Volume 6, Issue 2, February 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Assessment of Student Feedback from the Training Course and Instructor' Performance through the Combination of Clustering Methods and Decision Tree Algorithms

Elham Taherifar
Department of Computer Engineering,
Islamic Azad University,
Tehran, Iran

Toraj Banirostam
Department of Computer Engineering
Islamic Azad University,
Tehran, Iran

Abstract- Assessment of instructor by students, in the last decade has been widely used in many universities, and there is always great interest in various aspects of assessment. Using data mining methods on data obtained from the assessment, it is possible to extract patterns that affect the quality of education.In this study, data collected from the survey forms of Turkish university students, which includes 28 items on the school and teachers,was evaluated. Due to the large number of items selected data set, first, using principal component analysis algorithm, the number of items was reduced from 28 items to two factors. Then using the two-step and Kohonen clustering algorithms on the assessment results, and comparing the two methods on the Silhouette basis, two-step cluster with creating three clusters with labels (satisfied, neutral, dissatisfied) and silhouette of 0.6 was considered as the best cluster analyse, then, using Quest decision tree algorithm on the results of a two-step clustering, and review of laws, items that play the most important role in the students' satisfaction and performance of instructor teachers , were identified.Accordingly, Question 7 (The course allowed field work, applications, laboratory, discussion and other studies) is more important than any other questions in predicting student satisfaction and their rating to the educators. The accuracy of the algorithm was calculate as 0.93 which has the good quality of classification.

Key words- educational data mining, Students evaluation of teaching,two-step clustering,self organizing map(SOM)

I. INTRODUCTION

performance evaluation in education, determining the principles of Competences, value and importance of the learning process is by criteria arisen in each educational system. Due to the quality of education at universities and teaching quality assessment is very important for educational policy maker. Due to the rapid growth of university education, the universities' concern is that they can keep the quality of teaching and learning resources as updated.On the other hand, teachers play a big role in improving student performance and rising educational standard. Discovering of the factors affecting the quality of teaching is very difficult [1] Hence, evaluating the performance of professors is a critical step in improving the effectiveness of learning system.Universities make available survey forms to students in order to evaluate the performance of the professors at the end of each semester, and aware their views about the lessons provided and its professors. University administrators can testquality of their professors teaching by relying on these surveys [2].Data from student feedback contains valuable data and patterns that helps educators improve their performance. It seems that one of the most important applications of data mining is in higher education. Data mining knowledge, as a new and growing field with combination of artificial intelligence and statistical knowledge discovers the desired factorsfrom mass of data.

The main purpose of the application of data mining in this article is to explore the factors affecting the level of satisfaction of students with the performance of Teaching Professor, as well as coureses. Data of this article is collected from the UCI website which has been collected in 2013, which corresponds to the data from the survey forms distributed among the students of the GAZI University in Turkey [3].In this paper, in Section 2, the basic concepts are discussed. In section 3, the works related data set is evaluated. Then, in Section 4, step by step data mining process is described. In Section 5, data set studied is introduced. In Section 6, the proposed model is expressed, and in section 7, the results are evaluated.

II. BASIC CONCEPTS

A. Performance Appraisal System

Performance appraisal system is essentially a formal interaction between the employee and the supervisor or manager, which makes identifying the strengths and weaknesses of the employee. The purpose of these systems isstrengthening strengths and monitoring weaknesses of employees in order to improve it. The system in many organizations, including Education is possible in both traditional and online systems [4,5].

B. Student Evalution of Teaching

In recent years, assessment of professors teaching quality by students is considered in many universities. Evaluation score is the principal method of teaching assessment which is done as questionnaire or survey forms, which are answered by students at the end of each semester. The most common is Likert survey [6], which one of the following scores are given to each of the questions (strongly agree, agree, neutral, disagree, stronglydisagree). This assessment has important objectives that include: [4]

1. It is considered as the foundation for university reforms.
2. Identifying student feedback to improve teaching and improve the quality of education
3. Providing information to the students to choose course and professor according to their goals and needs
4. The monitoring of teaching and administrative decisions (Enterprise, etc.)
5. The formulation of strategic and efficient plans in order to guarantee the quality of professors and learning process

C. Educational Data Mining

In recent years, due to problems in the field of extracting knowledge from a giant database, and use them to good decision making, data mining techniques and their use in education is taken into consideration. This new research field is called educational data mining which develops knowledge discovery techniques from the learning environments data especially students [7].

