Data Mining Application in Advertisement Management of Higher Educational Institutes

Priyanka Saini
M.Tech(CS) Student, Banasthali University Rajasthan
Sweta Rai
M.Tech(CS) Student, Banasthali University Tonk,Rajasthan
Ajit Kuamr Jain
Assistant Professor, Computer Science Department Banasthali University
Tonk,Rajasthan

Abstract

In recent years, Indian higher educational institute's competition grows rapidly for attracting students to get enrollment in their institutes. To attract students educational institutes select a best advertisement method. There are different advertisements available in the market but a selection of them is very difficult for institutes. This paper is helpful for institutes to select a best advertisement medium using some data mining methods.

Keywords

Data Mining, KDD, Data Mining in Education, Association Rule Mining, Apriori Algorithm

1. INTRODUCTION

Data mining is a process of analyzing and extracting the hidden information from large amounts of data. Today educational data is increasing rapidly due to information and communication technologies but this data is not in proper format. It contains many hidden information so data mining is required here. Different data mining techniques are used to get hidden and important information from this data.KDD (Knowledge discovery in database) is also called as data mining that extract useful and unknown information from data [1]. Its application is in science, commerce field and other fields. Use data mining in education is an interesting research area and it is called Educational Data Mining (EDM).In EDM data mining techniques are applied on student data of educational institutes.EDM is a process of extracting unknown patterns from educational environment for improved performance [2]. EDM is linked with some fields like as Artificial Intelligence (AI), Machine Learning/Information Technology, Database Management System, Statics and Data Mining.EDM is an interdisciplinary area for research that provides teaching and learning process's knowledge [3].This study helps to find out the best method for advertising the educational institutes by which student attracted easily at the time of admission.

2. DATA MINING AND KDD PROCESS

Data mining refers to extracting or "mining" knowledge from large amounts of data [1]. Data mining is the powerful technology for analyzing & extraction of hidden predictive information from the data. It helps for making a better decision. Data mining is one of the steps in KDD process. Knowledge discovery (KDD) aims at the discovery of useful information from large collections of data [4] .Data mining is used for processing data & this data is applied in many areas to obtain useful knowledge from the data [5].

2.1 KDD

KDD refers to Knowledge discovery in database means a process of searching knowledge in large data. Data mining is used as a step in KDD process.KDD process is shown in figure1.KDD process includes following steps:
Step1: Data collection
Step2: Database cleaning
Step3: Integration of data
Step4: Selection and transformation of data
Step5: Use data mining techniques
Step6: Patterns are evaluated
Step7: Representation of knowledge

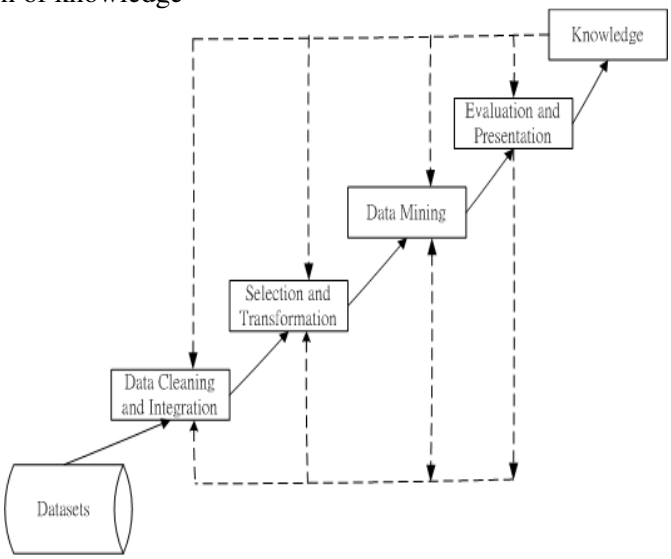

Figure 1: KDD Process

2.2 Association Rule Mining

It is used to discover relationships between attributes and items such as the presence of one pattern implies the presence of another pattern. Association Rule is a popular technique for market basket analysis because all possible combinations of interesting product groupings can be explored.

2.2.1Apriori Algorithm

Apriori algorithm is used for association rule's mining. It counts the support of item sets and generates candidate set using breadth-first search approach. It finds a frequent item sets by testing the groups of candidates against the data. When no further successful extensions are found the algorithm stops.

