, educational service, health and social services.

III. The Oromia Regional State

The State of Oromia has 12 administrative zones and 180 woredas. The major investment sectors in Oromia are Agriculture, hotels and restaurants, manufacturing, Health and social work, renting estate and business activities, wholesale and trade.

IV. The Regional State of Amhara

The State of Amhara consists of 11 administrative zones and 105 woredas. The major investment sectors in Amhara region are in the areas of agriculture, fish farm, hotel and tourism, real estate, educational service, health and social work services.

V. The Tigray Regional State

The State of Tigray consists of 4 administrative zones, one special zone, and 35 woredas . The major investment sectors in this region are on agriculture, hotel and tourism, social services, mining, construction and transport.

VI. The Regional State SNNPR

The State of SNNPR is administratively divided in to 9 zones and 77 woredas. The major investment sectors in these regions are agriculture, hotel and tourism, social services, construction and textile factory.

VII. The Regional State of Somalia

The State of Somalia has 9 administrative zones and 49 woredas. The state is very rich in livestock. Moreover, it is endowed with natural gum, natural salt (in Afdem zone), natural gas oil has high potential for investment. The Major investment sectors are in agriculture, in education service and hotels and tourism.

VIII. The Regional State of Harari

Harari has no administrative zones or woredas. The total number of kebeles of the region is 19. The major investment sectors are construction, construction machinery rental, dairy farming, real estate and hotel and tourisms.

IX. The Regional State of Gambella

The State of Gambella is composed of two administrative zones and eight woredas . Gambella has potential investment opportunity in the production of cotton, groundnut, sesame and other oil seeds. Fishing, mining gold and exploring petroleum, mineral water and construction materials are other important areas of investment in the state. So the major investment sectors in this region are agriculture and real estate, renting and business activity.

X. The Regional State of Benishangul Gumuz

The State of Benishangul Gumuz has 3 administrative zones and 19 woredas. Agriculture is the sector, which attracted majority of investors in this region. Farming and cattle breeding are predominant investment activity in the State.

XI. The Regional State of Afar

The State of Afar consists of 5 administrative zones and 29 woredas. Agriculture such as production of maize, beans, sorghum, cotton, papaya, banana, and orange is the main investment activity. And also business and commercial activity, especially salt production, is another area of investment.

3.3.1.2.5. Number of employee

Employee is the largest expenditure for many investment projects. An investment project should employ an accurate amount of workers with the necessary skills and credibility for their given field. If the investment project is overstaffed with many unproductive employees, they would be wasting money for unnecessary labor costs. If there are too few employees at hand, then the workload would be overwhelming for the reliable staff to handle, and as a result, the investment project overall performance may suffer. An investor should be aware of this employment challenge and create some sort of equilibrium in the workforce by delegating tasks to competent employees and to avoid hiring those who are not experienced. The table below shows the type of investors in investment project.

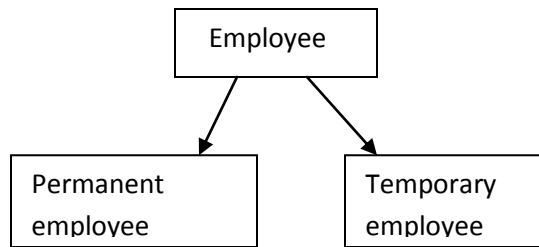


Figure 3.6: Type of employee

3.3.1.2.6. Types of investors

The Ethiopian investment regime identifies three types of investors:

I. Domestic investors

Domestic investors are investors that have Ethiopian nationality or live in Ethiopia and participate in different investment economic activity carried out with their own capitals or joint with the foreigner's capital. For domestic investor, there is no restriction of amount of capital needed to start investment and possible to invest in all investment activity except investment activity reserved to the government. For domestic investor there are different reserved investment areas (see these reserved areas in Appendix IV) [21].

II. Foreign investors

Foreign investor means investors have not Ethiopian nationality or nationality of other country. In this type of investment full ownership of an investment economic activity is carried out with the foreigner's capital. The minimum capital required by a foreign investor is US\$ 200, 000 per project. However the minimum entry capital required by a foreign investor investing in areas of engineering, architectural or other technical consultancy services, accounting and audit services, project studies or business and management consultancy services is US\$ 100,000 per project. Foreign investors can be investing in different areas except investment areas reserved for domestic .

III. Investors in joint ventures

The term "Joint venture" is commonly used to describe an economic activity carried out with the contribution of foreign and domestic capital in the form of partnership. Such form of joint venture is not separately regulated by the Ethiopian legislation. In joint venture the minimum capital contributed by the foreign partner(s) must not be less than US\$ 150,000. However, in areas of engineering, architectural or other technical consultancy services, accounting and audit services, project studies or business and management consultancy services or publishing must not be less than US\$ 50,000. The legal regime makes a distinction among the different classes with regard to areas of investment and capital requirements during licensing. In all other respects, the law treats all classes of investors in the same manner during and after licensing. Foreign investors that joint with the government or domestic investors have the right to invest in reserved areas for joint investors and other investment areas except reserved for domestic and government only (see appendix IV) [21].

3.3.1.2.7. Form of ownership

According to [11], there are three types of form of ownership. These are individual (wholly owned) investor, partnership investor and joint investor (domestic and foreign investors together). **Individual investor** means the owner of the investment project is single person either domestic or foreign investors or the responsibility of the investment is taken by that single person. **Partnership investor** means the owner of the investment project is either domestic investor together or foreign investors together invest in partnership, but not between domestic and foreign together. **Joint investors** means the owner of the investment project is on joint of foreign investors with domestic investor's or government.

Figures 3.8 below show that when form of ownership of investor is "partnership", the age and gender attribute value of investor are must be registered as plc. Also type of investor is registered as either domestic investors (having Ethiopian nationality) only in partnership or foreign investors (having other country nationality) only in partnership, but not foreign and domestic investors together in partnership.

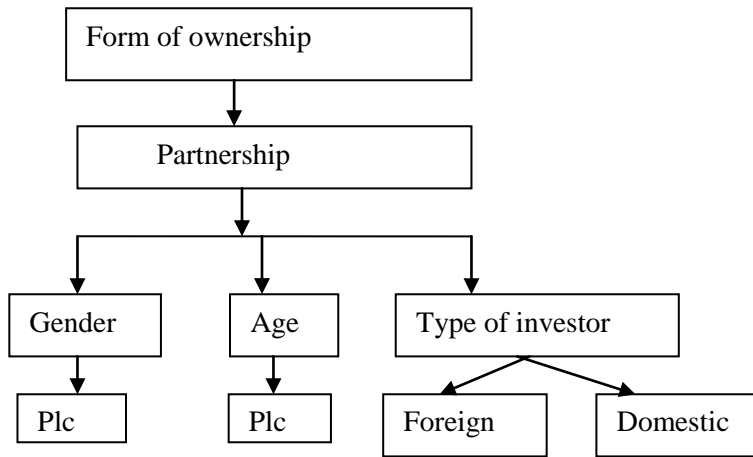


Figure 3.7: The hierarchy when form of ownership partnership.

Figure 3.9 below shows that when form of ownership of investor is “joint” the age and gender attribute value of investor are must be registered as “plc”. Also type of investor is registered as “joint” i.e. foreign investors and domestic investors together invest in joint form. So when form of ownership is joint, type of investor also “joint” because the type of investor is “joint” mean investors having Ethiopian nationality and investors having foreign nationality invest together in partnership.

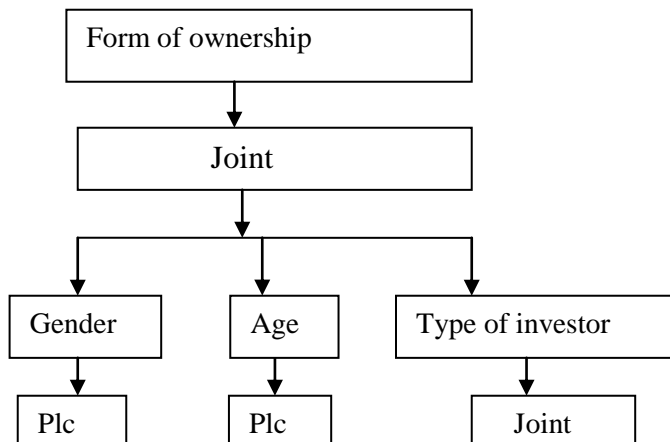


Figure 3.8: The hierarchy when the form of ownership is joint.

Figure 3.10 below shows that when form of ownership of investor is “individual” the age and gender attribute value of investor is the exact age and gender value of investors are registered. Also type of investor is registered as either domestic investors (having Ethiopian nationality) only in

partnership or foreign investors (having other country nationality) only in partnership, but not foreign and domestic investors together in partnership.

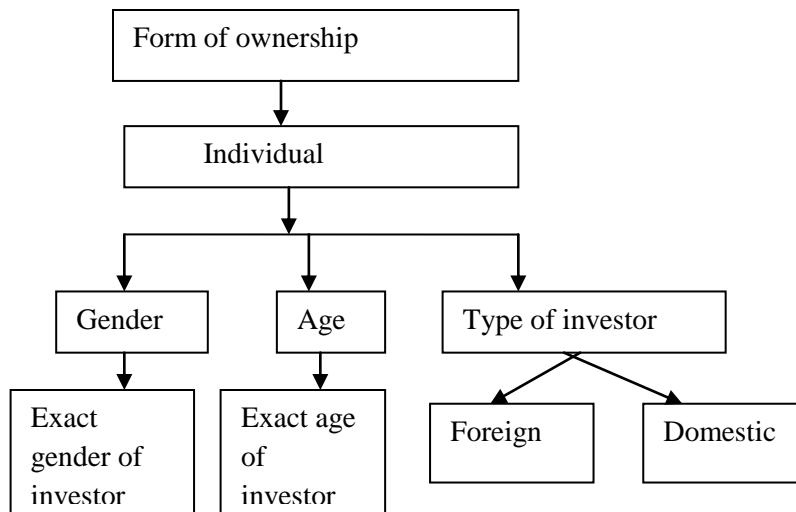


Figure 3.9: The hierarchy when the form of ownership is individual.

3.3.1.3. Structures of Investment Companies

- Wholly owned investment or individual (i.e. either male or female) – effective with full ownership of an investment by an individual.
- A private limited company (PLC) is a company whose partners are liable only to the extent of their contributions. The maximum number of the partners is fifty while the minimum is two. The company shall not issue transferable securities. Nowadays, most of the companies established in Ethiopia by foreign or domestic investors are private limited companies. The company has a minimum share capital of Birr 10,000, which must be paid up on registration.

3.3.1.4. Investment Areas

There are different areas with most promising key opportunities for potential investors exist in the country.

3.3.1.4.1. Investment sectors and their respective investment activity

+ List of investment sectors are

- ✓ Agriculture, hunting and forestry investment sector
- ✓ Construction investment sector
- ✓ Educational sector
- ✓ Health and social work sector
- ✓ Hotels and restaurants sector
- ✓ Manufacturing sectors
- ✓ Mining and quarrying sector
- ✓ Other community, social and personal service activities sectors
- ✓ Real estate, renting and business activities sector
- ✓ Transport, storage and communication sector
- ✓ Wholesale, retail trade & repair service sector

+ List of investment activities in each investment sectors are :

1. Agriculture, hunting and forestry investment sector

The major investment activity in this investment sectors are:

- ❖ Agricultural mechanization service
- ❖ Cash crop production (such as maize, Wheat and Barley Farming, Oil seeds and pulses, Pepper ,cotton, Teff, sorghum, gum, incense, Rice Farming, Coffee, Tea ,vegetables farm)
- ❖ Animal Raising and Livestock(such as Animal Fattening, animal breeding, Animal feed processing, Meat processing, milk processing, Dairy processing and Poultry)
- ❖ Fishery & Marine Cultivation (such as Fresh water fish farming, Crocodile farming and Aquaculture)
- ❖ Agricultural machinery rental service
- ❖ Agriculture & animal Husbandry Service
- ❖ Agro industry
- ❖ Agro Forestry
- ❖ Bee Keeping & honey production
- ❖ Horticulture & Floriculture farming (such as Fruits, vegetables & flowers, Apiculture, Integrated Forestry)

- ❖ Improved Seed Duplication
- ❖ Incense & Mucha Production
- ❖ Jatropha Plantation & Bio Fuel Production
- ❖ Mushroom farm
- ❖ Pig farm & processing

2. **Construction investment sector**

The major investment activity in construction investment sectors are the following:

- ❖ Building Construction
- ❖ Construction consultancy service
- ❖ Construction Machineries Rental and leasing
- ❖ Deep & Medium water Holes Drilling
- ❖ Electro mechanical Special Contractor
- ❖ General Construction
- ❖ Hole pole digging work Contractor
- ❖ Hydrological Survey, Water Well Drilling and Rehabilitation
- ❖ Pipe foundation contractor
- ❖ Road Construction
- ❖ Telecommunication Site Construction
- ❖ Warehouse construction

3. **Educational sector**

The following areas are some of the potential opportunities for investment:

- ❖ Kindergarten ,Primary ,Secondary schools and high school
- ❖ colleges/universities/institutes in different field
- ❖ ICT institutions or Basic computer skill training center
- ❖ Vocational training centers
- ❖ Language School (such as Chinese, Arabic, France, Italy, English etc.)
- ❖ Art & music training center
- ❖ Athletics Training Center

- ❖ Beauty and Hair Dressing Training Center

4. **Health and social work sector**

The major investment activities in this sector are:

- ❖ General & specialized clinics (such either medium or higher clinic of cancer, brain, eye, dental etc.)
- ❖ General & specialized hospitals (such as cancer, brain , eye , dental etc)
- ❖ Health Center & Physiotherapy Service
- ❖ Clinical laboratories and Diagnostic centers.
- ❖ Cultural Medicine Service
- ❖ Pharmacy (animal or human pharmacy)

5. **Hotels and restaurants sector**

The major investment activities in this sector are:

- ❖ Star hotels (such as 1 star, 2 Star, 3 Star, 4 star hotel etc.).
- ❖ Resorts & Lodges
- ❖ Modern Guest house, massage ,Gymnasium, steam bath service
- ❖ Café, Bar & Restaurant
- ❖ Specialized Restaurant (such as American, Arabian, Chinese, German, Russian etc.)
- ❖ Motel

6. **Manufacturing investment sectors**

The major investment activities in this sector are:

- ❖ **Textiles and clothing:** Spinning, weaving and finishing of textile fabrics and the production of garments;
- ❖ **Food and beverage products(agro- processing):** Processing and preserving of meat products, fish and fish products, and fruits and vegetables; integrated production and processing of dairy products; manufacture of starch and starch products; processing of animal

feed and processing and bottling of mineral water; sugar production; brewing and wine-making, etc.;

- ❖ **Tannery and leather goods:** Tanning of hides and skins up to finished level; manufacture of luggage items, handbags, saddle and harness items, footwear and garments, and integrated tanning and manufacturing;
- ❖ **Glass and ceramic:** Tableware and sanitary ware, sheet glass and containers;
- ❖ **Chemicals and chemical products:** Manufacture of basic chemicals based on local raw materials, including fertilizer, soda ash, rubber, PVC granules from ethyl alcohol; manufacture of caustic soda and chlorine-based chemicals; carbon and activated carbon; precipitated calcium carbonate; ballpoint ink; and tallow for soap;
- ❖ **Drugs and pharmaceuticals:** Manufacture of pharmaceutical, medicinal, chemical and botanical products in the form of tablets, capsules, syrups and injectables;
- ❖ **Paper and paper products:** Pulp from indigenous raw materials, paper and paper products;
- ❖ **Plastic products:** High-pressure pipes, pipe fittings, shower hoods, wash basins, insulating fittings, light fittings, office and school supplies, and fittings for furniture;
- ❖ **Building materials:** Manufacture of cement, lime, gypsum, marble, granite, limestone, ceramics, roofing tiles, corrugated sheets, tubes, pipes and fittings.