D .The Application of Data Mining in Teaching Performance Appraisal System

During the last century, student rating of professors has a special place the performance appraisal systems. Over the past seventy years, researchers have conducted more than 2,000 studies on this topic. Data from studies in turn are worthless. In some cases, the relationships among data analysis is so complex that traditional methods alone are not responsive.Over the past seventy years, researchers have conducted more than 2,000 studies on this topic. Data from studies in turn are worthless. In some cases, the relationships among data is so complex that traditional methods of analysis alone are not responsive. For this reason, in recent years, the relationship between students' evaluation of educational system performance and application of data mining in evaluating has been much discussed. Because, universities can obtain significant results in the extraction of data relationships using the results of applying data mining methods on data from evaluation forms. Different algorithms can be used to implement data mining in the field of education that include clustering, association rules, classification methods such as decision trees, neural networks and etc.

III. RELATED WORKS

Hajizadehet alidentified, the factors affecting not-selecting course selection by implementation of data mining methods on data, by the student. Three types of evaluation software Weka in this research has been done on the data, that in all three assessments on data are done author by software in this study (the number of times which a student has selected course) which is considered as purposefield, and the methods of association rules are used in order to identify the effectiveness of the role of professors and the course selected students success on passing the course.Also,difficulty level of course and survey questions are classified using Rep-Tree algorithm on the attendance fields, and results obtained from this assessment was that attendance Field and course difficulty level is directly related field withnb-repeat.Researcher by analyzing the survey questions concluded that the items of the readiness of the professorin the classroom, professor commitment to school and class, and the professor accountability to the questions of students, are more important than any other set of questions in determining the nb- repeat field [8].

Mohammad Abdollahet al, using k-means clustering algorithm on University survey data, the analysis of student feedbackis done in order to monitor the effectiveness of academic programs and improve the quality of education.It was done in such a way that clustering was performed on each of the questions individually and based on Euclidean distance criteria, andthree clusters was created according to the mean score, each cluster is labeled.The author achieved good resultson questions and relationship between them by examining each question clustered, while a few percent of students were satisfied with each of the items surveyed, or vice versa [9].

Necla Gündüzet al, reached the interesting patterns combining data mining methods (Supervised and unsupervised) with the assessment of survey fields such as the presence of students and how to answer the 28 questions.NeclaGündüz first classified data using two algoritms K-Means and hierarchical methods in R software, then Classified the clusters produced by the K-Means method according to the survey's items, so that the clusters were considered as objective variable and 28 questions regardless of the survey fields (attendance, course difficulty level, number of Number of times the student is taking this course) as the input fields.By using decision tree algorithm on data, clusters were classified based on the importance of the question. After analyzing the data, they concluded that 93 percent of data were classified correctly. The results of the algorithm werecompared with the results of algorithms of support vector machines, neural networks, Random Forest and Boosting in accordance with ROC curve [2].

Oyedotunet al, predicted the degree of success of students in passing the course selected using the methodology of neural network classification.For this purpose, two methods of propagation neural network and Radial Basis Function networks have been used to predict. The purpose of using neural networks in this article is relationships mapping between some training features of students, and the frequency of failure to pass the course.The results of running this algorithm shows that propagation neural network has had better performance than Radial Basis Function networks in terms of the error rate, and training time. [10]

IV. EDUCATIONAL DATA MINING PROCESS

Educational data mining process converts the raw data fromeducational systems into useful information, so that they will be usable by developers of educational software, professors, educational researchers and so on. Data mining process include the followingsteps: [11]

A. Preprocessing

In the first step, educational systems data must be converted to a suitable format for mining. One of the main tasks of this section, if a plurality of features, is extraction of important feature susing principal component analysis technique. In this way, by combining existing features, fewer features are obtained, so that these features have a large part of the information contained in the initial properties and are called component [12].

