

## Experiences in Mining Educational Data to Analyze Teacher's Performance: A Case Study with High Educational Teachers

Abdelbaset Al-Masri

*Department of Information Technology, Faculty of Engineering and Information Technology, Al Azhar University-Gaza, Palestine,  
abed\_massri@yahoo.com*

### **Abstract**

*Educational Data Mining (EDM) is a new paradigm aiming to mine and extract knowledge necessary to optimize the effectiveness of teaching process. With normal educational system work it's often unlikely to accomplish fine system optimizing due to large amount of data being collected and tangled throughout the system. EDM resolves this problem by its capability to mine and explore these raw data and as a consequence of extracting knowledge. This paper describes several experiments on real educational data wherein the effectiveness of Data Mining is explained in migration the educational data into knowledge. The experiments goal at first to identify important factors of teacher behaviors influencing student satisfaction. In addition to presenting experiences gained through the experiments, the paper aims to provide practical guidance of Data Mining solutions in a real application.*

**Keywords:** EDM, Knowledge, Survey, C4.5

### **1. Introduction**

Teaching-performance evaluations play important role in assessment of the quality of classroom instruction so most of educational institutions use Teacher Assessment Survey (TAS) to get student opinions measuring student satisfaction and to extract wide-range knowledge related to teaching behaviors in the courses they teach. Typical goals of the analysis of the TAS as follows [1,6]:

- What are the major teaching constructs that are satisfied (or dissatisfied) by students?
- How does the “dissatisfied” vary over student attributes or their combinations (Eg., across level, major, faculty, gender, *etc.*)? Are there any unusual variations; Eg., are there any subgroups of student in specific faculty, having a specific major, who are more dissatisfied as compared to similar combinations?
- Can dissatisfied students be partitioned into subsets, where students within each subset share lot of common characteristics?
- What are good predictors of student dissatisfaction?
- Identify “interesting” subsets of dissatisfied students.

Similar questions can be asked about satisfied students. The main goal of the paper is to build data mining models discovering teacher behaviors that are associated strongly with student satisfaction for example students of teachers who implement combination of teaching behaviors with specific rating scores (*e.g.*, Teachers' personality, scientific background, *etc.*) make considerable gains in student satisfaction. Data mining is fruitful for educational institutions in finding the

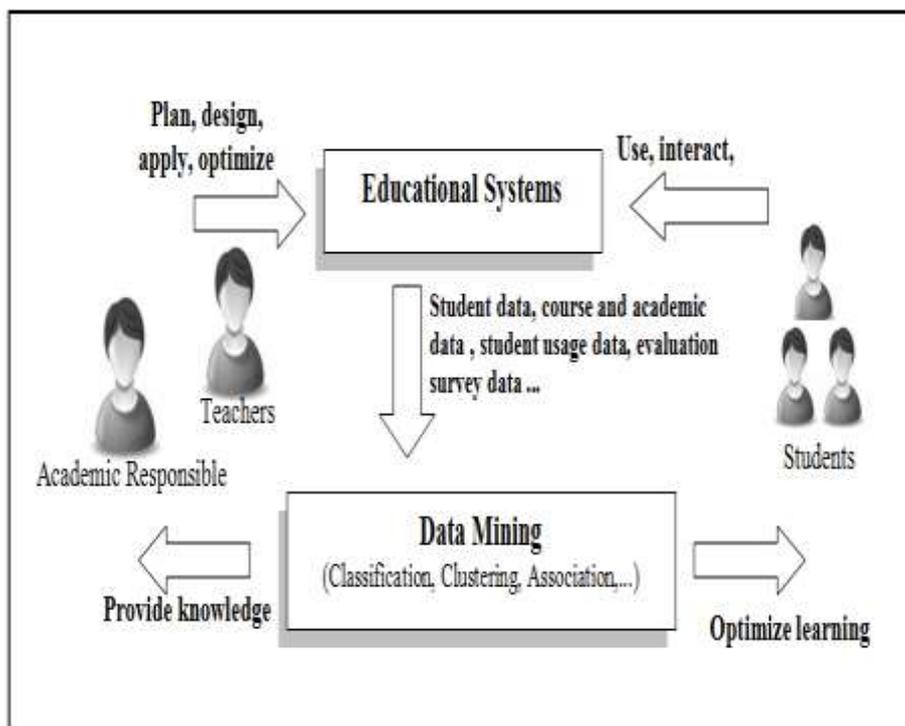
---

Received (March 27, 2017), Review Result (October 23, 2017), Accepted (October 31, 2017)