7. Mining and quarrying sector

The major investment activities in this sector are:

- ❖ Precious & Base minerals (such as mining of gold, tantalum, platinum, nickel, potash and soda ash)
- ❖ Industrial & Construction minerals (such as mining and exploration of marble, granite, limestone, clay, gypsum, gemstone , iron ore, coal, copper, potassium, silica, diatomite, betonies, and Stone Crashing etc.

8. Other community, social and personal service activities sectors

The major investment activities in this sector are the following:

- ❖ Dirty water wastage
- ❖ Female Beauty Salon
- ❖ Film and Music Production

- ❖ Laundry
- ❖ Library service
- ❖ Liquid & Solid Waste Disposal or cleaning Service
- ❖ Sport center (such as Fitness , Massage Center and Gymnasium Center)
- ❖ recreation center (such as Children , Cinema house, bathing, amusement park center)

9. Real estate, renting and business activities sector

The major investment activities included in this sector are:

- ❖ Construction Material Rental
- ❖ Advertisement and printing Service
- ❖ Agricultural Machineries Rental
- ❖ Aircraft Clearing Service
- ❖ Apartment
- ❖ Car Rent
- ❖ Business Center (market center or Super market)
- ❖ Telecommunication Site Construction
- ❖ Consultancy service (such as in Architectural and Engineering, agricultural, IT, business management and investment, mining, educational Consultancy etc.)
- ❖ Car washing, loader & parking
- ❖ Commercial Real estate
- ❖ Drug Sales
- ❖ Garage
- ❖ IT solution
- ❖ Tissue Culture Production

10. Transport, storage and communication sector

The major investment activities in this sector are:

- ❖ Tour operation service
- ❖ Magazine and newspaper distributer
- ❖ Storage & Ware House
- ❖ Travel agent and transportation (such as Dry Cargo transportation, Non – Scheduled Recreational Air Transport, Non-schedule Passenger Air Transport Service, Public Transportation Service).

11. Wholesale, retail trade & repair service sector

The major investment activities included in these sectors are the following:

- ❖ Alcohol and soft drink distributor
- ❖ Automotive Repair Service
- ❖ Chat, Coffee production and cattle export
- ❖ Drug and Medical Supply Whole Sale
- ❖ Oil Cereals Export
- ❖ Export of garment and textile products
- ❖ Export of meat and fish products
- ❖ Export of Leather Products & Incense
- ❖ Fuel station and Fuel Distribution
- ❖ Vehicle Spare part & Maintenance
- ❖ Trade center, Super market and shopping center
- ❖ Small, Medium & heavy vehicle Renovation and maintenance
- ❖ Import of LPG gas, Bitumen and blending of ethanol and petroleum
- ❖ Export of agricultural products (such as Crops, Flowers ,Fruits & Vegetables and Poultry)

3.4. Knowledge Representation

Knowledge representation is one of the basic steps in the process of case based recommender system development. Knowledge representation is the process of interpreting domain knowledge into computer understandable form using knowledge representation methods. Knowledge representation techniques include semantic network, logics, rules, case base and frames [53]. Among these, the researcher use case based representation method for this research.

The acquired cases are represented using one of the different case base representation methods that are appropriate for the researcher. Case base representation methods include feature-value case representation, relational database case representation; predicate based case representation and soft computing case representation methods [54].

For this research feature-value case base representation method is used. The reason for representing cases using feature-value case based representation is that this approach uses old experiences to understand and

solve new problems. It also reuses its solutions and lessons learned for future use. And also it represents cases in an easy way by using attribute and value pair representation [24, 25]. The algorithms used to calculate the similarity of cases in a case base representation for this research are nearest neighbor retrieval algorithm. The similarity function of nearest neighbor retrieval algorithm involves in computing the similarity between the stored cases in the case base and the new query. After that, selects the most similar stored cases to the query.

3.5. Investment sector and investment activity selection Case Structure

Investment sector and investment activity selection case structure has two parts. The first one is the problem (investor's cases) descriptions or situation and the second one is the solution.

Query Description/Situation: It is the part of the case structure that consists of attributes which describes about the investor.

Solution: This part of the case structure provides the recommended investment sector, investment activity and land size requirement given to investors based on the problem descriptions (investor's information).

Therefore, for this research the researcher identify different description and solution attributes with the help of investment expert and from recorded data set of previous investor's cases. But, there were different challenges during identification and representation of case structure. For instance, the first challenge was the attributes (the investor's information in this case) registered in EIA office was too many, so the researchers selects the most significant attribute that used to determine the selection of investment sector and investment activity selection decision using data mining attribute selection algorithm by using Weka software (see these important attributes in the table 3.1). The second challenge was some of previous investors information was missing or filled incorrectly or there was mismatch between attributes and values. So due to this the researcher selected 1344 correctly fill successful previous investors cases from 11500 successful previous investors cases and select 11 description attributes from total of 19 attribute to the development of case based recommender system.

The most important attributes that affects the selection of investment sectors and investment activity decision are listed below.

Attribute name	Parameter
Gender	Description
Age	Description
Capital	Description
Type of investor	Description
Form of ownership	Description
No of permanent employee	Description
No of temporary employee	Description
Region	Description
Zone	Description
woreda/town	Description
Interested investment Activity	Description
Recommended investment sector	Solution
Recommended investment activity	Solution
Recommended Land size requirement	Solution
Explanation	Solution

Table 3.1: Shows the Case Structure for investment sector selection

Short descriptions of attributes that are used for building the case structure are presented as follows:

Age: is the age of the investors. If investor’s form of ownership is either partnership or joint, the value of age is plc instead of put number.

Gender: is the term used to refer to a person’s self-representation as male or female. The values of this attribute are male or female or plc. If investor’s form of ownership is either partnership or joint, the value of gender is plc instead of male or female.

Capital: is the total amount of capital needed to start investment in the selected investment sector.

Number of permanent employee: is the number of persons who will benefit from the permanent employment opportunity created by the investment project.

Number of temporarily employee: is the number of persons who will benefit from the temporary employment opportunity created by the investment project.

Type of investor: indicates the nationality of investor's. The value of this attribute is domestic, foreign and joint. Domestic investor means investors having Ethiopian nationality and foreign means investors having other nationality (investor has not Ethiopia nationality). Joint investor means investors having foreign nationality and investors having Ethiopian nationality together.

Type of ownership: is indicates the ownership of an economic activity or investment. The value of this attribute is individual or joint. Individual means a person (either foreign or domestic investors) that invests individually and having wholly ownership of that investment. Partnership means either domestic investors together or foreign investors together are make partnership to invest but not partnership between foreign and domestic in this case. Joint means a foreign investor invests in a partnership with domestic investors, foreign investors or governments together and having joint ownership of that investment with other investors rather than wholly ownership.

Region: is the regional or city administration place where the investors plan to invest. The value of this attribute is Addis Ababa, Afar, Amhara, Benishangel gumze, Dire Dawa, Harari, Gambella, Oromia, SSNPR, Somali and Tigray.

Zone: is the zonal location of the investment. Since each region has its own zone , the investor selects specific zones based on the selected region.

Woreda/town: is the woreda or city location of investment that specifically facilitates service to investors. So the investor selects a specific woreda or city location of investment based on the selected zone.

Investor's Interest to investment: is the process of investor's interest of investment activity that wants to invest. So investors fill (enter) their own interest of investment activity as a query.

Recommended investment sector: it also a solution and provides a recommended investment sector based on the similarity of cases.

Recommended investment activity: it also a solution and provides a recommendation of investment activity. This recommended investment activity must be under the class of recommended investment sectors based on the similarity of cases.

Recommended land size required: it also a solution and provides a recommendation of land size required for implementing the investment activity based on the similarity of cases.

Explanation: is used to give explanation and description about the recommended investment activities.

CHAPTER FOUR

4.0. DESIGN AND IMPLEMENTATION OF THE PROTOTYPE

4.1. Introduction

The design and implementation part of this research involves the actual development of a prototype CBR system for investment sector and investment activity selection for new investors. Therefore, having all the necessary previous investor's cases from EIA and the knowledge from the domain expert and different relevant documents, the next task is coding the knowledge into computer using appropriate and efficient knowledge representation methods. After knowledge representation the next task is to develop the prototype recommender system for investment sector and investment activity selection. For this research, jCOLIBR 1.1 CBR frame work is used to develop the prototype recommender system. The retrieval algorithm used in this research is nearest neighbor retrieval algorithm. This is because jCOLIBRI uses this algorithm for retrieval task. Nearest neighbor retrieval algorithm is also suitable when there are attributes which have numeric (continuous) value [57].

4.2. Designing the Architecture of CBR system for investment sector and investment activity selection (CBRISAIAS)

The architecture of the CBRISAIAS system shown in figure 4.1 depicts how the prototype works during investment sector and investment activity selection. As the new query (problem) is entered, the prototype of the system matches the new case to the solved case in the case base of the system by using similarity measurement. If relevant cases are found within the case base, then the prototype system ranks the relevant retrieved cases based on their local similarity. Next, the prototype proposes a solution.

Building of case based recommender system was started by collecting the previously solved cases (i.e. previous investor's cases) from EIA consisting of investors who are successful in their investment activities (i.e. investors in the operational stage). Since previously solved cases having missing values and unnecessary information for this research, it need further processing in order to avoid missing values and remove unnecessary attributes for investment sector and investment activity selection process. After processing of cases and selecting the most important attributes, assigning weight and important parameters for each attribute was the next performed step. For the selection of important attributes that influence the recommendation of best investment sectors and investment activity, the researcher used data mining attribute

selection algorithm called attribute selection algorithm. The reason for using attribute selection mechanism is since all attributes are not equally important to recommend investment sectors and investment activity to new investors.

Once the case based recommender system is developed, users/investors can use the system easily to choose their investment sectors and investment activity based on the recommendation given by the system. When investors enter their query/case description through the user interface window, the system searches the best matching cases from the case base and retains the possible solution. If there is exact matching between the query and previous cases in the case base, the system recommends the most matched investment sector and investment activity for the investors. If the similarity between query and existing case is approximate, the proposed solution needs modification (adoption of solution) to fit the new case (query). At the end, the best modified solution should be stored into the case base for future use. The case base updates incrementally when the system learns from new case used by the investor.

The proposed solution can be derived directly from a retrieved case that matches exactly or partially to the problem of the new case. Partially match of retrieved cases means some attribute values of the existing case and new cases (query) are the same and some attribute values are different. Using the proposed solutions directly may have a risk because some attribute values need of editing (changing) based on different conditions. As a result the user of the system should have made an adaptation on the proposed solution having differences between the proposed case and the new case. In addition to adaptation, case contradictions are revised if there are situations where previous investor's cases attribute values are not similar with the new case (query) attribute values. There is no similarity between the existing case and new case means there are no previous stored cases having similarity with the new case (query) in all attribute values. Therefore if there is no similarity between the existing and new case, the proposed solution cannot give recommendation to new cases. So during this time, this new case or problem of investor can be revised and stored in the case base. Finally, the revised solution or stored cases is retained in the case base for problem solving in the next time.

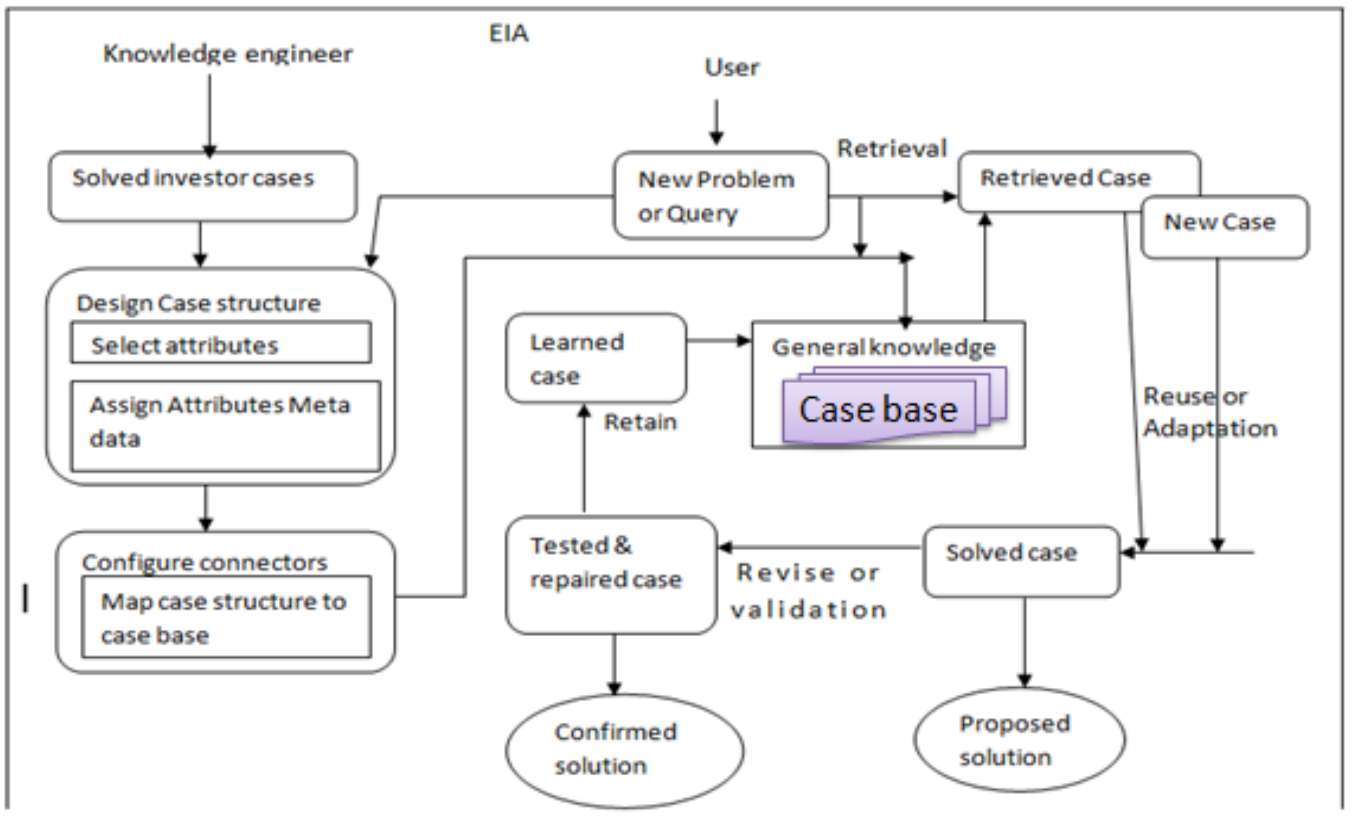


Figure 4.1: CBRISAIAS Architecture

4.3. Case-based Reasoning System for investment sector and investment activity selection

The development of CBR application involves a number of steps, such as collecting cases and background knowledge, modeling a suitable case representation, defining an accurate similarity measure, implementing retrieval functionality, and implementing user interfaces [58]. In this study, the researcher uses the main feature of jCOLIBRI to deliver the actual prototype. As [59] presented jCOLIBRI has been constructed as a core module to offer the basic functionality for developing CBR application. Implementing a CBR application from scratch remains a time consuming software engineering process and requires a lot of specific experience beyond pure programming skills [58].

jCOLIBRI can be started by clicking on exe file jCOLIBRIGUI.bat and then the GUI of jCOLBIRI will appear and it becomes ready for usage as shown in figure 4.2 . GUI of jCOLIBRI helps one to create new

CBR application with predefined task and methods. These predefined tasks and methods are represented in XML files that describe the tasks supported by the framework (tasks.xml) along with the methods for solving these tasks (methods.xml). The predefined tasks and methods stored in the framework of jCOLIBRI GUI for the purpose of configure the new system using (reusing) the tasks and methods that are predefined in the framework tasks.xml and methods.xml. Building a CBR system is a configuration process where the system developer selects the tasks the system must fulfill and for every task assigns the method that will do the task [59].



