International Journal of Advanced Research and Review

www.ijarr.in

DATA MINING IN EDUCATION: A COURSE EVALUATION WITH ASSOCIATION RULES

U. Tuğba Gürsoy

Istanbul University Faculty of Business Administration, Department of Quantitative Methods
tugbasim@istanbul.edu.tr, İSTANBUL, TURKEY

ABSTRACT

In this article data mining is applied in education field. Data from the students who take data mining class test scores, the level of the test and the distribution of the scores are examined. The association rules helped to determine the students answering the particular types of questions together.

With massive amounts of data continuously being collected and stored, many industries are becoming interested in mining such patterns from their databases. The discovery of interesting correlation relationships among huge amounts of business transaction records can help in many business decision-making processes, such as catalog design, cross- marketing, and customer shopping behavior analysis. Association rules are an important class of regularities in data. Mining of association rules is a fundamental data mining task. Its objective is to find all co-occurence relationships among data items. Association rules analysis is applicable in different application domains such as market basket analysis, education, bioinformatics, finance, health and web mining.

KEYWORDS: Data mining, Association Rules, GRI, Education

INTRODUCTION

Rapid advances in data collection and storage technology have enabled organizations to accumulate vast amounts of data. However, extracting useful information has proven extremely challenging. Often, traditional data analysis tools and techniques can not be used because of the massive size of a data set. Data mining is a technology that blends traditional data analysis methods for processing large volumes of data.

There is a huge amount of data stored in real-world databases, and this amount continues to grow fast. This creates both an opportunity and a need for methods that discover the knowledge “hidden” in such databases. If such knowledge discovery activity is successful, discovered knowledge can be used to improve the decision- making process of an organization. Data Mining and Knowledge Discovery is the name often used to refer to a very interdisciplinary field, which consists of using methods of several research areas to extract knowledge from real-world data sets (Freitas, 2002).

In data mining various different techniques are used: Query tools, statistical techniques, visualization, case-based learning, decision trees, association rules, neural networks, genetic algorithms (Adriaans and Zantinge, 1997). In this paper association rules analysis is used.

Data mining is a series of activities from defining objectives for evaluating results (Giudici, 2003):

- Definition of the objectives for analysis
- Selection, organization and pretreatment of the data
- Exploratory analysis of the data and subsequent transformation
- Specification of the statistical methods to be used in the analysis phase
- Analysis of the data based on the chosen methods
- Evaluation and comparison of the methods used and the choice of the final model for the analysis
- Interpretation of the chosen model and its subsequent use in the decision processes

LITERATURE REVIEW

Thomas F. Stahovich, Hanlung Lin (2016), worked on Enabling data mining of handwritten coursework. Data mining has become an increasingly important tool for education researchers and practitioners. However, work in this field has focused on data from online educational systems. Here, they present techniques to enable data mining of handwritten coursework, which is an essential component of instruction in many disciplines. Their techniques include methods for classifying pen strokes as diagram, equation, and cross-out strokes. The latter are used to strike out erroneous work. They have also created techniques for grouping equation strokes into equation groups and then individual characters. Their results demonstrate that their classification and grouping techniques are more accurate than prior techniques for this task. They also demonstrate applications of their techniques for automated assessment of student competence. They present a novel approach for measuring the correctness of exam solutions from an analysis of lexical features of handwritten equations. This analysis demonstrates, for example, that the number of equation groups correlates positively with grade. They also use their techniques to extend graphical protocol analysis to free-form, handwritten problem solutions. While prior work in a laboratory setting suggests that long pauses are indicative of low competence, their work shows that the frequency of long pauses during exams correlates positively with competence.

Barry Robson, Srinidhi Boray, (2016) studied data-mining to build a knowledge representation store for clinical decision support. Studies on curation and validation based on machine performance in multiple choice medical licensing examinations, extracting medical knowledge by structured data mining of many medical records and from unstructured data mining of natural language source text on the Internet will become increasingly important for clinical decision support. Output from these sources can be transformed into large numbers of elements of knowledge in a Knowledge Representation Store (KRS), here using the notation and to some extent the algebraic principles of the Q-UEL Web-based universal exchange and inference language described previously, rooted in Dirac notation from quantum mechanics and linguistic theory. In a KRS, semantic structures or statements about the world of interest to medicine are analogous to natural language sentences seen as formed from noun phrases separated by verbs, prepositions and other descriptions of relationships. A convenient method of testing and better curating these elements of knowledge is by having the computer use them to take the test of a multiple choice medical licensing examination. It is a venture which perhaps tells us almost as much about the reasoning of students and examiners as it does about the requirements for Artificial Intelligence as employed in clinical decision making. It emphasizes the role of context and of contextual probabilities as opposed to the more familiar

intrinsic probabilities, and of a preliminary form of logic that they call presyllogistic reasoning.

