A Study of Outstanding Instructors Selected from Teaching Evaluation Using the Data Mining Technique

Likit Neeranatpuree[1*] and Somchai Lekcharoen[2]

[1]Graduate student in Master of Science in Information Technology, College of Information and Communication Technology, Rangsit University, Phahonyothin Rd., Lak-hok, Patumtanee, Thailand 12000

[2]Lecturer in Master of Science in Information Technology, College of Information and Communication Technology, Rangsit University, Phahonyothin Rd., Lak-hok, Patumtanee, Thailand 12000

*Corresponding author, E-mail: likit.n@rsu.ac.th

Abstract

The objective of this research is to study the relationship among each of the evaluated questions Rangsit University used for selecting an "Outstanding Instructor" along with the tendency of being nominated for this valuable position. Several techniques, including Rough Set, K-Means, and Association Rules, were used. The samples were the teaching average points of 1018, instructors in 2014 gained from 13 evaluated questions. 10 Association Rules with a minimum support of 15% and a confidence level of 1 were selected. It was found that the evaluated question no. 12 (The teacher uses measurement and evaluation techniques in congruence with the course objectives) was the most useful question. Next, the evaluated question no. 1 (The teacher regularly teaches in class and completely covers the contents) and the evaluated question no. 13 (The teacher spares time in and outside the classroom for consultancy as needed by students) were determined to be the second useful questions. However, the evaluated question no. 5 (The teacher appropriately introduces and empowers students' ethics and morale while teaching) and the evaluated question no. 8 (The teacher implements learning activities that motivate students to do group works) were considered not to be the main factors.

Keywords: *Rough set, K-Means, Association Rules, teaching evaluation*

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาความสัมพันธ์ระหว่างข้อคำถามที่ใช้ในการประเมินอาจารย์สอนดีเด่นกับแนวโน้มของการได้เป็นอาจารย์สอนดีเด่น โดยใช้เทคนิค Rough Set เทคนิค K-Means และเทคนิค Association Rules มีกลุ่มตัวอย่างเป็นผลคะแนนเฉลี่ยประเมินการสอนของอาจารย์จำนวน 1,018 คนในปีการศึกษา 2557 และพิจารณาจากข้อคำถามที่ใช้ในการประเมินทั้งหมด 13 ข้อ ผลการวิจัยที่ได้คือกฎความสัมพันธ์ที่เลือกมาทั้งหมด 10 กฎโดยมีค่าสนับสนุน เท่ากับ 15% และค่าความเชื่อมั่นเป็น 1 สรุปได้ว่าข้อคำถามที่ 12 คือ อาจารย์ใช้วิธีการวัดประเมินผลที่สอดคล้องกับวัตถุประสงค์รายวิชา มีความสำคัญมากที่สุด และข้อคำถามที่ 1 คือ อาจารย์สอนสม่ำเสมอ ตรงต่อเวลา และครบถ้วนตามเนื้อหารายวิชา กับข้อคำถามที่ 13 คือ อาจารย์มีเวลาให้คำปรึกษาแก่นักศึกษาทั้งในและนอกชั้นเรียน มีความสำคัญรองลงมา ส่วนข้อคำถามที่ 5 คืออาจารย์สอนสอดแทรกและส่งเสริมเรื่องคุณธรรมจริยธรรมแก่นักศึกษา กับข้อคำถามที่ 8 คืออาจารย์มีกิจกรรมการเรียนที่ส่งเสริมให้นักศึกษามีการทำงานร่วมกัน ทั้งสองข้อนี้อาจบอกได้ว่าไม่มีความสำคัญกับการเป็นอาจารย์สอนดีเด่น

คำสำคัญ: *ราฟเซ็ต เค-มีนส์ กฎความสัมพันธ์ ประเมินการสอน*

1. Introduction

The Office of the Higher Education Commission has an enforcement of teaching efficient evaluation voted by students based on the Thai Qualifications Framework for Higher Education in 2009. For more concordance with Quality Assurance in Education Mechanism, according to the National Education Act B.E. 2542 (1999) and Amendments (second National Education Act B.E.2545 (2002)), an academic vice chancellor of Rangsit University appointed directly to ISDC (Instructional Support and Development Center)

to be in charge of an accumulation of teaching scores of all faculty instructors before concluding the results on the faculty to the university level.