Figure 4.2: The Main Window of jCOLIBRI

After running jCOLIBRI GUI as shown in fig 4.2, the next task was **creating new CBR applications**. Fig 4.2, Shows the main window of jCOLIBRI with upper toolbar consisting of 4 menu lists, namely, file, CBR, Evaluation and Help. New CBR application can be developed step by step through this GUI. To develop new CBR application select “new CBR system” on the CBR toolbar. After that the box of entering new application name are displayed as shown fig 4.3.

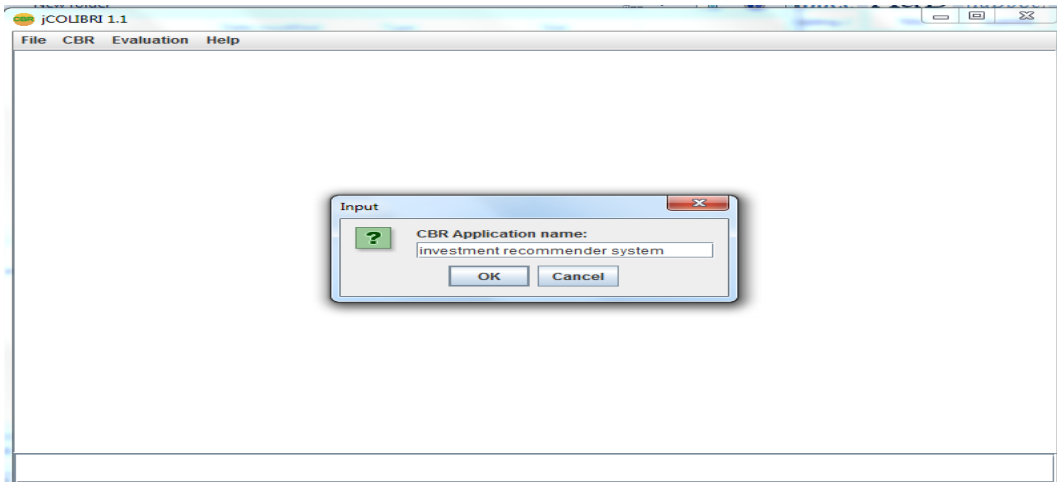


Figure 4.3: Creating new CBR Application

Then, a new window will appear to select one extension out of five as shown in fig 4.4. The five extensions are [59]: Core extension, Case Retrieval Nets Extension, Description Logic Extension, Textual Extension. and User Components Extension.

Core extension contains basic components of jCOLIBRI. Case Retrieval Nets Extension support case retrieval nets. In Description logic extension, description logic of jCOLIBRI is supported. Textual base CBR components are supported in textual extension. If user wants to define his/her own components he/she can use User components Extension. For this research the researcher uses core extension because it contains all basic components of jCOLIBRI that are needed to make case based recommender system.

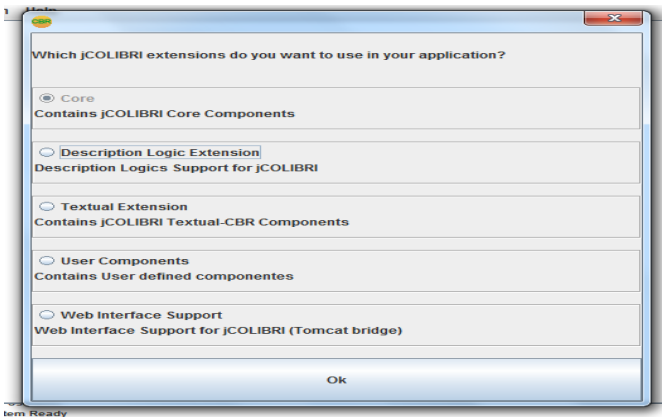


Figure 4.4: Types of jCOLIBRI extensions

After selecting the core extension as shown in fig 4.4, the main CBR application window that contains preCycle, CBRcycle and post Cycle applications are displayed as shown fig 4.5.

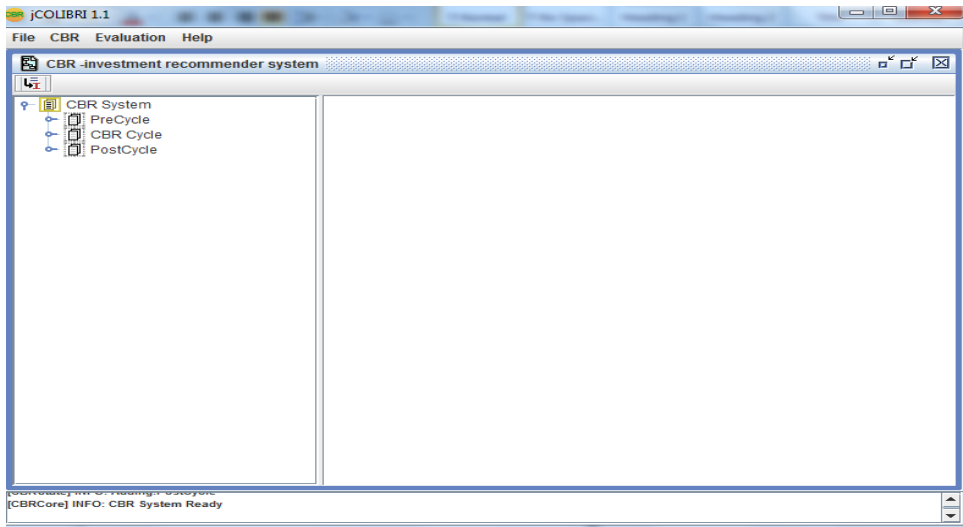


Fig 4.5: CBR applications

After displaying of the main CBR application, since the development of CBR system is very complex, the development of the CBR system for investment sector and investment activity selection for this research is divided into the following subsections which enable to achieve the objectives of this research.

4.3.1. Building the Case Base

One of the objectives on building the case base is collecting previous investor cases in order to build a case base and represent the cases using the appropriate case representation method. So, the researcher collected previous investor's cases from EIA. The acquired cases are used to build investment sector and investment activity selection CBR system that can offer decision support to investment experts, investors and other entrepreneur professionals. All the acquired cases are stored as plaintext file in a feature-value representation format. Feature value representation means each attribute has its own value in a column and row format.

The case base is represented as a plaintext feature value representation comprising of n columns representing case attributes (A1, A2, A3... An) and each m rows representing individual cases C ({C1, C2, C3, ..., Cm}). Each attribute has a sequence of possible k values associated to each column attribute $A=\{V1, V2, V3, \dots, Vk\}$. The reason for representing cases using feature-value representation is that this approach supports nearest neighbor retrieval algorithm and it represents cases in an easy way [24].

4.3.2. Case Representation

Case representation in case-based reasoning (CBR) makes use of familiar knowledge representation formalisms from AI to represent the experience contained in the cases for reasoning purposes. A large variety of representation formalisms have been proposed. However, three major types of case representation have arisen: feature value (or propositional) cases, structured (or relational) cases, and textual (or semi-structured) cases. For this research the researcher uses feature-value case representation because feature value case representation represent a case as attribute-value pairs, similar to the propositional representations used in Machine Learning (ML), that support k-nearest neighbor matching and instance-based learning [52]. Case representation is one of the main components in case based system, because case representation is the process of represent the case in the way the programming language (jCOLIBRI) easily understands or interpret.

Designing a case structure helps to define easily the features available in the case and to measure the similarity between existing case and new cases (query). Hence, the overall application of this research is to retrieve similar cases from the case base that can show future reasoning, problem solving, transforming a solution retrieved into a solution appropriate to the current problems (i.e. to retrieve similar cases to the query from the case base that guide investors), and making a recommendation in investment sector and investment activity selection process.

The collections of cases are represented in the feature-value representation to make efficient retrieval process. This is done through case indexing process in the jCOLIBRI programming tool. Indexing refers to assigning index to cases for retrieval by comparing the existing case and the query given by the user. [60].

4.3.3. Description of CBRISAIAS Case Attributes

A case is composed of three components: description (describes the problem), solution (represents a possible solution approach) and result (reveals if the proposed solution is able to solve the problem). Description and solution are collections of simple or compound attributes, permitting us to build a hierarchical case structure.

Defining case structure in jCOLIBRI are done by using simple manage case structure window. Description of attribute is the way of describe attributes or manage the case structure that used for the recommendation of

investment sector or investment activity. Description of attributes are done by add simple or compound description attributes in description case structure and set properties of attributes or metadata of attributes for each description attributes. Metadata of attributes are including weight of attribute, data type of attribute and similarity function.

Before creating CBR application we need to configure the case structure. Then use the toolbar menu button CBR and select option mange case structures. After that the new windows will appear for configuration of case structure as shown in the figure 4.6. In the figure, the left side's options are case structure description, solution and result. If we want to add Description cases in application, we can add by using “add simple” button. Add simple button is used to add simple attributes only. “Remove button” can be used to remove the description attribute if it's not necessary .When select one of description attributes, for instance select on Gender, then its properties will appear on the right side. Name, type, weight and similarity of case are properties of description attribute. “Apply changes” button was used to change the properties of decription attribute Local similarity is used for computing the similarity of each attributes. After defining the structure of cases, case structure is saved in xml file. During configuration of case structures, jCOLIBRI creates codes automatically and saves in xml file format.

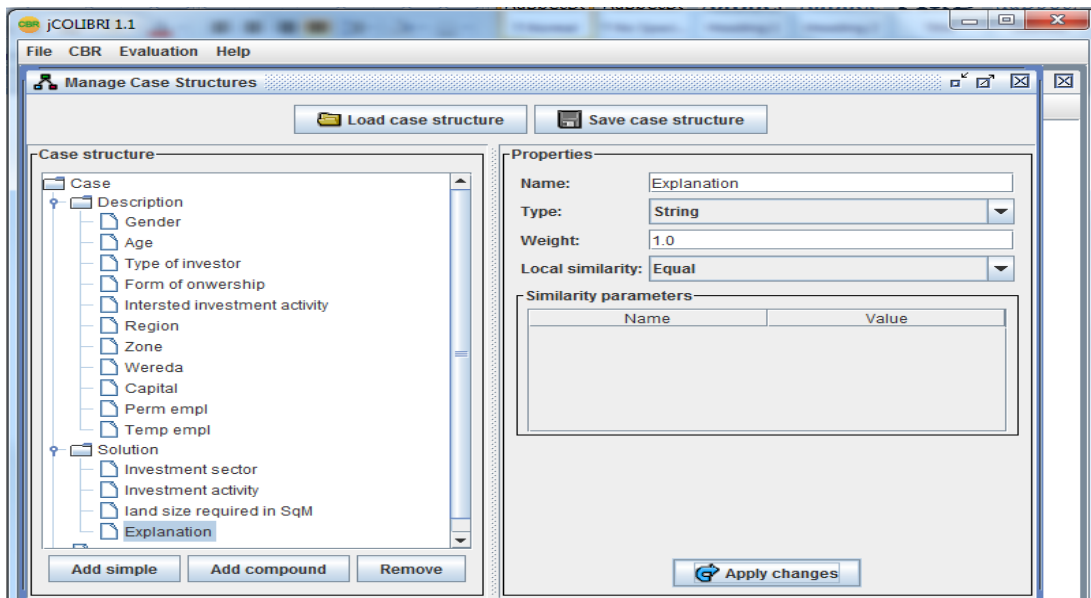


Figure 4.6: Defining Case Structures and similarity

As shown in fig 4.6 the case structure in this research consists of eleven description attributes that served to contain descriptions of the problem needed to make decision by the system and four solutions attributes.

Since the goal of this research is to give recommendations or solutions to investors on investment sector and investment activity selection, the developed CBRISAIAS system gives recommendation or solution. Solution attribute is assigned to the new case (investor) after investor input the value of all description attributes and measuring the similarity between the existing cases attribute value and new cases attribute value. For this research the solution attributes include recommend investment sectors, recommended investment activity, recommended land size requirements and the explanation facility about the recommended investment activity.

Table 4.1 shows the description of case attributes and solution attribute regarding name, data type, weights, local and global similarity.

Significant attributes			
Attribute Name	Data type	Weight	Local similarity
Gender	String	1.0	Equal
Age	String	1.0	Max string
Type of investor	String	1.0	Equal
Form of ownership	String	1.0	Equal
Region	String	0.9	Max string
Zone	String	0.9	Max String
Wereda	String	1.0	Max String
Capital in birr	String	0.95	Max string
No. of permanent employee	Integer	0.7	Interval
No. of temporarily employee	Integer	0.7	Interval
Interested investment Activity	String	1.0	Max String
Solution			
Investment Sector	String	1.0	Equal
Investment activity	String	1.0	Equal
Land size required in Sq.M	integer	1.0	Equal
Explanation	String	1.0	Equal

Table 4.1 Descriptions and Weight of the Selected Attributes

Table 4.1 shows the general description of attributes consisting of attribute name, data type, weight and local similarity. The most significant attributes to the problem domain are having the highest weights value of 1.0

as shown in table 4.1. These attributes are the most relevant to investors to select an investment sector and investment activity that matches with their personal and socio-economical characteristics. Next to these, attributes like Capital, Region, Zone, No of permanent employees and No of temporary employees have the weight value of 0.95, 0.9, 0.7 and 0.7 respectively. The assignments of weights to each attribute indicates that attributes having high weight is the most relevant to the investors in the selection process of investment sector and investment activity. The weight value of each attribute has been assigned by using information gain attribute selection algorithm and domain experts. The local similarity of most description attributes is maximum string. This is due to the similarity between query and cases can be calculated with maximum string length. Few attributes such as gender, form of ownership, and type of investor have equal similarity weight since local similarity needs exact match of existing cases and new case (query).