B. Data mining

At this point, data mining techniques are applied on data pre-processed. In this paper, two two-step clustering and the Kohonen algorithmsare used. In Two-step cluster analysis algorithm,it is assumed that the variables areindependent in the clustering model, degree of Likelihood is considered as the entities distance. It is assumed that, each continuous variable has normal distribution and discrete variablehas multivariate normal distribution.Two-step clustering algorithm is carried out in two stages. At first, based on cumulative hierarchical clustering approach, each record is considered as a cluster. All records one by one, and on the basis of criteria are intended and for existences distance, decisions are made about its integration with the cluster before or beginning of a new cluster.Then, in the next step, sub-clusters of the first phase clusters are considered as inputs, and based oncumulative hierarchical clustering, they are converted to the desired number of clusters [13,14].Clustering algorithmis a Kohonen self-organizing map based on an unsupervised neural network, which is formed by nerve cells in a regular grid structure with low dimension. SOM architecture is very simple which is shown in Figure 1.Basic units are neurons which are organized in two layers: the input layer and output layer (output map). All input neurons are connected to all output neurons, and these connections have strengths or relevant weights.During training, each neuron fights with all other neurons to win.Neurons are connected to each other by a neighborhood function. In each phase of the training, each input vector of set of input data is randomly and based on maximum likelihood,it enables neuron in the output layer that is called winner cell. The similarity is usually measured by the Euclidean distance between two vectors. [15,16]

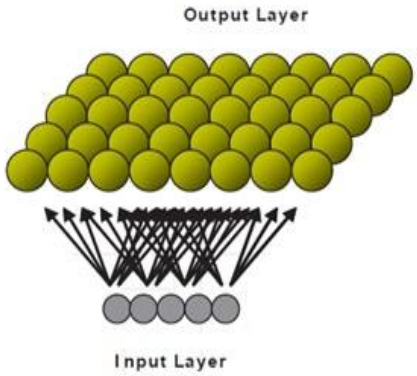


Fig. 1 kohonen Network

Decision tree technique data is one of the data mining common methods used for classification and prediction. The structure of this technique is based on flowchart, which explains its prediction as set of rules.Based on the training set, a tree is created in which each internal node shows a test on an attribute, each branch represents the conclusion of the test, and each leaf holds a class label. [17]

C. Post Processing

The final step involves evaluating the results obtained and calculation of the accuracy of the model which is interpreted and used for educational decisions. Partition node is the most appropriate way to assess and validate the model.The model is created based on a part of the sample, and the model is tested based on other sections. Accuracy evaluation index is considered in the Silhouette clustering algorithms, 18] and examination of accuracy criteria and F-score are considered in classification methods. [19]

V. DATA SET

Data set studied is collected from 5820 students scoring to evaluation form items of Turkish Gazi University professors. Items on this data set includes 5 fields (Instructor's identifier, Course code, the Number of times the student is taking this course, student attendance and course difficulty level) and 28 questions raised about the course and relevant professor.Survey forms are distributed in 12 classes, students can choose one of the numbers of 1 to5 to answer questions and provide their opinions, choose. (1: strongly disagree, 2: disagree, 3: neutral, 4: agree, 5: Completely agree). 12 initial questions are related to course and 14 next questions are related to the instructer' characteristics.