Sultan M. Al-Daihani, Alan Abrahams (2016), made a study on Text Mining Analysis of Academic Libraries' Tweets. This study applies a text mining approach to a significant dataset of tweets by academic libraries. The dataset for this research was collected from the complete Twitter timelines of ten academic libraries. The total dataset comprised 23,707 tweets with 17,848 mentions, 7625 hashtags, and 5974 retweets. Academic libraries from the dataset have typically posted fewer than 50 tweets per month, though tweet volume grew rapidly in late-2013 through 2014. The results show variance between academic libraries in distribution of tweets over time. The most frequent word was "open," which was used in a variety of contexts by the academic libraries. It was noted that the most frequent bi-gram (two-word sequence) in the aggregated tweets was "special collections". The most frequent tri-gram (three-word sequence) was "save the date". The most frequent word categories in the semantic analysis for most libraries were related to "knowledge, insight, and information concerning personal and cultural relations". The most common category of the tweets was "Resources" among all the selected academic libraries. These findings highlight the importance of using data- and text-mining approaches in understanding the aggregate social data of academic libraries to aid in decision-making and strategic planning for patron outreach and marketing of services.

Jieun Kim, Mintak Han, Youngjo Lee, Yongtae Park (2016) analyzed Futuristic data-driven scenario building: Incorporating text mining and fuzzy association rule mining into fuzzy cognitive map. Fuzzy cognitive maps (FCMs) are one of the representative techniques in developing scenarios that include future concepts and issues, as well as their causal relationships. The technique, initially dependent on deductive modeling of expert knowledge, suffered from inherent limitations of scope and subjectivity; though this lack has been partially addressed by the recent emergence of inductive modeling, the fact that inductive modeling uses a retrospective, historical data that often misses trend-breaking developments. Addressing this issue, the paper suggests the utilization of futuristic data, a collection of future-oriented opinions extracted from online communities of large participation, in scenario building. Because futuristic data is both large in scope and prospective in nature, they believe a methodology based on this particular data set addresses problems of subjectivity and myopia suffered by the previous modeling techniques. To this end, text mining (TM) and latent semantic analysis (LSA) algorithm are applied to extract scenario concepts from futuristic data in textual documents; and fuzzy association rule mining (FARM) technique is utilized to identify their causal weights based on if-then rules. To illustrate the utility of proposed approach, a case of electric vehicle is conducted. The suggested approach can improve the effectiveness and efficiency of scanning knowledge for scenario development.

Kaur, Singh, and Josan (2015) worked on the classification and prediction based data mining algorithms to predict slow learners in education sector. They mentioned that educational data mining field concentrate on prediction more often as compare to generate exact results for future purpose. In order to keep a check on the changes occurring in curriculum patterns, a regular analysis is must of educational databases. The paper focus on identifying the slow learners among students and displaying it by a predictive data mining model using classification based algorithms. Real world data set from a high school is taken and filtration of desired potential variables is done using WEKA an Open Source Tool. The dataset of student academic records is tested and applied on various classification algorithms such as Multilayer Perception, Naïve Bayes, SMO, J48 and REPTree using WEKA. As a result,

statistics are generated based on all classification algorithms and comparison of all five classifiers is also done in order to predict the accuracy and to find the best performing classification algorithm among all. In the paper, a knowledge flow model is also shown among all five classifiers. The paper showcases the importance of prediction and classification based data mining algorithms in the field of education and also presents some promising future lines.

Nurbiha A. Shukor, Zaidatun Tasir, Henny Van der Meijden (2015) made an Examination of Online Learning Effectiveness Using Data Mining. Online learning has become increasingly popular due to technology advancement that allows discussion to occur at distance. Most studies report on students' learning achievement as a result of effective online learning while assessment on the learning process is also necessary. It is possible by applying data mining technique where students' online learning experiences can be assessed based on their log files. This study found that students could also perform well by being a silent learner during online learning. However, students have to spend more effort to be a successful silent learner as suggested by the produced predictive model.

