International Journal of Technical Research and Applications e-ISSN: 2320-8163,
www.ijtra.com Volume 2, Issue 3 (May-June 2014), PP. 25-28

RESULT MINING: ANALYSIS OF DATA MINING TECHNIQUES IN EDUCATION

Jignesh Doshi
Associate Professor,
Research Scholar, Ahmedabad

***Abstract*— Data mining or Knowledge discovery (KDD) is extracting unknown (hidden) and useful knowledge from data. Data mining is widely used in many areas like retail, sales, e-commerce, remote sensing, bioinformatics etc. Student's performance has become one of the most complex puzzle for universities and colleges in recent past with the tremendous growth. In this paper, authors deployed data mining techniques like classification, association rule, chi-square etc. for knowledge discovery. For this study, authors have used data set containing Approx. 180 MCA (post graduate) students results data of 3 colleges. Study found that one can apply data mining functionalities like Chi-square, Association rule and Lift in Education and discover areas of improvement.**

***Key words*— Data mining, KDD Apriori Algorithm, Decision Tree, Association Rule, Chi-Square Test, LIFT.**

I. INTRODUCTION

We are data rich but information poor - with advent of Information Technology (IT), our ability of data storing, accessing and managing have increased. However, our ability to interpret data has decreased [1].

Data mining or Knowledge Discovery (KDD) methodologies are used successfully to extract hidden knowledge in healthcare, telecommunications, financial, Customer relationship management etc.

The universities are facing big challenge for improving performance and reducing drop outs. With the increase in number of institutes and students in last 5 years[5], It costs more to re-evaluate the failures. As a result Universities spends more efforts and energy on failures.

The main objective of the paper is to use data mining methodologies to study the student's failures using Data mining techniques. Data mining techniques are used successfully in other areas. The study goals are: 1) to focus on areas causing failures 2) to improve quality of students and 3) reduce drop out ratio.

For the study, we have used data of post graduate branch of one of the leading university of Gujarat. The

data set used comprise of 180 students data from 3 colleges. The data set selected was similar to what is

used for data mining in other areas with attributes relevant for analysis. Various classification techniques, e.g., decision trees, naïve Bayesian classification and SVM were tried. However, the results were unsatisfactory. After several demonstrations and interaction with domain experts, we finally designed and implemented an effective approach to perform the task. The final system is based on class association rules, general impressions and visualization. The system has been deployed at one of the college of MCA.

Due to confidentiality, we will not give data specifics but only present a general discussion about the results.

II. RELATED WORK

A. Introduction : Data Mining

Data Mining, refers to extracting of "mining" knowledge from large amounts of data. Data mining is also popularly known as Knowledge Discovery in Database and are frequently treated as synonyms[2][3][4],

Data mining techniques are used to operate on large volumes of data to discover hidden patterns and relationships helpful in decision making.

Data mining is used in a vast array of areas, and numerous commercial data mining systems are available. Some application domains are: Biomedical

& DNA data analysis, Financial data analysis, retail industry, telecommunication Industry and so on[ ].

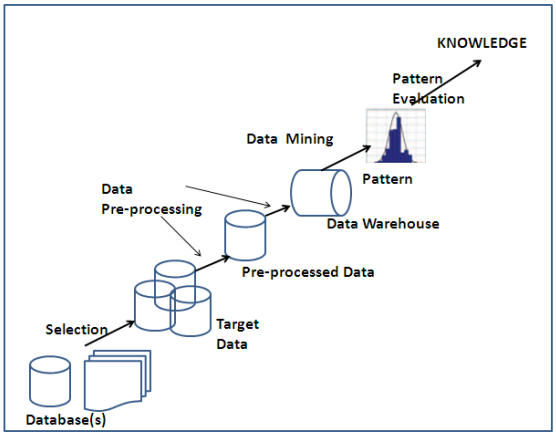

Figure: 1 KDD Process

Various algorithms and techniques like Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor method, etc., are used for knowledge discovery from databases.

B. Related Work :Education Mining

Mining in educational environment is called Educational Data Mining.

As per Alaa el-Halees , Data Mining can be used in educational field to enhance and evaluate the learning process students. .

Pandey and Pal explored the use of Bayes Classification on 17th attributes to found whether new comer students will be performer or not.

Hijazi and Naqvi applied simple linear regression analysis to discover that the factors like mother's education and student's family income were highly correlated with the student academic performance or not.

Using data analysis, Galit tried to predict the results and to warn students at risk before their final exams.

Decision tree model was used by Al-Radaideh, et al to predict the final grade of students. Three different classification methods namely ID3, C4.5, and the NaïveBayes were used. The outcome of their results indicated that Decision Tree model had better prediction than other models.

Pandey and Pal applied association rule and try to find the inter stingness of student in opting class teaching language.

As per Ayesha, Mustafa, Sattar and Khan,we may use k-means clustering algorithm to predict student‟s learning activities.

The applications of ata mining technique may be helpful for University, Faculty, Colleges as well as for students.