3. DATA MINING IN EDUCATION

Now-a-days there are growing research field in educational data mining. Educational Data Mining is a research region with the application of data mining in which information is generated from educational institutes. This research field continues to grow, and many data mining techniques have been applied in it. The goal of each research is to convert raw data into a meaningful form to make a better decision. Educational data mining explores the educational data to better understand the students[6]. It uses many techniques such as decision tree, neural networks to discover knowledge [7].The educational data mining research is increasingly growing, starting by organizing workshop since 2004.There are many research papers on educational data mining in which higher education related problems are discussed with their possible solutions using data mining techniques. In 2007 many literature reviews of the EDM are provided by Romero and Ventura [7].Educational data mining (EDM) helps the student for getting better result. It helps also management for maintaining education infrastructure, increase in the area of interest & courses. The educational systems face many problems that overcome by data mining techniques. It helps to the institution by giving the student guidance and helps to management in increasing the performance by using data mining techniques. It can help to provide a better quality of education.

4. LITERATURE SURVEY

Data mining is helpful to finding hidden information from student database. There is a lot of research papers regarding educational data mining field. Yadav and Pal [8, 9] give a study on educational data mining and provide data mining technique's list.

Yadav, Bharadwaj and Pal [10] obtained the students data such as attendance, seminar, assignment marks and class test to predict the end semester performance using three algorithms ID3decision tree, C4.5 and CART and result shows that CART gives better result for classification of data.

Pandey and Pal [11] show their study using association rule analysis to find the student interest of choosing class language. In this paper they use seven different interestingness measures. Their result concluded that student has shown their interest in mix mode class language.

Bharadwaj and Pal [12] use the classification decision tree technique to evaluate student' end semester performance, this study helps to identify the dropouts and students who require special attention and teacher advising.

AI-Radaideh. et al [13] presents a classification based model for student performance prediction using ID3 algorithm,C4.5 and Naïve Bayes algorithm but decision tree had better results.

K.S. Priya and A.V.S. kumar [14] use a classification approach that extracts the knowledge from student end semester marks. Bhise R.B,Thorat S.S and Supekar A.K. [15] performed data mining process on the student database using clustering-K-mean algorithm. Abeer Badr El Din Ahmed and Ibrahim Sayed Elaraby [16] use decision tree technique for predicting the student final grade.

5. EXPERIMENTAL WORK

In this study the data is collected through the questionnaire survey at the institute. There are 500 questionnaires are collected. This questionnaire includes student personnel and academic information. This study is helpful for finding a better advertisement for institutes so this questionnaire includes a important question "How They knows about this university"?Table1 shows different advertisement methods that are available in questionnaire's field. Table 2 shows different combination of advertisements occurrences that occur while responding the questionnaire field.

Table1: Different advertisement methods

Advertising methods	**Codes**	**Answers**
Friends	**FR**	**309**
Family	FA	17
Internet Search	IS	24
Online Advertising	OA	8
Newspaper	NE	8
Others	OT	5

Table2: Different combination of advertisements

Relations	**Occurrences**
Friends and Family(FRFA)	10
Friends and Internet Search(FRIS)	6
Friends and Online Advertising(FROA)	2
Friends and Newspaper(FRNE)	3
Friends and Others(FROT)	0
Family and Internet Search(FAIS)	0
Family and Online Advertising(FAOA)	0
Family and Newspaper(FANE)	7

Family and Others(FAOT)	0
Internet Search and Online Advertising(ISOA)	4
Internet Search and Newspaper(ISNE)	2
Internet Search and Others(ISOT)	0
Online Advertising and Newspaper(OANE)	2
Online Advertising and Others(OAOT)	0
Newspaper and Others(NEOT)	0
Total counts(support)	**36**

Analysis of Support and Confidence Value

The support value of a data item indicates the percentage of database transaction in which that data item appears. Confidence value measures the rule's strength. A small support and large confidence value are meaningful. Support and confidence analysis is shown in Table3.Support and Confidence values are calculated as:

Support= Occurrences/Total Support
Confidence= Support(X,Y)/Support(X)

Table 3: Support and confidence analysis of different relations

Relations	**Support =occurrences/Total Support**	**Confidence =support(X,Y) / support(X)**
FA->FR	0.28	0.59
IS->FR	0.17	0.25
OA->FR	0.06	0.25
NE->FR	0.08	0.375
Friends and Others	0	-
Family and Internet Search	0	-
Family and Online Advertising	0	-
NE->FA	0.19	0.875
Family and Others	0	-
IS->OA	0.11	0.17
NE->IS	0.06	0.25
Internet Search and Others	0	-
OA->NE	0.06	0.25
Online Advertising and Others	0	-