factors that affect student satisfactions strongly and how these factors related to other. Data mining is one of the rapidly growing fields according to the huge amount of data accumulated by running institutions to its business [2]. New methods for data mining have been studied, which describe data exploring and knowledge-extracting processes including data preprocessing, data analysis, and knowledge representation. The common tasks of data mining include induction of classification models [12], association rules [13], evolution and deviation analysis and making clustering for similar data objects [2]. To make data suitable for mining, preparing methods should be applied to it for cleansing and transforming data to a format ready for the mining [2]. Educational Data mining [8] is a novel research area offering solid ground for applications interested for educational environment. Educational data mining can mine educational data to extract knowledge related to learning activities. Figure 1 demonstrates how the data mining could strongly contribute in providing the knowledge necessary to educational responsible for making the correct decisions to optimize the educational systems and shows how the usage of the data mining in educational institutions forms an interactive cycle for learning improvement.

The main objective of this study is to use data mining techniques to improve student achievement through the followings:

- a) Get a detailed understanding of the current situation of teaching behaviors in the classroom.
- b) Discover the teaching behaviors that are associated strongly with student satisfaction and can be used as significant predictors for the teacher performance.
- c) Design a future plan for achieving specific improvements based on the findings in (a, b).



**Figure 1. The Cycle of Employing Data Mining in Educational Institutions**

## 2. Related Work

Taherifar and Banirostam [5] used data mining techniques on data collected from survey forms of Turkish university students. In their work, they used Principle component analyses to reduce the data set then they applied and compared two-step and Kohonen clustering algorithms. Then, they used Quest decision tree algorithm on the results of a two-step clustering and extracted the important predictors that identify student satisfactions.

Hamada and Abadi [16] analyzed opinion of students about their teachers in Teacher Evaluation system. In their paper, they showed an application of data mining and presented analysis of the obtained result using WEKA tool.

Hemaid and El-Halees [3] investigated teaching performance factors using data mining. In their work, they proposed a model to evaluate teacher performance through the use of data mining techniques like association, and classification rules and also they applied these techniques using WEKA tool on real data collected for teachers from the Ministry of Education and Higher Education in Gaza City.

Ajay and Saurabh [17] used data mining techniques to evaluate performance of university teachers. They used four classification techniques which are Naive Bayes, ID3, CART and LAD tree. The Naïve Bayes classifier was the best algorithm having lowest average error compared to others.

Palshikar *et al.*, [6] introduced how survey responses can be analyzed and processed using data mining techniques and described a tool called QUEST for analyzing survey responses. They presented a real-life case-study where QUEST was used to analyze responses from a real-life employee satisfaction survey in an IT company.

Barracosa and Antunes [4] proposed a new methodology for predicting teachers performance based on the analysis of educational surveys. In their methodology, they use classification and sequential pattern mining for identify and discovering meta-patterns describing frequent teacher behaviors.

Abu Naser *et al.*, [18] developed Artificial Neural Network model for predicting a sophomore student performance. They tested the model and showed that the model was able to predict the performance of more than 80% of prospective students.

## 3. Data Set Description

This paper inspects real data collected from an educational database system and via on-line Teacher Assessment Survey (TAS) in a higher education institution. The institution conducts survey for each course they teach in the last of each semester. This survey aims to examine issues viewed as essential to students by seeking their opinion on a number of factors related to teaching, assessment and support provided by their course teachers in a classroom. The conducted survey contains 20 of structured questions offering fixed options to a student, who chooses one from them. For example, “*answer students' questions clearly: Excellent, Good, Average, Poor, Very poor*”. The TAS questions are grouped into 4 categories as shown in Table 1 and the questions within each category gather responses about a specific aspect of a teacher behavior in a classroom.

**Table 1. TAS Questions**

Categories	Questions
Personal Characteristics	1. has strict and an influential figure. 2. the overall appearance stylish and decent. 3. is committed to the dates of the lectures. 4. treats students with humility and respect
Scientific Background	5. proficient in the scientific materials. 6. answers students' questions clearly 7. widely acquaintance in diverse areas of knowledge. 8. presents the material in suitable way to student levels. 9. presents the material in a coherent and sequential displays 10. covers course topics during the semester.
Professional Skills	11. enriches the material by examples. 12. uses methods growing the student thinking. 13. grows positive attitudes among students towards the Specialization. 14. invests time in the lecture presentation of the material and the scientific activities. 15. develops research skills by different activities. 16. encourages students to use a variety of knowledge sources.
Assessment	17. uses a variety of questions in the exam. 18. covers the most scientific topics in the exam. 19. the number of questions are proportional to the exam time. 20. assesses duties and activities objectively.

