International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 19 (2018) pp. 14431-14437

Design and Development of Recommender System for Target Marketing of Higher Education Institution Using EDM

S. JothiLakshmi [1], Dr. M.Thangaraj [2]

[1]Assistant Professor, Department of Computer Science, Sourashtra College, Madurai, India.
[2]Professor&Head, Department of Computer Science, Madurai Kamaraj University, Madurai, India.

Abstract

A recommender system is a special type of information filtering system and it is very popular in e-commerce, entertainment, social networking, higher education sector etc. Target marketing involves breaking down the entire market into various segments and planning marketing strategies in the business world. Educational data mining is a powerful tool to extract knowledge from educational information system. This paper provides a unique design of recommender system that recommends different solutions for the effort of target marketing of higher educational institution.

Keywords: Data mining, Target Marketing, Recommender, Agent, EDM

INTRODUCTION

Huge volume of educational data is generated by various higher education sectors. These massive data cannot be handled by traditional learning system. They need to be handled and data being utilized using mining technologies [14].

The main objective of the recommender system [6] is to prioritise information about items such as music, news, books, image or web page to user with respect to their interest. The selection is based on the knowledge of user behaviour or the knowledge of all items in the database. Each of these systems employs algorithms which identify user with similar preference. Recommender systems used in education sector to generate different recommendations for students, teachers and educational institution [14].

Data mining is a technique of extraction of hidden information from large amount of database. This powerful technology is being used in universities or institutions to predict future trends and behaviour pattern. EDM[12] covers DM methods with respect to the structure of educational data. EDM deals with analysis of study related data as to understand student behaviour and applied to provide effective learning environment. The educator used EDM techniques to design the structure of course. Nowadays the researchers utilize EDM technique to guide the student learning environment, develop or refine student model, measuring the effect of individual interventions, improve teaching support or predict student performance and behaviour.

RELATED WORK

This section provides a study of all the previous EDM Researches and also review the available educational recommender system. The previous predictive model based on only the student performance. This Paper[7] reviews the student weak point and provide recommendation to the student in the higher education institutions. They do not focus on the other Academic stages. Bravo Leila shatli developed a recommender system[5], for University to help the students to take decisions on their academic itineraries. They consider only students performance. The other aspects of higher Education were not studied.

This work [14] reviews the collaborative filtering based recommendation techniques to recommend elective courses to students depending upon their grade points. The other courses are left out by their study. The Amer AI-Badarnah,Jamal Alsakaran designed an automated recommenders system for course selection using with collaborative filtering systems[2]. They focus only on the course selection with limited data set. Maria Iuliana Dascalu had developed recommender system for educational materials and tools[10].

The previous recommended system, most of them are student oriented and they predicted the performance of a specific subject. The existing systems mostly focused on predicting the student's performance in a particular course. They are not focused on the Distance learning, course material, faculty evaluation, institution services, socio economic back ground of the student, access to the institution and location of the student. The other models have not taken the above factors all together. The proposed system takes all these factors here into the proposed recommender system.

PROPOSED FRAME WORK

The proposed recommender system IREDM **(Integrated Recommender Educational Data mining)** which considers attributes in Distance learning programmes such as course material, student performance, faculty evaluation, institution services, socio economic back ground of the student, access to the institution and location of the student to take decision in Distance Learning programmes.

IREDM architecture has a layered design in order to organize its components. Task and responsibilities of the IREDM are distributed. So it is easy to maintain. The architecture of educational recommender systems is shown in Fig. 1. Each layer has a function explained as follows:

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 19 (2018) pp. 14431-14437

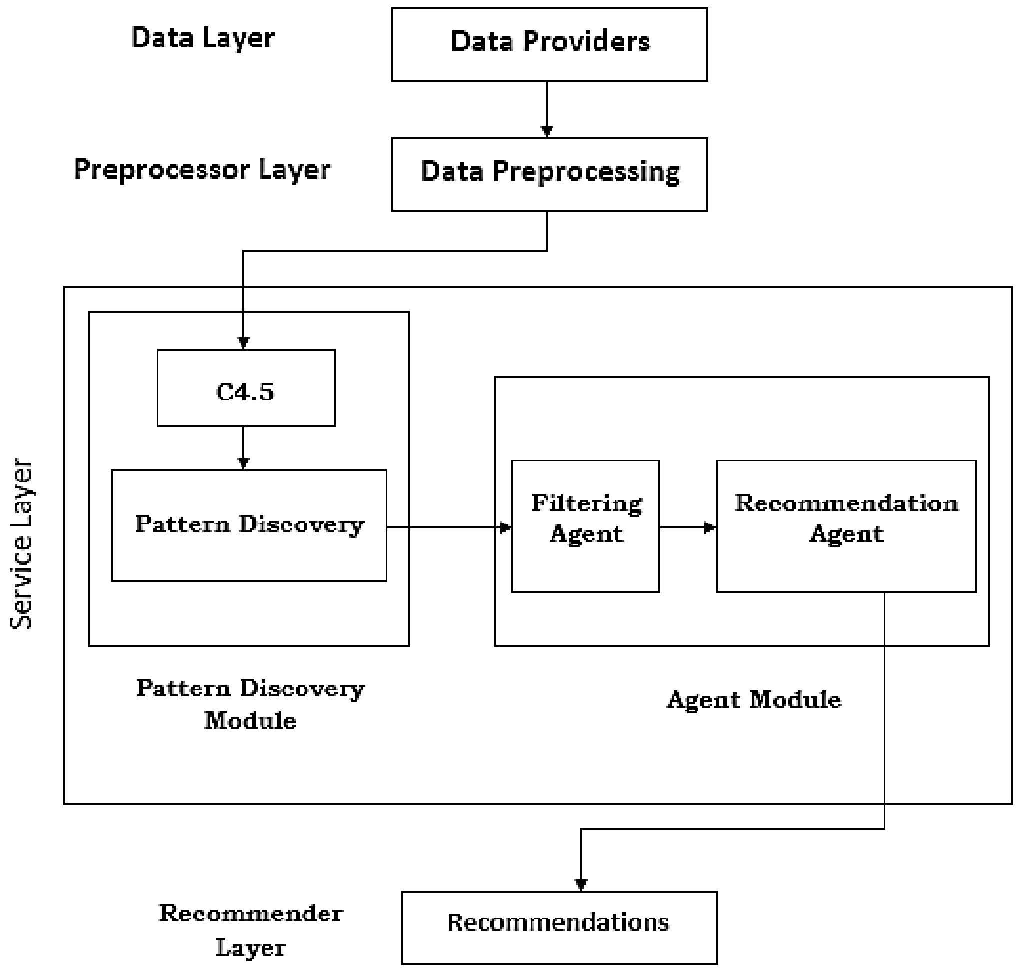


Figure 1: IREDM architecture

Data Layer

Data layer stores the information about various educational resources provided by the educational institutions. This involves integrating information from Colleges, Universities, technical and vocational institutions. Moreover this Data layer contains the different sources of information where educational resources may be located. Here data layer integrates various data bases in the area of distance learning sector.

Pre-processing Layer

Data pre-processing is a data mining technique that involves transforming raw data into readily pre-processable data format. Data is originated from different databases. It is susceptible to noise, missing value and inconsistency. In order to get correct result it is required to pre-process the data. Data cleaning, convert common log format, user identification, session identification, stop word removal, stemming process, white space removal, identify user request tasks performed at data pre-processing stage. The Architecture uses the CLEALG(Udb) cleaning algorithm proposed in [3].The purpose of the algorithm is to find the valid html. Files with long user sessions are deleted using the algorithm.

Service Layer

There are two modules in this layer, Pattern discovery module and Agent module.

Pattern Discovery Module

Pattern discovery module consists of two component, C4.5 and pattern discovery. C4.5 is a decision tree algorithm commonly used to generate decision tree since it has a high accuracy in decision making. C4.5 component uses the historical database of student and results obtained by them, with a goal of obtaining the rule. It uses the training data (pre-processing and filtering prepares the training data) as input data for generating the decision tree[5]. This tree is used by pattern discovery component to generate rules for recommendations.

Table 1: C4.5 Algorithm

```
C4.5 Algorithm
{
Input:an attribute-Valued dataset D
Output: A Decision tree.
Tree={}
If D is "pure" OR other stopping criteria met than
terminate
end if
for all attribute a £ D do
Compute information-theoretic criteria if we split on a.
end for
a_best =Best attribute according to above computed criteria
Tree=Create a decision node that tests a_best in the root
Dv=Induced sub-datasets from D based on a_best
for all D_v do
Treev=c4.5(D_v)
Attach Tree_v to the corresponding branch of Tree
end for
return Tree
}
```

Agent Module

There are two components in agent module.