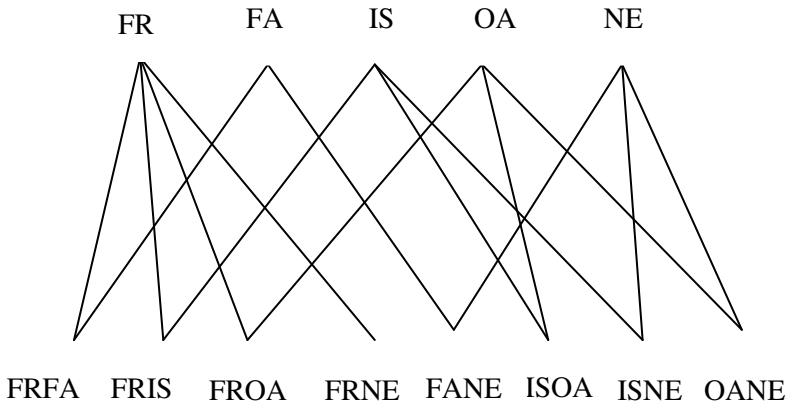

Figure 2: Advertisement method's links

Cosine Value Analysis

Cosine value occurs between 0 and 1.If its value is close to 1 then a good correlation occurs between X and Y. Cosine analysis formula defined as:

Cosine(X->Y) = p(X,Y) / sqrt(p(X)*p(Y))

Table 4: Cosine analysis of different relations

Relations	**Cosine value=(p(X,Y)/sqrt(p(X)*p(Y)))**
FA->FR	0.13
IS->FR	0.06
OA->FR	0.04
NE->FR	0.06
NE->FA	**0.60**
IS->OA	0.28
NE->IS	0.14
OA->NE	0.25

From table 3 result it can conclude that 28% students have answered with family and friends. Since 28% is largest support then it is a better advertisement method and its confidence value is 59% shows that if 59% of the time family occurs then friends also occurs in response.Table3

and Figure 2 concluded that the all advertisement methods are linked to either friend advertisement or newspaper advertisement medium.

In figure 2 it is shown that friends and newspaper have five edges. In table 4, cosine values are calculated for all advertisement mediums and it concludes that **newspaper → family** has good relation to each other but others relations are not so close to each other.

Apriori algorithm implementation

Table 5: Advertisement Methods data set

Relations	**Code**	**Occurrences (Support count)**
Friends	FR	309
Family	FA	17
Internet Search	IS	24
Online Advertising	OA	8
Newspaper	NE	8
Others	OT	5
Friends and Family	FRFA	10
Friends and Internet Search	FRIS	6
Friends and Online Advertising	FROA	2
Friends and Newspaper	FRNE	3
Family and Newspaper	FANE	7
Internet Search and Online Advertising	ISOA	4
Internet Search and Newspaper	ISNE	2
Online Advertising and Newspaper	OANE	2

Step 1: In this step the table 5 is scanned by algorithm and obtained information generates candidate item set C_1 (shown in table 6).
Step 2: In this step, compare C_1 candidate item set with minimum support count that generate frequent item set L_1

Apriori algorithm C_1 Join L_1 transformation

Table 6: Generated Candidate item set C_1

Code	Occurrences (Support count)
FR	309
FA	17
IS	24
OA	8
NE	8
OT	5

= >Minimum support (5)

Table 7: Generated frequent item sets L_1

Code	Occurrences (Support count)
FR	309
FA	17
IS	24
OA	8
NE	8
OT	5

Minimum support count=5

Step 3: In step 3, L_1 Join L_1 provides L_2 frequent item set and candidate item set C_2 it is called joining process.