At present, students are required to access the university's online evaluation system in order to view all their subjects that have been registered in each semester including the first, the second, and the summer semester. 20 evaluated questions are classified into 3 main topics: 2 items in student's perception, 13 items in quality of teaching (Lecture), and 5 items in quality of teaching (Laboratory). Subsequently, ISDC is responsible for concluding the score results of each instructor in the faculty. The scores are then ranked from the highest to the lowest ones.

Every year, an outstanding teaching instructor is granted an award. This conventional practice has been held since 2005 up to the present. The selection is based on instructors' average scores gained from their teaching quality, subjects, and groups throughout the three semesters. To An instructor with outstanding teaching quality must have an average score of at least 4.50 calculated from all the 13 questions. Moreover, those who are awarded every 3, 6, and 9 years, are granted an extra one.

Rough Set is a promising technique for data mining and knowledge discovery from the set theory. The theory consists of a lower approximation, the case of certainty and an upper approximation, the case of all possibilities). The difference of a lower approximation and an upper approximation is called the boundary region, which means the case is ambiguous. Rough Set is one of the best techniques used for solving complicated data. Therefore, the researchers were interested in bringing Rough Set to find absolutely outstanding instructors.

K-Means is one of the easiest clustering techniques. When the number of groups that user wants is set, its algorithms will calculate the similarity of data based on the distance of values. In the lower approximation, it was found that the range of scores was very close, e.g. 4.25, 4.26, 4.27... 4.94. Therefore, K-Means algorithm was used to record the groups of intriguing scores in order before using the next technique, Association Rules.

Association Rules were selected to find the relationship among each of the questions on teaching evaluation from only outstanding instructors.

Although some of the instructors had similar evaluation scores, the evaluation results of outstanding instructors were different. This is why Rough Set was used to find absolutely outstanding instructors. Also, K-Means was used after the grouping of the score range of absolutely outstanding instructors. Since the range of scores was very close, the results showed a lot of useless data when using Association Rules.

2. Objectives

To study the trend of an outstanding instructor selected from the teaching evaluation by using Rough Set, K-Means, and Association Rules

3. Materials and methods

3.1 Population and sample groups

The data used in this research were from the online database, http://ev.rsu.ac.th. The scores belonged to 1,018 instructors, were classified based on their total point averages accumulated from all the three semesters, in 2014.

3.2 Materials

3.2.1 Data Accumulating Materials

There are 13 questions in a part of the Quality of Teaching (Lecture) as follows:

Q1 = The teacher regularly teaches in class and completely covers the content.

Q2 = The teacher has clearly explained the objectives and the lesson plan of the content.

Q3 = The teacher uses various teaching techniques to keep the class motivated.
Q4 = The teacher recommends and uses appropriate and adequate printed or electronic teaching media
Q5 = The teacher appropriately introduces and empowers students' ethics and morale while teaching
Q6 = Teacher encourages the students to use information technology as tools of self-learning.
Q7 = The teacher implements learning activities that motivate students' self-learning enquiries.
Q8 = The teacher implements learning activities that motivate students to do group work.
Q9 = The teacher provides students the opportunities to discuss, express opinions, and share what they have learned.
Q10 = The teacher is open to discussion and provides advice and learning assistance for students.
Q11 = The teacher regularly marks assignments, monitors students's learning outcome, and provides them with useful advice when needed.
Q12 = The teacher uses measurement and evaluation techniques in congruence with the course objectives
Q13 = The teacher spares time in and outside the classroom for consultancy as needed by students

Each of the evaluated questions is relied on the 5 Rating Scale, where 5 is a maximum point and 1 as a minimum point of the student's perception. The points were calculated from all subjects in the academic year 2014 for a total average point. (Rangsit University, 2014)

3.2.2 Rough Set, K-Means, and Association Rules were used as the association tests.

3.3 Data Storage

The researchers collected the score data of the evaluation made by students in all faculties via the online evaluation database in 2014 found a total of 414,539 records. Then, the data in each evaluated question were calculated. The researchers found that 1,018 instructors in Rangsit University were listed on the online evaluation system. To become an outstanding teaching instructor, he or she must be rated with an average point of at least 4.50 from 13 questions. Moreover, the board of committee and the dean of each faculty are responsible for approving the appropriation and validity of the scores in the final process.

