



CASE-BASED REASONING SYSTEM TO RECOMMEND DESTINATION COUNTRY FOR HUMAN MIGRANTS

Baye Yemataw, Information Systems, Debre Berhan University

Prof. A. N. Mohamad, Department of Mathematical Modeling, Debre Berhan University

Abstract: *Human migration is the challenge process in Ethiopia, special illegal human migrants they across different border countries without any basic information or knowledge. Due to this, there is difficulty of getting vast amount of information to know the destination countries, simple they go without any information, also wastes more time, cost, violet their human right and also it danger for their life.*

Hence, the objective of this thesis is to build case based automatic recommender (CBR) system in case of human migration in Ethiopia. The methodologies employed in this study are data collection tools like interview, questionnaire and assessment of existing documents. The system is tested in terms of performance the system and its user acceptance. From these perspectives, the average performance, recall and precision of the system is score 92% (4.60), 0.91 (91%) and 0.63 (63%) respectively from domain experts and system performance. Developing a recommender system in this study therefore helps migrants to have a right and clear decision about the destination countries for human migrants.

Keywords: *Migrant, Semi-Permanent, Destination Country, CBR-System*

1.1. BACKGROUND OF THE STUDY

According to IOM[1], “no commonly conventional definition for migration exists”. In reality, there are a lot of definitions of human migration, and it defines in a different way. It can be define as “a process of moving, either across different border countries, or within a state; it is a peoples movement, nearby any kind of movement of peoples, whatever its length, composition, and causes; it includes migration of refugees, displaced persons, uprooted people, and economic migrants”[1]. On the other hand, Encyclopedia Britannica defines the term ‘migration’ as being “a permanent change of residence by an individual or group; it excludes such movements as migrant labor, travel and tourism, all of which are transitory in nature”[2]. Finally,[1], defines the migration as being a long-term movement of an individual



people or group of peoples outside their place of migrant's source. In general these processes are it may be illegally or legally.

As noted by [1], "since the dawn of human progress, more than 70,000 years ago, humans have migrated across different border countries to search for food, shelter, safety, and hospitable climate". The movement is continuing in the beginning of 21st century [2], but new reasons for human migration are arising too, such as job related movement[3]. Due to the great technological advances in electronic communications and transportation, the scale of modern human migration is even bigger and more dynamic than ever before. It is a global problem of huge strength, involving people from various countries, of different age, gender, and race, belonging to various nationalities and religions, and having various social, economic and political background problems [4].

As stated in [5], illegal migration front multiple challenges or problems to countries of source, bridge and destination, as well as to migrants themselves. Migrants in illegally situation are particularly open to bias, exploitation and violence. Such migrants are also in danger of being exploited by crime organizations involved in human trafficking and migrant smuggling crimes that constitute a serious violation of the human rights of its sufferers[5].

1.2. STATEMENT OF THE PROBLEM

The current practices to control illegal migration of Ethiopia Government made to control illegal migration the first option was to prevent human trafficking, by coordinating both regional and national policy to make awareness the dangers of illegal migration in Ethiopia, by create TV program. The government monitored the activities of labour staffing agencies and closed in different agency that were facilitating the migration for danger condition. And also, make labor migration right discussion in the previous reporting period with Jordan, Kuwait, and Qatar stay in place, the government discuss new agreements in 2013 with the Governments of Djibouti, Sudan, the UAE, and Kenya. However, these agreements did not explicitly contain provisions to protect workers, such as by outlining mandatory rest periods, including grounds for filing grievance, and prohibiting recruitment fees.

One of the challenging problems in the world to control illegal migrants is the absence of [5] recorded information for analysis purposes. Ethiopia also faces this problem, since most young Ethiopians migrant follows illegal system. As a result, the country loses its young



human power every year. Therefore law enforcement bodies like police need to learn the factors that constitute human migration. To control migration, there should be a need for careful migration prevention strategies and policies.

The challenging problem in Ethiopia is that most of the migrants cross the border of the countries in illegally system. They move and across different life cycles to arrive the destination countries[5]. In the midst of illegal migration, with the hope of finding better jobs, a considerable pack of nationals leave Ethiopia with the help of human traffic, and their preferred destinations are European countries.

Since, in Ethiopia there is no local research to recommend destination countries for human migrants to give basic information about the migrant where they go and also make awareness for illegal migrants. So, we design a prototype CBR system to solve those problems in Ethiopia.