Before analyzing the data, the satisfaction index (SI) is computed for each TAS question as indicator that assesses the overall student satisfactions for that question intention (aspect of teaching behavior). To establish a SI of  $i$ th question ( $Q_i$ ) answered by  $N$  students, the student answer for  $Q_i$  is mapped to number value ( $v$ ) on a scale from 0 to 4 (Excellent{4}, Good{3}, Average{2}, Poor{1}, Very poor{0}), where 0 is the worst value while 4 is the best value. The satisfaction index of  $Q_i$  whose fixed domain  $D_i$  of possible answers ( $0..|D_i|-1$ ) is calculated as shown in equation 1 ( $n_{iv}$ = no. of students that selected answer  $v$  for  $Q_i$ ) [5]. If all students answer 0 to a question  $Q_i$ , then  $S(Q_i) = 0\%$ . If all students answer  $|D_i| - 1$  to a question  $Q_i$ , then  $S(Q_i) = 100\%$ .

$$S(Q_i) = 100 \times \frac{\sum_{v=0}^{|D_i|-1} v \times n_{iv}}{(|D_i| - 1) \times N} \quad 0 \leq S(Q_i) \leq 100.0 \text{ for each question } Q_i. \quad (1)$$

We also computed the overall SI for each category (related questions of a specific concern) as the average of the SI  $S(C_j)$  for  $j$ th category ( $C_j$ ) containing  $N$  questions (see equation 2) [5].

$$S(C_j) = \frac{\sum_{i=1}^N S(Q_{ij})}{N} \quad (2)$$

After processing the collected data, the data comprises 608 records, each record consists 29 attributes describing a course and student satisfactions for the course-teaching aspects. Table 2 presents the attributes and their description as taken from the source database and after calculating the satisfaction measures.

**Table 2. Data Set Attributes**

Attribute	Attribute Description
Teacher_id	The ID number of teacher
Qualification	The qualification of teacher; Values: BA, MA, Ph.D.
Course_id	The ID number of course
Faculty	College name
Major	The specific of the course
Course_level	The course level in a curriculum Values:1..5
No_Students	The total number of student in the course
Satisfaction index scores of 20 questions in the survey	Satisfaction index scores of 20 questions in the survey of studies. Values: 0..100
Total_SI	The total average of the SI for all questions
Achievement_ average	The total average of the student marks in the course taught by the teacher Values: 0..100

## 4. Experiments

As we mention before, using data mining on performance data will be fruitful in building classification and predictive models to know the well-defined teaching performance indicators influencing the student satisfaction. So we will discuss later our experiences in mining educational data and migrating these data to knowledge.

### 4.1. Experiment 1: using only responses to predict student satisfaction

In the first experiment, we used only the responses to the questions to build a predictive model to predict student satisfaction for a teacher performance without using any other data (Eg., teacher qualification, course level, no. of student in the course *etc.*). in this experiment, we used the SI attributes for the TAS categories with making the attribute values to be class label (poor, average, good), First we discrete the total average of the SI for all questions to make it class label (poor, average, good).

**Table 3. Data Set Attributes of Experiment1**

Fields	Description	Values Domain	Direction	
PersChar	The overall SI of Personal-Characteristics category	>= 80 79-65 < 65	good average poor	input
ScBackground	The overall SI of Scientific-Background category	>= 80 79-65 < 65	good average poor	input
ProfSkills	The overall SI of Professional-Skills category	>= 80 79-65 < 65	good average poor	input
Assessment	The overall SI of Assessment category	>= 80 79-65 < 65	good average poor	input
TechPerfAvg	The total average SI of the teacher performance	>= 80 79-65 < 65	good average poor	Output (Target )

After preparing the data set to make it suitable for mining as shown in Table 3, we applied the c4.5 classification algorithm which is a tree-based classification and prediction method which uses recursive splitting for the training data set into subsets with similar target field values. The c4.5 examines the input fields to find the best split, measured by the reduction in an impurity index that results from the split. The split defines multiple subsets, each of which is subsequently split into more subsets and so on until one of the stopping criteria is triggered [14, 15]. Figure 2 and Figure 3 illustrate the result of the classification on the TechPerfAvg (The total average SI of the teacher performance) as a target class.