[1] EXPERIMENT: Result Mining : Analysis of Data Mining Techniques in Education

3.1 Data Extraction, Transformation and Load (ETL)

In this paper, authors have selected data set from one of the Largest university “Gujarat Technological University (GTU)” of Gujarat. The data sample picked up was from Post graduate course “Master of Computer Application”.

Following steps were carried out as a part of ETL:

1) Data Selection
Data of 3 Institutes out of 40
Total 180 students data collected out of 1228

2) Data collection
University Result data was collected form results ie. GTU Web portal (link; www.gtu.ac.in).

3) Data cleaning , Integration and Trasformation
Missing values: In case of non availability of data students mark sheets were used to collect data. Result was initially entered into excel file.

Data removed which were meeting following criteria: a) Data of left out students b) students having ATKT in all subjects of semester c) Marks of practical subjects are not included in analysis

4) Data Load
Various excel files created were used for application of data mining techniques. The data were converted into required format on need basis.

3.2 Data Mining Technique:Chi-square (χ^2)

In probability theory and statistics, The chi-square test will be used to test for the "goodness to fit" between observed and expected data according to a specific hypothesis.

χ^2 = SQRT((observed – expected) ^{2} / expected)

Data used for this technique was for one institute as below:

Hypothesis: there is no difference in institute result and university result.

	Institute	University	Total
Fail	43 (30)	369(382)	412
Pass	47 (60)	769(756)	816
Total	90	1138	1228

Figures in brackets indicate expected value (e) and is derived for each cell e.g. E11=90*412/1228 = 30.195

For the above data the value we get for **χ^2 = 9.11**

To test hypothesis, let us find out Chi-square value from table with K level of significant.

degree of Freedom = (r-c) (r-c) = (2-1)(2-1) = 1
We will use pre-defined significance (alpha) as 0.05 (5%) and value returned from table or degree of freedom as 1 is 3.84.

Here, Chi-square value 9.11, which is greater than 3.84. So, the hypothesis is accepted i.e. there is no difference between institute and university result.

3.2 Data Mining Technique Apriori Algorithm

Data considered for Applying apriori algorithm was of 42 failed students out of 120 students.

Case1: The minimum support considered as 4 failures.

Step 1 : The L1 found as individual items(subjects).
Here, all counts are above in. support >= 4

Subjects	Total Fails
{FON}	11
{WTAD}	28
{OR}	20
{MIS}	12
{DC -1}	07

Step 2; L1 x L1 resulted as
Here, after applying min. support >=4 the L2XL2 is

Subjects	Count
{FON,WTAD}	7
{FON,OR}	7
{FON,MIS}	6
{FON,DC-1}	2
{WTAD,OR}	10
{WTAD,MIS}	8
{WTAD,DC-1}	4
{OR,MIS}	7
{OR,DC-1}	3
{MIS,DC-1}	4

Subjects	Count
{FON,WTAD}	7
{FON,OR}	7
{FON,MIS}	6
{WTAD,OR}	10
{WTAD,MIS}	8
{WTAD,DC-1}	4
{OR,MIS}	7
{MIS,DC-1}	4

Step 3: L3 is as below

Subjects	Count
{FON,WTAD,OR}	4
{FON,WTAD,MIS}	5
{FON,OR,MIS}	5
{WTAD,OR,MIS}	6
{WTAD,MIS,DC-1}	4

Step 4: L4 is as below

Subjects	Count
{fon,wtad,or,mis}	4
{fon,wtad,mis,dc-1}	2

After applying min. support we get 4-itemset frequent pattern for failure as { FON,WTAD,OR,MIS}.

Case 2: If select Minimum support as 20% i.e. count as 8 count of 42, then it will return Frequent pattern as
[1] WTAD,MIS and ii) WTAD,OR

3.3 Data Mining Technique: Association Rule

To build association rules, we will derive confidence. Let us try to build association rules Based on WTAD subject as WTAD is the subject having maximum number of failures.

Confidence (A->B) = P(B/A)=P(ANB)/P(A)

	Support of LHS and RHS	Frequency of LHS	Confidence
WTAD→ OR	10	28	36%
WTAD→MIS	10	28	36%
FON→ WTAD, OR ,MIS	4	11	36%
FON→WTAD	7	11	64%
WTAD→FON	7	28	25%

From above data, we can make out that FON->WTAD is the strong rule and association rule can be written as

Fails(X,FON)=>FAILS(X,WTAD) (supp= 25% , conf=64%)

3.4 Data Mining Technique : Lift

Support and confidence measures are insufficient at filtering out uninteresting association rules. Lift (co-relation) measure is used to tackle this weakness.

Lift of item set A is independent of the occurrence of itemset B if P(A [B) = P(A)P(B); otherwise, itemsets A and B are dependent and correlated as events.

The lift between the occurrence of A and B can be measured by computing

LIFT(A,B)=P(AUB)/P(A)*P(B)

Here,

P(WTAD,FON) = 7/42 = 0.166
P(WTAD) = 7/42 = 0.66
P(FON) = 0.25

LIFT(WTAD,FON) = P(WTAD,FON) / P(WTAD)*P(FON)
= 0.166 / (0.66 * 0.25)
= 1.05 > 1

Conclusion:

- WTAD and OR failures are dependent
- As Lift > 1 , There is positive co-relation between OR and WTAD.