Manolis Chalaris, Stefanos Gritzalis, Manolis Maragoudakis, Cleo Sgouropoulou, Anastasios Tsolakidis, (2014) worked on improving Quality of Educational Processes Providing New Knowledge Using Data Mining Techniques. One of the biggest challenges that Higher Education Institutions (HEI) faces is to improve the quality of their educational processes. Thus, it is crucial for the administration of the institutions to set new strategies and plans for a better management of the current processes. Furthermore, the managerial decision is becoming more difficult as the complexity of educational entities increase. The purpose of this study is to suggest a way to support the administration of a HEI by providing new knowledge related to the educational processes using data mining techniques. This knowledge can be extracted among other from educational data that derive from the evaluation processes that each department of a HEI conducts. These data can be found in educational databases, in students' questionnaires or in faculty members' records. This paper presents the capabilities of data mining in the context of a Higher Education Institute and tries to discover new explicit knowledge by applying data mining techniques to educational data of Technological Educational Institute of Athens. The data used for this study come from students' questionnaires distributed in the classes within the evaluation process of each department of the Institute.

Renza Campagni, Donatella Merlini, Renzo Sprugnoli, Maria Cecilia Verri, (2015) used Data mining models for student careers. This paper presents a data mining methodology to analyze the careers of University graduated students. They present different approaches based on clustering and sequential patterns techniques in order to identify strategies for improving the performance of students and the scheduling of exams. They introduce an ideal career as the career of an ideal student which has taken each examination just after the end of the corresponding course, without delays. They then compare the career of a generic student with the ideal one by using the different techniques just introduced. Finally, they apply the methodology to a real case study and interpret the results which underline that the more students follow the order given by the ideal career the more they get good performance in terms of graduation time and final grade.

Amirah Mohamed Shahiri, Wahidah Husain, Nur'aini Abdul Rashid, (2015) made a review on Predicting Student's Performance Using Data Mining Techniques. Predicting students performance becomes more challenging due to the large volume of data in educational

databases. Currently in Malaysia, the lack of existing system to analyze and monitor the student progress and performance is not being addressed. There are two main reasons of why this is happening. First, the study on existing prediction methods is still insufficient to identify the most suitable methods for predicting the performance of students in Malaysian institutions. Second is due to the lack of investigations on the factors affecting students achievements in particular courses within Malaysian context. Therefore, a systematical literature review on predicting student performance by using data mining techniques is proposed to improve students achievements. The main objective of this paper is to provide an overview on the data mining techniques that have been used to predict students performance. This paper also focuses on how the prediction algorithm can be used to identify the most important attributes in a students data. They could actually improve students achievement and success more effectively in an efficient way using educational data mining techniques. It could bring the benefits and impacts to students, educators and academic institutions.

Twijri and Noaman (2015) worked on A New Data Mining Model Adopted for Higher Institutions. In the current times, one of the biggest challenges that educational institutions face in both the developing and developed nations is the explosive growth of educational data and to use this data to improve the quality of managerial decisions. Data mining techniques are considered as an analytical tool can be used to extract meaningful knowledge from these large data sets. The quality of higher education institutions implies providing the services, which most likely meet the needs of students, academic staff, and other participants in the education system. The paper proposes a new Data Mining model to be applied in higher education institutions. They suggested model assists in decision making process in the strategic levels of higher institutions as well as regulates the disciplines of students' admission.

In their paper Gobert, Kim, San Pedro, Kennedy and Betts (2015) studied many national policy documents underscore the importance of 21st century skills, including critical thinking. In parallel, recent American frameworks for K-12 science education call for the development of critical thinking skills in science, also referred to as science inquiry skills/practices. Assessment of these skills is necessary, as indicated in policy documents; however, this has posed a great challenge for assessment researchers. Recently, some science learning environments seek to assess these science skills. These systems log all students' interactions within the given system, and if fully leveraged, these logs provide rich assessments of inquiry skills. Here, they describe their environment Inq-ITS (inquiry intelligent tutoring system), that uses educational data mining to assess science inquiry skills, as described as 21st century skills. Additionally they describe how they measure students' skills at designing controlled experiments, a lynchpin skill of inquiry, in the context of complex systems. In doing so, their work addresses 21st century skill assessment in two ways, namely of inquiry (designing and conducting experiments), and in the context of complex systems, a key topic area of 21st century skills. They use educational data mining to develop their assessment of this skill for complex systems.