Based on the above stated problems, the study addresses the followings research questions:

- ☞ What are factors of human migration in Ethiopia?
- ☞ How to identify variables that help to recommend destination country?
- ☞ How to recommend?
- ☞ To what extent the recommender system gets users acceptance?

1.2.1. General Objective

The general objective of this research is to develop a knowledge based recommender system for human migrant to enhance the selection of destination countries.

1.2.2. Specific Objectives

To achieve the general objective, this study targets on the following specific objectives:

1. To identify important attributes that help in constructing a model to recommend the destination country for migrants.
2. To design architecture of the recommender system and select suitable techniques.
3. To develop a prototype recommender system.
4. To create similarity functions using CBR.
5. To evaluate the performance of the prototype.



1.3. SCOPE AND LIMITATION OF THE STUDY

The main aim of this thesis is to construct knowledge-base model for develop a prototype system of recommend destination countries for migrants. This work is conducted based on the data obtained from the record office of central statically agency and MoLSA, comprising 84, 000 and 459, 810 datasets from 2000-2014 years. The selection of datasets was using clustering and random sampling techniques.

The main limitation of this study of was the availability of the data in unorganized. To organize the data in suitable format for experimentation, it took a lot of time. Due to the time and financial limitations available to complete the study, the researcher is forced to use data from a single source: central statically agency found in Debre Berehan and MoLSA found in Addis Ababa. And also the system recommend only for the new migrants, is not recommend user selection destination countries.

1.4. RELATED WORKS

Here in this section, the researcher discusses different research papers that are related to case based reasoning in different areas conducted locally in Ethiopia which helps to understand on how to solve and develop case based reasoning to recommend destination countries for human migrants.

Biazen [6], has done his research with title of “application of case based recommender system in field of study selection in the case of higher education in Ethiopia”. The objective of the study was to develop a prototype case base recommender system that assists the students in their field of study selection process. That helps and recommends to the students based on previously solved cases and new query given by the users/students. With the collection of 105 cases that are collected from successful students. For the implementation the author uses JCOLIBRI as an implementation tool to develop the prototype recommender system. Queries from users 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 helps the students to make decisions easily. 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 was 77.2% and 80.2% by the domain experts and students respectively.

This study is based on a data collected from central statically agency and MoLSA office. Before implementing the Knowledge-base (CBR) system, the researcher clearly understands the migrant's factors, and has worked closely with the two dataset source.

To understand the migrant's factors to migrate legally, the first step is identifying the migrant's causes to understanding the characteristics of migration that determines migrant's goals and DM goals were set to identify and prepare data required for the study. Next, CBR system task is ready for retrieve, retain, revision and also to recommend destination country for human migrants in Ethiopia.

1.5. OVERVIEW OF MIGRATION

Migration can be defined as a process of moving, either across in different border countries, or within a State. Any kind of migration in the world and also in Ethiopia, people, whatever their pulls and push causes or factors to leave their lands [3]. Migration is part of the human history since its very beginning. People have migrated from one continent to the other, from country to country or internally, inside the same country. Currently, IOM states that there are "about one billion migrants around the world" [1]. This number includes "214 million international migrants and 740 million internally displaced persons" [1].

1.5.1. Illegal Migration in the world

The range of illegal migration and forced migration in the Horn of Africa region and neighbouring countries is significant. It is estimated that there are [37]: 37 In "2014, 91,500 illegal migrants arrived in Yemen after crossing the Red Sea from Djibouti or Somalia", the vast majority were Ethiopians intending to migrate the destination country to Saudi Arabia, some part of Somalias migrant stay in Yemen and also in "2014, around 2,000 Eritrean illegal migrants" arrived in neighbouring Ethiopia and Sudan every month.

1.5.2. Types of Migration

Migration especially in Ethiopia is the illegal movement of people from one place in the world to another for the purpose of taking up "permanent or semi-permanent residence", usually across to change their life [38]. 38 An example of "semi-permanent residence" legal



or illegal migrants would be living a short time movement of migrant until they get their needs.

1.5.3. Causes for Migration

In the world in general and in Ethiopia, peoples across different borders of the countries legally or illegally for variety of reasons. The migrants they consider the “advantages and disadvantages of staying versus reasoning, as well as factors such as the distance destination countries, travel costs, travel time, transportation system, security, and cultural barriers” [2].

Push Factors: Reasons for migration because of a difficulty.

Pull Factors: Reasons for migration in different countries because of something advantageous. Several types of push and pull factors may influence people in their movements, including: environmental, political economic and cultural [2]. These could be the reasons of people to be motivated for migration especially in Ethiopia.