Filtering Agent and Recommendation Agent

- **Filtering Agent**

Filtering Agent uses the item–based collaborative filtering algorithm. In item-based collaborative filtering the recommendations are based upon preference of similar users. Item will be recommended to the user based on the preference of other similar users for the same item. If set of users of have strongest correlation in the past, they will be identified as nearest neighbour. Score of the new items will be predicted based upon the scores of nearest neighbour [15]. In collaborating filtering Pearson correlation or Log-likelihood ratio can be used to identify preferred items for the user. Pearson correlation can be calculated as following

$$\mathrm{Sim}\ (i, j) = \frac{\sum_{u \in U} (R_{u,i} - \bar{R}_i) - (R_{u,j} - \bar{R}_j)}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_i)^2}\ \sqrt{\sum_{u \in U} (R_{u,j} - \bar{R}_j)^2}}$$

where $R_{u,i}$ is the rating given user u to item i; $\bar{R}_i$, $\bar{R}_j$ is the average rating of i^{th} and j^{th} item. Item based recommendation of collaborating filtering is purely based upon ranking given by a user to a particular item where as content based recommendation are focused on attributes of the items as well as attributes of user.

- **Recommendation Agent**

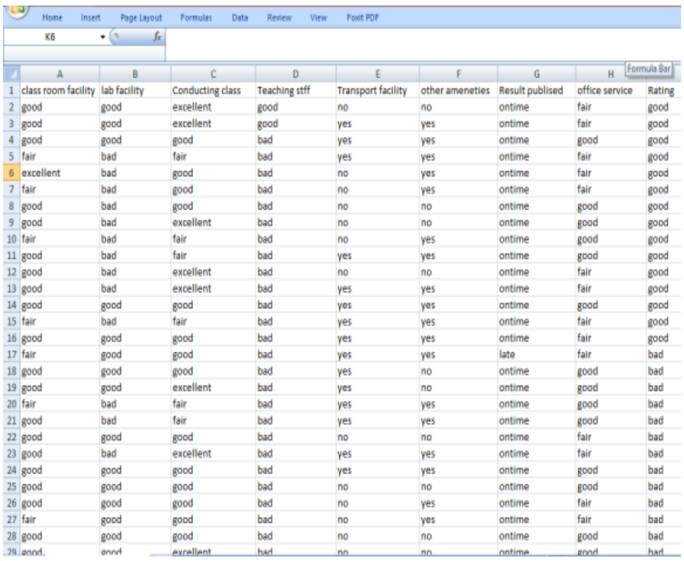
Recommender agent implements advance data analysis techniques to help users find the items based on the interest of the users score or a list of top-N recommended items for a given active user. The recommendation of the item can be made using different methods. Each recommendation method is having different results. Collaborative filtering based algorithms provide item recommendations or predictions based on the opinions of other like-mined. [9]

Recommender Layer

The main Function of the Recommendation layer is to produce recommendations using with recommendation agent that will be used to give positive solution to higher education institutions

EXPERIMENTAL SETTINGS

IREDM has been implemented with JAVA language using NetBean version 7.3 as JAVA environment. All experiments were done Intel core i3 2 .10GH 4 GBRAM, running Windows 8. For Input Data Set Indian university Distance learning Database collected for past 10 years. WEKA java API is used to implement the data mining techniques and machine learning algorithm. The experiments were performed on Indian University distance learning dataset. From the Data set various Decision tree generated in the classified area such as centre-service, social-economic back ground, Faculty Evaluation, course Book Evaluation, Student performance and various classification rules found out from Decision tree is used to Generate recommendations. Study centre data is as follow.



