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

Yibeltal Chanie Manie, Abebe Belay Adege

Abstract— Investment is a commitment of funds, directly or indirectly, to one or more assets with the expectation to enhance future wealth. Each 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 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, 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. The acquired knowledge was represented using feature value case base representation and implemented using jCOLIBRI programming tool. 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, 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 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. And also further research can be done by developing the case based recommender system using different local languages for investors can easily communicate with the system by using their own languages.

Index Terms— Knowledge-based system, Case Based Reasoning, PROLOG, Investemnt area, investment sector

1 INTRODUCTION

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. 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 frame-

- ***Yibeltal Chanie Manie*** *is currently pursuing PHD degree program in electric engineering and computer scienece in National Taipei university of Technology, Taiwan, PH-0966647602. E-mail: yibeshmamaru@gmail.com*
- ***Abebe Belay Adege*** *is currently pursuing PHD degree program in electric engineering and computer scienece in National Taipei university of Technology, Taiwan, PH-01123456789. E-mail: abbblybelay@gmail.com*

work for investment [13].

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]. 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].

2 STATEMENT OF THE PROBLEM

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 can give advice on investment sector and investment activity selection more suitable to each investor's.

Difficulty of getting investment advice is a critical issue for investors since knowing the right investment area is a key factor and knowing the right company to invest is another factor to consider to new investors [21]. 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.

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.

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.

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 identify and collect the previous cases, facts, insights and rules that new investors need to know 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 develop a prototype case based recommendation system to investors on the selection of investment Sector and investment activity that best matches with their characteristics.

Methodology

1344 cases represented as case base in a CSV format that are used as previously solved cases. 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. Other case representation methods like relational database case representation, predicate based representation and soft computing case representation methods have their own advantages and disadvantages. Feature-value case representation method is a process of representing a case as a vector of attribute-value pairs, like 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]. 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 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 12 intends to be a standard interchange language of data between computers, not to be managed directly by humans. [55].

The developed prototype case based recommender system is tested and evaluated to ensure the performance of the system whether meeting the objective or not. 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]. 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 researcher also made an experiment on case similarity testing to know how new cases are matched with the cases from the case base.

KNOWLEDGE ACQUISITION AND MODELING

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 model-

ing them in using hierarchical structure. In this research, Primary sources of knowledge are collected from domain experts in EIA, investors and from Relevant Document

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. 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. In the following hierarchical structure figure 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.