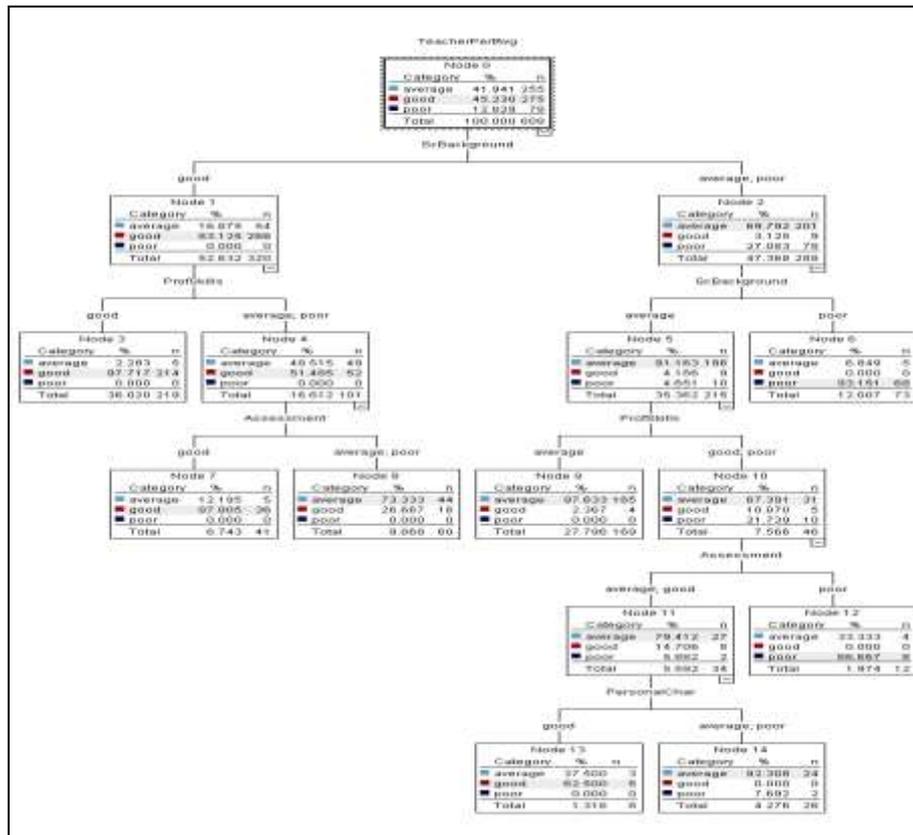


Figure 2. Teaching-behavior Classification Tree in Experiment 1

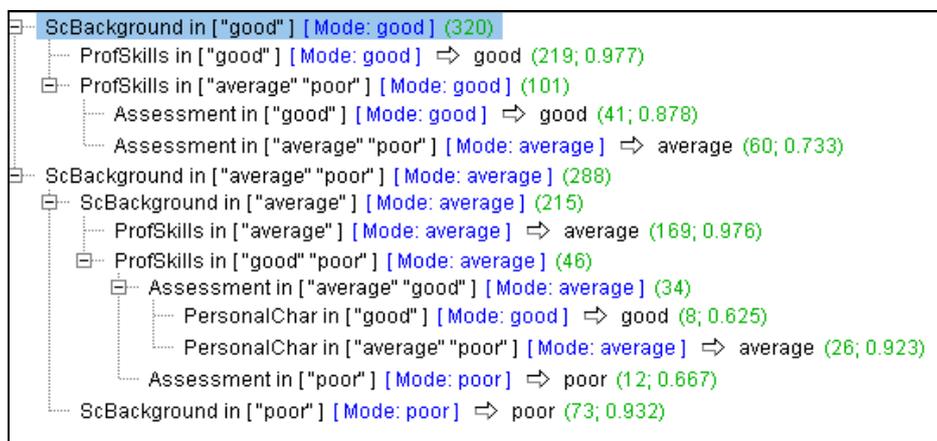


Figure 3. Teaching-behavior Classification Rules in Experiment 1

From the classification result, we found that the scientific background of the teacher is the best predictor for teacher performance.

#### 4.2. Experiment 2: using responses and achievement average

In this experiment, we use the all of the question responses and course achievement average to classify the teacher performance. Figure 4 shows the resulted classification tree with high accuracy rate 94.2%. Figure 5 shows rule based view for the classification result. We will explain some of the interested rules:

Rule1: if(Q12 in [poor, average] and Q9 in [poor]) then teacher performance is poor.

Rule2: if(Q12 in [good] and Q18 in [good] and Q1 in [good]) then teacher performance is good.

From these rules, we found that the Q12 attribute concerning about how the teacher uses methods growing the student thinking plays important role in classifying the teacher performance.