Description of the Initial Data

The data collected from MoLSA and central statically agency offices are using the same standard paper format and they are not directly collected for the purpose of this study. Initially the data were in record file format. The data in record file format are then exported to Microsoft Excel for further preparation and ready for as input to WEKA 3.6.9 data mining tool. This initial dataset is described and visualized using Microsoft excel. Then data preprocessing is performed to verify the quality of the dataset such as missing values, error values and to obtain high level information regarding the data mining questions. Table 3-1 shows the dataset information which is collected from the MoLSA and central statically agency office.

Table 1-1: Description of initial dataset

Data source	Categories	Total
MoLSA office	No of Instances	459, 810
	Number of Attributes	11
Central statically agency office	No of Instances	84,000
	Number of Attributes	10

As shown in Table 1-1 the initial data collected from MoLSA and central statically agency office has 11 and 10 attributes respectively.



The following Figure 1-1 shows that the selected attributes for this study after discussed with the two data source domain experts.

1	Gender	Age	LevelOfQualification	Marital_status	Region	Zone	YearOfMigration	IncomeLevel	Destination
2	Male	young	College Level	Unmarried	Amhara	South Gondar	2000	Low	Israel
3	Male	young	College Level	Unmarried	Amhara	South Gondar	2000	Low	Israel
4	Male	young	College Level	Unmarried	Amhara	South Gondar	2000	Low	Israel
5	Male	young	College Level	Unmarried	Amhara	South Gondar	2000	Low	Israel
6	Male	young		Unmarried	Amhara	South Gondar	2000	Low	Israel
7	Male	young	College Level	Unmarried	Amhara	South Gondar	2000	Low	
8	Male	young	College Level	Unmarried	Amhara	South Gondar	2000	Low	Israel
9	Male		College Level	Unmarried	Amhara		2000	Low	Israel
10	Male	young	College Level	Unmarried	Amhara	South Gondar	2000	Low	Israel
11	Male	young	College Level	Unmarried		South Gondar	2000	Low	Israel
12	Male	young	College Level	Unmarried	Amhara	South Gondar	2000	Low	Israel
13	Male	young	College Level	Unmarried	Amhara	South Gondar	2000	Low	
14	Male	young	College Level	Unmarried	Amhara	South Gondar	2000	Low	Israel
15	Male	young	College Level		Amhara	South Gondar	2000	Low	Israel
16	Male	young	College Level	Unmarried	Amhara	South Gondar	2000	Low	Israel
17	Male	young		Unmarried	Amhara	South Gondar	2000	Low	Israel
18	Male	young	College Level	Unmarried		South Gondar	2000	Low	Israel
19	Male	young	College Level	Unmarried	Amhara	South Gondar	2000	Low	
20	Male	young	College Level	Unmarried	Amhara	South Gondar	2000	Low	Israel
21	Male	young	College Level	Unmarried	Amhara	South Gondar	2001	Low	Israel
22			College Level	Unmarried	Amhara		2002	Low	Israel
23	Male	young	College Level		Amhara	South Gondar	2003	Low	Israel
24	Male	young	College Level	Unmarried	Amhara	South Gondar	2004	Low	Israel

Figure 1-1: Dataset before Experimentation

1.5.4. Data Cleansing

There were numerous inconveniences in the original dataset which need additional pre-processing the migrants' data. These are:

- ☞ Attributes have so many missing values. For example, the attribute migrants region has a missing value of 51187, 8426 from MoLSA and central statically agency office respectively. The researcher removes each instance from the dataset, because had enough instance in the dataset.
- ☞ Discretize or organize an instance filter that discretizes a range of numeric attributes in the dataset into nominal attributes.
- ☞ In the original dataset, there was an error on changing the attribute values from attribute to attribute. For instance, some records had age value in sex attribute and sex value in age attribute. Like the attribute age had a value of "M" at some place "F" in another place. Whereas, the attribute sex had a value of number in some places.



☞ Another problem in the dataset was that there was inconsistency or irregularly in writing, way of attributes i.e. some attribute were written in abbreviation format and some attribute in standard form. For example, in the attribute migrants region, one of its attribute regions was written as reg. in some places and the full name region in other places. These are simple tasks for the researcher to full fill these problems.