Peña-Ayala (2014) studied on the educational data mining. They did a survey and a data mining-based analysis of recent works. In that review they pursued a two goals, the first is to preserve and enhance the chronicles of recent educational data mining (EDM) advances development; the second is to organize, analyze, and discuss the content of the review based on the outcomes produced by a data mining (DM) approach. Thus, as result of the selection and analysis of 240 EDM works, an EDM work profile was compiled to describe 222 EDM approaches and 18 tools. A profile of the EDM works was organized as a raw data base,

which was transformed into an ad-hoc data base suitable to be mined. As result of the execution of statistical and clustering processes, a set of educational functionalities was found, a realistic pattern of EDM approaches was discovered, and two patterns of value-instances to depict EDM approaches based on descriptive and predictive models were identified. One key finding is: most of the EDM approaches are ground on a basic set composed by three kinds of educational systems, disciplines, tasks, methods, and algorithms each. The review concludes with a snapshot of the surveyed EDM works, and provides an analysis of the EDM strengths, weakness, opportunities, and threats, whose factors represent, in a sense, future work to be fulfilled.

Mohamad and Tasir (2013) studied on the topic Educational Data Mining. They made a review on Procedia - Social and Behavioral Sciences. In their study they mentioned that Data Mining is very useful in the field of education especially when examining students' learning behavior in online learning environment. This is due to the potential of data mining in analyzing and uncovering the hidden information of the data itself which is hard and very time consuming if to be done manually. The purpose of this review is to look into how the data mining was tackled by previous scholars and the latest trends on data mining in educational research. Several limitations of existing research are discussed and some directions for future research are suggested.

Sen and Ucar (2012) studied the evaluation of the achievements of computer engineering department of distance education students with data mining methods. They said recently, the internet technology has become an indispensable part of life, a very useful application that cannot be earlier have made it possible. One of these is distance learning technologies. Due to limitations of traditional learning-teaching methods in classroom activities and practitioners who intend to conduct training activities in the absence of the possibility of communication and interaction among learners with special education units are prepared and provided a wide range of media center through a certain method of teaching. According to a further recognition of distance education, although far away from each other with the student who teaches the same time (synchronous) or different time (asynchronous) communications with a tool as training system established. The aim of the study is to compare the achievements of Computer Engineering Department students in Karabuk University according to criteria such as age, gender, type of high school graduation and whether the students studying in distance education or regular education using data mining techniques. Also discussing the differences of the techniques according to the results and to make suggestions for which technique would be more effective.

Şen, Uçar and Delen (2012) worked on predicting and analyzing secondary education placement-test scores with a data mining approach: Understanding the factors that lead to success (or failure) of students at placement tests is an interesting and challenging problem. Since the centralized placement tests and future academic achievements are considered to be related concepts, analysis of the success factors behind placement tests may help understand and potentially improve academic achievement. In their study using a large and feature rich dataset from Secondary Education Transition System in Turkey they developed models to predict secondary education placement test results, and using sensitivity analysis on those prediction models they identified the most important predictors. The results showed that C5 decision tree algorithm is the best predictor with 95% accuracy on hold-out sample, followed by support vector machines (with an accuracy of 91%) and artificial neural networks (with an accuracy of 89%). Logistic regression models came out to be the least accurate of the four with and overall accuracy of 82%. The sensitivity analysis revealed that previous test

experience, whether a student has a scholarship, student's number of siblings, previous years' grade point averages are among the most important predictors of the placement test scores.

ASSOCIATION RULES ANALYSIS

Data mining is commonly defined as the process of discovering useful patterns or knowledge from data sources. The patterns must be valid, potentially useful, and understandable. Data mining is a multi-disciplinary field involving machine learning, statistics, artificial intelligence, and visualization. There are many data mining tasks. Some of the common ones are classification, clustering, association rule mining and sequential pattern mining. In this study, association rules analysis is applied. (Liu, 2008)

Association Rules are derived from a type of analysis that extracts information from coincidence. The methodology allows discovering correlations, or co-occurences of transactional events (Westhpal and Blaxton, 1998).

Association rules are an important class of regularities in data. Mining of association rules is a fundamental data mining task. It is perhaps the most important model invented and extensively studied by the database and data mining community. Its objective is to find all co-occurence relationships, called associations, among data items. (Liu, 2008)

Association analysis is useful for discovering interesting relationships hidden in large data sets. The uncovered relationships can be represented in the form of association rules or sets of frequent items.