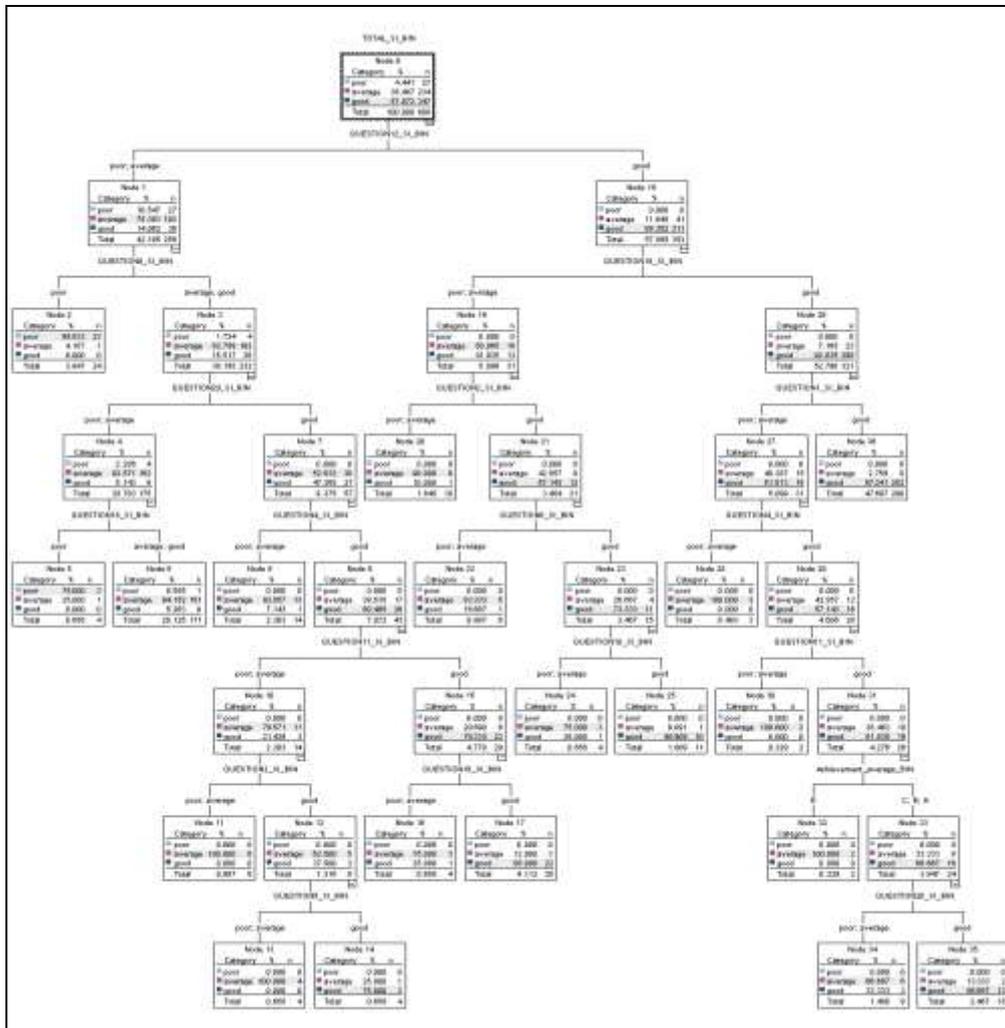
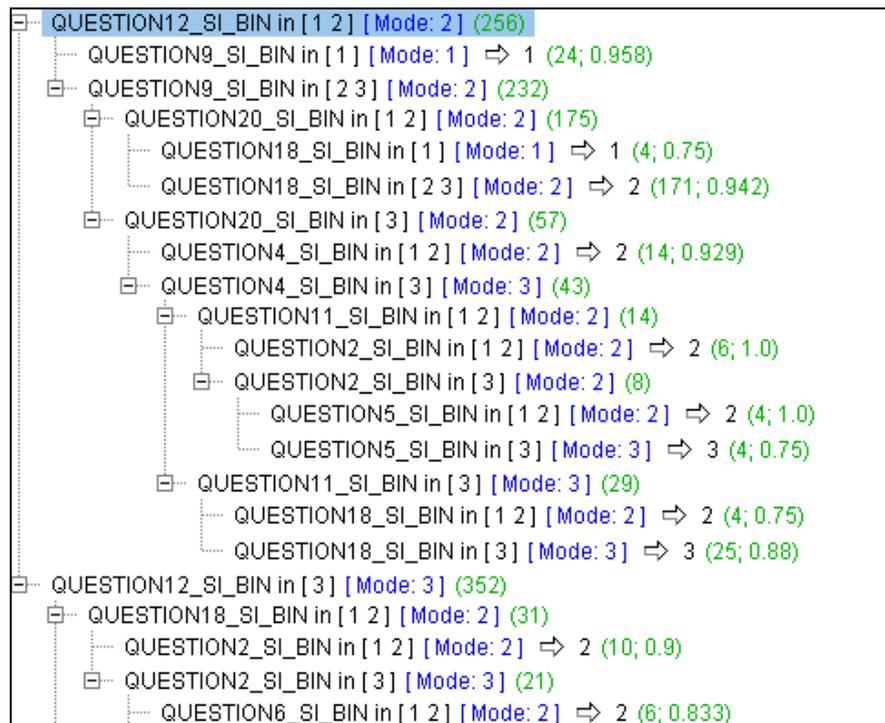