Code	Occurrences (Support count)
FR	309
FA	17
IS	24
OA	8
NE	8
OT	5

L_1 Join L_1

m set C_2 generation by L_1 Join L_1 tr

Table 8: Generated Candidate set C_2

L_1*L_1	L_1-1	L_1-2
FR,FA	FR	FA
FR,IS	FR	IS
FR,OA	FR	OA
FR,NE	FR	NE
FR,OT	FR	OT
FA,IS	FA	IS
FA,OA	FA	OA
FA,NE	FA	NE
FA,OT	FA	OT
IS,OA	IS	OA
IS,NE	IS	NE
IS,OT	IS	OT
OA,NE	OA	NE
OA,OT	OA	OT
NE,OT	NE	OT

Step 4: In this step the table 9 is scanned by algorithm and compa minimum support count that generate frequent item set L_2

Table 9: Candidate set C_2 with Support Count

Code	Support Count
FR,FA	10
FR,IS	6
FR,OA	2
FR,NE	3
FR,OT	0
FA,IS	0
FA,OA	0
FA,NE	7
FA,OT	0
IS,OA	4

= > Minimum

Table 10: Generated frequent item sets L_2

Code	Support count
FR,FA	10
FR,IS	6
FA,NE	7

IS,NE	2
IS,OT	0
OA,NE	2
OA,OT	0
NE,OT	0

Step 5: In step 5, L_2 Join L_2 provides L_3 frequent item set and candidate item set C_3 (shown in table 11).

Code	Support count
FR,FA	10
FR,IS	6
FA,NE	7

L_2 Join L_2

Table 11: Generated Candidate set C_3

L_2*L_2	L_{2-1}	L_{2-2}
FR,FA,IS	FR,FA	FR,IS
FR,FA,NE	FR,FA	FA,NE
FR,IS,FA,NE	FR,IS	FA,NE

Candidate item set C_3 generation by L_2 Join L_2 transformation

Step 6:

- The subsets of (FR,FA,IS) are (FR,FA),(FR,IS)and(FA,IS).(FA,IS) is not a member of L_2. It is not a frequent item set. Therefore, remove (FR,FA,IS) from C_3.
- The subsets of (FR,FA,NE) are (FR,FA),(FA,NE)and(FR,NE).(FR,NE) is not a member of L_2. It is not a frequent item set. Therefore, remove (FR,FA,NE) from C_3.
- The subsets of (FR,IS,FA,NE) are (FR,IS),(FR,FA),(FR,NE),(IS,FA),(IS,NE)and(FA,NE) .(FR,NE),(IS,FA),(IS,NE) are not a members of L_2. It is not a frequent item set. Therefore, remove (FR,IS,FA,NE) from C_3.
- These item sets are pruned according to pruning property because its item sets are not frequent ,C_3 is null, so final frequent item set is L_2.If the minimum confidence is 50% then Confidence (FA->FR) and Confidence (NE->FA) is maximum so these are the frequent item sets (shown in table 13).

Table 12: Frequent item set L_2

Code	Support Count	Confidence
FRFA	10	0.59(59%)

Table 13: Frequent item sets

Code	Confidence
FRFA	0.59(59%)
FANE	0.875(87%)

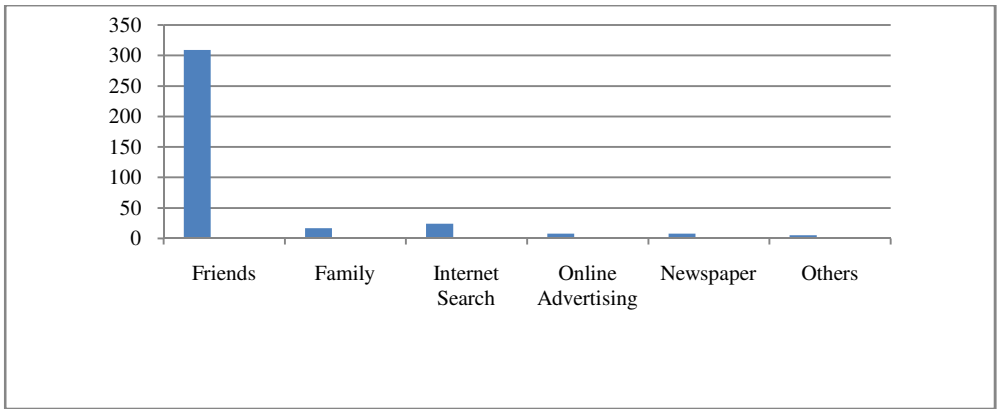

Figure 3: Analysis chart that shows different advertisement methods(answered given in questionnaires)

