JOURNAL OF MECHANICS OF CONTINUA AND MATHEMATICAL SCIENCES
www.journalimcms.org

ISSN (Online) : 2454 -7190, Special Issue, No.-7, February (2020) pp 95-107 ISSN (Print) 0973-8975

AN AGENT BASED RECOMMENDATION ENGINEFOR COURSE SELECTION USING EDUCATIONAL DATA MINING

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

[1]Asst. Professor, Dept. of Computer Science, Mannar Thirumalai Naicker College Tami Nadu, India

[2]Professor & Head Dept. of Computer Science, Madurai Kamaraj University Madurai, India

https://doi.org/10.26782/jmcms.spl.7/2020.02.00007

Abstract

In a higher education system, student faces a difficulty in choosing a right course from the large pool of courses in the institution. The course recommender framework provides necessary guidance to student network to choose a course in their scholarly schedule. This paper explores the potential of Educational data mining for course selection recommendation by predicting student's course selection which involves analysing admission data of student in the institution. The proposed framework was designed as agent based recommender system to improve the efficiency of recommendations. There are three agents in this model, Pattern discovery agent generates the course selection pattern which is filtered by filtering agent. The recommendation agent provides recommendation. The proposed model was analyzed and tested using admission data collected from the higher educational institution. More specifically the model is applied on 10000 student admission data in the distance learning programme. The model is evaluated by three experiments, the experimental results indicates that this recommender system can more accurate predictions of course selections.

Keywords: Data mining, classification mining, collaborative recommendation, course selection, EDM.

I. Introduction

The higher educational institutions are placed in a very highly aggressive situation. They are targeting to get more viable advantages over the opposite business competitors. Today, one of the highest complications that instructive establishments confront is the hazardous progress of useful information and to utilize this information to enhance the nature of administrative choices. Students pursing higher education degrees are faced with two challenges: one is the selection of the course and another one is lack awareness of the courses.

The Paper Presented at 14th International Conference on Intelligent System and Control (ISCO'20) Organized by The Department of Computer Science and Engineering, Karpagam College of Engineering, Coimbatore, India

Data mining techniques is employed by the education sectors to extract hidden information from great deal of knowledge. For education establishments whose objective is to feature to the enhancement of nature of upper education, the accomplishment of creating of human Capital is that the subject of nonstop investigation. Therefore, the prediction of students' success is crucial for education institutions, because the standard of teaching process is that the ability tosatisfy students' needs. Educational institute seeks more efficient technology to raised manage and support deciding procedures or assist them to line new strategies and plan for a far better management of the present processes.. The educator used EDM techniques to design the structure of course. The EDM techniques are used by the researches 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.

Recommender system provides information to the user to find the items of 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 frameworks utilized in training division to produce diverse suggestions for understudies, instructors and instructive establishment [IX]. Recommender system implements filtering techniques to find the recommendations. Collaborative filtering-based, content based filtering and hybrid filtering algorithms are used to find the user choice.

II. LiteratureReview

This Literature Reviews presents the comprehensive study of earlier research work on Course Selection.

Amer AI Badrenah ,Jamal Alsakaan has proposed a frame work for course selection for university course using datamining technique and also tried to extract the hidden information. Gerasimos,Spanakis hasdesigned a model to predict the course selection of graduaes in the institution using graduate attributes. The authors applied EDM technique to improve the effectives of course selection

JinijiaoLin, HailaPu has applied mining technique to design a model named as Intelligent recommendation system for course selection.It is used to provide Top-N recommendation of course to the appropriate students in the institution. Kokangchu, Maiga, Chang has designed course selection recommender system for course selection of student in the E-learning system.KoKang Chu, Maiga Chang has designed a framework for course selection recommender system in web based education system. They demonstrated how educational data mining is used to achieve the academic guidance,

E.K. Subramanian, Ramachandern has desigined a student carrier guidance system for providing recommendation to the student to select the course in the institution.

They observed the defect of the use of EDM model in the higher educational institution. Surabhi Dwivedi, Kumari Roshiniapplied collbrative filtering technique to design a model using Big data which is used to generate recommendation for course selection in the higher educational institution. Kuahjuta Bhumichitrpropsed a model for selecting university elective course.