Figure 4. Teacher-performance Classification Tree in Experiment 2



**Figure 5. Teacher-performance Classification Rules in Experiment 2**

**5.3. Experiment 3: Using responses and student data to predict Student satisfaction for teaching performance.**

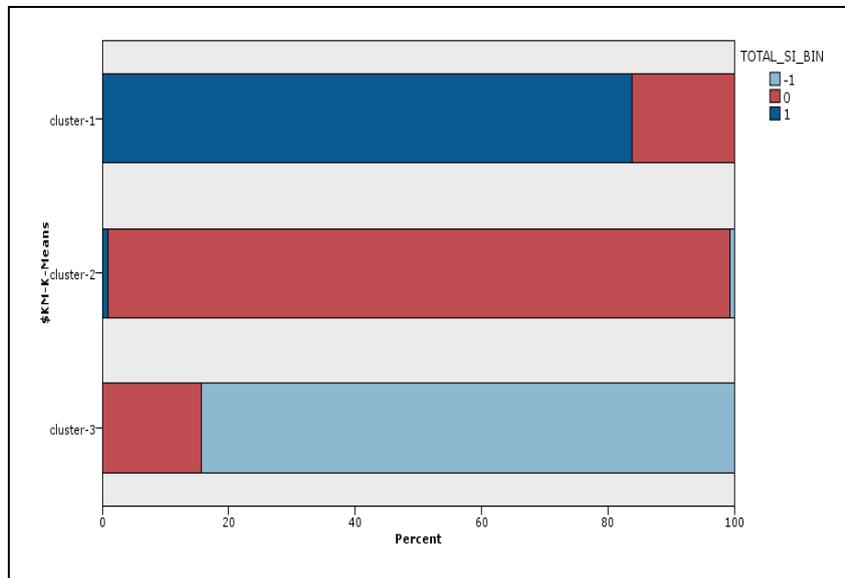
The goals of this experiment are to extract the important factors that identify the student satisfaction for their teacher performance and to build the classification model for predicting the student satisfaction. Later we illustrate the mining experience starting from the preparation of the data to the application of the mining process and its evaluation.

**Data preprocessing:** We discrete the TAS questions SI values and number of student enrollment in a course into categories based on the values of the mean and the standard deviation of the values distribution. Table 4 presents the attributes and their description that exists in the data set.

**Data mining functionality:** (clustering, classification): In this case, we firstly segment the course satisfaction data into 3 cluster by applying k-means clustering algorithm [11]. The clustering process groups the data according to their similarity. The input data for the clustering process is shown in Table 4. The output clusters classify the satisfaction data into three groups: cluster-1, cluster-2 and cluster-3. Figure 6 shows a graph of the three clusters plotted and colored according the total overall SI percent. We observe that the cluster-3 presenting at most the data of the dissatisfied courses.

**Table 4. Data Set Attributes of the Clustering Process**

Fields	Description	Values Domain		Direction
Faculty	The name of the faculty of the course			input
Question1_SI_Bin . . . Question20_SI_Bin	categories based on the values of the mean and standard deviation of the distribution of the field.	$x < (\text{Mean} - \text{Std. Dev})$  $(\text{Mean} - \text{Std. Dev}) \leq x \leq (\text{Mean} + \text{Std. Dev})$  $x > (\text{Mean} + \text{Std. Dev})$	-1(poor)  0(average)  1(good)	input
No_Students_Bin	The number of students enrollment in the course	$x < (\text{Mean} - \text{Std. Dev})$  $(\text{Mean} - \text{Std. Dev}) \leq x \leq (\text{Mean} + \text{Std. Dev})$  $x > (\text{Mean} + \text{Std. Dev})$	-1(small)  0(average)  1(large)	input