Gender Nominal	Age Nominal	LevelOfQualification Nominal	Marital_status Nominal	Region Nominal	Zone Nominal	YearOfMigration Numeric	IncomeLevel Nominal	Destination Nominal
Male	Adult	College Complete	Married	Amhara	East G...	2010.0	Low	Saudi Arabia
		College Complete	Married	Amhara	East G...	2010.0	Low	Saudi Arabia
Male	Adult	College Complete	Married	Amhara	East G...	2010.0	Low	Saudi Arabia
Male	Adult		Married	Amhara	East G...	2010.0	Low	Saudi Arabia
Male	Adult	College Complete	Unmarried	Amhara	East G...	2010.0	Low	Saudi Arabia
	Adult	College Complete	Married	Amhara	East G...	2010.0	Low	Saudi Arabia
Male	Adult		Married	Amhara	East G...	2010.0	Low	Saudi Arabia
Male	Adult	College Complete	Unmarried		East G...	2010.0	Low	
	Adult		Unmarried	Amhara	East G...	2010.0	Low	Peru
Male	Adult	College Complete	Unmarried	Amhara	East G...	2010.0	Low	Saudi Arabia
	Adult	College Complete	Married	Amhara	East G...	2010.0	Low	Dubai
Male	Adult		Married	Amhara	East G...	2010.0	Low	Saudi Arabia
Male	Adult	College Complete	Married	Amhara	East G...	2010.0	Low	Saudi Arabia
Male	Adult	College Complete	Unmarried		East G...	2010.0	Low	
	Adult			Amhara	East G...	2010.0	Low	Saudi Arabia
Male	Adult	College Complete	Married	Amhara	East G...	2010.0	Low	Saudi Arabia
Male	Adult	College Complete	Married	Amhara	East G...	2010.0	Low	Peru
Male	Adult		Married	Amhara	East G...		Low	Saudi Arabia
	Adult	College Complete	Married	Amhara	East G...	2010.0	Low	Saudi Arabia
Male		College Complete	Married	Amhara	East G...	2010.0	Low	
Male	Adult		Unmarried		East G...	2010.0	Low	Saudi Arabia
Male	Adult	College Complete	Married	Amhara	East G...	2010.0	Low	Saudi Arabia
Male	Adult	College Complete	Married	Amhara	East G...	2010.0	Low	Saudi Arabia
Male	Adult	College Complete	Unmarried	Amhara	East G...	2010.0	Low	Saudi Arabia
Male	Adult	College Complete	Unmarried	Amhara	East G...	2010.0	Low	Saudi Arabia
Male	Adult	College Complete	Unmarried	Amhara	East G...	2010.0	Low	Saudi Arabia
Male	Adult	College Complete	Married	Amhara	East G...	2010.0	Low	Saudi Arabia

Figure 1-2: Before Preprocessing Some Missing attribute value MoLSA Office Dataset

Some attribute names are different in different places. For instance, the attribute name “sex” in some places is and “gender” in another place. In Table 1-2 below, the total missing instance from Ethiopian MoLSA dataset. The next task of the study is to rewrite each instance. It must be reprocessing in manually system by filtering each missing attribute values, in Table 1-2 it show the number of missing after ignoring missing values. After removing, then the researcher imported the data to Microsoft Office Excel 2007 and changed to with appropriate ways as indicated in Figure 3-3.



Table 1-2: Missing Values in the Dataset

No	Station	Number of records	Numbers of records missing values	Number of records after Ignoring missing values
1	MoLSA office	459, 810	51187	354623
2	Central statically agency office	84,000	8426	75574

For this reason, the study takes time to correct and normalize the above stated problem of the migrants' dataset sequentially by manually system, using Microsoft Excel. After removing the missing values in the dataset, the dataset MoLSA decreased from 405, 810 rows data to 354623 rows and also decreased the dataset central statically agency from 84,000 instances to 75574 instances as indicated in above Table 1-2.

1.5.4.1. Random Sampling

In this study, the researcher already know that the number of cluster and also we, fixed in each instance to use numbers of sample in each destination country can be applied random sampling techniques. As shown in Figure 1-3, for example to select the number of sample from destination country using Microsoft Excel by selecting the total migrants and also by selecting random option and typing the number of sample we take the number of sampling as shown in Figure 1-3 organize the sample instance in each countries ready for experimentation. But if there is reputation from the sampling instance, we apply sample again and again to make clear the redundancy of the dataset. As shown in Figure 1-4, the researcher take total instance in each destination country by applying random sampling techniques, finally, after integration the two data source we used 75,920 instances or rows for experimentation.