After identifying relevant attributes of the case, the next task is definition of appropriate similarity measure in JCOLIBRI. JCOLIBRI follows both local and global similarity measures.

- I. Local similarity:** Local similarity measure divides the similarity definition into a set of local similarity of each attribute. There are three types of local similarity measurement :
 - A. Equal:** If we select equal local similarity for each attribute, then our input and value of case base must be exact match. If the value between attribute are exactly match, the system gives (assign) a solution or recommended investment sectors and investment activity to new cases. Otherwise matches are a failure and have no any solution or recommendation to the new query.
 - B. Interval:** When we select similarity interval and adjust interval value, then, jCOLIBRI matches value keeping in mind that interval. Exact value match is not compulsory in interval local similarity.
 - C. Max string:** if we select the max string local similarity, the system matches by using the maximum string length.
- II. Global Similarity** is linked with compound attributes and used to get similarity of collected attributes in unique similarity value. Global similarity calculates the final similarity measure. The type of global similarity used in this research is average similarity.
 - **Average:** It is a type of global similarity that considers the average of all attribute of local similarity values.

The local similarity of all case attributes which have string data type have either equal and Max String similarity value .The Global similarity of all case attributes which have any data type have average similarity value.

4.3.4. Managing Connectors

Once case structures are configured in jCOLIBRI, CBR systems must access the stored cases in an efficient way from the case base. So, managing connector performs the task of configuring the connector that is going to load the case base. JCOLIBRI supports both SQL database and plain text file to store its cases base. jCOLIBRI splits the problem of case base management in two separate although related concerns: persistency mechanisms through connectors and in-memory organization.

Cases are often derived from legacy databases, thereby converting existing organizational resources into exploitable knowledge. To take advantage of these previously existing resources, facilitate intelligent access to existing information, and incorporate it as seed knowledge in the CBR system (the case base), jCOLIBRI offers a set of connectors to manage persistence of cases.

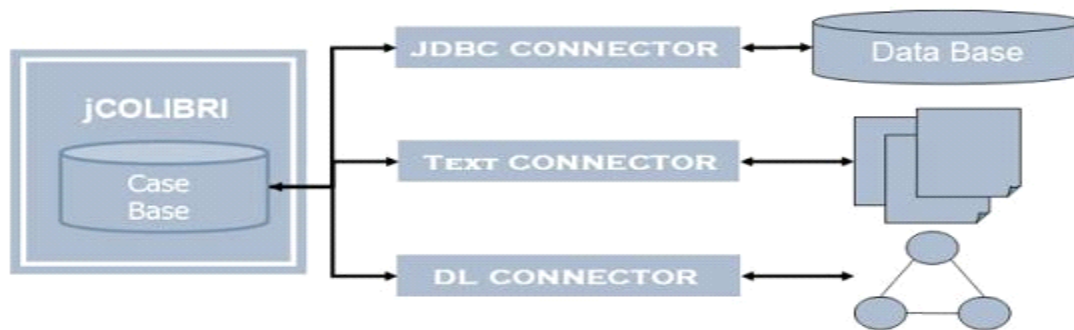


Figure 4.7: jCOLIBRI Connector Schema

Connectors are objects that know how to access and retrieve cases from the storage media and return those cases to the CBR system in a uniform way. Therefore connectors provide an abstraction mechanism that allows users to load cases from different storage sources in a transparent way. As shown in figure 4.7, jCOLIBRI includes connectors that work with plain text files, relational databases and Description Logics systems.

For the implementation CBRISIAS prototype, the researcher used plaintext connector because investor's cases are stored in plaintext file format as shown fig 4.8. Plaintext file case base connector is used for persistence of cases. In this connector, the researcher has to specify the path of case structure and also path of text file. All the attributes of a case should be mapped. The case structure path is used to access and match

attributes from case structure and file path is used to specify the .txt file that contains the case base. Delimiter of this connector uses comma (,) to separate value of each attribute in the case. This is connector's responsibility to retrieve data from case base and return it back to GUI. Like that of case structure, connector is also saved in xml format.

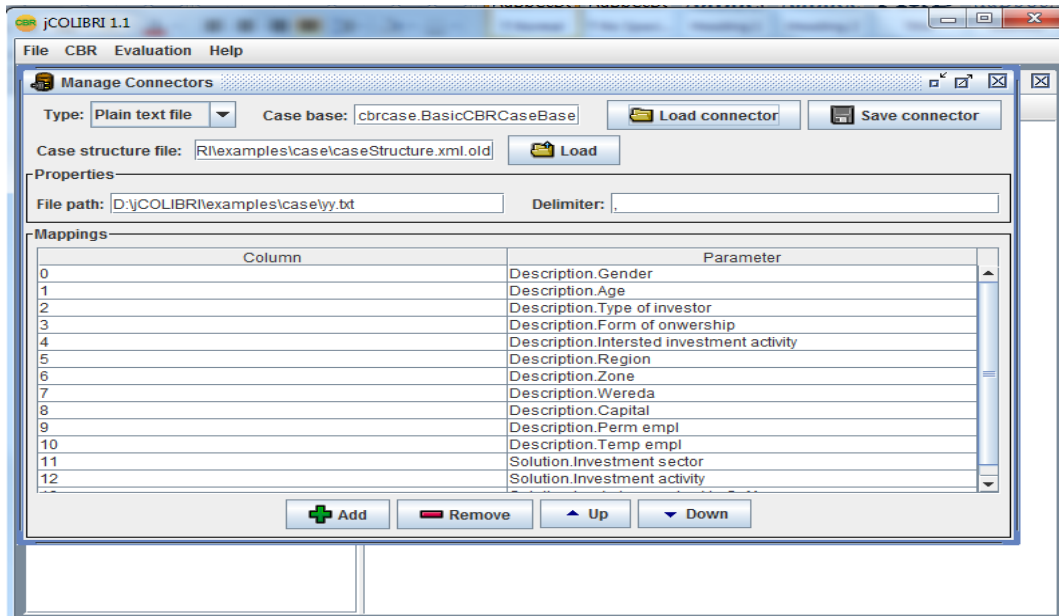


Figure 4.8: Managing Connector Configuration

4.3.5. Managing Tasks and Methods

JCOLIBRI is organized into packages. These packages can perform and execute tasks and methods of decomposition process. For the development of case base recommender system prototype, the researcher used core package task. The detail of each tasks and methods can be discussed separately as follows.

4.3.5.1. Managing Tasks / CBR application

After configuring the connector and case structure, the next task is selecting tasks and methods of application. jCOLIBRI has two types of task packages, namely, Core packages and User defined package tasks. For the development of CBRISAIA prototype, the researcher used core package tasks. A core package contains all classes that represent core functionality of a CBR application such as the domain model, case bases, similarity functions and retrieval algorithms. Core packages also have predefined tasks and methods that used to configure new system by reusing the tasks rather than using tasks or methods defined by the

system developer itself like user defined packages, because defined tasks and methods by user itself for every system is time taking and complex. Different core packages are available in JCOLIBRI. The main components of Core packages which are used in CBRISAIAS prototype development are PreCycle, main CBR cycle and PostCycle. The component of core packages is the final and important step for creating a new application where the CBR application is configured. The left side of Figure 4.9 shows PreCycle, CBR Cycle and PostCycle.

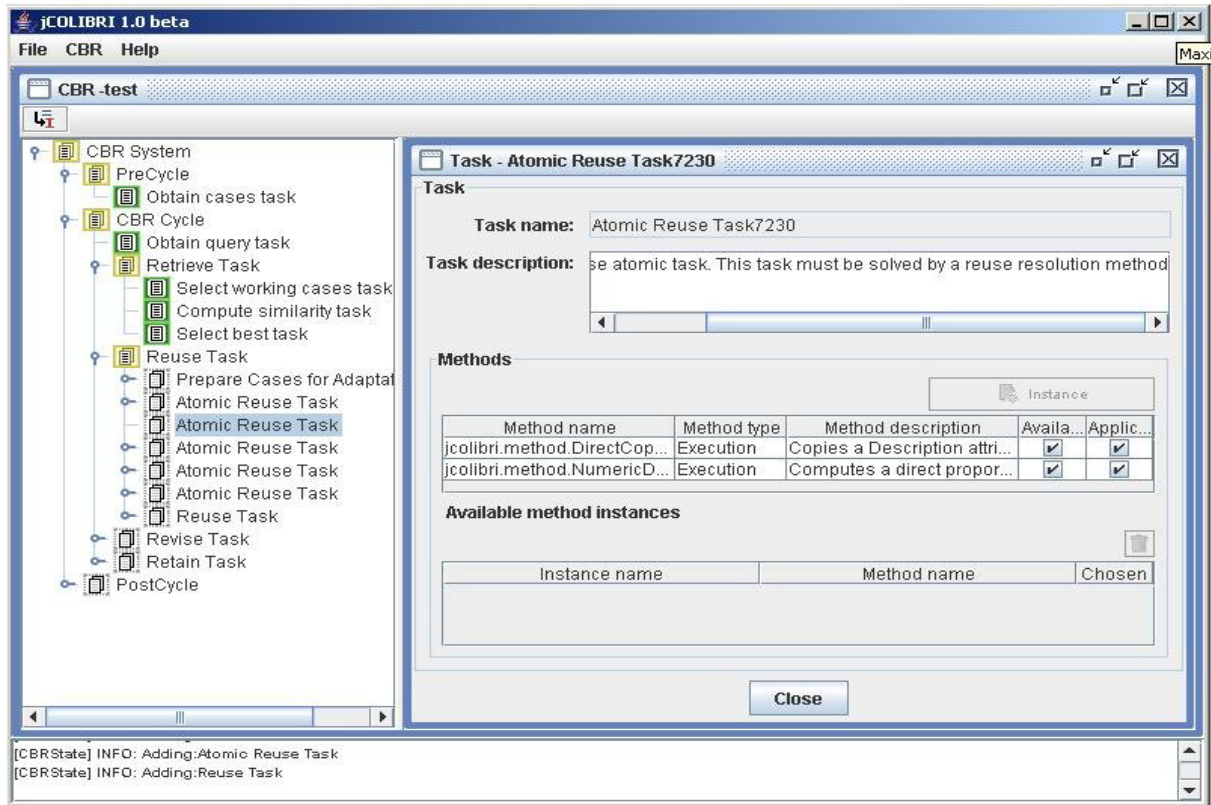


Figure 4.9: Configure the CBR Application

From the above figure, the main tasks and activates on each CBR systems or components of core packages can be describe as follows:

- + **PreCycle task:** from the component of the core packages the researcher start with “PreCycle” in order to loads the cases from data sources (case base). In preCycle tasks are solved once before the main cycle, like computing the index structure or processing texts in textual CBR. Therefore to load the cases from the case base it is necessary to define the path of the connector on subtask of Precycle called “obtain cases task” and make “instance” to instance the tasks and methods. In Precycle task there is only

one subtask called “obtaining case task”. Obtain case task is used retrieve (load) cases from investor’s case base before the execution of the main CBR cycle.

✚ **Main CBR cycle** is the main task of CBR cycle and it also has sub tasks. The developer has to give path of case structure that is saved in xml format in Main CBR cycle sub task called “obtain query task”. “Obtain query task” is used to know the number of investor’s case attributes that are available after he path is assigned. In addition to obtaining query task, there are other significant tasks under the main CBR cycle. These are retrieve tasks, reuse tasks, revise task and retain tasks.

- **Retrieve tasks** is used to retrieve case(s) from the stored case base. Retrieve tasks is also decomposed into different subtasks. These subtasks include select working cases task, compute similarity task and select the best case. “Select working case task” selects cases from case base and stores them into current context. “Compute similarity task” compute similarity of the stored cases with the case entered by the user using the query window. “Select best case” shows the best matched case(s) after computing the similarity of stored cases against the new case. It means that the number of best matched case(s) is shown to the user depending on the method used and the threshold.
- **Reuse/Adaptation tasks** enable to reuse previously stored cases. It has three subtasks. These subtasks are: prepare cases for adaptation task, atomic reuse task and reuse task. “Prepare cases for adaptation task” selects cases from case base and stores them into context. Here also specifying the path of case structure in this method is needed to “instance” the tasks and methods. “Atomic reuse task” should be resolved by reuse resolution method. After the process of the two subtasks “Reuse task” generates the proposed solution for the problem based on similarity. But there are situations where previous investor’s cases are not similar with the new case or the problem of investor, so during this time, this new case or problem of investor can be stored in the case base and will be reused by other investors for the next time. The system can learn at every entry of new case and new users adopt this knowledge for investment sector and investment activity selection process.
- **Revise task** is the evaluation and correction stage about the recommended solution in reuse phase. As shown in fig 4.10(a) after selecting the most similar cases from the retrieved results, the solution for the problem should be confirmed and validated before the solution is stored for future use.
- **Retain tasks** is also used to CBR case retention on a persistence layer. It has also its own subtasks like “select cases to store task” and “store cases task” . “Store case “ was used to type a new case name as shown in fig 4.10 (b). Select cases to store task give authentication to the user for storing case. The store cases task enables to store case(s) into the case base. Retain task is performed after having

confirmation in revision phase. So after the evaluation and correction of retrieved cases in revise task the problem together with its solution will be stored in case base.

- ✚ **PostCycle:** is the last task in managing tasks in jCOLIBRI. PostCycle task have only one sub task called “close connectors task” which is usually executed after the main CBR cycle. Its main task is to close a connection between case base and GUI.