The classic application of association rule mining is the market basket data analysis, which aims to discover how items purchased by customers in a supermarket or a store are associated. Retailers can use this type of rules to help them identify new opportunities for cross-selling their products to the customers. Besides market basket data, association analysis is also applicable to other application domains such as bioinformatics, medical diagnosis, web mining, education, finance and scientific data analysis. (Tan et al., 2006)

GRI algorithm is used for the analysis. In the Generalized Rule Induction task, the goal is to discover rules that predict the value of a goal attribute, given the value of other attributes. However, the goal attribute can be any attribute not occurring in the antecedent of the rule. (Freitas, 1997)

APPLICATION

Problem Definition

Association rules that is one of the Data mining tools, used for the analysis. The aim is to identify the success of the students who took "Data Mining and Knowledge Discovery" elective course. There are 59 students in the class. 49 of them took the exam. They answered 30 questions. If the student answer the question wrong, it is coded as "0", if it is right it is coded as "1". This can be seen in Table 1.

In Table 2 data preparation phase can be seen. Questions (Q1,Q2, …,Q30) typed as "flag" and direction is "both" for association rules analysis.

Modeling

In modeling stage appropriate technique for data set is chosen. The built model for web graph can be seen in Figure 1.

Figure 2 shows the web graph. There are links between the questions. If the students answer both question 27 and 13; there is a line. The thickest line shows the most frequently answered questions together.

All of the students answered Q3, Q11, Q27, Q29 and Q30 correctly. All of them coded as "1". This is a problem for the analysis. By using filter node these questions were eliminated.

The GRI rulesets can be seen in Table 3.

- 100% of students who answered questions 8,14 and 23, also answered question 18 with a support rate of 12.24%.
- 92.31% of the students who answered questions 4,13 and 15, also answered question 1, with a support rate of %26,53

Rule support and confidence are two measures of rule interestingness. The strenght of an association rule can be measured in terms of its support and confidence. Support determines how often a rule is applicable to a given data set, while confidence determines how frequently items in Y appear in transactions that contains X. (Tan, et.al, 2006) They respectively reflect the usefulness and certainty of discovered rules.

Model Evaluation

In Table 4, the number of right and wrong answers and the scores of the students could be seen.

In the Table 5 the ranges and frequencies of students' scores could be seen. 42.22% (19 students) of the students got scores between 70-79 while 40 % of students (18 students) got scores between 80-89. Those scores might show that the difficulty of the questions are relatively easy.

In Table 6 the difficulty of the questions could be seen. In the exam there are 30 questions, 3-5-6-7-9-11-20-22-27-28-29-30 questions are the ones that are easier than the others, the rate is above 95% percentage. The red numbers show that how any students answer the question correctly.

CONCLUSION

In this study data mining technique is used for identifying under graduate class success assessment. The students who took "Data Mining and Knowledge Discovery" elective course were analyzed. The evaluation of difficulty of the questions and student success is aimed.

At the end of the analysis the questions that are answered together are identified by the association rules and web graph. For each question it is found that how many students answered the question correctly. The frequencies shows that the exam questions are found to be easy.

Association rules is used for the analysis. This technique is used to discover patterns that describe strongly associated features in the data. The discovered patterns are typically represented in the form of rulesets. These rulesets and such valuable information can be used to support a variety of business-related applications such as CRM, inventory management, catalog design and marketing promotions.

The results of this paper can be used to improve the quality of the course, and the style of the questions.

APPENDICES

Table- 1 Data set

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE
1	Students	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21	Q22	Q23	Q24	Q25	Q26	Q27	Q28	Q29	Q30
2	1	1	1	1	1	1	1	1	0	1	0	1	1	1	0	1	1	1	0	1	1	1	1	0	0	0	1	1	1	1	1
3	2	1	1	1	1	1	0	1	1	1	0	1	1	1	0	1	0	1	0	1	1	1	1	1	0	0	1	1	1	1	1
4	3	1	1	1	1	0	1	1	0	1	1	1	1	1	0	1	1	0	1	0	1	1	1	1	0	1	0	1	1	1	1
5	4	0	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1
6	5	0	1	1	1	1	1	1	0	0	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1
7	6	0	0	1	0	0	1	1	0	1	0	1	0	0	1	0	1	1	0	1	1	1	1	0	0	0	0	1	1	1	1
8	7	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
9	8	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	0	1	1	1	1	1	0	0	1	1	1	1	1
10	9	0	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	0	1	0	1	0	1	1	1	1	1
11	10	0	1	1	0	1	1	0	1	1	1	1	1	1	0	1	1	0	0	0	1	1	1	1	0	0	1	1	1	1	1
12	11	0	1	1	1	1	1	1	1	1	1	1	0	1	0	0	1	1	0	1	0	0	1	0	1	0	1	1	1	1	1
13	12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1	0	1	0	1	1	1	1	1
14	13	0	1	1	0	1	1	1	1	1	0	1	1	1	0	1	0	1	0	1	1	0	1	0	0	0	0	1	1	1	1
15	14	1	0	1	0	1	0	1	0	1	1	1	1	0	1	0	1	1	0	1	1	1	1	0	0	1	1	1	1	1	1