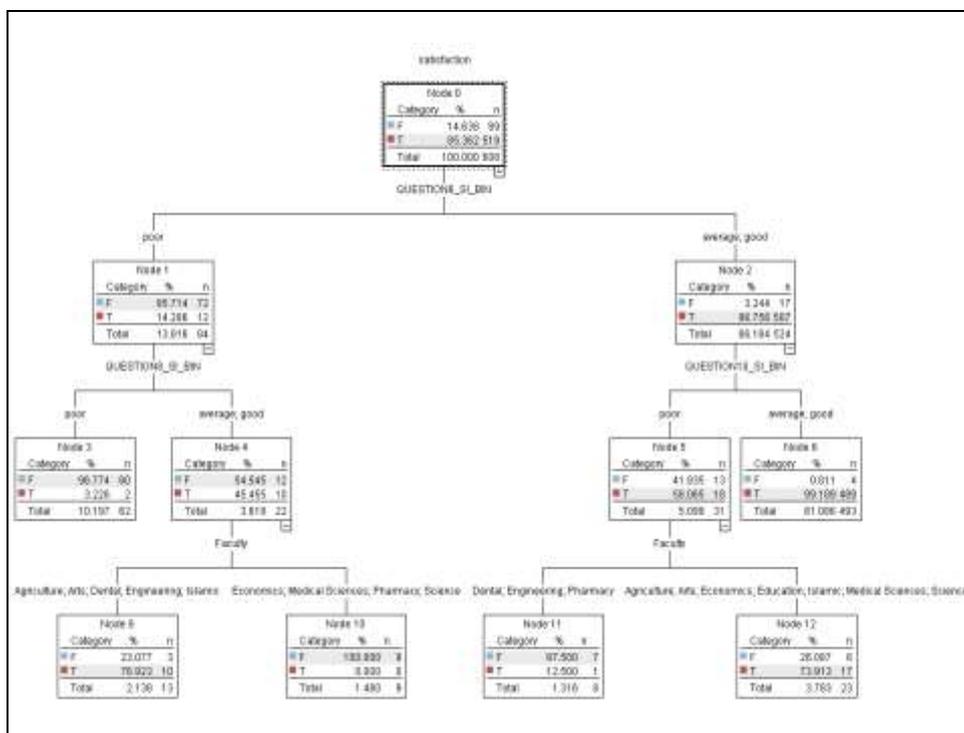


**Figure 6. Satisfaction Clusters Graph of Experiment 3**

In the second step, the output resulted from the clustering process is used to drive new class field named 'Satisfaction'. The Satisfaction field is flag attribute having true when 'course data' doesn't belong to "cluster-3". Then We apply the classification algorithm to build a classification model and identify the most important factors determining the satisfaction of the students to their course's teacher performance. Table 5 shows the data fields used to build the model. Figure 7 shows the resulted classification tree.

**Table 5. Data Set Attributes of the Classification Process**

Fields	Description	Values Domain	Direction
Faculty	The name of the faculty of the course	Agriculture, Arts, Dental, Economics, Education, Engineering, Islamic, Law, Medical Sciences, Pharmacy, Science	input
Question1_SI_Bin . . . Question20_SI_Bin	categories based on the values of the mean and standard deviation of the distribution of the field.	$x < (\text{Mean} - \text{Std. Dev})$ $(\text{Mean} - \text{Std. Dev}) \leq x \leq (\text{Mean} + \text{Std. Dev})$ $x > (\text{Mean} + \text{Std. Dev})$	-1(poor) 0(average) 1(good)
Satisfaction	Student Satisfaction	True False	output



**Figure 7. Classification Tree of Experiment 4**

**Evaluation:** A total of 608 course-teaching records were used in the experiment. Of these records, 592 records are classified correctly with high accuracy of 97.37%. After this mining, we observe that there are some of interesting teacher performance factors. These factors are important in determining and classifying the course's teacher performance. The first importance factor is question6 that concerns about how the teacher answering students' questions clearly. The second factor is question10 concerning about covering course topics during the semester. The third factor is question8 that concerns on how the teacher presenting the material in suitable way to student levels. We observed that the faculty attribute plays role in classifying the student satisfaction about the teacher performance. We discovered that most of the scientific colleges interested in the teaching construct concerning about how the curriculum coverage during the semester.

## 5. Conclusion

This study presents the importance of data mining techniques for exploring and discovering educational data. This study examines teaching constructs that are influencing on student satisfactions and indicates the important predictors for teacher performance. We applied several data mining techniques like data preprocessing techniques, C4.5 classification algorithm and K-means clustering algorithm. This study shows how the data of survey responses can be processed and mined. The study also shows the potential of the data mining for predicating student-satisfaction factors concerning about their teacher performance. We have met our objective which is to examine data of student satisfaction by data mining techniques. On working on these data, many attributes have been tested, and some of them are found effective on the performance prediction. The teaching construct that goals to growing the student thinking was important predictor for the student satisfaction. The teacher answering students' questions clearly, covering course topics during the semester and the presenting the material in suitable way play important roles in classifying the teachers' performance.