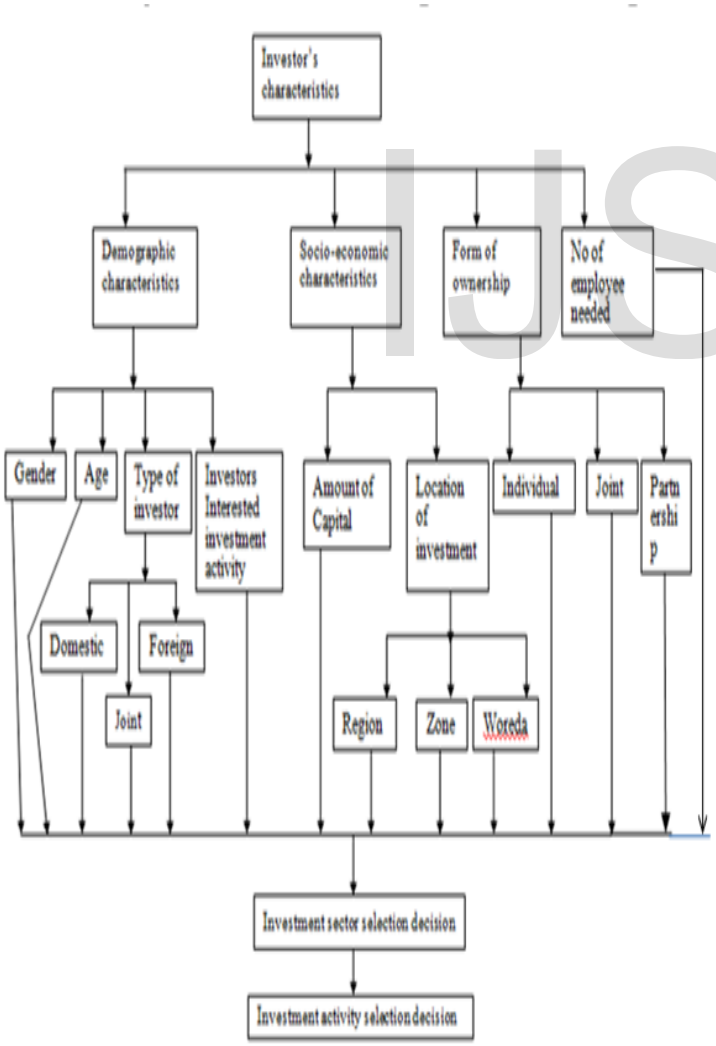

Figure 1: Hierarchical structure of investment sector and investment activity selection

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

The architecture of the CBRISAIAS system shown in figure 2 depicts how the prototype works during investment sector and investment actvity 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

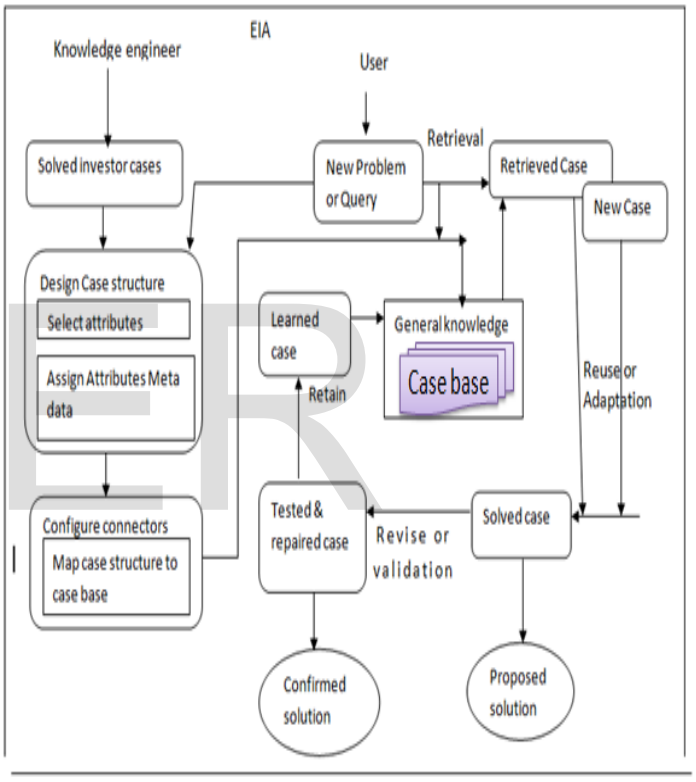

Figure 2: CBRISAIAS Architecture

Description of CBRISAIAS Case Attributes

case is composed of three components: description (describes the problem), solution (represents a possible solution approach) and result (reveals if the proposed solution can solve the problem). Description and solution are collections of simple or compound attributes, permitting us to build a hierarchical case structure. After that the new windows will appear for configuration of case structure as shown in the figure 3. In the figure, the left side's options are case structure description, solution and result.