Table I Data Set

instr	Instructor's identifier{1,2,3}
class	Course code{1-13}
repeat	Number of times the student is taking this course{0,1,2,3,...}
attendance	Code of the level of attendance{0, 1, 2, 3, 4}
dificulty	Level of difficulty of the course as perceived by the student{1,2,3,4,5}
Q1	The semester course content, teaching method and evaluation system were provided at the start
Q2	The course aims and objectives were clearly stated at the beginning of the period.
Q3	The course was worth the amount of credit assigned to it.
Q4	The course was taught according to the syllabus announced on the first day of class.
Q5	The class discussions, homework assignments, applications and studies were satisfactory.
Q6	The textbook and other courses resources were sufficient and up to date.
Q7	The course allowed field work, applications, laboratory, discussion and other studies.
Q8	The quizzes, assignments, projects and exams contributed to helping the learning.
Q9	I greatly enjoyed the class and was eager to actively participate during the lectures.
Q10	My initial expectations about the course were met at the end of the period or year.
Q11	The course was relevant and beneficial to my professional development.
Q12	The course helped me look at life and the world with a new perspective
Q13	The Instructor's knowledge was relevant and up to date.
Q14	The Instructor came prepared for classes.
Q15	The Instructor taught in accordance with the announced lesson plan.
Q16	The Instructor was committed to the course and was understandable
Q17	The Instructor arrived on time for classes.
Q18	The Instructor has a smooth and easy to follow delivery/speech.
Q19	The Instructor made effective use of class hours.
Q20	The Instructor explained the course and was eager to be helpful to students.
Q21	The Instructor demonstrated a positive approach to students.
Q22	The Instructor was open and respectful of the views of students about the course
Q23	The Instructor encouraged participation in the course.
Q24	The Instructor gave relevant homework assignments/projects, and helped/guided students.
Q25	The Instructor responded to questions about the course inside and outside of the course
Q26	The Instructor's evaluation system (midterm and final questions, projects, assignments, etc.) effectively measured the course objectives.
Q27	The Instructor provided solutions to exams and discussed them with students.
Q28	The Instructor treated all students in a right and objective manner.

VI. PROPOSED MODEL AND ANALYSIS OF RESULTS

The hybrid model of this study is shown in Fig. 2 First, Two-Step and Kohonen algorithms runs on the 28-item dataset. Then, using PCA and extracted two components of the 28 items, clustering algorithm runs again, and their results are compared. The best method of cluster is chosen, and by running decision tree algorithm on the results of the clustering, we can predict satisfaction of students with professors and training period.

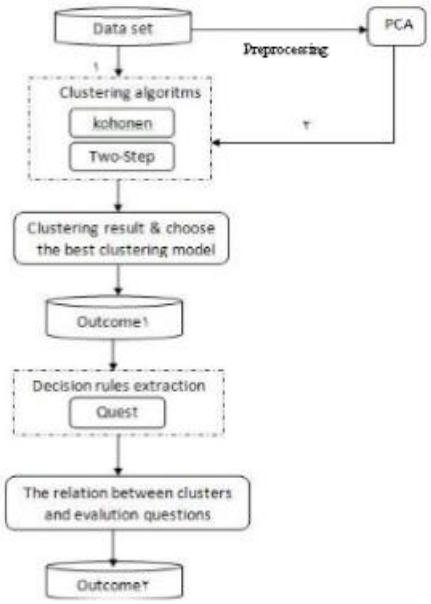


Fig 2. proposed model

A. Conclusions obtained from the Two-Step and Kohonen algorithms

In Fig.3, Two-Step and kohonen will be shown in at best mode of silhouette. In addition to the Silhouette value, the number of clusters, largest cluster size, smallest cluster size are displayed. According to the Silhouette value, software, Two-Step Model software with 4 clusters Si = 0.574 was chosen as the first model. Also, Kohonen model with 9 clusters and Si = 0.446 has worse performance than the previous model.

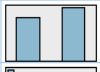
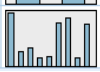
Use?	Graph	Model	Build Time (mins)	Silhouette	Number of Clusters	Smallest Cluster (N)	Smallest Cluster (%)	Largest Cluster (N)	Largest Cluster (%)	Smallest/ Largest
☐		Tw...	< 1	0.574	4	630	14	1559	35	0.404
☐		Tw...	< 1	0.572	3	920	21	1756	40	0.524
☑		Tw...	< 1	0.565	2	1945	44	2414	55	0.806
☐		Ko...	< 1	0.451	9	127	2	957	21	0.133

Fig. 3 Conclusions Obtained from Clustering Without PCA

B. Data clustering with two Kohonen and Two-Step algorithms by applying PCA

Because, the numbers of survey items are 28, and surely whatever number of inputs are more, the problem will be more complex with lower quality, and is provided with more time, we can reduce the complexity of model by applying the method of PCA and reducing the number of items (columns), and Silhouette of methods used in the previous step can be improved. To do so, instead of taking the 28 inputs to node,we can enter Auto-cluster node of two factors extracted from the PCA phase.Table II shows the eigenValues of two factors, which the first factor covers the most dispersion.