In this paper the keypurpose of the proposed system is to create an agent based recommender system that will help the higher educational systems to identify the similarities of course selection data from the institution. It scrutinizes eachusercourse selection data to find the similarity with target user. It differs from the previous stated system that applied only EDM technique. In addition EDM technique and collaborative technique is integrated in proposed system to enhance the accuracy of the recommendations.

III. Proposed Recommendation Engine

In order to address the course selection of the higher education, this work proposed aarchitecture, named as Course Selection Recommender Educational Data mining (CSREDM). CSREDM is an agent based recommender system. The Functionalities of the CSREDM are pattern discovery and generate Recommendations.

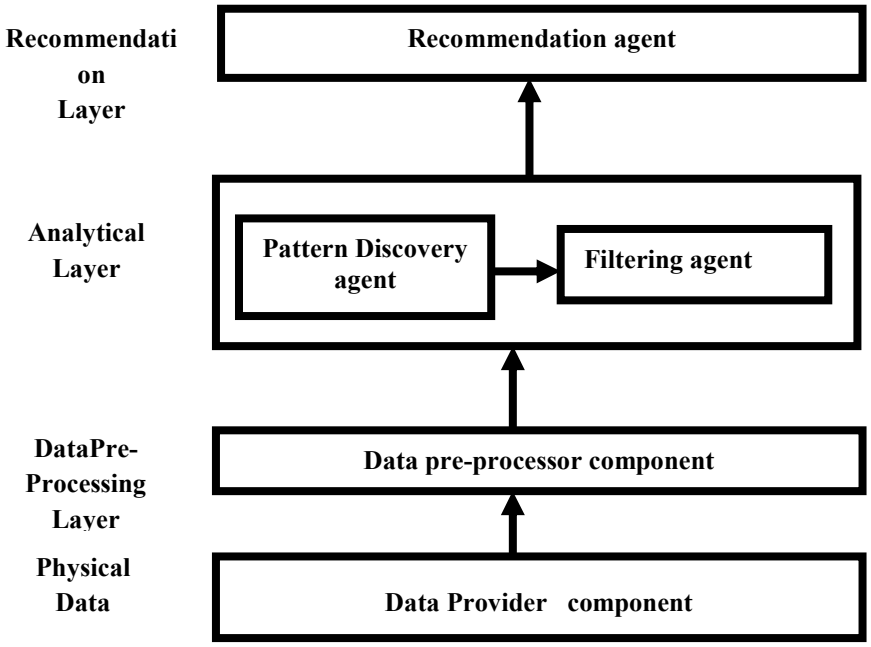

Fig.1:CSR-EDM

A high level view of the overall system is given in Figure1 .The architecture has a layered design.The various layers are Physical Data layer, DataPre-processing layer, Analytical layer Recommendation layer and the components in this system are Data

provider, Datapre-processor,Pattern discoveryagent, Filterin agent, Recommendation agent. The various functions of the propose recommendation engine are distributed in various layers so it is easy to maintain.

The flow of information in this architecture are controlled both horizontally and verticallyto assure properties with the help of the components of the system.The overall scenario utilizes the various attributes of the course enrolment data of the student to identify the similar patterns which are used to take decision in the colleges and universities. The result of the recommendations will be used in the development of the institution

Physical Data Layer

The Physical Data **layer** gathers and supplies the information about course enrolment data provided by the educational institutions. It acts as the data source of CSREDM.It involves integrating information from various databases. This layerhavethe of information course enrolment whichare collected from various educational resources may be situated in different places. A total number of 10000 data are collected from distance learning programme in the university from various courses.

Datapreprocessing Layer

DataPre-Processor Layer makes the data for mining. It is liable for performing the required data cleaning activity in the recommendation process. The function of this layercontainstechniques for data cleaning and prepares the data which are used in the pattern discover process.The data provider provides heterogeneous data. These data tend to be noisy, inconsistent, and incomplete in nature. After the step of data gathering from multiple educational data sources, this can be put together into a one single database. The academic data are in the form of files. Sometimes the academic data are noisy and inconsistency data.