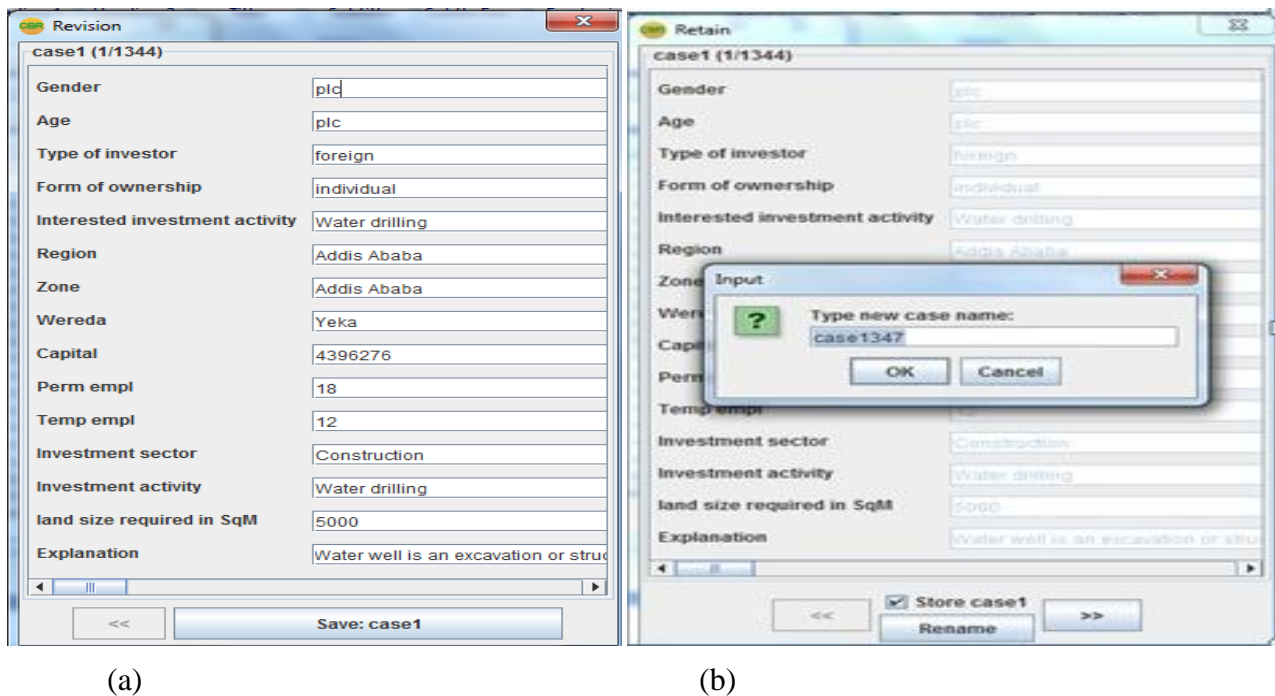


Figure 4.10: Revision and retain tasks

4.3.5.2. Case Similarity, Matching and Ranking

One of the primary goals of CBR system is to retrieve best similar cases by using some similarity assessment of heuristic functions. The similarity function involves computing the similarity between the stored cases in the case base and the new cases (query), and selects nearest similar cases to the query. Therefore, jCOLIBRI uses the nearest neighbor algorithm as a case retrieval technique. Nearest neighbor algorithm is used to measure the similarity between the stored (existing) cases and the new cases (queries), and return the search results within their ranked order. For each attribute in the query and case, local similarity function measures the similarity between each and every simple attribute values in the case base with new case queries. Based

on the matching weighted sum features from those simple attributes, the similarity score between the queries and stored cases for each simple attribute is assigned.

Finally, the average score (global similarity) of each attribute between the existing case and the query are computed and the result is assigned to the object (the similarity between the stored case and the query). And then the maximum degree of similarity among the retrieved cases is displayed according to their ranked order.

4.3.5.3. Managing Methods

The managing method library stores classes that actually resolve the task. These classes can resolve the CBR cycle using in programming or using GUI. All tasks that are mentioned in 4.5.5.1 should have their own methods to be assigned in order to achieve the tasks goal. The following is a list of methods which are used to solve tasks for this CBRISAIAS application.

LoadCaseBaseMethod: This method returns the whole available cases from the case base to designer. This method uses connector to retrieve case base.

ConfigurQueryMethod: This method resolves obtain query task. By receiving case structure as input parameters, it displays a GUI window so that user can enter query to retrieve cases from the case base.

SelectAllMethod: This method allows displaying all the available cases from the case base to the result window.

SelectSomeMethod: This method resolves to select best task by choosing the “n” number of nearest exact similar cases from the returned cases. The “n” indicates since there is more than one similar (relevant) case to the new case (query), the system retrieve “n” number of similar cases in ranking order of from highest similarity to lowest similarity. These method requests the users enter the value of each query as input. Then the system measuring the similarity between the new queries input value and the existing case value. Finally the system gives a recommendation on the type of investment sector and investment activity that best match with the requested input.

NumericSimilarityComputationalMethod: this is used to calculate similarity between the query and cases that are stored in the case base.

NumericProportionMethod: it is the sub method of reuse task which involves in computing numeric proportion between the description attributes and solution attributes.

ManualRevisionMethod: Manual revision method permits users to modify cases in the query window as they need.

RetainChooserMethod: This method allows the user to choose the method. Chosen method will store case base. User can choose if he/she want this method to store in case base.

In general, these are some of the methods discussed and used for this research. But, there are many other methods available in jCOLIBRI method library. It is the task of the knowledge engineer to choose the most appropriate methods during designing CBR application. Figure 4.11 shows the configuration of tasks and methods. The left side shows the tasks and subtask and the right side shows the methods.

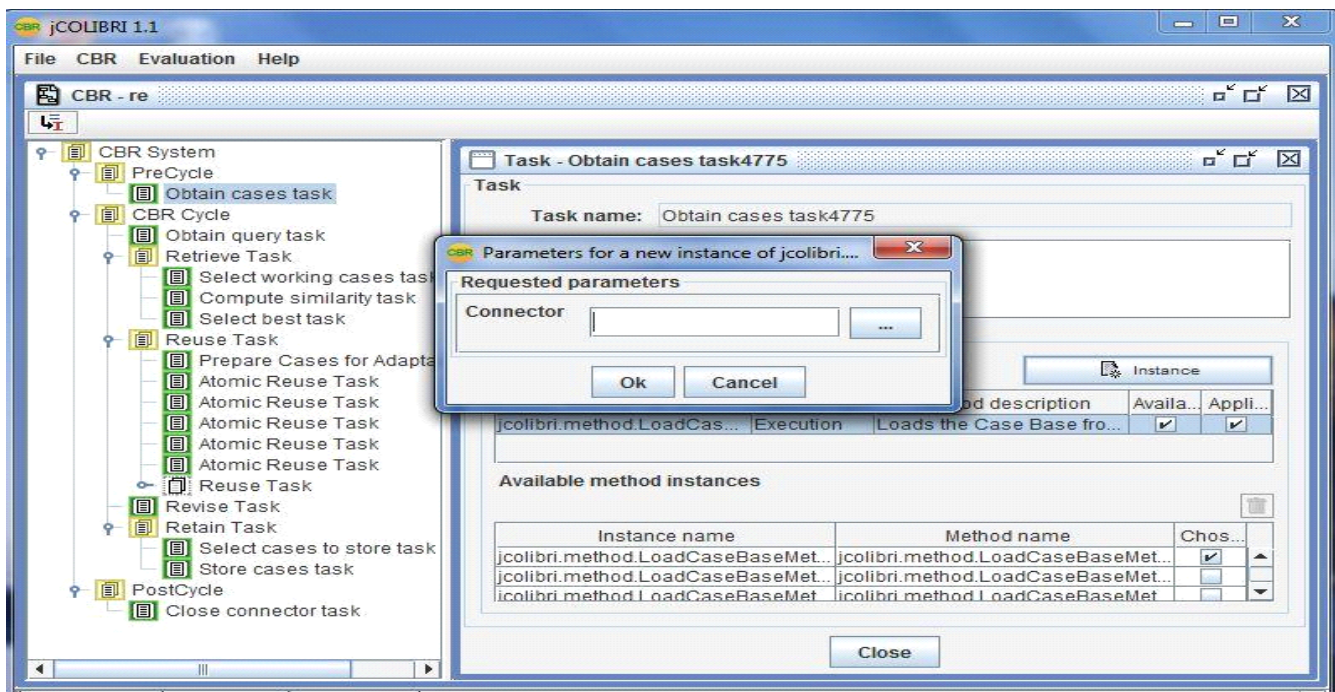


Figure 4.11: Tasks and Methods Configuration

4.3.5.4. Deploy the case base recommender system

After defining and configuring all the necessary steps required in designing case base recommender system in JCOLIBRI, new case (query) entry application for new investors is the next step as shown in figure 4.12.

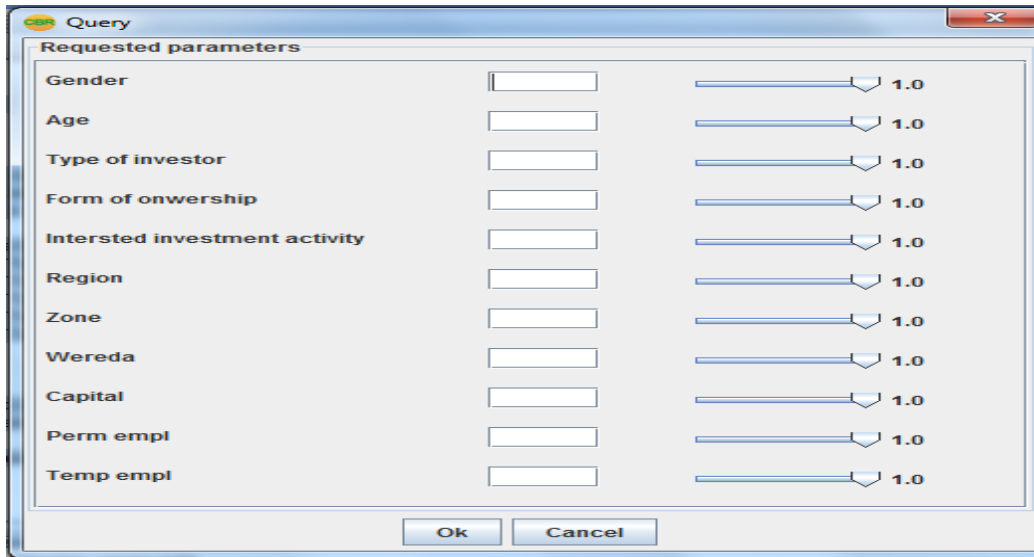


Figure 4.12: Window for Case Entry into the Case Base

In the above figure investors are required to enter the query to each requested parameters or attributes in the space provided. After entering the query, at the bottom of the screen they will see the results of similar previous investor cases and the recommended investment sectors, investment activity, land size requirement and explanation facility about investment activity on the execution log.

For instance, in the “form of ownership” box, investors are required to enter the query of individual, partnership or joint. And also in the “type of investor” box investors are required to enter the query foreign, domestic or joint. Investors who fill “plc” in gender must fill “plc” in age also because in it difficult to put the age of different investors or groups together. Investors who fill “joint” in type of investor must fill also “joint” in form of ownership because if investors are from different countries the investment’s form of ownership is joint also. In “Interested investment activity” box investors enter the investment activity that is want to invest. The “location of investment” is also entered by investors such as regions of investment location, Zone of investment location, wordeda of investment location in different box. Other attributes also require the same step to the pervious attributes.

4.4. Explanation Facilities

One of the more interesting features of knowledge based systems is their ability to explain themselves. The explanation facility in this study is used to give explanation about the recommended investment activity after decision or recommendation is made by the system. As shown in fig 4.13, once the system reaches final

decision on the recommendation of investment sector and investment activity, the user may not understand the recommended investment activity. In this case the system gives explanation facility about the recommended investment activity in addition to recommend investment sectors, investment activity, land size requirement .Then the system gives more descriptions about the investment activity such as the definitions, steps of manufacturing, or raw material needed to invest in that area etc.

As shown fig 4.13 the system provides a recommendation or solution to new investors after enter queries. Then the system gives the solution investment sector and investment activity and at the last the system gives explanation about the recommended investment activity. For instance in the fig 4.13 shown the system recommends to invest in Animal fattening and Dairy farm and the system also gives explanation about what is animal fattening and dairy fattening, the benefit of animal fattening, what type of animals are best for fattened etc are given in the explanation . There is a limitation in this explanation facility because the explanation system does not give response based on investor’s questions. The explanation facility gives explanation about only the solution investment activity but does not give any explanation during query entrance system or at any time the user wants explanation. For instance, if one investor not clear about form of ownership, there is no a possibility to ask explanation.

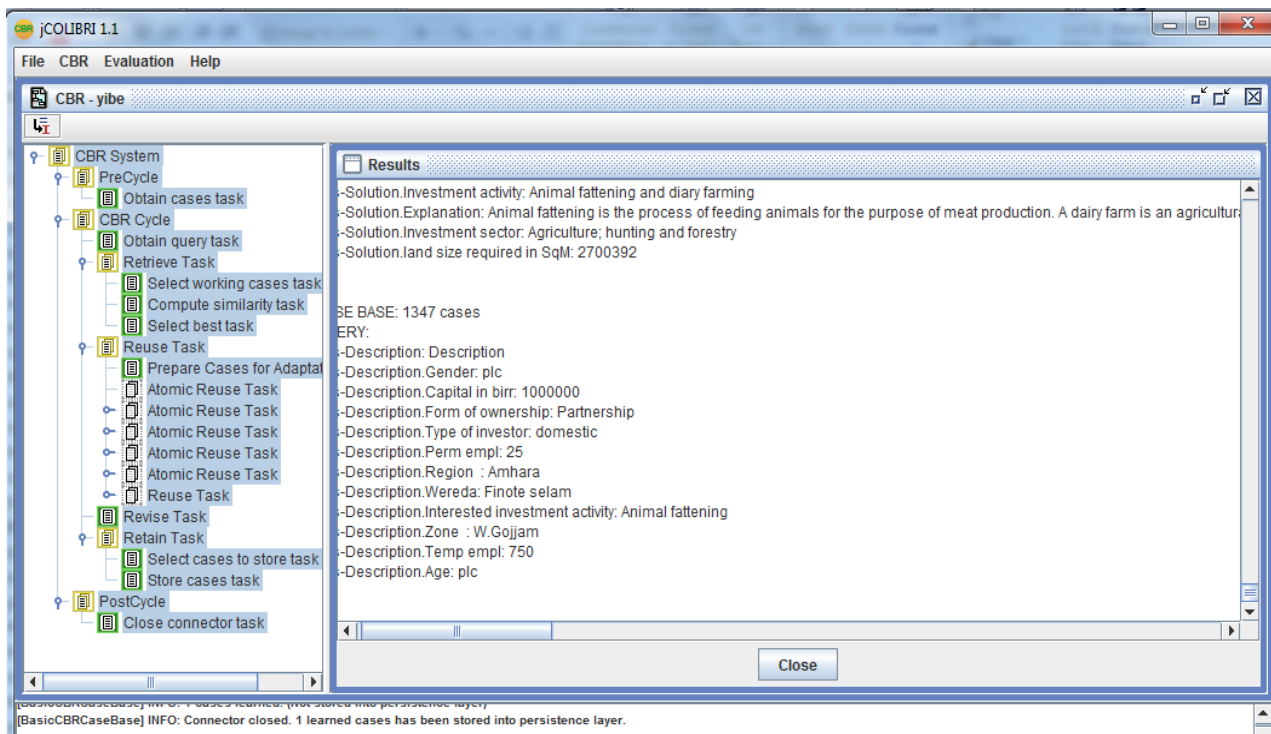


Figure 4.13: Explanation facility on investment activity.