According to the result, only 281 instructors were selected to be outstanding teaching instructors whereas 737 instructors were not. The data are shown in Table1 below.

Table 1. Total average points in each evaluated questions and status taps

Instructor	Question numbers													
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Status*
T_1	4.44	4.55	4.47	4.50	4.59	4.53	4.35	4.67	4.58	4.40	4.57	4.44	4.67	YES
T_2	3.98	4.21	4.17	4.50	4.03	4.24	4.35	4.29	4.30	4.40	4.25	4.23	4.13	NO
T_3	4.44	4.55	4.61	4.39	4.51	4.58	4.50	4.67	4.58	4.46	4.49	4.61	4.53	YES

*An outstanding instructor.

3.4 Data Analysis

Data analysis process consists of steps as follows:

3.4.1 Data Analysis by Rough Set (Lekcharoen, S. & Pankham, S., 2016)

The majority of evaluation scores were similar. From the table above, T1 and T2 had the same scores of 4.50, 4.35 and 4.40 on Q4, Q7, and Q10, respectively. However, T1 was classified as an outstanding instructor where T2 was not. This shows an uncertainty of the data known as the Boundary Region Set, as shown in Table2 below.

Table 2. The same evaluation scores of T1 and T2 in 3 questions where the status of T1 and T2 are different

Instructor	Question numbers													
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Status*
T_1	4.44	4.55	4.47	**4.50**	4.59	4.53	**4.35**	4.67	4.58	**4.40**	4.57	4.44	4.67	**YES**
T_2	3.98	4.21	4.17	**4.50**	4.03	4.24	**4.35**	4.29	4.30	**4.40**	4.25	4.23	4.13	**NO**

*An outstanding instructor.

In another case, T1 and T3 received scores of 4.44, 4.55, 4.67, and 4.58 on Q1, Q2, Q8 and Q9, respectively. Also, Table2 shows that both of them were classified as outstanding instructors. Therefore, they were considered as truly outstanding instructors, also known as the Lower Approximation Set, as shown in Table3 below. The researchers developed Rough Set algorithm with PHP and data in MySQL database.

Table 3. Scores and statuses of the truly outstanding instructors

Instructor	Question numbers													
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Status*
T_1	**4.44**	**4.55**	4.47	4.50	4.59	4.53	4.35	**4.67**	**4.58**	4.40	4.57	4.44	4.67	**YES**
T_3	**4.44**	**4.55**	4.61	4.39	4.51	4.58	4.50	**4.67**	**4.58**	4.46	4.49	4.61	4.53	**YES**

*An outstanding instructor.

3.4.2 Grouping the score by K-Means (Jamesmanoharan, J., et al, 2014)

K-Means is a technique particularly used for making a group where users can define preferred groups. Seeking an appropriate group by considering from the data distance and the centroid, users can then identify a new center from the current data. The activation should be repeated until all data are unchangeable.

The K-Means technique was used for grouping the score range of instructors certainly chosen to be outstanding instructors. Following the technique, 5 groups were defined to divide the score range of within different questions. Each question was given a variety of score ranges. For example, Q1 in group1 consisted of the score range starting from 4.25 to 4.46. In contrast, the score range of Q2 in group1 started from 4.23 to 4.41. The reason why these scores were different was each question had different maximum and minimum scores., the researchers developed K-Means algorithm with PHP and data in MySQL database.

3.4.3 Association Rules techniques (Poomithes, C., 2012, Jetpenpas, V., 2012)

Association Rules with Apiori algorithm, a technique for creating data mining, are mostly used for searching the simultaneous data. A support value and a confidence value are defined. A support value informs the relationship rule that occurs in the database whether it is more or less, where a confidence value informs the probability of the rule.