The pre-processor layer performs various pre-processing technique to reduce the noise and eliminate the unwanted information in the data. Thepre-processor layer performs the following preprocessing techniques.

i)Lexical analysis ii) Stop Word Removal iii) Stemmingiv) Selection of index termsv)Data Cleaning

Analyticallayer

The Pattern Discovery layer plays an important role in this architecture. It aims at providing middleware that serves value added service to the higher layer. It encapsulates with two agent components named as pattern discovery agent and filtering agent. This layer receives the pre-processed course enrolment data from the lower layer which is used by pattern discovery agent to discover the pattern. These patterns are utilized by filtering agent to provide recommendation to the higher education institutions

Pattern Discovery Agent

It uses the C4.5 algorithm to make a decision tree using the data received from the pre-processor Layer.The development of decision trees from data is an ancient discipline. In machine learning, most investigations rely upon information theories. It is conventional to cite the ID3 Quinlan technique (Induction of Decision Tree - Quinlan 1979). There are different changes in ID3 algorithm, C4.5, algorithm is a fundamental technique in the advancement of decision tree.

The algorithm calculates the information gain of each attribute set of data. The high value of information is selected as the test attribute of the data set S. A tree node is created and labelled with the testattribute, branches of the binary tree are created the remaining attribute of the data set and it is partitioned.The decision tree is generated by service layer using C4.5 algorithm.This agentgenerates various user defined rule-based patterns.

Table1:C4.5 Algorithm

Parameters :V,S, S_{best} ,$V_{t.}$ V - * Set of attributed valued dataset *.,S - *attribute list* S_{best-} -* Attribute with maximum information gain *.,V_t-* Sub data set of V * N - * Decision tree *
Output: A Decision tree N

```
Procedure:C4.5(V)
begin
N=Tree;
N=empty;            /* Tree Node initially empty*/
S_best=empty;       / *Attribute for storing max.Information gain*/
If V is "pure" OR other stopping criteria met then
terminate
end if
for all attribute s £ V do          / * calculate information criteria * /
Calculate information-gain if we split on s.
end for
S_best=attribute with MAX_INFO gain;
X=MIN_INFO gain;
If(S_best>X) then
N= S_best ;   /*create  S_best as root node * /
V_t=Induced sub-datasets from V based on S_best/ * partitioning the data set * /
for all V_t do
N_t=C4.5.(V_t)
Attach N_t to the corresponding branch of N
end for
return N
end
```

Filtering Agent

The filtering agent has received the pattern from the course selection pattern from pattern discovery agent. This agent filters this pattern using collaborative filtering technique. The collaborative filtering system uses the item clustering algorithm which is used to identify the neighbourhoods. The clustering algorithm may make equal sized clusters or based on some similarity threshold it may generate a requested number of clusters of varying size.

There are lots of algorithms that can be utilized to create item clustering. The proposed system chooses the k- means algorithm as the basic clustering algorithm. The number k is an input to the algorithm that specifies the required number of clusters. First, it randomly selects k of the item, each of which initially represents a cluster mean or center. For each of the remaining items, the most similar item is assigned, based on the distance between the item and the cluster mean. The algorithm

computes the new mean for each cluster, this procedure repeats until the criterion function coverage.

The Proposed system uses the pearson's correlation formula in Equation to measure the linear correlation between two vectors of ratings as the target item ***t*** and theremaing item r.

$$sim\ (t,r) = \frac{\sum_{i=1}^{m}(A_{it} - B_t)(A_{ir} - B_r)}{\sqrt{\sum_{i=1}^{m}(A_{it} - B_i)^2 \sum_{i=1}^{m}(A_{ir} - B_r)^2}} \quad (1)$$

Where A_{it} is the rating of the target item t by user i, A_{ir} is the rating of the remaining item r by user i, B_t is the average rating of the target item t for all the co-rated users, B_r is the average rating of the remaining item *r* for all the co-rated users, and m is the number of all rating users to the item t and item *r*.

Table 2: Source Selection Pattern

User/Pattern	P1	P2	P3	P4	P5	P6	P7
U1	0	1	1	0	1	?	?
U2	1	0	0	1	0	0	1
U3	0	1	1	0	1	1	0
...	....	....	.....	.....	.....	.....	.....

Accordingto the table2, both the pattern p6 and p7 are not yet taken by user U1 and U2 and consider whether he should select them or not based on the user U3.First we find the similarity between the user U1 and user U3.The user U1 selects the pattern P6.From the similarity of the user U1 and U3 it is found that the user U1 does not select the patternP7.