Table- 2 Data preparation

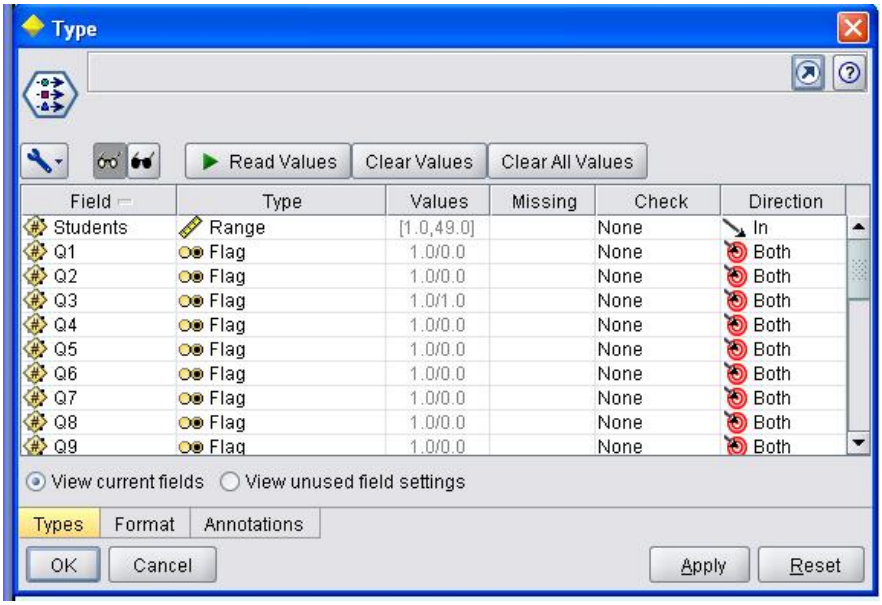

Table- 3 GRI Rulestes

Consequent	Antecedent	Support %	Confidence %
Q18	Q8 Q14 Q23	12,24	100,0
Q11	Q7 Q14 Q23	12,24	100,0
Q1	Q7 Q11 Q15	26,53	92,31
Q11	Q14 Q23 Q28	16,33	87,5
Q18	Q14 Q23 Q26	16,33	87,5
Q18	Q14 Q23	16,33	87,5
Q18	Q14 Q19 Q23	16,33	87,5
Q18	Q14 Q17 Q23	16,33	87,5
Q18	Q14 Q16 Q23	16,33	87,5
Q11	Q14 Q15 Q23	16,33	87,5
Q18	Q7 Q14 Q23	16,33	87,5
Q11	Q6 Q14 Q23	16,33	87,5

Table- 4 The scores of the students

	Number	Name		Empty	Correct	Wrong	Scores
2	100112017		A	0	22	8	73
3	100112008		A	0	19	11	63
4	100112030		A	0	22	8	73
5	100112011		A	0	25	5	83
6	100112031		A	0	22	8	73
7	100112006		A	0	22	8	73
8	100112000		A	0	23	7	77
9	100111002		A	0	24	6	80
10	100112009		A	0	23	7	77
11	100111013		A	0	24	6	80
12	100112009		A	0	24	6	80
13	100112029		A	0	22	8	73
14	100112030		A	0	24	6	80
15	100112031		A	0	21	9	70
16	100112017		A	0	27	3	90
17	100108037		A	0	23	7	77
18	100108012		A	0	21	9	70
19	100112019		A	0	25	5	83
20	100110022		A	0	25	5	83
21	100112032		A	0	25	5	83
22	100103058		A	0	26	4	87
23	100112031		A	0	22	8	73
24	100112018		A	0	26	4	87
25	100112028		A	0	25	5	83
26	100112029		A	0	25	5	83
27	100112021		A	0	27	3	90
28	100111011		A	0	23	7	77
29	1001138		A	15	12	3	40
30	502110299		A	0	26	4	87
31	100112028		A	0	24	6	80
32	100112027		A	0	20	10	67
33	100112002		A	0	25	5	83
34	100112037		A	0	21	9	70
35	100112034		A	0	21	9	70
36	100112019		A	0	26	4	87
37	100112014		A	0	23	7	77
38	100112023		A	0	21	9	70