CHAPTER FIVE

5.0. TESTING AND PERFORMANCE EVALUATION OF THE PROTOTYPE

5.1. Introduction

Testing and evaluation of the prototype case based recommender system is the final step that helps the knowledge engineer to measure whether the system achieves the proposed objectives or not. This chapter presents performance evaluation of the prototype system. For the performance evaluation this research has conducted case similarity testing, retrieval performance evaluation using recall and precision, evaluation of the reuse process, evaluations of learning mechanisms, and user acceptance testing of the prototype.

5.2. Case similarity testing

An experiment is made to know how new cases are matched with the cases from the case base. For this experiment, the researcher uses three experimental groups. The first group is made up of cases from the case base. The second group consists of cases which are made by modifying one of the attribute values of the case from the case base, while the third group is made up of cases which have two modified attribute values. Each test case is presented to the system individually to evaluate the performance of the similarity measures .Table 5.1 below shows the sample of queries that are used in this experiment with their values.

	Gender	Age	Type of investor	Form of ownership	Interested investment activity	Region	Zone	Wereda/city	Capital	Permanent employee	Temporary employee
Query 1	Plc	plc	Foreign	Partnership	Water drilling	Addis Ababa	Addis Ababa	Yeka	4396276	18	12
Query 2	M	46	Domestic	Individual	Agricultural development	B.Gumuz	Metekel	Guba	426000	15	150
Query 3	M	52	Domestic	Individual	Primary school	Amhara	E.Gojjam	D/Markos	265000	10	5

Table 5.1: Sample of queries that are used in this experiment with their values

Based on the above attributes, the next step is doing the experiments for the three groups. The number of cases used to check this experiment is three and make in to three different queries. After the query is provided to the system the similarity of the query with respect to the case are generated as shown in table 5.2.

Query	Description of Query	With respect to case	Degree of similarity
Query 1	The same value for all attributes	Case 1	1.0
Query 2	A value of attribute “type of investor” is changed.	Case 1	0.82
Query 3	Values of attribute “type of investor” and “woreda/city” is changed.	Case 1	0.75
Query 4	The same value for all attributes	Case 30	1.0
Query 5	A value of attribute “Gender” is changed.	Case 30	0.91
Query 6	Values of attribute “gender” and “Age” is changed.	Case 30	0.82
Query 7	The same value for all attributes	Case 1000	1.0
Query 8	A value of attribute “form of ownership” is changed.	Case 1000	0.91
Query 9	Values of attribute “form of ownership” and “capital” is changed.	Case 1000	0.86

Table 5.2: query similarity with their corresponding cases from the case base.

The case similarity test result of this experiment shows that when the test case has attributes value the same as a case stored in case base, the degree of similarity(global similarity) becomes 1.0(i.e. exact match) as in query 1, query 4, and query 7 as shown in table 5.2. On the other hand, the degree of similarity decreases when there is a change in one or more attribute values of the test case as compared to a case from the case base. When attribute values with higher weight value is changes the degree of similarity highly decreases.

5.3. Evaluations of the retrieval and reuse process by using statistical analysis

The statistical analysis uses sample of 52 cases that make up the case base as training and testing data that have been collected from EIA. The reason for selecting this sample case for testing is, since the researcher uses leave-one-out cross validation testing and it needs experimentation for each case, which is time consuming. The statistical evaluation uses a leave-one-out cross validation testing proportion, where each case in turn is left out , and the learning method is trained on all the remaining cases, i.e. the evaluation is

done for all cases by making one of the cases as a testing data and the rest of the cases as a training data(case base). The main reason that the researcher uses leave-one-out cross validation is that it is common evaluation strategy in case based reasoning [22] and it provides almost unbiased estimate of generalization performance.

The researcher conduct 52 experiments for both retrieval and reuse task of the case based reasoning system. The retrieval performance of the case based reasoning system is measured by using recall and precision, and the performance of the reuse task is measured by using accuracy.

5.3.1.Evaluation of the retrieval process

The first task of CBRISAIAS is to retrieve cases that are relevant to the new investor cases, so as to enable users to manage the new investor case by analyzing the retrieved case. In this research, the effectiveness of the retrieval process of the CBRISAIAS is measured by using recall and precision. According to [22], recall and precision values are the most commonly used measures of the performance of the retrieval process in CBR system. Recall measures the proportion of relevant cases to a given new cases (query) that have been retrieved from all the relevant cases in the case base. Precision on the other hand, measures the proportion of relevant cases to a given new case (query) from those that have been retrieved. Both recall and precision, being ratios, give values between 1 and 0.

To do this evaluation, for each test case the relevant investor cases from the case base should be identified. Due to this, test cases are given to the domain experts in order to assign possible relevant cases from the case base to each of the test case. The domain expert uses the value of recommendation attribute of the investor's cases as the main concept to assign the relevant cases to the queries, i.e. investor cases that have similar solution (recommendation) are relevant to each other. Based on this concept recall and precision are calculated.

Table 5.3 below shows sample test case with their corresponding relevant investor cases that are assigned by the domain expert from the case base.

Test case	Relevant cases from the case base
Case 35	Case1323,case1082,case14,case200,case569
Case 108	Case 483,case 1114,case165,case98,case601
Case256	Cse357,case198,case200
Case 449	Case 39,case770
Case465	Case1282,case 1313,case438,case1235,case204
Case489	Case490, case642
Case682	Case406,case764,case490,case624,case1207,case256,case577,case685
Case835	Case834,case1020,case1032
Case1293	Case1224,case642,case1217,case660,case489,case72,case1278,case1267
Case 1313	Case1216,case1217,case 465

Table 5.3: relevant cases assigned by domain experts for the sample test case

After relevant cases are identified and assigned to the test cases the next step is calculating the recall and precision value of the retrieval performance of the CBR system with a threshold interval.

As [29], indicated in his research, there is no standard threshold for the degree of similarity that has been used for retrieving relevant cases in CBR. Different CBR researchers use different case similarity threshold. Both [31] and [29] used a threshold level of [1.0, 0.8) i.e. this means cases with global similarity score greater than 80% are retrieved. In this research, the threshold is set by the researcher. So since the two researchers are satisfied in the threshold value of [1.0, 0.8], for this research the researcher used the threshold value of [1.0, 0.8].

The researcher conducting Fifty two (52) experiments to measure recall and precision by using a leave-one-out cross validation testing proportion and [1.0, 0.8] threshold interval.

The table below shows the precision and recall value of each test case.

Test case	Recall	Precision
Case 35	0.80	0.40
Case 108	0.80	0.67
Case256	0.67	0.67
Case 449	1.00	0.50
Case465	0.80	0.57
Case489	1.00	1.00
Case682	0.88	0.70
Case835	0.67	0.50
Case1293	0.88	0.78
Case 1313	1.00	0.60
Average	0.85	0.64

Table 5.4: recall and precision results for the sample test case

As shown in table 5.2 both recall and precision results are above average which is a good result. The average recall and precision results 85% and 64% respectively which is also a promising result. As seen in the table 5.4, for every test case more than average is registered both recall and precision. In terms of recall this research achieved a very good result. But, precision is somewhat lower compared to the average recall. This is because of the tradeoff between precision and recall.

5.3.2. Evaluation of the reuse process

The goal of reuse process in this research is to evaluate the proposed system whether recommend investment sectors and investment activity correctly for new investor cases, i.e. to solve the problem correctly. The performance of the reuse process is measured by using accuracy. Accuracy is one of the useful measurements in case based reasoning [22]. This measurement had been used by [31] on their research. Accuracy is defined by the percentage of the number of correctly recommended cases [23].

Since the research uses a leave-one-out cross validation testing proportion, 52 experiments are conducted to evaluate the performance of reuse process of the case based reasoning system prototype for investment sector

and investment activity selection. The result in table 5.5 shows that the reuse process also registers above average, which is a good result.

Total number of tested case query	Total number of correctly recommended cases	Accuracy
52	45	87%

Table 5.5: Accuracy value of the reuse process

5.3.3. Comparison of the Performance of CBRISAIAS with Previous CBR Systems

The performance of the system is compared with the previously conducted thesis research. The previous thesis research focus on retrieval phases of the case based reasoning system but the performance of the reuse phase is not evaluated by others except [31]. Thus, the recall and precision value of retrieval performance of the systems developed by the earlier thesis research are compared as shown in table 5.6 below.

Domain area and researcher	Programming tools used	Retrieval task		Reuse task
		Recall	Precision	Accuracy
AIDS Alemu(2010)	JCOLIBRI	72%	63%	Not evaluated
Hypertention Henok(2011)	Python	86.1%	60%	88.89%
Mental health Getachew(2012)	JCOLIBRI	82%	71%	Not evaluated
Field of study selection Biazen(2013)	JCOLIBRI	85%	55%	Not evaluated
Investment	JCOLIBRI	85%	64%	87%

Table 5.6: a comparison of CBRISAIAS system with the previous CBR systems

As indicated in table 5.6, the result of the recall value of the system is nearly the same with the recall value of Hencok [31], Getachew [29] and Biazen [28], while the value of precision shows an improvement from the recall value of Biazen [28] and Henock [31]. On the other hand the accuracy of the reuse performance of the system is nearly the same as that of Henock [31].

One of the main objectives in this research is to investigate the applicability of the case based reasoning system in recommending the appropriate solutions to the investor cases. This deals with the reuse task. And more than an average accuracy is registered which is a promising result. This task is not evaluated in Biazen [28], Getachew [29] and Alemu [30] research.

Since there is no global and local researches was done in CBRISAIAS. As a result, the researcher’s uses related thesis research that was done by Biazen[28] on the title of case based recommender system for field of study selection to compare the performance of this research from Biazen sresearch. Based on that,some of the problems of Biazen [28] research were overcome by this research. The main differences of this research from Biazen’s [28] research are summarized as follows.

- The prototype developed system in this research has the capability of giving explanation facility about the recommended investment activity. But in Biazen [28] research, the developed prototype system does not give explanation facility about the recommended field of study to students.
- Selection of main attributes that are important to case base recommender system is a challenging task. For this research the researcher uses attribute selection algorithm by using WEKA 3.7 software in order to select the most important attributes to the recommendation process. Biazen [28] performs the selection process manually with the help of domain experts by considering which attributes are more important for decision making in selecting field of study.

5.4. Testing the learning mechanism

The main aim of testing of the performance of learning mechanism is to test the learning performance of the prototype from solved investor cases for future use. This part of the prototype is tested with cases that are collected from new investors in EIA. For the experiment a new investor case is provided to the prototype. The sample of investor case that is used for the testing purpose is as follows:

Gende r	Ag e	Type of investo r	Form of ownershi p	Interested investmen t activity	Regio n	Zone	Woreda/ci ty	Capital	Permanen t employee	Temporar y employee
plc	Plc	Foreign	Individua l	Water drilling	Addis Ababa	Addis Ababa	Yeka	43962 76	18	12

Table 5.7: sample of new investor case

After the problem description of the above new investor case is fed for the prototype, the prototype computes the similarity of the new with the old cases from the case base. The prototype retrieves cases which are considered as a relevant case, rank ordered. Then the prototype recommends a solution to the new case from the retrieved relevant cases. The proposed solution for the new investor case is “construction” as an investment sector and “water drilling” as an investment activity. As this solved investor case is new for the CBRISAIAS, the revise process of the prototype proposes the new solved investor case to be verified by domain experts. The revise process is depicted in the figure 5.1

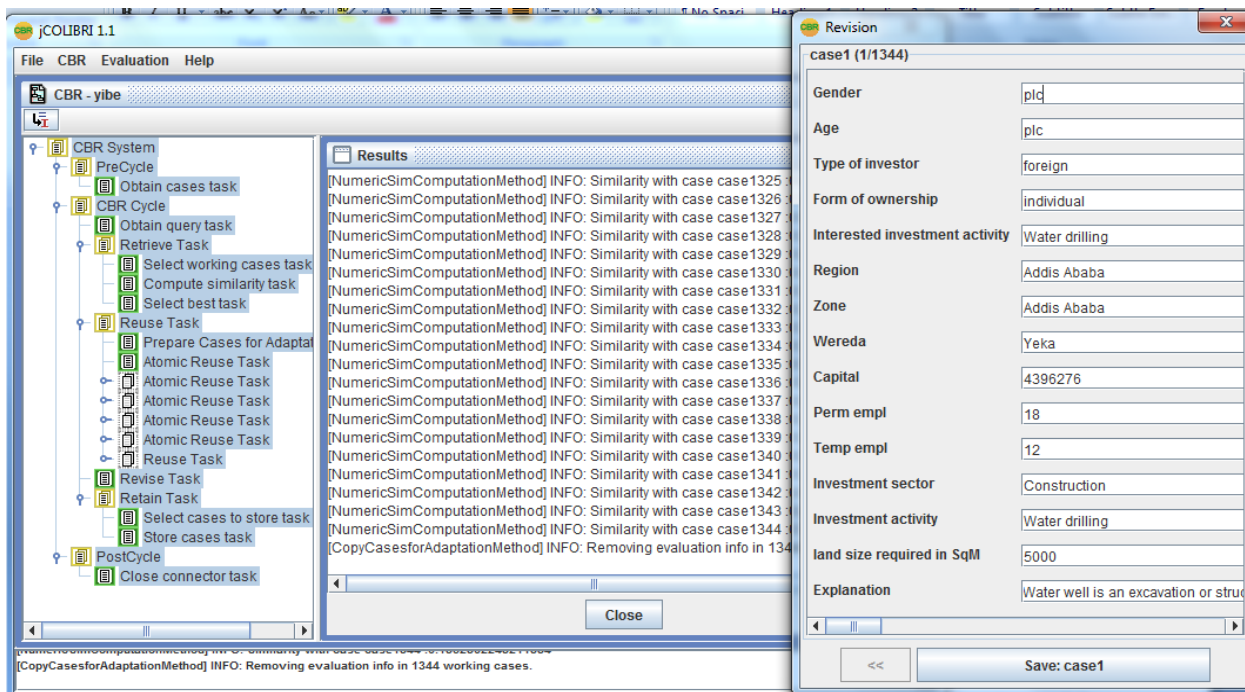


Fig 5.1: Revise process for the newly solved cases

The domain expert knows the validity of the proposed recommendation. As the domain expert confirms that the new solved investor cases is valid, the new solved investor case is retained for future use by the retain process of the prototype. So the new cases are stored in the existing case base and used in future recommendations. This is shown in fig 5.2 below:

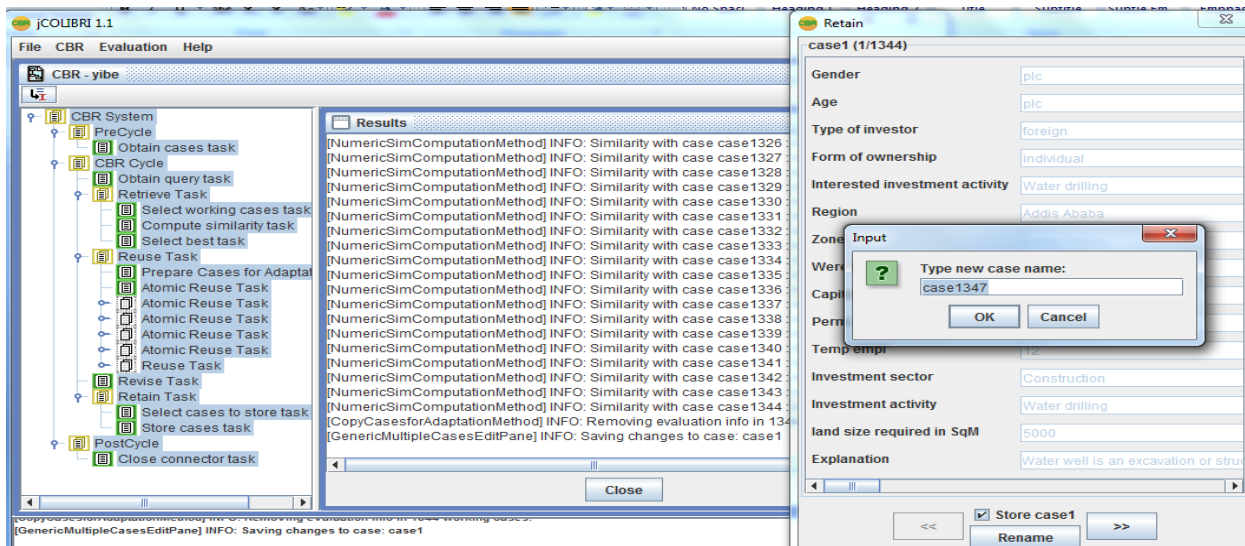


Fig 5.2: Retaining process of the newly solved case for the future use

In order to test the learning mechanism of the prototype, the problem description of the above investor case is again fed to the prototype. The prototype proposes a solution for the case, but the revise task doesn't propose the solved investor case to be verified. This shows that the CBRISAIAS system has learnt from successfully solved investor cases and uses it in solving other investor cases.