Recommender Layer

In the proposed system, recommender layer plays a major role in the educational institution.The filtering component of the agent module filters the pattern. This component can be automatically created a cluster based on the similarity of the user.The recommendation agent component discovers the recommendations

Recommendation Agent

The main function of the recommender layer provides recommendation to the domain. It is found out the recommendation based on these ratings the recommender layer tries to find either user that are most similar to the active user or items that are similar to the item for which the user's rating is predicted.Since membership of item is available, calculation of the weighted average of neighbour's ratings, weighted by their similarity to the target item is possible. The rating of the target user ***u*** to the target item ***t***is an shown in Equation

$$P_{ut} = \frac{\sum_{i=1}^{c} R_{ut} \times sim(t,i)}{\sum_{i=1}^{c} sim(t,i)} \quad (2)$$

Where R_{ui} is the rating of the target user u to the neighbour item *i*, sim(t, i) is the similarity of the target item t and the neighbour it user i for all the co-rated items, and m is the number of all rating users to the item ***t*** and item ***r.***

IV. Experimental Setup

The proposed system has been executed with java language using netbeen version7.3 as java platform. All the experiments of the proposed system are run on windows-based PC with Intel core 3.2 .The speed of the processor is of 3.20GHz and storage capacityis 4GB.For Input Data Set Indian University Distance learning student enrolment Database collected for past 10 years.The data mining techniques and machine learning algorithm are applied on WEKA tool. The experiments were executed on real data set. The Decision tree isgenerated from the given data set.

Table3: Used variables in Course Enrolment Data set

Parameter	Possible Values
Course Name	{UG,PG}
Gender	{MALE,FEMALE}
Course Category	{arts, science}
Distance of the Centre from home	{<50KM,> 50 KM}
Area of the location	{rural,urban}
Easy Access Location	{yes,no}

The decision tree and recommendation rules are generated by this model which is used to give course selectionrecommendation to the higher education institution.Fig 4.shows the Decision tree for course selection. The informationgain formula given in the equation 1.Information gain of attribute in the data set is calculated as follows.

Gain Ratio(D,S)= Gain(D,S) /Split INFO (3)

where Split INFO=- $\left(\sum_{i-1}^{s} \frac{Di}{D} \log 2 \frac{Di}{D} \right)$

Where D is Data set.

The highest information gain is selected as the root node and branches are portioned.The decision tree is shown in the fig.3

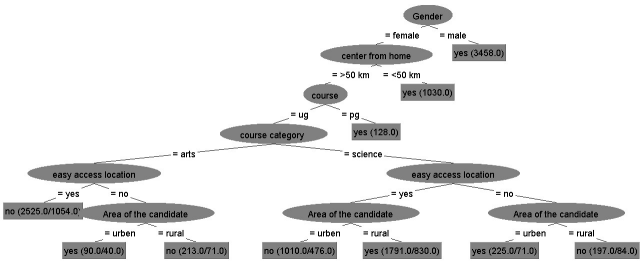

Fig. 2:ADecision Tree obtained by pattern discovery Agent

The table 4 shows the recommendation rules that are generated from the decision tree for a student who took the course in the distance learning system

Table 4:Recommendation Rules

Rule1	If Gender=”male” then course selection=”yes”
Rule2	If Gender=”female “and centre from home= <50km”then course selection=”yes”
Rule3	If Gender=female” and centre from home= >50km course=”PG” then join course=”yes”
Rule4	If Gender=female” and course=UG and category =”science “course selection=”yes”
Rule5	If Gender=”female” and course category=”science” and easy access location=”yes” course selection=”yes”
Rule6	If Gender=”female and course category=”science” and easy access location=”yes” area of the candidate=”rural”
Rule7	If Gender=”female “and centre from home= >50km”then course selection=”no”
Rule8	If Gender=”female “and centre from home= >50km” and course category=”arts” then course selection=”no”
Rule9	If Gender=”female “and centre from home= >50km” and course category=”science” and easy access location=”no” then course selection=”no”
Rule 10	If Gender=”female “and centre from home= >10km” and course category=”arts” then course selection=”no”
Rule11	If Gender=”female “and centre from home= >10km” and course category=”arts” and area of the candidate = “rural” then course selection=”no”

V. Practical Implementation

The Quality of the recommender system is measured using various kinds of metrices. Accuracy of the recommender engine is the ratio of correct recommendations out of total possible recommendations. The proposed CSREDM is evaluated with three metric essuchasas Precision, Recall and F-measure. These metrics help the user in choosing items that are very high quality out of the possibleset of items. The practical calculation of the proposed system has been done by the calculating the values of accuracy metrics.