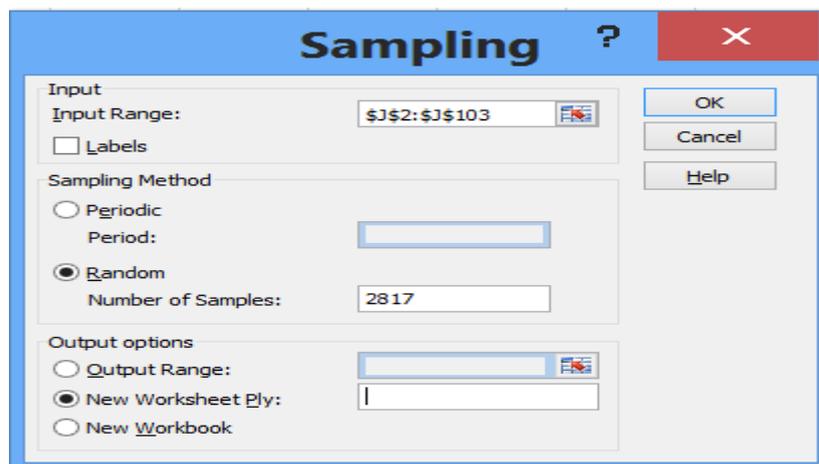


Figure 1-3: To select Number of sampl from destination country India



```
=== Attribute Selection on all input data ===  
  
Search Method:  
    Attribute ranking.  
  
Attribute Evaluator (supervised, Class (nominal): 9 Destination):  
    Information Gain Ranking Filter  
  
Ranked attributes:  
    1.896    6 Zone  
    1.461    7 YearOfMigration  
    1.417    3 Educational Level  
    1.241    2 Age  
    1.207    5 Region  
    0.475    1 Gender  
    0.419    8 IncomeLevel  
    0.418    4 Marital_status  
  
Selected attributes: 6,7,3,2,5,1,8,4 : 8
```

Figure 1-5: Attribute selection using by Information Gain (WEKA)

1.6. CONCLUSIONS

The objective of this research was to solve the problems faced for human migrants to know basic information human migrant for destination countries. In this case, a case based reasoning would be more suitable. A case based system stores previously solved and successful cases, reasons by comparing a problem case to solved cases and retrieve human factors or causes. The solution of a CBR system can be revised by testing its success in the real world or by giving to domain experts. It also increments its knowledge by learning from its own solving experience. Recommender systems are tools and techniques providing suggestions aimed at supporting users in decision making. In this research, a case based recommender system which uses case based reasoning to suggest the destination countries for human migrants.

To undertake the from section stated problems, the researcher initiated to conduct a research having the main goal of developing a prototype CBR system to recommend destination country for human migrants.

The CBR system uses the well-known CBR cycles (Retrieval, Revise and Retain) to perform different tasks. In this recommender system, the first task is retrieval of cases by entering a new problem description (case) by using the query window. Next case similarity computation is performed and retrieves most similar cases. The retrieval task of the prototype used in this study is Nearest Neighbour retrieval algorithm.



Regarding to the evaluation process of the system, the recommender system registers encouraging retrieval performance which is an average value of 91% recall and 63% precision. The system was also evaluated from the users' side which is called user acceptance testing. Then domain experts the average user acceptance evaluation registered 92% performance by domain experts.

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

- ✓ The major attribute that have more influence to migrate: "ZONE", "YEAROFMIGRATION", "EDUCATIONAL LEVEL", "AGE", "REGION", "GENDER", "INCOMELEVEL" and "MARITAL STATUS".
- ✓ 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. And also the proposed system contributes in the recommendation of destination country.
- ✓ We design prototype system to make simple select optional attribute for end-users for recommend destination country human migrants in Ethiopia.

1.7. RECOMMENDATIONS

From the result of the study, the researcher is recommendations are the following:

- ☞ For this study, the researcher collected the relevant attribute from Ethiopian MoLSA office. To analysis the migrants' destination Countries and to know the factors of migrants, it needs further research by adding other important attributes such as family history, their spoken language, push and pull factors.
- ☞ It is known that most of the young Ethiopians people go across different Arabic countries to make their life better economically. So, further research must be needed to create awareness for the society by integrating Knowledge-base system into mobile application.
- ☞ Some manual processes are needed to determine human migration status in Ethiopia. So further research is needed to create a web based destination countries' recommendation system is very important to come up illegal migrants into legal



migrants and for decision making, by applying a hybrid DM model to implement a recommender system prototype for analyze different factors from migrants.

☞ To make the system effectiveness and accurateness, it is important integrating CBR with simple k-means DM algorithm techniques for further research.