5.5. User Acceptance Testing

Case based system user acceptance evaluation method allows users (domain expert and investors) to directly interact with the system and evaluate the performance of the case based system from the users' point of view. User acceptance testing helps to ensure the performance of the prototype by assessing the feedback acquired from the domain expert and investors towards the developed prototype system.

This research uses questionnaires adapted from [28] and [29] to evaluate user acceptance of the CBRISAIAS prototype system. To achieve the goals of user acceptance evaluation of the prototype system, twelve domain experts from EIA and twelve investors who are participating in different investment sectors in the country were purposely selected. During the case based recommender system development these domain experts were actively involved in the different stages of the study, including knowledge acquisition and prototype development. Before starting the evaluation process of the system using the questionnaire, the researcher first gave explanation about the system to domain experts and investors in EIA. This explanation helped the

experts and investors to avoid the variation of awareness among them about the prototype case based recommender system.

Then after, the domain experts and investors were allowed to interact with the system by running number of cases having similar parameter with the facts incorporated in the case base. After the consultation of the system, to assess the user acceptance of the prototype case based recommender system, close-ended questionnaires were distributed to domain experts and investors. The questionnaire has nine close ended questions.

The first three questions are on the user interface design aspect which is basic for users interface satisfaction. These questions assessed whether the user interface of the system is easy to use, attractiveness and time efficiency of the system. The rest of the questions are used to evaluate the prototype's adequacy and clarity, relevancy of retrieved cases, relevance of the attributes used, clarity of the explanation facility, problem solving ability and significance of the prototype knowledge based system in investment sector and investment activity recommendation system. All these nine closed ended questions answered as excellent, very good, good, fair and poor. For the ease of analyzing the performance of the system based on user's feedback, the researcher assigned numeric values to the five options as follows: excellent=5, very good= 4, good=3, fair=2, poor=1. The system evaluators give the value for each closed ended questions.

The Table below indicates the feedbacks obtained from the domain experts (evaluators) on systems interaction as calculated based on the given scales.

No	Evaluation criteria	Performance value						
		1	2	3	4	5	Average	%
1	Easy to use of the recommender system				7	5	4.4	88
2	Is the system efficient in time				6	6	4.5	90
3	Is the user interface interactive			4	6	2	3.8	76
4	Adequacy and clarity of decision support			3	7	2	3.9	78
5	Relevancy of the retrieved case in the decision making			4	6	2	3.8	77
6	Fitness of the final solution to the new case			3	6	3	4.0	80
7	Relevancy of the attributes in representing investors case			5	5	2	3.8	76
8	Does the explanation facility give brief description about the recommended investment activity			3	6	3	4	80
9	Rate the significance of the system in the domain area				4	8	4.7	94
		Total average					4.1	82%

Table 5.8: the CBRISAIAS system performance evaluation by the domain experts.

As shown in table 5.8, 58% of the respondents rate the ease of use of the recommender system as very good and the remaining 42% of the respondent's rate it as excellent. Similarly, efficiency in terms of time is rated very good by 50% of the respondents whereas the remaining 50% of the respondents rate it as excellent. In the case of user interface interactivity of the prototype, 50% of the respondents rate the prototype as very good and 33% as good and 17% as excellent. 25% of the respondent rate adequacy and clarity of the system as good and in the same way 58% and 17% of the respondent's rate as very good and excellent respectively. The relevancy of the retrieved case in the decision making is also rated by 50% of respondents as very good, 33% as good and 17% as excellent. The fitness of the final solution to the new case, 25% of the respondents rate as good and in the same way 50% and 25% of the respondent's rate it as very good and excellent respectively. The relevancy of the attributes in representing investor's case is also rated by 42% of respondents as good, by 42% as very good and by 16% as excellent. In the case of explanation facility, 25% of the respondents rate explanation facility as good, 50% as very good, and 25% as excellent. Finally, 67% of

the respondents rate the applicability of the prototype in their domain area as excellent and the remaining 33% of the respondent's rate as very good.

On the other hand, table 5.9 below shows the performance evaluation of the prototype by the investors.

No	Evaluation criteria	Performance value						
		1	2	3	4	5	Average	%age
1	Easy to use of the recommender system			1	8	3	4.2	84
2	Is the system efficient in time				7	5	4.3	86
3	Is the user interface interactive			2	6	4	4.2	84
4	Adequacy and clarity of decision support			1	9	2	4.1	82
5	Relevancy of the retrieved case in the decision making			1	8	3	4.2	84
6	Fitness of the final solution to the new case			2	6	4	4.2	84
7	Relevancy of the attributes in representing investors case			2	7	3	4.1	82
8	Does the explanation facility give brief description about the recommended investment activity				7	5	4.2	84
9	Rate the significance of the system in the domain area				4	8	4.7	94
		Total average					4.2	84%

Table 5.9: the CBRISAIAS system performance evaluation by the investor's

As shown in table 5.9, 8% of the respondent's rate ease of use of the recommender system as good and the remaining 67% and 25% of the respondent's rate as very good and excellent, respectively. Similarly, efficiency in terms of time is rated as very good by 58% of the respondents whereas the remaining 42% of the respondents rate it as excellent. In the case of user interface interactivity of the prototype, 50% of the respondents rate the prototype as very good, 17% as good and 33% as excellent. In terms of adequacy and clarity of the system, 8% of the respondents rate the prototype as good, 75% as very good and 17% as excellent. The relevancy of the retrieved case in the decision making is also rated by 67% of respondents as

very good, 8% of respondents as good and 25% as excellent. Regarding fitness of the final solution to the new case, 17% of the respondents rate the prototype as good, 50% as very good, and 33% excellent. The relevancy of the attributes in representing investor’s case is also rated by 58% of respondents as very good and the remaining 17% is rate it as good and 25% rate it as excellent. While 58% of the respondent rate explanation facility gives brief description about the recommended investment activity as very good and the remaining 42% rate it as excellent. Finally, 67 % of the respondents rate the applicability of the prototype in their domain area as excellent and the remaining 33% of the respondent’s rate as very good.

5.6. Discussion on user acceptance and system performance using recall and precision, case similarity and reuse process

The evaluation and testing procedures help to address the question of user acceptance and accuracy of the CBRISAIAS prototype. Visual interaction and questionnaire methods are used to assess user’s acceptance issues and applicability of the prototype. Based on the evaluation results obtained using closed ended questions none of the evaluators respond as poor or fair both in the domain expert and investors side. The following table 5.5 summarizes the results obtained on close ended questions.

Respondents who responded as	Total number of respondents for each option		Percentage (%)	
	Domain expert	Investor	Domain expert	Investor
Poor(1)	0	0	0	0
Fair(2)	0	0	0	0
Good(3)	18	9	16.7%	8.3%
Very good(4)	52	62	48.1%	57.4%
Excellent(5)	38	37	35.2%	34.3%
Total average	4.1	4.2	82%	84%

Table 5.10: Domain experts and investors feedback on closed ended questions

As shown in the above table 5.10 none of the evaluators (domain experts and investors) respond as poor or fair. On the other hand domain experts and investors rate the prototype as very good fifty two times (48.1%) and sixty two times (57.4%) respectively. Also domain experts and investors reply as very Excellent thirty eight times (35.2%) and thirty seven times (34.3%) respectively. Lowest value of responses is rated by domain experts and investors as the prototype is good eighty times (16.7%) and nine times (8.3%) respectively. Also as shown in the table the overall average user acceptance evaluation of the prototype case

based recommender system is 82% and 84% by domain experts and investors respectively; which means the prototype is accepted by 82% of domain experts and 84% of investors. Therefore, above 82% of domain experts and 84% of investors are satisfied with the easiness, attractiveness, speed, adequacy and clarity of problem solving ability, explanation facility about the recommended investment activity and the significance of the prototype case based recommender system in the domain area. This implies that the prototype modeled relevance and satisfactory domain knowledge in useful way and it performs well in making right decisions on the recommendation of investment sectors and investment activity.

Finally, when the researcher compares the results responded by domain experts and investors, the performance of the case based recommender system responded by investors 84% (4.2) and domain experts 82% (4.1) which is above average in acceptance of the system in users. This shows that the developed CBRISAIAS system is more acceptable and applicable in the domain area.

Additionally, the testing procedure by using test cases helped to analyze the performance of the prototype knowledge based system. The result obtained using test cases indicate that the prototype has recall performance of 85% and precision performance of 64%. And also the accuracy of the prototype system for reuse process achieves 87%, which is above average. As the goal of reuse process in this research is to recommend correctly for investor cases, i.e. to solve the problem correctly, the performance of the reuse process is measured by using accuracy. As a result of accuracy result the developed prototype system has the capability to advise and recommend in investment sector and investment activity selection correctly.

The case similarity test result of this experiment shows that when the test case has attributes value the same as a case stored in case base, the degree of similarity(global similarity) becomes 1.0(i.e. exact match). On the other hand, the degree of similarity decreases when there is a change on one or more attributes value of the test case as compared to a case from the case base.

Based on the above evaluations the main strength of the system and its applicability in the domain area are:

- The prototype case based recommender system helps to solve problems in the areas where experienced and skilled investment experts are unavailable.
- Have a learning capacity when the new cases coming
- Consistent answers for repetitive decisions, processes and tasks
- More accurate and Easy to use based on the result of user acceptance evaluation and the accuracy of the system.

- Applicable anywhere, even at home we can advise ourselves by having the software
- Based on the user acceptance evaluation on the explanation facility the percent of responds by both investors (84%) and domain experts (80%) is a good result. As a result explanation facility giving brief explanation about the recommended investment activity and to make clear.
- Encourage self-advising

Generally, all the evaluation and testing results of the prototype show encouraging finding for further research work to fully implement and apply case based recommender systems technology in recommending investment sector and investment activity in Ethiopia.

Chapter six

6.0. Conclusion and Recommendations

6.1. Conclusion

CBR is a part of AI which enables us to design an intelligent agent that makes decision from the past solved cases i.e. a new problem is solved by finding a similar past case and reusing it in the new problem situation. As compared to rule based reasoning, CBR can work with new cases that match partially to the case from the case base. However, rule based reasoning cannot solve a problem that doesn't exactly match with the rule of the system. This shows that rule based reasoning works in closed assumption where every fact are known and represented.

In Developing country like Ethiopia the advising system on investment remain at lower stage. Different factors affect the investment advising system in Ethiopia. These factors include lack of guide line or criteria to assign investors in different investment sectors, shortage of skilled manpower in the area, lack of consistency of investment experts in advising, and lack awareness of investors about the purpose of advising systems for the selection of investment sector and investment activity.

To address the above problems, the main goal of this research is to develop a prototype CBRISAIAS system. The system aims to assist both the domain experts and investors in the processes of making proper investment sector and investment activity selection decisions from already solved investor cases.

The relevant knowledge required for the development of CBRISAIAS system is acquired from domain experts, investors and document analysis. During the prototype development, previous investor cases are collected from EIA .The relevant knowledge acquired from domain experts, investors and secondary document is conceptually modeled using hierarchical structure conceptual modeling method. The Case representation method that is used in this study is feature value case representation method. Feature value case representation is applied to represent the knowledge before it has been codified using the jCOLIBRI tool. The prototype of CBRISAIAS is developed by using jCOLIBRI 1.1 Programming tool.

CBRISAIAS system uses the prominent CBR cycles (Retrieval, Reuse, Revise and Retain) to perform different tasks. In CBRISAIAS, the first task is retrieval of cases by entering a new problem description

(case) by using the query window. Then, similarity computation is performed and retrieves most similar cases. After retrieval of similar cases, reusing the previously solved cases from the case base is performed and followed by manual revision of cases to fit the problem at hand by investment experts. The last task is storing the revised case in the case base for future use. The retrieval task of the prototype uses Nearest Neighbor retrieval algorithm.

The evaluation result shows that CBRISAIAS system is encouraging as retrieval performance of the prototype registers an average value of 85% recall and 64% precision, while its reuse performance registers an average value of 87% accuracy. The user acceptance evaluators (both domain experts and investors) assign more than average value for all parameters that are used in the user evaluation form for the prototype. So the average user acceptance evaluation achieved 82% and 84 % performance by domain experts and investors respectively.

Furthermore, the following conclusions are drawn from the finding with regard to the research questions:

- The major attribute that have more influence in investment sector and investment activity selection are age, gender, location of investments (such as region , zone, woreda), form of ownership, type of investor ,capital and investors interested area of investment.
- The type of investment sectors open to both foreign and domestic investors are Agriculture, construction, manufacturing, hotel and tourism, educational service, health and social work, mining, fishing, renting and business activity, wholesale and trade service, communication and transport service and other community service. For each investment sector there are different investments activities open for investment.
- The applicability of case based recommender system for investment sector and investment activity selection haven been proved.
- The result of system performance indicated that users are satisfied with proposed system and the performance of the system validation result showed the system recommends highly acceptable investment sector and investment activities to investors.
- In the proposed case based recommender system learning is made for new cases by dynamically updating in the existing case base for the purpose of future use as a case base.

- CBRISAIAS contribute in the selection or recommendation of investment sectors and investment activity by assisting investment experts and investors in the domain area. The proposed prototype system actually contributes a lot, especially for those less experienced experts or in the absence of expert.