Task 1: Precision

Fig 4.is drawn using the data set of course Enrolment Dataset .X-axis represents the varied data size of course Enrolment data set and the precision value mentioned in the Y-axis. From the figure 4 it is found that that the there is a significant improvement in precision value of proposed CSR-EDM system is enhanced by 10% value. This result shows that precision value is improved in the proposed system.

$$\text{Precision(p)} = \frac{True\ positive}{True\ positive + False\ postive} \quad (4)$$

Table 5: Precision Measure

Entrolment data set	2K	4K	6K	8K	10K
CSR(precision)	0.776	0.78	0.783	0.789	0.795
CSREDM(Precision)	0.876	0.881	0.885	0.891	0.894

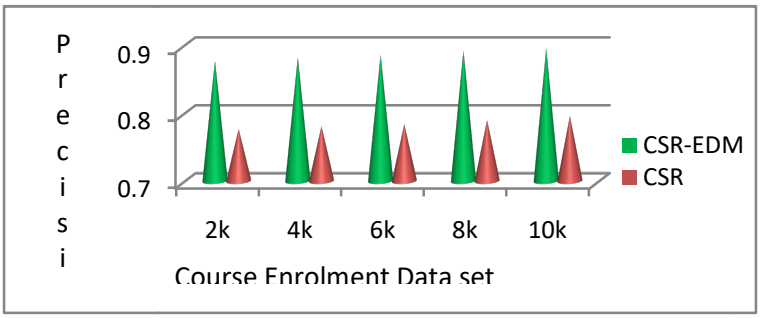

Fig.3: Course Enrolment Data set

Task 2: Recall Measure

Recall Measure is the ratio of correctly predicted positive values to the actual positive values

$$\text{Recall(R)} = \frac{True\ positve}{true\ postive + False\ Negative} \quad (5)$$

The comparison recall value of proposed system and existing system mentioned in the table 6..From the result it is found that recall value is increased by 9%..

Table 6:Recall Measure

Entrolment data set	2K	4K	6K	8K	10K
CSR(Recall)	0.653	0.662	0.668	0.675	0.681
CSREDM(Recall)	0.855	0.861	0.865	0.866	0.879

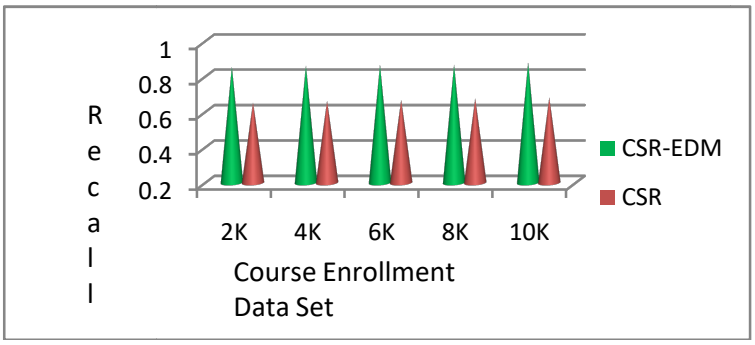

Fig.5: Recall Measure

Task3. F-measure

F-Measure is defined as follows. In table 7 the comparison value of F-measure mentioned The Fig.6shows the comparison value of f-measure value. From the result the proposed system's F-measure value is increased by of 13%.The proposed system achieved significant improvement in the F-Measure value.

$$\text{F-measure}=\frac{2*\boldsymbol{precision}*recall}{\boldsymbol{Precision}+Recall} \quad (6)$$

Table 7: F-Measure

Entrolment data set	2K	4K	6K	8K	10K
CSR(F-Measure)	0.62	0.628	0.639	0.645	0.649
CSREDM(F-Measure)	0.763	0.769	0.772	0.776	0.782