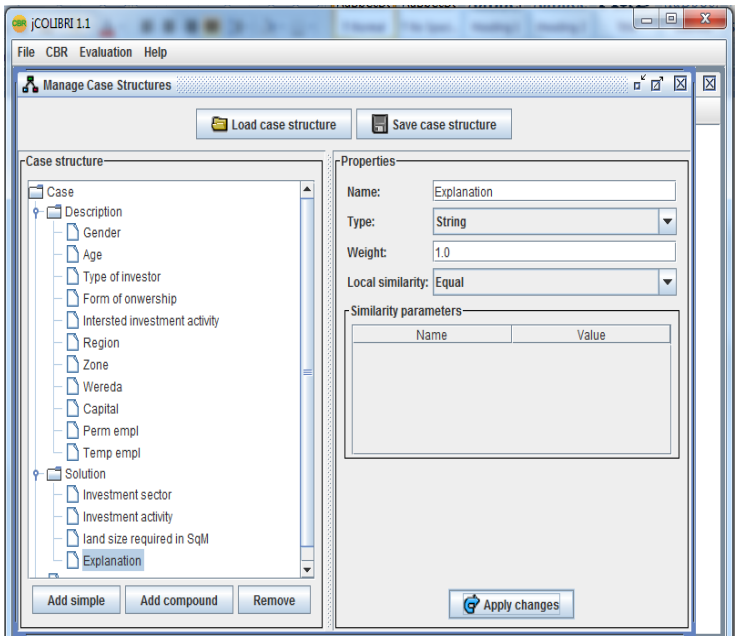

Figure 3: Defining Case Structures and similarity

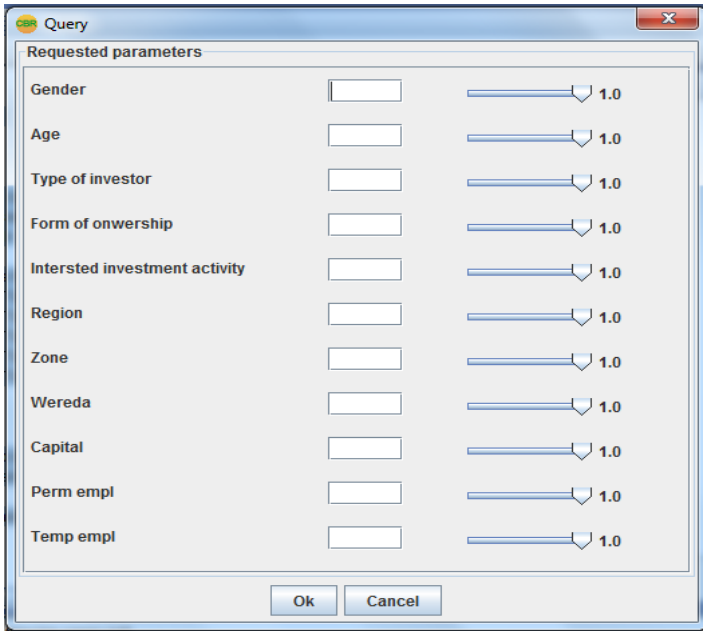

Figure 4: Window for Case Entry into the Case Base

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.

Case Similarity, Matching and Ranking

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.

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. 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.

TESTING AND PERFORMANCE EVALUATION OF THE PROTOTYPE

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.

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 1.

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 1: Query similaritywith their corresponding cases from the case base

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. The average recall and precision results 85% and 64% respectively which is also a promising result. In terms of recall this research

achieved a very good result. 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. Because of accuracy result the developed prototype system has the capability to advise and recommend in investment sector and investment activity selection correctly.

Then after calculating the similarity measurement and calculating recall and precision the next step is evaluate the user acceptance testing. This research uses questionnaires to evaluate user acceptance of the CBRISAIS 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. 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 user's interface satisfaction. 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. 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.

As shown in table 2, 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.

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 2: the CBRISAIAS system performance evaluation by the investor

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 3: the CBRISAIAS system performance evaluation by the domain expert

On the other hand, table 3 below shows the performance evaluation of the prototype by the investors.

As shown in table 3, 8% of the respondent's rate eases 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.

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. 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.

Conclusion

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 112 (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 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 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 average user acceptance evaluation achieved 82% and 84 % performance by domain experts and investors respectively, this shows the result of system performance indicated that users are satisfied with proposed system

Future works

Therefore, the researcher recommends the following issues as a future research direction based on this study.

- 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
- Hybrid strategy approaches should be investigated which combines rule based reasoning and case based reasoning.
- In this study the explanation facility given by the proposed system is not user interactive.
- Further investigation can be conducted by developing a CBRISAIAS system in different local languages.

ACKNOWLEDGMENT

First and foremost, we would like to give the almighty God who provided me everything to finish this thesis. Next to this, we gratefully thanks to my advisor Dr. Gashaw Kebede, for his valuable commitment, patience reading for 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. We would like to thank Ethiopian investment agency domain experts of each investment sector and also Mr. GirumTadesse (investment promotion senior expert), Yonas Latamo (promoation 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. 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.

REFERENCES

[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.