Table II Components Extracted from the PCA

Component	Initial EigenValues		
	Total	% of variance	cumulative
1	23.041	82.289	82.289
2	1.253	4.475	86.764

Fig.4 shows clustering models run by two factors, the number of clusters and Silhouette in both clustering methods. As it is clear, like the Phase before (without PCA result), the Two-step method with three clusters and Silhouette=0.623 has better performance than Kohonen method with 12 clusters and Silhouette =0.597. The models are shown based on Silhouette priority.

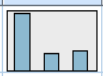
Use?	Graph	Model	Build Time (mins)	Silhouette	Number of Clusters	Smallest Cluster (N)	Smallest Cluster (%)	Largest Cluster (N)	Largest Cluster (%)	Smallest/ Largest
☑		Tw...	< 1	0.623	3	747	17	2686	61	0.278
☐		Tw...	< 1	0.615	4	733	16	1551	35	0.473
☐		Ko...	< 1	0.597	12	68	1	892	20	0.076

Fig. 4 Conclusions obtained from clustering with PCA results

C. Evaluation of clustering results

According to Table III, it is important to mention that the use of Dimensionality dimension reduction PCA was effective in both methods, and led to no only reduce the complexity of the problem and time of running by reducing the dimensions from 28 items to 2 items, but the result of clustering is also somewhat improved in both methods. By comparing the results of both methods:

Two-step method in terms of the number of clusters obtained, and Silhouette has better performance than kohonen method, and as is chosen a better clustering model for further actions.

Table III . Results clustering

Clustering algoritms	Clustering without preprocessing		Clustering with preprocessing			
			Train data		Test data	
	Silhouette	Number of Cluster	Silhouette	Number of Cluster	Silhouette	Number of Cluster

kohonen	0.448	9	0.597	12	0.56	12
Two-Step	0.574	4	0.623	3	0.6	3
Execution Time	48 Seconds		10 Seconds			

1) Analysis of the results of PCA + Two-Step as better clustering algoritm: Two-Step clustering results show that:
Cluster 1: 61.6% of all evaluators (2686 people) are in cluster 1. 90% of evaluators get the course for the first time, and 32 percent of them have been present in all classes held. The cluster consists of students who are satisfied with their instructer and course and have given points 3 and 5 to questions from 1 to 5.
Cluster 2: 21.2% of all evaluators (926 people) are in cluster 2. 87% of evaluators get the course for the first time, and 38% of them had missed three sessions. The cluster consists of students who are satisfied with their instructer, and but are no satisfied with getting course and curriculum.
Cluster 3: 17.1% of all evaluators (747 people) are in cluster 3.84% of evaluators get the course for the first time, and half of them had Absence Over three sessions. The cluster consists of students who are not satisfied with their instructer and course and have given points 3 and 5 to questions.
2) *Labeling the clusters according to analysis results*: According to the results of clustering model Two-Step, Three clusters have been achieved. Most students who are at first cluster, have been satisfied with their instructer and course. Most students who are at third cluster, have been dissatisfied with their instructer and course .Cluster 2 represents students who have gotten the course, are dissatisfied, but are satisfied with their dissatisfied.