Table- 5 Ranges and frequencies

Ranges	Number of the students	%
0-9	0	0
10-19	0	0
20-29	0	0
30-39	0	0
40-49	1	2,22
50-59	1	2,22
60-69	3	6,67
70-79	**19**	**42,22**
80-89	**18**	**40**
90-100	3	6,67

Table- 6 Difficulties of the questions

Questions	Empty	A	B	C	D	E	Difficulty
1	0	21	1	1	2	**20**	0,44
2	0	2	**35**	1	3	4	0,78
3	1	0	0	**44**	0	0	0,98
4	0	2	**17**	13	7	6	0,38
5	0	1	0	0	0	**44**	0,98
6	1	**43**	0	0	1	0	0,96
7	0	0	0	0	**43**	2	0,96
8	1	**33**	2	4	1	4	0,73
9	1	0	1	**43**	0	0	0,96
10	0	4	4	8	**26**	3	0,58
11	0	0	0	**45**	0	0	1,00
12	0	1	**38**	0	6	0	0,84
13	0	4	0	0	5	**36**	0,80
14	0	1	**25**	6	2	11	0,56
15	1	2	0	4	**34**	4	0,76
16	1	1	0	**40**	0	3	0,89
17	1	**42**	0	0	1	1	0,93
18	1	0	0	8	**12**	24	0,27
19	0	2	2	**41**	0	0	0,91
20	0	0	**44**	0	1	0	0,98
21	0	2	**27**	8	1	7	0,60
22	1	0	0	**43**	1	0	0,96
23	1	**12**	12	0	20	0	0,27
24	1	7	**27**	7	3	0	0,60
25	1	36	5	0	**3**	0	0,07
26	0	**40**	1	2	0	2	0,89
27	1	**44**	0	0	0	0	0,98
28	1	0	0	**43**	0	1	0,96
29	1	0	0	0	**44**	0	0,98
30	0	0	**45**	0	0	0	1,00