4 Conclusion

From the study, authors found out that:

1. From Chi-Square test, the conclusion is that we use to verify the performance of institutes with Top most institute results.
2. From Association Rule, the conclusion is that using the rules, we can predict number of failures of subject.
3. From Apriori Algorithm, the conclusion is that we used to choose frequent pattern (subjects in which students are having most of the time failures).
4. From LIFT, the conclusion is that we can detect interesting ness among the subjects,.

In this paper, we analyzed the technique for result mining. This study will help the institutes and faculties to

1. Focus on critical areas like subject(s), students etc.
2. In decision making while subject allocation and elective selection
3. In improving performance of students

Using the technique discussed in this paper, one can access the performances in a particular subject which is accessed by a teacher.

III. FEATURE WORK

The focus of current work is to experiment the application of Data mining techniques in Education. To apply various data mining functionalities with automated tools and prepare detail framework for result mining.

IV. ACKNOWLEDGEMENTS

I express deep sense of gratitude to Dr. Pradeep Jha for giving me an insight on Data Analysis. I am thankful to Dr. A R Prasanna for their guidance. I am thankful to Ms.Tejas Mehta for providing me feedback.

REFERENCES

[1] Paul G. Kaplan, Christopher A. Rautman Sandia National Laboratories Albuquerque, NM 87185-0716, USA: Data Rich, Information Poor

[2] Han, J. & Kamber, M. (2006), Data Mining: Concepts and Techniques, 2nd edition, Morgan Kauf-mann.

[3] G.K.Gupta, 2011, Introduction to Data Mining with case studies, 2nd edition, PHI , ISBN 978-81-203-4326-9.

[4] David Hand, heikki Mannila, Padhraic smyth, "Principles of Data Mining", PHI, ISBN 978-81-203-2457-2.

[5] All India Council for technical Education (AICTE) approval process handbook (2012-13), page 4.-7.

[6] R. Agrawal, T. Imielinski, and A. Swami, "Database Mining: A Performance Perspective," IEEE Trans. Knowledge and Data Eng.,vol. 5, no. 6, Dec. 1993.

[7] Alaa el-Halees, "Mining students data to analyze e-Learning behavior: A Case Study", 2009.

[8] U . K. Pandey, and S. Pal, "Data Mining: A prediction of performer or underperformer using classification", (IJCSIT) International Journal of Computer Science and Information Technology, Vol. 2(2), pp.686-690, ISSN:0975-9646, 2011.

[9] S. T. Hijazi, and R. S. M. M. Naqvi, "Factors affecting student's performance: A Case of Private Colleges", Bangladesh e-Journal of Sociology, Vol. 3, No. 1, 2006.

[10] Kumar, V. (2011). An Empirical Study of the Applications of Data Mining Techniques in Higher Education. IJACSA - International Journal of Advanced Computer Science and Applications, 2(3), 80-84. Retrieved from http://ijacsa.thesai.org.

[11] B.K. Bharadwaj and S. Pal. "Data Mining: A prediction for performance improvement using classification", International Journal of Computer Science and Information Security (IJCSIS), Vol. 9, No. 4, pp. 136-140, 2011.

[12] J. R. Quinlan, "Introduction of decision tree: Machine learn", 1: pp. 86-106, 1986.

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

[14] Agrawal, R. & Srikant, R. (1994), Fast algorithms for mining association rules in large databases, in 'International Conference on Very Large Data Bases',Santiago de Chile, Chile, pp. 487–499.

[15] Agarwal R., Mannila H., Srikant R., Toivonan H., Verkamo, "A Fast Discovery of Association Rules," Advances in Knowledge Discovery and Data Mining, 1996.

[16] Bing Liu, Wynne Hsu, Yiming Ma, "Integrating Classification and Association Rule Mining," In Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98, Plenary Presentation), New York, USA, 1998.

[17] A Data Mining view on Class Room Teaching Language, by: Umesh Kumar Pandey, S. Pal, Research Scholar, Singhania University, Jhunjhunu, Rajasthan, India. Dept. of MCA, VBS Purvanchal University, Jaunpur – 222001, Uttar Pradesh, India.

[18] Al-Radaiedh O A, A;-Shawakfa E M and Al-najjar M I, "Mining student data using Decision tree.", International Arab Conference on information technology (ACIT 2006), Yarmouk University, Jordan, 2006.

[19] Bharati S., Ramagaeri, "Data Mining Technique and Application", IJCSE vol 1 no.4 301-305.

[20] Cristobal Romero, Sebastian Ventura Ekaterina vasilyeva and Mykola pechenizkiy, "Class asoociation rule mining from students' test data.

[21] Hijazi S T and Naqvi R S M M, " Factors affecting Students' performance: A case of private colleges", Bangladesh e-journal of sociology Vol. 3 no. 1 2006.2002.