After grouping the score range by K-Means, Apiori is needed to find the association of the simultaneous questions such as the score of Q1 shown in Group 3, the score of Q2 shown in Group 4, as well as Q5 score shown in Group 3. The Weka3.6 program was used to find association rules.

4. Results

4.1 To identify an outstanding instructor using Rough Set

From the total data of 1,018 instructors, 281 outstanding instructors were considered based on their average scores gained from all questions. When using Rough Set to find outstanding instructors, the researchers found that 1,018 instructors were e identified as Upper Approximation. 281 instructors were identified as Lower Approximation, and none of them were identified as Boundary Region. Lower

Approximation was found to be the number of possible truly outstanding instructors. 13 questions led into the finding of outstanding instructors. For example, if instructors gain 4.66, 4.61, and 4.54 from Q1, Q2, and Q5, respectively, then they are truly outstanding instructors.

4.2 Grouping the score range of evaluated questions by K-Means

After using Rough Set, each question showed a difference in its maximum and minimum values. Also, each of them had different mean scores as shown in Table4 below.

Table 4 Average scores from 281 outstanding instructors in each question

	Question numbers												
	Q1	**Q2**	**Q3**	**Q4**	**Q5**	**Q6**	**Q7**	**Q8**	**Q9**	**Q10**	**Q11**	**Q12**	**Q13**
	4.25	4.23	4.10	4.25	4.11	4.23	4.24	4.20	4.26	4.23	4.22	4.25	4.21
	4.33	4.27	4.12	4.26	4.12	4.25	4.25	4.22	4.32	4.29	4.31	4.26	4.22
	4.39	4.35	4.13	4.33	4.26	4.28	4.34	4.23	4.37	4.34	4.35	4.32	4.28
	4.41	4.37	4.17	4.34	4.29	4.31	4.39	4.24	4.38	4.38	4.37	4.33	4.34
	4.42	4.40	4.29	4.35	4.30	4.33	4.40	4.25	4.40	4.41	4.40	4.35	4.36
	...	...	...	...	...	...	...	...	...	...	...	...	...
	4.94	4.88	4.75	4.79	4.76	4.77	4.89	4.74	4.86	4.91	4.88	4.85	4.87
		4.89	4.76	4.80	4.77	4.78		4.75	4.88		4.92	4.92	4.88
		4.91	4.77	4.81	4.78	4.79		4.76	4.89				4.92
			4.78	4.82	4.79	4.80		4.77					
			4.79	4.83	4.80	4.81		4.78					
			4.80	4.86	4.82	4.83		4.79					
			4.81	4.87	4.85	4.86		4.80					
			4.82		4.87			4.88					
			4.83										
			4.87										
			4.89										
Max	**4.94**	**4.91**	**4.89**	**4.87**	**4.87**	**4.86**	**4.89**	**4.88**	**4.89**	**4.91**	**4.92**	**4.92**	**4.92**
Min	**4.25**	**4.23**	**4.10**	**4.25**	**4.11**	**4.23**	**4.24**	**4.20**	**4.26**	**4.23**	**4.22**	**4.25**	**4.21**

From the table above, the scores were totaled and classified into 5 groups, by range, in each question using K-Means. The result of grouping is shown in Table5 below.

Table 5 Groups of classification by range of scores in each question

Group	**Question Numbers**												
	Q1	**Q2**	**Q3**	**Q4**	**Q5**	**Q6**	**Q7**	**Q8**	**Q9**	**Q10**	**Q11**	**Q12**	**Q13**
1	4.25-4.46	4.23-4.41	4.10-4.32	4.25-4.42	4.11-4.36	4.23-4.39	4.24-4.38	4.20-4.34	4.26-4.43	4.23-4.46	4.22-4.43	4.25-4.41	4.21-4.41
2	4.47-4.57	4.42-4.55	4.33-4.48	4.43-4.53	4.37-4.50	4.40-4.51	4.39-4.52	4.35-4.48	4.44-4.56	4.47-4.58	4.44-4.55	4.42-4.54	4.42-4.54
3	4.58-4.70	4.56-4.68	4.49-4.62	4.54-4.64	4.51-4.62	4.52-4.63	4.53-4.65	4.49-4.59	4.57-4.67	4.59-4.68	4.56-4.66	4.55-4.66	4.55-4.66
4	4.71-4.82	4.69-4.79	4.63-4.75	4.65-4.74	4.63-4.73	4.64-4.73	4.66-4.75	4.60-4.71	4.68-4.77	4.69-4.79	4.67-4.77	4.67-4.76	4.67-4.78
5	4.83-4.94	4.80-4.91	4.76-4.89	4.75-4.87	4.74-4.87	4.74-4.86	4.76-4.89	4.72-4.88	4.78-4.89	4.80-4.91	4.78-4.92	4.77-4.92	4.79-4.92