Table IV Labeling the clusters

Cluster 3	Cluster 2	Cluster 1	Cluster Number
dissatisfied	Neutral	satisfied	Label

D .Forecast of student satisfaction with Quest decision tree algorithm
One advantage of decision tree algorithms is allowing for better understanding of important fields, because in a decision tree which automatically the most important fields are transferred in the upper decision tree nodes, and also discarded minor decision tree fields.It helped that, in this study a decision tree is used to identify critical questions of the 28 questions raised about the instructor which has the most important role in the classification of them based on the label dedicated in the clustering.This makes good view of the importance of each item to be achieved. For this purpose, Quest algorithm is used for classification f obtained clusters from the Two-Step algorithm in the previous step that shows the most ideal possible clusters.
1) How to run the QUEST algorithm: To run this algorithm, we select QUEST node. Questions Q1 ... Q28 are input variables and the target variable (predictor variable) is Cluster Field ($ T-Two-Step2), which is obtained from the previous stage. Fig. 5 shows a view of the model after the run.

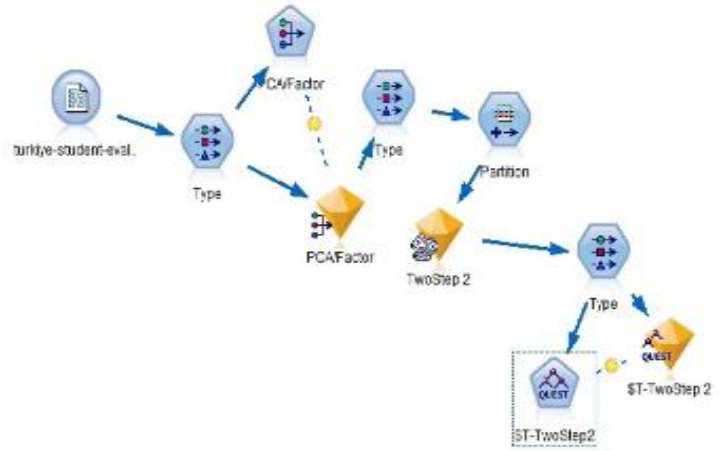


Fig.5 Execution Flow

Table V shows rules of discovered that are obtained from the decision tree. These rules help to identify items that affect the educators scoring.

Table V Classification Rules Obtained from Quest algoritm

The Rules Tree
If(Q7<3 AND Q22<3) then degree of Student satisfaction = **dissatisfied**
If (Q7<3 AND Q22>2) OR (Q7>2 AND Q1=1) then then degree of Student satisfaction = **neutral**
If(Q1>3 AND Q7>3) then

then degree of Student satisfaction = **satisfied**

E .Analysis of the Results of Classification Clusters

According to the label of the clusters, which was mentioned in section 4-6, it can be concluded that the majority of students who have given a low point to Question 7(The course allowed field work, applications, laboratory, discussion and other studies.) and Question 22(The Instructor was open and respectful of the views of students about the course), are in the dissatisfied cluster. If they have given low point to Question 7 and high point to Question 22, are classified in cluster 2, and have abstention about the performance of their course and instructor. If most students have given points 3, 4 or 5 to question 7, if give point 1 to question 1(The semester course content, teaching method and evaluation system were provided at the start), are in cluster 2, and if give Points above 2 to this question,are in the satisfied cluster. Based on the results, we can say that, two items of the study and laboratory aspects and being project of the course, and the determination of curricula, and teaching methods from the beginning of the semester have the greatest impact in predicting students' satisfaction with their course and instructor.Similarly, how they respond to items of the studies, laboratory aspects and being project of the course and the importance of instructor to student feedback have a significant role on their dissatisfaction.

It is important to note that this classification is formed according to the majority of students, and may be classified according to the percentage of students who are in clusters other than the cluster expressed according to this classification. Given the importance of this issue, and to what extent the model is accurate and reliable, Analysis Node is used to evaluate the accuracy, given that, at the stage of data Clustering, 75% of data is used for training and 25 percent to test the model, by performing analysis node, the number of right and wrong records in both training and testing, the confusion table and accuracy can be seen.Percentage of right and wrong records of training and testing data is as follows. Number of training and testing records is 4359 and 1461, respectively, which Analysis Accuracy of this model is obtained 93.77%. That means in accordance with this classification, about 94 percent of the records are in the predicted cluster, and the second factor, which includes questions related to the course is more important to classification clusters

Table VI . Right and wrong records predicted

'Partition'	1_Training		2_Testing	
Correct	4,132	94.79%	1,370	93.77%
Wrong	227	5.21%	91	6.23%
Total	4,359		1,461	

F. Confusion Matrix

According to Table VII, 2584 records of test data from cluster 1, 709 records of cluster 2 and 839 records of cluster 3 are correctly predicted.