	A	B	C	D	E	F	G	H	
1	class room facility	lab facility	Conducting class	Teaching stff	Transport facility	other ameneties	Result publised	office service	Rating
2	good	good	excellent	good	no	no	ontime	fair	good
3	good	good	excellent	good	yes	yes	ontime	fair	good
4	good	good	good	bad	yes	yes	ontime	good	good
5	fair	bad	fair	bad	yes	yes	ontime	fair	good
6	excellent	bad	good	bad	no	yes	ontime	fair	good
7	fair	bad	good	bad	no	yes	ontime	fair	good
8	good	bad	good	bad	no	no	ontime	good	good
9	good	bad	excellent	bad	no	no	ontime	good	good
10	fair	bad	fair	bad	no	yes	ontime	good	good
11	good	bad	fair	bad	yes	yes	ontime	good	good
12	good	bad	excellent	bad	no	no	ontime	fair	good
13	good	bad	excellent	bad	yes	yes	ontime	fair	good
14	good	good	good	bad	yes	yes	ontime	good	good
15	fair	bad	fair	bad	yes	yes	ontime	fair	good
16	good	good	good	bad	yes	yes	ontime	fair	good
17	fair	good	good	bad	yes	yes	late	fair	bad
18	good	good	good	bad	yes	no	ontime	good	bad
19	good	good	excellent	bad	yes	no	ontime	good	bad
20	fair	bad	fair	bad	yes	yes	ontime	good	bad
21	good	bad	fair	bad	yes	yes	ontime	good	bad
22	good	good	good	bad	no	no	ontime	fair	bad
23	good	bad	excellent	bad	yes	yes	ontime	fair	bad
24	good	good	good	bad	yes	yes	ontime	good	bad
25	good	good	good	bad	no	no	ontime	good	bad
26	good	good	good	bad	no	yes	ontime	fair	bad
27	fair	good	good	bad	no	yes	ontime	fair	bad
28	good	good	good	bad	no	no	ontime	good	bad
29	good	good	excellent	bad	no	no	ontime	good	bad

Figure 2: Study Centre

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 19 (2018) pp. 14431-14437

Decision Trees

Figure 3: Study Centre

Production rules

Production rules are generated from the Decision trees that are used by recommendation agent to provide the recommendation.

Table 2: Production rules

If Class room facility=excellent and providing staff=good Then the state Centre service=good	IF Gender=female and centre in your palace=no then join course=no
If class room facility=good and providing staff=bad and lab facility=bad Then ==centre service =bad	If Gender =female and course fee=medium then join course =yes
If course fee=high and parent occupation=driver join course=no	If general proficiency=yes and ITG=good and LW=good and result=pass
If course fee=high and parent occupation=others then join course=yes	If general proficiency=good and ITG=bad and ATT=bad and result=fail
If Course fee=high and parent occupation=driver and centre in your place=yes hen join course=yes	If ITG=good and LW=bad and ATT=bad then result=good
If Gender=female and centre from home=10km then join course=yes	If ITG=bad and ATT=bad and LW=bad then result=fail
If Gender=female and centre from home= >10km then join course=no	If complete the syllabus=yes and delivery of notes=high and communication skill=yes then faculty=good
If complete the syllabus=no and delivery of notes=low and communication skill=no then faculty=bad	

EXPERIMENTAL EVALUATION

The Quality of a recommendation algorithm can be evaluated using different types of measurement. Accuracy is the fraction of correct recommendations out of total possible recommendations The proposed system is evaluated with Decision support accuracy that are popularly used are Precision, Recall and F-measure. These metrics help the user in selecting items that are very high quality out of the available set of items.

Experiment 1: Precision

Precision is the fraction of recommended item that is actually relevant to the user.[9]

$$\text{Precision(p)} = \frac{True\ positive}{True\ positive + False\ postive}$$

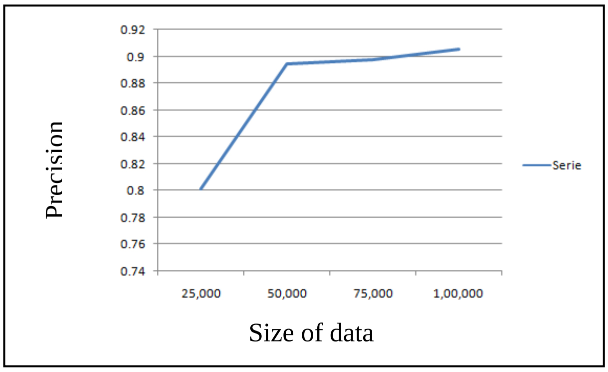


Table 3: Precision

Data set	25,000	50,000	75,000	1,00,000
Precision	0.801	0.894	0.897	0.905

This precision Graph is created using with Value of study centre data set. In this graph X-axis shows the various range of data set in the study centre and Y-axis shows the corresponding precision values. The high value of precision is 0.905 reached with the Dataset of 1,00,000. This metrics highlights the correct positive predictions out of all positive predations. High precision indicates low false value.