6.2. Recommendation

Although the results of this study are promising, there are problem areas that need further investigation for future work. Therefore, the researcher recommends the following issues as a future research direction based on this study.

- The attributes used for this research are collected from the previous investor cases from EIA. These attributes are not enough for the selection of investment sector and investment activity decision. So further research can be conducted by adding other important attributes for investment sector and investment activity selection such as level of education, marital status, and level of risk taker by making a direct survey of successful investors.
- To enhance the performance of the prototype case based recommender system the hybrid strategy approaches should be investigated which combines rule based reasoning and case based reasoning. The inclusion of rule based reasoning in this research helps the proposed system to give explanation facility when the user wants explanation and used to represented fact and rules extracted from the domain expert.
- In this study the explanation facility given by the proposed system is not user interactive. Not user interactive means the explanation is given only once when the system assigns investment sectors and investment activity to the new case. The explanation facility gives explanation only about the recommended (assigned) investment activity, but does not give explanation at any time the user needs explanation. So further research can be done to add explanation facility that can give explanations any time the user wants explanations in addition to explanation of the recommended investment activity.
- It is proven that CBRISAIAS system helps in the investment sector and investment selection decision based on the performance measure of the system. But the system was developed in English language and is difficult to understand by some investors. Further investigation can be conducted by developing a CBRISAIAS system in different local languages. This helps investors communicate using their own language with the case based recommender system.

Reference

- [1]. Dereje W. (2012). Role of Financial Institutions in the Growth of Small and Medium Enterprises in Addis Ababa, Addis Ababa University, Ethiopia.
- [2]. Abiyu J. (2011). Factors constraining the growth and survival of micro and small enterprises in burayu, Addis Ababa University, Ethiopia.
- [3]. Justin (2012). Factors Affecting International Investment, İstanbul, Turkish.
- [4]. EIA (2012). Business Operation, Addis Ababa, Ethiopia.
- [5]. Mellese D., et al (2008). Overview of Environmental Impact Assessment in Ethiopia, Gaps and Challenges, Addis Ababa, Ethiopia.
- [6]. Finke, et al (2003). Financial risk and tolerance and wealth, Journal of Family and Economic Issues.
- [7]. Marla, et al (2011). Studies Orientation and Recommendation System (SORS): Use Case Model and Requirements.
- [8]. Sharon (2009). Individual investment behavior: A brief review of research, Personal Finance Research Centre, University of Bristol.
- [9]. Guy, et al (2010). Evaluating Recommendation Systems: Introduction to Recommender System.
- [10]. Halil, et al (2010). The analysis of factors affecting investment choices of households in turkey with multinomial logit model, Department of Economics, Faculty of Economics, İstanbul University , İstanbul, Turkey.
- [11]. UNCTAD) (2000). An investment guide to Ethiopia, opportunity and conditions, Ethiopia, Addis Ababa.
- [12]. de Janeiro (2008). Supporting entrepreneurship at the bases of the pyramid through business linkages, report of a roundtable dialogue, Brazil.
- [13]. EIA (2011). Investing in Ethiopia and a guide for new investors, Ethiopia trade and investment, Addis Ababa, Ethiopia.
- [14]. UNCTAD (2002). Investment and Innovation Policy Review, Addis Ababa, Ethiopia.
- [15]. Trade and Development office (1997). Micro and Small Enterprises Development Strategy, Addis Ababa, Ethiopia.
- [16]. Burke R (2006). Knowledge based recommender systems, University of California, Irvine.
- [17]. Schafer, et al (2007). Collaborative Filtering Recommender Systems. Department of Computer Science, University of Northern Iowa.
- [18]. Burke R (2007). Hybrid Web Recommender Systems, University of California, Irvine.

- [19]. Shimazu (2002). A Conversational Case-based Reasoning Tool for Developing Salesclerk Agents in E-Commerce Web shops. *Artificial Intelligence Review* 18 (3-4), 223–244.
- [20]. Fabiana, et al (2003). Case-based recommender systems: A unifying view, in *Intelligent Techniques for Web Personalization, IJCAI 2003 Workshop*, p. 2-10.
- [21]. EIA (2012). *Manuals of Ethiopian investment guide*, Addis Ababa, Ethiopia.
- [22]. McSherry (2001). Precision and Recall in Interactive Case-Based Reasoning. In *Case Based Reasoning Research and Development (ICCBR), Lecture notes in Artificial Intelligence*, pp. 392-306.
- [23]. Junker, et al (1999). On the Evaluation of Document Analysis Components by Recall, Precision and Accuracy. *Proceeding of the fifth International Conference on Document Analysis and Recognition*, Berlin, pp. 713-716.
- [24]. Salem, et al (2005). A Case Base Experts System for Diagnosis of Heart Disease. *International Journal on Artificial Intelligence and Machine Learning*, 5(1), pp. 33-39.
- [25]. Bergmann, et al (2005). Representation in Case-Based Reasoning. *The Knowledge Engineering Review*, 20, pp. 1-4.
- [26]. Investment advising (<http://www.investopedia.com/terms/i/investment-advice.asp>), Access on Tuesday 12 march 2013 @ 3:15 Am.
- [27]. Lipe, et al (1998). Individual investors' risk judgments and investment decisions: the impact of accounting and market data, *Accounting, Organizations and Society*, pp. 625-640.
- [28]. Biazen G. (2013). Application of case based recommender system in field of study selection, Addis Ababa University, Ethiopia.
- [29]. Getachew W. (2012). Application of case-based reasoning for anxiety disorder diagnosis, Addis Ababa University, Ethiopia.
- [30]. Alemu J. (2010). A Case-Based Approach for Designing Knowledge Base System for Addis Resource Center (ARC): The Case of Warmline Clinician Consultation Service, Addis Ababa University, Ethiopia.
- [31]. Henok B. (2011). A Case-Based Reasoning Knowledge Based System for Hypertension Management, Addis Ababa University, Ethiopia.
- [32]. Akerkar, et al (2010). *Knowledge Based Systems for Development*. pp 1 – 11.
- [33]. Fogel, et al (2006). *Defining Artificial Intelligence: Evolutionary Computation*. The Institute of Electrical and Electronics Engineers, Inc, Page 1.

- [34]. Saxena, et al (2011). Architecture for Medical Diagnosis Using Rule-Based Technique. The First International Conference on Interdisciplinary Research and Development. Thailand: Dayalbagh Educational Institute.
- [35]. Aref, et al (2000). A Knowledge based system for comfort analysis of internal environment of hotels.
- [36]. Alechina, et al (2012). Knowledge acquisition, representation, and reasoning, United Kingdom.
- [37]. Tehrani, et al (2009). A Conceptual Model of Knowledge Elicitation, College of Business, Technology and Communication, pp. 2.
- [38]. Ranjan, et al (2006). Knowledge Acquisition, Representation, and Reasoning. pp. 69–79.
- [39]. Wang, et al (2011). Knowledge elicitation approach in enhancing tacit knowledge sharing. Emerald, pp. 1039-1064.
- [40]. Gau, et al (1990). Knowledge Acquisition for a Diagnosis-Based Task. NATO ASI Series.
- [41]. Osuagwu, et al (2006). The Underlying Issues in Knowledge Elicitation. Interdisciplinary Journal of Information, Knowledge, and Management.
- [42]. Chen, et al (2007). Knowledge Representation and Reasoning Methodology based on CBR Algorithm for Modular Fixture Design. Journal of the Chinese Society of Mechanical Engineers, pp.593-604.
- [43]. Plaza, et al (1994). Case-Based Reasoning: foundational issues, methodological variations, and system approaches. AI Communications, pp. 39-59.
- [44]. Shiu, et al (2004). Case-Based Reasoning: Concepts, Features and Soft Computing. Applied Intelligence, pp 233–238.
- [45]. Abbas, et al (2008). A Hybrid Model for Knowledge Acquisition using Hierarchical Cluster Analysis. International Arab conference of e-Technology, IACe-T. Egypt, pp.1-14.
- [46]. Sagheb (2009). A Conceptual Model of Knowledge Elicitation: College of Business, Technology and Communication, pp.24-30.
- [47]. Birmingham, et al (2009). The Knowledge acquisition tool with explicit problem solving models. Cambridge Journal Online, 8(1), 5-25.
- [48]. Makhfi (2011). Introduction to knowledge modeling and neural network. Available at http://www.makhfi.com/KCM_intro.htm accessed date [Accessed March 1, 2013].
- [49]. Mason, et al (1997) .The business angel’s investment decision: an exploratory analysis. Entrepreneurship in the 1990s, Paul Chapman Publishing, London, pp 29-46.
- [50]. Morin, et al (1983). Risk aversion revisited. Journal of Finance, 38, pp.1201–1216.
- [51]. Michael, et al (2003). An empirical investigation of personal financial risk tolerance.

- [52]. Yemisrach (2009). Application of Case-Based Reasoning System in legal Knowledge-Based Systems: A Prototype in Children Criminal Cases in Ethiopia, Addis Ababa University, Ethiopia.
- [53]. Chakraborty (2010). Artificial intelligence; Knowledge Representation, Issues, predict logic, rules, Available at:http://myreaders.info/html/artificial_intelligence.html. Accessed date (March 17, 2013).
- [54]. Robert (2000). WUENIC, A Case Study in Rule-based Knowledge Representation and Reasoning, World Health Organization (WHO).
- [55]. Juan, et al (2009). jCOLIBRI 1.0 in a nutshell. A software tool for designing CBR systems, Dep. Sistemas Informáticos y Programación Universidad Complutense, De Madrid, Spain.
- [56]. Powell, et al (1997). “Gender Differences in Risk Behavior in Financial Decision-Making: An Experimental Analysis”, Journal of Economic Psychology, vol. 18, no. 6, pp. 605–628.
- [57]. Fag H, et al (2007). Case Based Reasoning for Logistics Outsourcing Risk Assessment Model. Proceeding of International Conference on Enterprise and Management Innovation, pp. 1133-1138.
- [58]. Stahl, et al (2008). Rapid Prototyping of CBR Applications with the Open Source Tool myCBR. German Research Center for Artificial Intelligence (DFKI) GmbH Image Understanding and Pattern Recognition Department (IUPR), German.
- [59]. García, et al (2008). JCOLIBRI2 Tutorial Document version 1.2, Group for Artificial Intelligence Applications Universidad Complutense, De Madrid, Spain.
- [60]. Lenz, et al (1998). Case-based Reasoning Technology: From Foundations to Applications. LNAI: State of the Art.s

Appendix I

Interview questions to domain experts

After introducing the objective of the study and requesting the respondents' participation in the study the interviewer records their answers for the following interview questions.

1. What type of advising system is given when investors come to invest?
2. Which type of investment activity or investment sector is more risky?
3. Which investment activity or investment sector is less risky?
4. What are the criteria to consider assigning investors in different investment sector?
5. What attributes are given more consideration in order to identify investment activity that best mach with the investor?
6. Which investment sectors are more preferable to female investors and which one is for male investors?
7. Which type of investment sector or investment activity it needs long time horizon to get benefit?
8. Which type of investment sector or investment activity it needs short time horizon to get benefit?
9. How much capital is needed to start investment?
10. What is the influence of investment location to be success in investment project?
11. Which investment activities are reserved for government, reserved for domestic investor, reserved for foreign investor?
12. What are the main reason for majority of investors are not successful?

Appendix II

Interview questions to investors

After introducing the objective of the study and requesting the respondents' participation in the study the interviewer records their answers for the following interview questions.

1. How to get advising systems from domain experts
2. What are the problems in advising systems of EIA
3. Is there any experienced investment experts that gives a brief advice on how to invest and where to invest?
4. How o select the investment sectors and investment activity?
5. Have you get any guidance from experts to select investment sectors and investment activity?

Appendix III

Prototype Evaluation form for the Domain Expert and investors

This is an evaluation form to be filled by anxiety disorder diagnosis experts in order to evaluate the applicability of the case-based reasoning system in anxiety disorder diagnosis. I thank you in advance for your willingness and valuable time.

Description of the parameter values are as follows.

Performance value	1	2	3	4	5
Description	Poor	Fair	Good	Very good	Excellent

Instruction: please assign (X) on the appropriate value for the corresponding parameter of evaluation questions of the case based recommender system in investment sector and investment activity selection.

No	Parameters of evaluation criteria	Performance value				
		1	2	3	4	5
1	Easy to use of the recommender system					
2	Is the system efficient in time					
3	Is the user interface interactive					
4	Adequacy and clarity of decision support					
5	Relevancy of the retrieved case in the decision making					
6	Fitness of the final solution to the new case					
7	Relevancy of the attributes in representing investors case					
8	Does the explanation facility give brief description about the recommended investment activity					
9	Rate the significance of the system in the domain area					

Appendix IV

Reserved investment sectors to investments

The following areas are reserved for domestic investors:

1. Areas of investment sector exclusively reserved for the Government:

- Postal services except courier services;
- Transmission and supply of electrical energy through the Integrated National Grid System; and
- Passenger air transport services using aircraft with a capacity of more than 20 passengers.

2. Areas of investment sector reserved for foreigners or domestic joint-venture investment with the government:

- Production of weapons and ammunition
- Telecommunication services

3. Areas of investment sector exclusively reserved for domestic investors:

- Export of raw coffee, chat, oil seeds, pulses, hides and skins bought from the market, and live sheep, goats and cattle not raised or fattened by the investor;
- Import trade (excluding LPG, bitumen and, upon approval from the Council of Ministers, material inputs for export products);
- Retail trade and brokerage; and
- Wholesale trade (excluding supply of petroleum and its by-products as well as wholesale trade by foreign investors of their locally produced products).
- Agriculture, including agribusiness and processing for exports;
- Bakery products and pastries for the domestic market;
- Barber shops, beauty salons, and provision of smith, workshops and tailoring services except by garment factories;
- Building maintenance and repair and maintenance of vehicles;
- Car-hire and taxi-cab transport services;
- Commercial road transport and inland water transport;
- Construction companies excluding those designated as Grade 1;

- Customs clearance;
- Grinding mills;
- Hotels (excluding star-designated hotels), motels, pensions, tea rooms, coffee shops, bars, night clubs and restaurants (excluding international and specialized restaurants);
- Museums, theatres and cinema hall operations;
- Printing industries;
- Saw milling and timber-making;
- Tanning of hides and skins up to crust level; and
- Travel agency, trade auxiliary and ticket selling services.

4. Areas of investment sector that foreign investors are possible to invest but these areas are not reserved for foreign investors only it is also for domestic investor:

- Manufacturing industries (including food, beverages, chemicals and pharmaceuticals, plastics, metallic and non-metallic products, paper products, leather and leather products, textiles and garments);
- Agriculture, including agribusiness and processing for exports;
- Real-estate development;
- Education and health services;
- Grade 1 construction contracting;
- Mining and quarrying of gold, marble and granite; and
- Engineering and management consultancy.

DECLARATION

This thesis is my original work and has not been submitted as a partial requirement for a degree of master in any other university.

Yibeltal Chanie

June, 2013

Advisor: Gashaw Kebede (PHD)