Table VII . Confusion matrix

'Partition' = 1_Training	cluster-1	cluster-2	cluster-3
cluster-1	2,584	72	30
cluster-2	38	709	0
cluster-3	61	26	839
'Partition' = 2_Testing	cluster-1	cluster-2	cluster-3
cluster-1	891	28	17
cluster-2	16	213	0
cluster-3	19	11	266

G. Evaluation of the results with previous works

Chart 1 shows a comparison between the proposed method and the obtained results of the article [2]. After calculating the F-Measure, Recall and Precision for each model (Article [2] and the proposed algorithms). It is concluded that the accuracy of the second model is reached to 0.937, and classifier error is reduced to 0.062, and it shows improving classification accuracy than previous methods. But, the indices F-Measure, Recall and Precision Performance have declined compared to the previous work. Table VIII illustrates this issue.

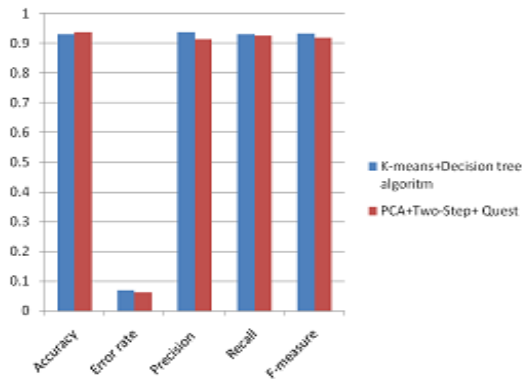


Chart 1 Comparison of F-Measure, Recall and Precision Results of the proposed model with article [2]

Table VIII. Comparison of evaluation results with related work

Criterion	K-Means+ D-Tree algoritm	PCA+ Two-Step Quest+
Accuracy ✓	0.93	0.937
Error Rate ✓	0.07	0.062
Precision	0.937	0.915
Recall	0.932	0.926
F-Measure	0.934	0.920

VII. CONCLUSIONS AND RECOMMENDATIONS

With clustering of results of the 5820 evaluators of teaching and courses taught at the university, three clusters have obtained, which the first cluster represents the people satisfied, second cluster represents people dissatisfied with the abstentions and the third cluster represents people dissatisfied with the course and the teaching quality of their instructor.According to the label that is assigned to each cluster, in the next stage of the project, each cluster is considered as a class field (goal) and evaluation questions are considered as input of QUEST classification algorithm. The purpose of classification of clusters is to discover important questions and more effectiveness in Label clusters and rating instructor.The rules of this tree indicate that questions Q7 (The course allowed field work, applications, laboratory, discussion and other studies), Q22 (The Instructor was open and respectful of the views of students about the course), and Q1 (The semester course content, teaching method and evaluation system were provided at the start.), are the most effective factors on students' satisfaction and performance of professors. According to results conducted in this study, other questions have little importance in view of the quality of the education process.In this study, clustering is done on the 28 items (question posed in the survey), so that the algorithms used don't considerthree fields ofnb-repeat and attendance and difficulty. Each of these items in turn, can affect the response of the students on the questions.

For future research, we can examine the role of fields ofcourse difficulty,the number ofthe presence of students in classes, as well as the number of courses in responding to the questions.The use of association rules algorithmon the assessment results can be appropriate to discover effective relationships between items, for example, the relationship between the difficulty of the course and the number of courses taken or the relationship between the students' participation in class, and his points to each of the survey questions.