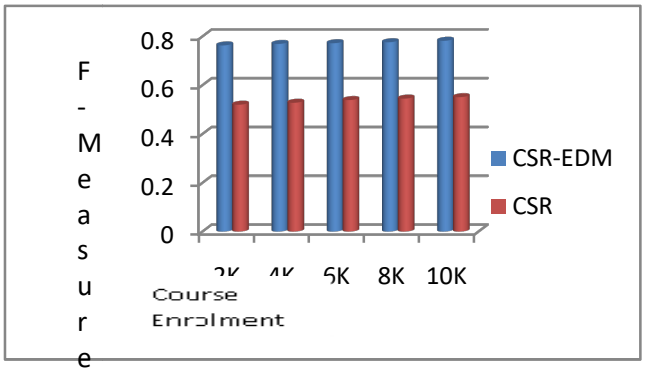

Fig. 6: F-Measure

VI. Result And Discussion

The mainobjective of this work has been presenting a CSR-EDM framework which is recommendation engine that integrates datamining technique and collaborative filtering technique to provide course selection recommendation in the

university. Decision tree is generated by CSR-EDM, from the decision tree various course selection patterns are discovered and it is filtered using collaborative filtering technique. The filtered patterns are used as recommendations.

The functional work of the model is evaluated using three different experiments. e student enrolment data collected from distance learning programme in the university for validation of the CSREDM.Form the threedifferent experiments the accuracy of the proposed system is measured. For model prediction accuracy measure, three criteria are adopted: Precision,Recall, and F-measure.In all the experiments, data size is started from 2000 course enrolment data and it is increased up to 10000 data set.

Precision,Recall and F-measure value of the CSR-EDM iscalculated for each data size and compared to existing system namely CSR system.The data size is stared from 2000 student course enrolment data and it is increased upto 1000 data. From the result,it is found that there is 10% improvement in precision value. The Recall value is improved by 9%. The improvement of F-Measure is 13 %.The result of various experiments shows that CSR-EDM outperforms the previous work and it generates various rules for course selection decision support system.

From the study and analysis, the following recommendations of course selection of the student were identified from the recommendation rule: The suggested Recommendation of Course Selection of Student in a higher educational institution as follows

Male students are selected the course.

Female students are selecting the course in the institution from nearby home.

Female Students are selecting science course.

The student prefers easy access location.

The Minimum course fee is selected by student.

The student prefers job-oriented course.

VII. Conclusion

In this paper an agent-based Recommender system has designed for course selection of the higher educational institution. This system is based on course selection of the student using appropriate educational data mining technique and recommender system technique. The proposed system can make more accurate prediction of student course selection. The result of the various experiments shows that the proposed system outperforms the existing system by integrating EDM technique and collaborative filtering technique. The rules are generated by decision tree provides the useful information to the educational institution in order to make good decisions for the course selection.Allthough experiment results of the proposed system is very good, the performance of the proposed system will be fine-tuned with other EDM .Hybrid model of the recommender system may be the future work.

References

I. Amer Al-Badarenah, Jamal Alsakran, "An Automated Recommender System for Course Selection" *(IJACSA)* International Journal of Advanced Computer Science and Applications, Vol. 7, No. 3,pages 163-175, 2016

II. Cesar Vivalardi, Javier Bravo, L. S. A. O. "Recommendation in higher education using data mining techniques". Master's thesis, Universidad Autonoma de Madrid,2009

III. E K Subramanian, Ramachandran "Student Career Guidance System for Recommendation of Relevant Course Selection" International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7, Issue-6S4, April 2019

IV. Gerasimos (Jerry) Spanakis, Samira Elatia, "Course Recommender System based on Graduating Attributes" conference paper Research gate 2017

V. Goga, M., Kuyoro, S., and Goga, N. "A recommender for improving the student academic performance". Procedia-Social and Behavioral Sciences",180:1481–1488.2015

VI. Ipperciel, D. and ElAtia, S. "Assessing graduate attributes: Building a criteria-based competency model". International Journal of Higher Education, 3(3):p27., 2014

VII. JinjiaoLin, HailaPu,yibinLi,JianLian "Intelligent recommendation system for course selection in smart Education" " International Conference on Identification, Information and Knowledge in the Internet of Things",2017

VIII. Kiratijuta Bhumichitr, Songsak Channarukul Recommender Systems for University Elective Course Recommendation, IEEE explore,2017

IX. Ko Kang Chu, Maiga Chang, Y.-T. H. "Designing a course recommendation system on web based on the students' course selection records. Master's thesis, Chung-Yuan Christian University,pages493-4962009

X. S. Jothilakshmi&, Dr. M. hangaraj "a survey on mining algorithms and techniques used in target marketing of higher educational institutions" International Journal of Computer Engineering and Applications, Volume XII, Issue VII, ISSN 2321-3469 .2018

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