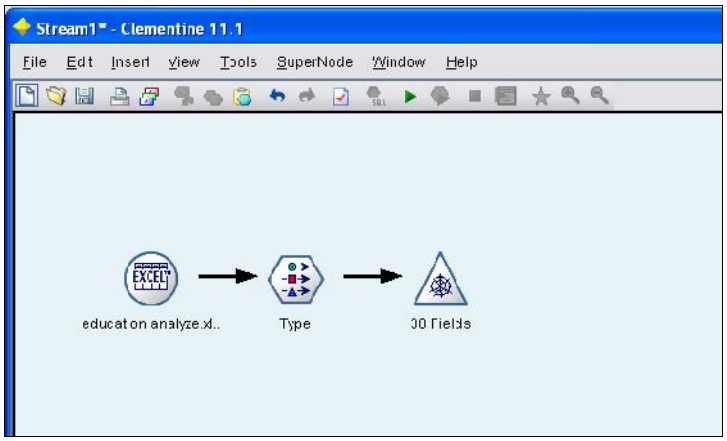

Figure-1 Model

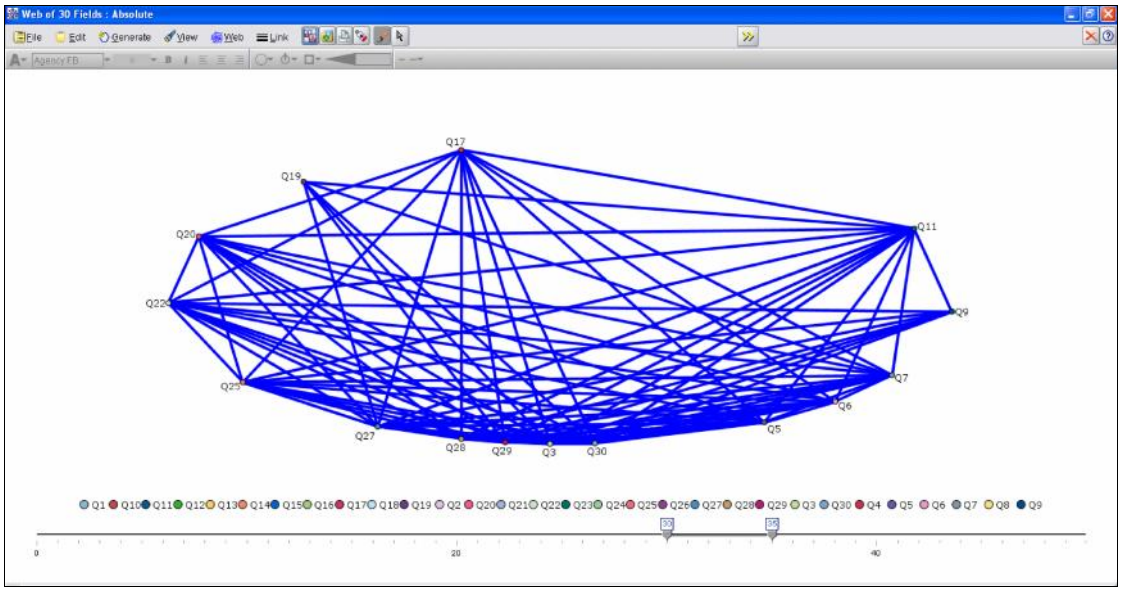

Figure- 2 Web Graph

REFERENCES

Adriaans, P., Zantinge, D. 1997. Data Mining. Addison-Wesley. p.47.

Al-Daihani, S., Abrahams, A., 2016. A Text mining analysis of academic libraries' tweets. The Journal of Academic Librarianship. Vol. 42, Issue 2, 135-143.

Al-Twijri, M., Noaman, A., 2015. A new data mining model adopted for higher institutions. Procedia Computer Science. Vol. 65, 836-844.

Campagni, R., Merlini, D., Sprugloni, R., Verri, M.C., 2015. Data mining models for student careers. Expert Systems with Applications. Vol. 42, Issue 13, 5508-5521.

Chalaris, M., Stefanos, G., Maragoudakis, M., Sgouropoulou, C., Tsolakidis, A., 2014. Improving quality of educational processes providing new knowledge using data mining techniques. Procedia - Social and Behavioral Sciences. Vol. 147, 90-397.

Freitas, A. 1997. A genetic programming framework for two data mining tasks: Classification and generalized rule induction, Proc. of the 2nd Annual Conference.

Freitas, A. 2002. Data Mining and Knowledge Discovery with Evolutionary Algorithms, Springer, p.1.

Giudici, P., 2003. Applied Data Mining, Wiley, p.6-7.

Gobert, J., Kim, Y.J., Sao Pedro, M., Kenedy, M., Betts, C., 2015. Using educational data mining to assess students' skills at designing and conducting experiments within a complex systems microworld. Thinking Skills and Creativity. Vol. 18, 81-90.

Han, J., Kamber, M., 2006. Data Mining Concepts and Techniques. Morgan Kaufmann. 229-230.

Kaur, P., Singh, M., Singh Josan, G., 2015. Classification and prediction based data mining algorithms to predict slow learners in education sector. Procedia Computer Science. Vol. 57, 500-508.

Khadijah M.S., Tasir, Z., 2013. Educational data mining: A review. Procedia - Social and Behavioral Sciences. Vol. 97. 320-324.

Kim, J., Han, M., Lee,Y., Park, Y., 2016. Futuristic data-driven scenario building: Incorporating text mining and fuzzy association rule mining into fuzzy cognitive map. Expert Systems with Applications. Vol. 57, 311-323.

Liu, B., 2007. Web Data Mining. Springer. 13-15-16.
Pena-Ayala, A. 2014. Educational data mining: A survey and a data mining-based analysis of recent works. Expert Systems with Applications. Vol.41, Issue 4, 1432-1462.

Robson, B., Boray, S., 2016. Data-mining to build a knowledge representation store for clinical decision support. Studies on curation and validation based on machine performance in multiple choice medical licensing examinations. Computers in Biology and Medicine, Volume 73, 1, 71-93.

Sen, B., S., Ucar, E.,2012. Evaluating the achievements of computer engineering department of distance education students with data mining methods. Procedia Technology. 1, 262-267.

Sen, B., Uçar, E., Delen, D., 2012. Predicting and analyzing secondary education placement-test scores: A data mining approach. Expert Systems with Applications. Vol. 39, Issue 10, 9468-9476.

Shahiri, A.M., Husain, W., Rashid, A.N., 2015. A review on predicting student's performance using data mining techniques. Procedia Computer Science. Vol. 72, 414-422.

Shukor, N., Tasir, Z., Van der Meijden, H., 2015. An examination of online learning effectiveness using data mining. Procedia - Social and Behavioral Sciences. Vol. 172, 555-562.

Stahovic, T., Lin, H., 2016. Enabling data mining of handwritten coursework.
Computers & Graphics. Vol. 57, 31-45.

Tan, P.N., Steinbach, M., Kumar, V. 2006. Introduction to Data Mining. Pearson International Edition. 328.
Westhpal, C., Blaxton, T., 1998. Data Mining Solutions. Wiley. p.186.