REFERENCES

[1] Huabin Qu, Xueqing Li, " *Application of Data Miningin the Assessment of Teaching Quality*", Frontier and Future Development of Information Technology in Medicine and Education Lecture Notes in Electrical Engineering ,Volume 269, pp 1813-1819,2014.

[2] N.Gündüz, E.Fokoue, "*Data Mining and Machine Learning Techniques for Extracting Patterns in Students*' Evaluations of Instructors",Rochester Institute of Technology ,The John D. Hromi Center for Quality and Applied Statistics (KGCOE) ,PP.1-28, 2013.

[3] G. Gunduz &E. Fokoue, UCI Machine Learning Repository[https://archive.ics.uci.edu/ml/datasets/Turkiye+Student+Evaluation]. Irvine, CA: University of California, School of Information and Computer Science ,visit 2013

[4] S.Mardikyan, B.Badur, " *Analyzing Teaching Performance of Instructors Using Data Mining Techniques*", Informatics in Education, Vol. 10, No. 2, PP. 245–257, 2011.

[5] A. F. Ola, Prof. S.Pallaniappan, "*A data mining model for evaluation of instructors' performance in higher institutions of learning using machine learning algorithms*", International Journal of Conceptions on Computing and Information Technology, Vol. 1, Issue 2, pp.17-22, Dec 2013,

[6] R. LIKERT, " *A Technique for the Measurement of Attitudes*", Archives of Psychology, 22(140), pp.1-55, 1932.

[7] C.Romero. And S.ventora ,"*educational data mining: A survey from 1995 to 2005*", expert systems with applications, pp.135-146,2007.

[8] Hajizadeh,N., Ahmadzadeh,M., "*Analysis of factors that affect students' academic performance - Data Mining Approach* ", International Journal of advanced studies in Computer Science and Engineering IJASCSE, Volume 3, Issue 8,pp.1-4, 2014.

[9] Anwar Muhammad Abaidullah, Naseer Ahmed, and Edriss Ali, " *Identifying Hidden Patterns in Students Feedback through Cluster Analysis*", International Journal of Computer Theory and Engineering, Vol. 7, No. 1, February 2015.

[10] K. Oyedotun,N. Tackie, O.Ebenezer "*Data Mining of Students' Performance: Turkish Students as a Case Study*"، Intelligent Systems and Applications, 2015, Published Online August 2015 in MECS (http://www.mecs-press.org/).

[11] C.Romero,S.Ventura & De Bra P. , "*Knowledge discovery with genetic programming for providing feedback to courseware author*". User Modeling and User-Adapted Interaction: The Journal of Personalization Research, 14(5), pp.425-464,2004.

[12] M.Ringnér, " *What is principal component analysis?* ", Nature Publishing Group, 2008.(http://www.nature.com/naturebiotechnology)

[13] J.Han and M.Kamber,"*Data Mining: Concepts and Techniques*", The Morgan Kaufmann Series in Data Management Systems The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Edito ,2006

[14] D.Şchiopu et al.,"*Applying Two-Step Cluster Analysis for Identifying Bank Customers' Profile*".[online], p66-75. Sep2010 (http://citeseerx.ist.psu.edu/index)

[15] J .Vesanto, "*Clustering of the self-organizing map*", Neural Networks, IEEE Transactions on (Volume:11 , Issue: 3) ,pp.586-600.,2010.

[16] M.Huliane, A.Silva Cícero, and L.Flavius Gorgônio, " *A Self–Organizing Map Based Strategy for Heterogeneous Teaming*", computer and information science.[online],November2012 (http://www.intechopen.com/)

[17] J.Han and M.Kamber "Data Mining: Concepts and Techniques", Morgan Kaufmann, 2001

[18] P. J.Rousseeuw, " *Silhouettes: a graphical aid to the interpretation and validation of cluster analysis*", Journal of Computational and Applied Mathematics 20, pp.53-65,1987

[19] R.H. Higa; C. L. Tozzi, "*Prediction of binding hot spot residues by using structural and evolutionary parameters*", Genetics and Molecular Biology, vol.32, no.3, pp.626-633 , 2009