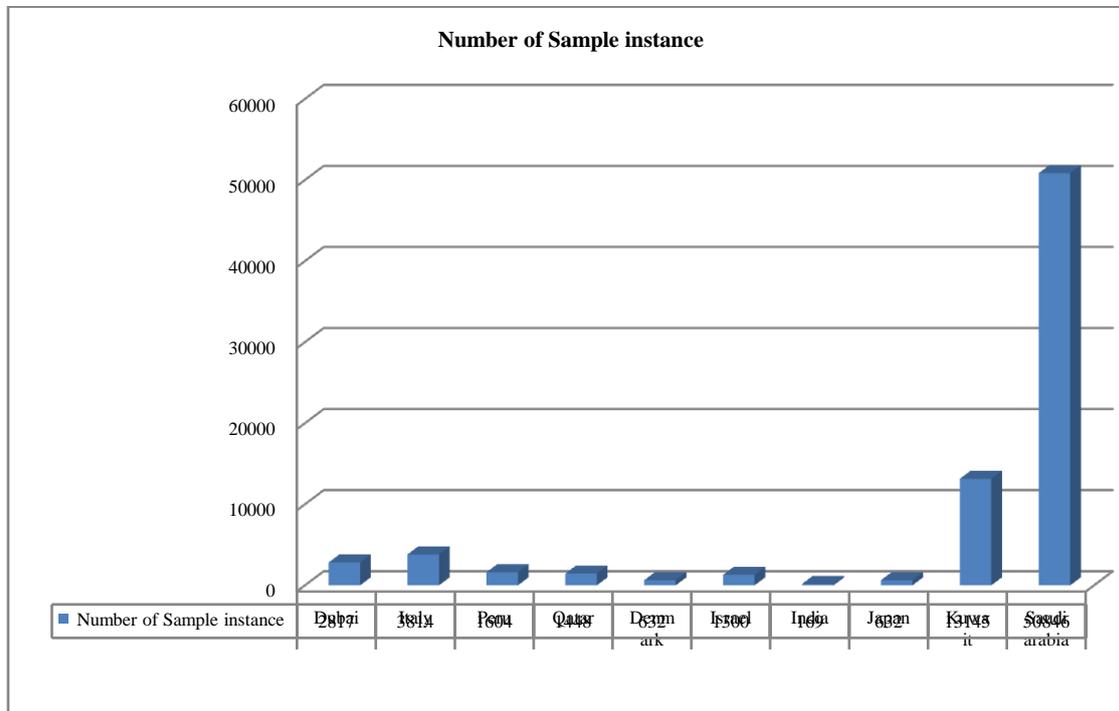


Figure 3-7: Number of Sample instance in each Country

REFERENCES

- [1] R. Perruchoud. (2015, May) from International Organization for Migration (IOM). [Online]. http://publications.iom.int/bookstore/free/IML_1_EN.pdf Oct. 02, 2014 [Jun. 29, 2015].
- [2] Ilija Hristoski Kosta Sotiroski, "Conceptual Data Modelling of Modern Human Migration," *Jornal of Migration and Human security*, vol. Vol.7, no. Human Migration, pp.12-20, October 2012.
- [3] H. Boubakri, "Revolution and international migration in Tunisia. Migration policy Centre research: European University Institute.," in Florence, Robert Schuman Centre for advanced studies., 2013.
- [4] Alexander Betts. (2015, October) the guardian. [Online]. <https://www.theguardian.com/commentisfree/2015/sep/20/migrants-refugees-asylum-seekers-21st-century-trend> Jun. 05, 2013 [May. 29, 2015].



- [5] Yemisrach Feraw, "EXPERIENCE OF EIGHT DEPORTED WOMEN ETHIOPIANS WHO RETURNED FROM KINGDOM OF SAUDI ARABIA," Addis Ababa University , Addis Ababa, JULY, 2015.
- [6] Biazen Getnet., "Application of case based recommender system to advise students in field of study selection at higher education in Ethiopia," Addis Abeba University, Addis Abeba, Master's thesis.
- [7] Christopher Horwood, "Irregular Migration Flows in the Horn of Africa: Challenges and implications for source, transit and destination countries," Australian Government, Australian, 2015.
- [8] Population Reference Bureau. [Online]. [http://www.prb.org/Content/NavigationMenu/PRB/Educators/Human_Population/Migration2/MJun. 10, 2012](http://www.prb.org/Content/NavigationMenu/PRB/Educators/Human_Population/Migration2/MJun.10,2012) [Sep. 29, 2016].