## References

- [1] R. A. Berk, "Survey of 12 strategies to measure teaching effectiveness", *International Journal of Teaching and Learning in Higher Education*, vol. 17, no. 1, (2005), pp. 48-62.
- [2] J. Han, M. Kamber and J. Pei, "Data Mining: Concepts and Techniques", Morgan Kaufmann Publishers, San Francisco, (2011).
- [3] R. K. Hemaidd and A. M. Halees, "Improving Teacher Performance using Data Mining", *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 4, no. 2, (2015), pp. 407-412.
- [4] J. Barracosa and C. Antunes, "Anticipating Teachers' performance", *Proceedings of the KDD Workshop: Knowledge Discovery in Educational Data*, (2011), pp. 77-82.
- [5] E. Taherifar and T. Banirostam, "Assessment of Student Feedback from the Training Course and Instructor Performance through the Combination of Clustering Methods and Decision Tree Algorithms", *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 6, no. 2, (2016), pp. 56-64.
- [6] G. K. Palshikar, S. S. Deshpande and S. S. Bhat, "Quest: Discovering insights from survey responses", *Proceedings in the Eighth Australasian Data Mining Conference*, vol. 101, (2009) December 1, pp. 83-91.
- [7] S. Abu Naser, A. Al-Masri, Y. Abu Sultan and I. Zaqout, "A Prototype Decision Support System for Optimizing the Effectiveness of E-learning in Educational Institutions", *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, vol. 1, no. 4, (2011).
- [8] R. Llorente and M. Morant, "Data Mining in Higher Education", *Kimito Funatsu*, (2011).
- [9] P. N. Tan, M. Steinbach and V. Kumar, "Introduction to Data Mining", Addison-Wesley, (2005).
- [10] J. Luan, "Data mining, knowledge management in higher education, potential applications", *Proceedings of workshop associate of institutional research international conference*, Toronto, (2002), pp. 1-18.
- [11] A. Likas, N. Vlassis and J. J. Verbeek, "The global k-means clustering algorithm", *Pattern recognition*, vol. 36, no. 2, (2003), pp. 451-461.
- [12] M. Kamber, L. Winstone, W. Gong, S. Cheng and J. Han, "Generalization and decision tree induction: efficient classification in data mining", *Proceedings of the 7th International Workshop on Research Issues in Data Engineering (RIDE '97) High Performance Database Management for Large-Scale Applications*, (1997), pp. 111.
- [13] R. Agrawal, T. Imielinski and A. Swami, "Mining association rules between sets of items in large databases", *Proceedings of the ACM SIGMOD Int'l Conf. on Management of Data (ACM SIGMOD '93)*, Washington, USA, (1993).
- [14] R. Kohavi and J. R. Quinlan, "Data mining tasks and methods: Classification: decision-tree discovery", *Proceedings of handbook of data mining and knowledge discovery*, Oxford University Press, Inc, (2002) January, pp. 267-276.
- [15] S. Ruggieri, "Efficient C4. 5 [classification algorithm]", *IEEE transactions on knowledge and data engineering*, vol. 14, no. 2, (2002), pp. 438-444.
- [16] F. Ahmadi, M. E. Abadi, "Data Mining in Teacher Evaluation System using WEKA", *International Journal of Computer Applications*, vol. 63, no. 10, (2013), pp. 12-18.
- [17] A. K Pal and S. Pal, "Evaluation of Teacher's Performance: A Data Mining Approach", *Journal of Computer Science and Mobile Computing*, vol. 2, no. 12, (2013), pp. 359-369.

- [18] S. Abu Naser, I. Zaqout, M. Abu Ghosh, R. Atallah and E. Alajrami, "Predicting Student Performance Using Artificial Neural Network: in the Faculty of Engineering and Information Technology", Journal of Hybrid Information Technology, vol. 8, no. 2, (2015), pp. 221-228

### Author



**Abdelbaset Rajab Almasri**, was born in AL-Maghazi, Gaza, Palestine, in 1978. He received the B. S. in Computer Science from Al Azhar University, Palestine in 2000 and M.S. in Computer Science from Vrije University Brussel in 2006. He was appointed the head of the Department of Information Technology in the Faculty of Engineering and Information Technology Al Azhar University during the years 2010-2013. He has been working as lecturer in the Department since 2006. He has worked as A Teaching Assistants in Al Azhar University-Gaza between the years 2000 and 2005.