Experiment 2: Recall

Recall is the ratio of correctly predicted positive values to the actual positive values. [9].

$$\text{Recall(R)} = \frac{\textit{True positve}}{\textit{true postive+False Negative}}$$

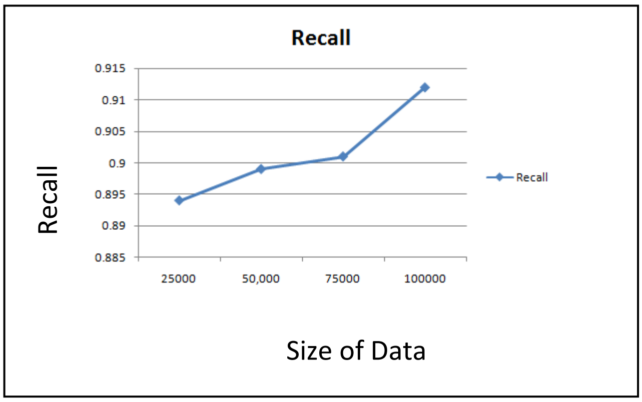


Table 4: Recall

Data set	25,000	50,000	75,000	1,00,000
Recall	0.894	0.899	.0901	0.912

This Recall precision Graph is created using with Value of study centre data set. In this graph X-axis shows the various range of study centre and Y-axis shows the corresponding Recall values. The high value of precision is 0.912 reached with the Dataset of 1,00,000. This metrics highlights the sensitivity of the algorithm. Dataset of 1,00,000 consider to be the best since it has the highest recall.

Experiment 3: F-Measure

F-measure defined below helps to simplify precisions and recall in to single metric [9].They are computed as

$$\text{F-measure} = \frac{2*precision*recall}{Precision+Recall}$$

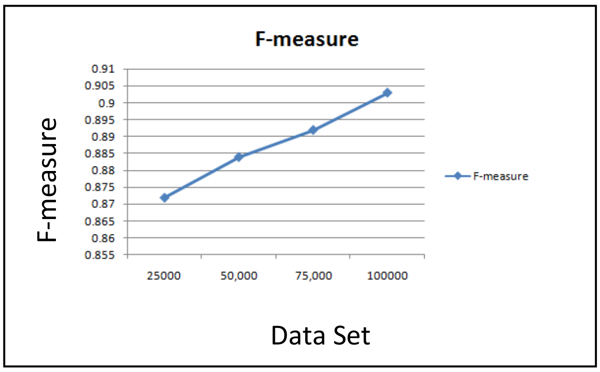


Table 5: F-Measure

Data set	25,000	50,000	75,000	1,00,000
F-.Measure	0.872	0.884	0.892	0.903

This F-measure Graph is created using with Value of study centre data set. In this graph X-axis shows the various range of study centre and Y-axis shows the corresponding F-measure values. From this graph the high value of F-measure is 0.903reached with the Dataset of 100000. High value of F-measure indicates the relevant result to the data items.

Experiment 4:Accuracy of Various Classifier

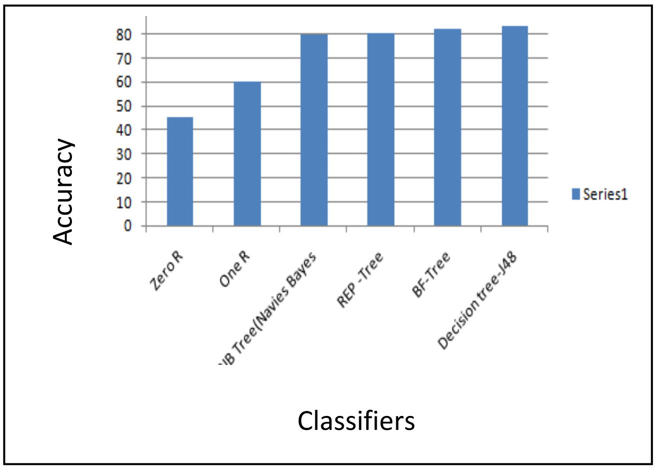


Table 6: Accuracy of Various classifiers.

Name of the classifier	Accuracy(%)
ZeroR	45.4
One R	60.2
NB Tree(Navies bayes)	79.61
REP –Tree(Reduced error pruning)	80.64
BF-Tree(Best First)	82.51
Decision tree-J48	83.69

Table 5shows the overall classification accuracy of the various Classifier such as Zero R, One R,NB Tree,REP-Tree,BF-Tree,Deciosion-J48.One can observe that the Decision Tree-J48 classifier gives the maximum classification accuracy for the proposed problem.