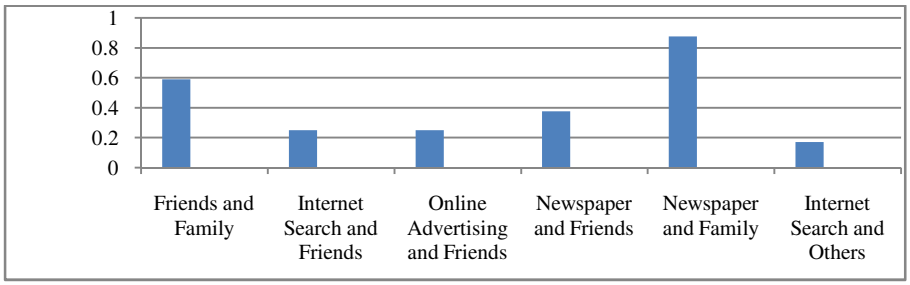

Figure 4: Analysis chart that shows different advertisement methods (according to data mining analysis)

6. CONCLUSION

This study looks for a better advertisement method. Fig.3 shows that most students have answered with friends (309 times) in the questionnaire but in this study data mining techniques are applied and this analysis concluded that the Family and Newspaper could be the best advertising method(shown in Fig.4). Table 1 shows that most students have answered with friends (309 times) but with data mining technique's analysis it is concluded that the Friends and Family, Family and Newspaper could be the best advertisement methods because these advertisement methods are linked with all advertisement methods. After applying apriori algorithm on the data set, Table 13 shows frequent item sets (FRFA,FANE) that concluded that Friends and Family, Family and Newspaper is the best advertisement methods. This study is helpful for educational institutes to making a good advertisement strategy that attracts student effectively.

7. REFERENCES

[1] Han, J., Kamber, M.: Data Mining: Concepts and Techniques,Morgan Kaufmann Publishers USA,(2000).

[2] A.Y.K. Chan, K.O. Chow, and K.S. Cheung*",Online Course Refinement through Association Rule Mining"*Journal of Educational Technology Systems Volume 36, Number 4 / 2007-2008, pp 433 – 44, 2008.

[3] Romero, C., and Ventura S.(2013), *" Data Mining in Education"*. WIREs Data Mining and Know.Dis.,Vol.3,pp.12-27.

[4] Heikki,Mannila, .Data mining: machine learning, statistics, and databases., IEEE, 1996.

[5] Klosgen, W. & Zytkow,J. (2002), Handbook of data mining and knowledge discovery , Oxford University Press, New York.

[6] International Educational Data Mining Society *www.educationaldatamining.org*.

[7] C. Romero and S. Ventura, *"Educational data mining: A survey from 1995 to 2005"*.Expert Systems with Applications 33 (2007) 135-146.

[8] S.Pal., "Mining Educational Data to Reduce Dropout Rates of Engineering Students", I.J. Information Engineering and Electronic Business
(IJIEEB), Vol. 4, No. 2, 2012, pp. 1-7.

[9] C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005", expert system with applications 33 (2007) 135-146.

[10] S. K. Yadav, B.K. Bharadwaj and S. Pal, "Data Mining Applications: A comparative study for Predicting Student's Performance", International Journal of Innovative Technology and CreativeEngineering (IJITCE), Vol. 1, No. 12, pp. 13-19, 2011.

[11] U. K. Pandey and S. Pal, "A Data Mining view on Class Room Teaching Language", IJCSI, Vol 8 issue 2 Maerch 2011 pg 277-282, ISSN 1694-0814.

[12] B.K. Bharadwaj and S. Pal. "Mining Educational Data to Analyze Students' Performance", International Journal of Advance Computer Science and Applications (IJACSA), Vol. 2, No. 6, pp. 63-69, 2011.

[13] Q. A. AI-Radaideh, E. W. AI-Shawakfa, and M. I.AI-Najjar*,"Mining student data using decision trees"*,International Arab Conference on Information Technology (ACIT'2006), Yarmouk University, Jordan, 2006.

[14] K.Shanmuga Priya and A.V.Senthil Kumar," *Improving the Student's Performance Using Educational Data Mining"*, Int. J. Advanced Networking and Applications Volume: 04 Issue: 04 Pages:1680-1685 (2013) ISSN : 0975-0290.

[15] Bhise R.B, Thorat S.S, Supekar A.K, *"Importance of Data Mining in Higher Education System",* IOSR(Journal Of Humanities And Social Science),Volume 6, Issue 6 ,PP 18-21, (Jan. - Feb.2013).

[16] Abeer Badr El Din Ahmed and Ibrahim Sayed Elaraby, *"Data Mining: A prediction for Student's Performance Using Classification Method"*, World Journal of Computer Application and Technology 2(2): 43-47, 2014.