DISCUSSION

From the study and analysis, the following issues were identified: The suggested Recommendation of Target Marketing areas of Distance learning sectors as follows

Course Admission in Study centre

- Improve the Class Room facility, lab facility and Teaching staff in learning centres
- Centre reach will be within 10 KMS
- Reasonable course Fee
- Standard Course material

New Centre Establishment

- Availability of potential students in particular Area
- Tap the rural area students
- Facilitate easy transport

New course Introduction

- Identify the course demand in particular area
- Identify the employability of the course.

Student Result

- Focus on the student result in Internal Test, Lab Work, Attendance and Previous SEM result

CONCLUSION AND FUTURE WORK

Recommendation system can be proved to be very helpful to the higher education system. This paper Proposes a Recommender system using with rule based system and collaborative filtering system that analyse the centre service, Faculty evaluation, course Book evaluation, Socio economic Background of the student, student performance in distance learning, to point out the target marketing areas of Distance learning sector and provide appropriate recommendations. The suggestions generated by recommenders system can be useful to find out marketing areas of distance learning. Hybrid model of recommendation, Analysis of very large volume of data (Big data) may be the future.

REFERENCES

[1] Aleksandra Klasnja Milicevic, Boban Vesin, Mirjana Ivanovic, Zoran Budimac: E-Learning personalization based on hybrid recommendation strategy and learning style identification. Computers & Education, 56, 885–899 (2011).

[2] Amer AI-Badarnah,Jamal Alsakaran"An Automted Recommender System for course selection.

[3] Arumugam, G., and Suguna, S. Predictive Prefetching Framework Based on New Preprocessing Algorithms Towards Latency Reduction, Asian Journal of Information Technology, Medwell Journals.ISSN: 1682 -3915.,2008.

[4] Carlos Cobos, Orlando Rodriguez, Jarvein Rivera, et al.: A hybrid system of pedagogical pattern recommendations based on singular value decomposition and variable data attributes. Information Processing and Management, 607–625 (2013).

[5] Cesar Vialardi,Javier Bravo Leila Shafti Alvaro Ortigosa Recommendation in higher Education using Data mining Techniques, HADA project TIN - 64718, 2007.

[6] Dietmar Jannach, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich. Recommender Systems An Introduction. Cambridge University Press, 2011

[7] Ghadeer Mobasher, Almed Shawish and Osman Ibrahim Educational Data mining Rule based Recommender systems in processdings of the 9th International conference on computer supported Education.

[8] Hsua Chia Cheng, Chena Hsin Chin, Huangb Kuo Kuang, Huang Yueh Min: A personalized auxiliary material recommendation system based on learning style on Facebook applying an artificial bee colony algorithm. Computers and Mathematics with Applications, 64, 1506–1513 (2012).

[9] F.O.Isinkaye, Y.O.Folajimi, B.A.Ojokoh Recommendation systems: Principles, methods and evaluation Production and hosting by Elsevier2015.

[10] Maria Iuliana Dascalu, Constanta Nicoleta Bodea, Alin Moldoveanu, Anca Mohora, Miltiadis Lytras, Patricia Ordoñez de Pablos: A recommender agent based on learning styles for better virtual

collaborative learning experiences. Computers in Human Behavior, 243–253 (2015).

[11] Pavel zezula,Giuseppo Amato,vlastislav dohnal,and Michal Batco, similarity search the metric space Approach,volume 32 of advances in database system springers, ISBN 0-387-29146-6, 2006

[12] C.Remo and ventura Educational data mining: A survey from 1995 to2005 Expert syst Appi.33(1):135-146,.ISSN 0957-4174, 2007.

[13] J.B Schafer.The application of data mining to recommender system.J.WANG(ED) Encyclopedia of data warehousing and miningpp(14-48)Hershey PA: idea Group Reference.

[14] Surabhi Dwivedi,Dr.Kumari Roshini ,Recommender system for Big Data in Education VS 5th National Conference on E-Learning & E-Learning Technologies (ELELTECH), 2017

[15] Simbi,P:course recommender,Senior Thesis of ,Princeton University(2003).

[16] Salehi Mojtaba: Application of implicit and explicit attribute based collaborative filtering and BIDE for learning resource recommendation. Data & Knowledge Engineering, 87, 130–145 (2013).

[17] Salehi Mojtaba, Pourzaferani Mohammad, Amir Razavi Seyed: Hybrid attribute-based recommender system for learning material using genetic algorithm and a multidimensional information model. Egyptian Informatics Journal, 67–68 (2012).