After completely grouping, the original scores were replaced by group numbers in each question as shown in the Table6 below.

Table 6 Examples of how to replace a score with grouping by K-Means

Instructor	Question numbers				
	Q1	Q2	Q3	...	Q13
T_1	Group 1 (the original score is 4.44)	Group 2 (the original score is 4.55)	Group 2 (the original score is 4.47)	...	Group 4 (the original score is 4.67)
T_3	Group 1 (the original score is 4.44)	Group 2 (the original score is 4.55)	Group 3- (the original score is 4.61)	...	Group 2 (the original score is 4.53)

4.3 Association of Question By Apiori

According to 281 outstanding instructors, as being represented in groups by K-Means and Association Rules by Apiori, Weka program was used for selecting 10 best rules with a minimum support value of 15% and the Association Rules with a confidence value of 1 as shown in Table7 below.

Table 7 10 Association Rules

Ranks	Left Hand Side	Quantities Transaction	Right Hand Side	Quantities Transaction	Confidence
1	Q1=3 Q11=3 Q12=3 Q13=3	53	Q2=3	53	1
2	Q1=3 Q6=3 Q12=3 Q13=3	49	Q2=3	49	1
3	Q1=3 Q3=3 Q12=3 Q13=3	48	Q2=3	48	1
4	Q1=3 Q7=3 Q11=3 Q12=3 Q13=3	48	Q2=3	48	1
5	Q6=3 Q9=3 Q11=3 Q12=3	46	Q7=3	46	1
6	Q2=3 Q4=3 Q10=3 Q11=3	45	Q12=3	45	1
7	Q1=3 Q4=3 Q6=3 Q12=3 Q13=3	45	Q2=3	45	1
8	Q1=3 Q4=3 Q7=3 Q12=3 Q13=3	45	Q2=3	45	1
9	Q1=3 Q6=3 Q7=3 Q12=3 Q13=3	45	Q2=3	45	1
10	Q1=3 Q6=3 Q11=3 Q12=3 Q13=3	45	Q2=3	45	1

From Table7 above, using Ranks no. 6 as an example, it shows the relationship among questions 2, 4, 10, and 11 with a score range of 3.

By using Rough Set, K-Means, and Association Rules to find the relationship of questions, the results of 10 Association Rules are shown in Table8 below.

Table 8 Shows the relationship of questions by the association rules

Question ranks	Quantities	
	Left Hand Side	Right Hand Side
Q12= 3	9	1
Q1 = 3	8	0
Q13 = 3	8	0
Q6 = 3	5	0
Q11 = 3	5	0
Q4 = 3	3	0
Q7 = 3	2	1
Q2 = 3	1	8
Q3 = 3	1	0
Q9 = 3	1	0
Q10 = 3	1	0
Q5	0	0
Q8	0	0

From Table8, questions 12, 1, 13, 6, 11, 4, 7, 2, 3, 9, and 10 were found to be relevant to the evaluation to find outstanding instructors. This is because there were numbers appearing on the Left Hand Side and the Right Hand Side. On the other hand, for question 5 and 8, no numbers appeared. This means questions 5 and 8 were not relevant to the evaluation.

It can be concluded that questions 1, 2, 3, 4, 6, 7, 9, 10, 11, 12, and 13 were important for the evaluation to find outstanding instructors. Also, for question 2, there was a relationship that occurred in a way as the numbers on the Right Hand side were high. This could mean, apart from questions 12, 1, 13, 6, 11, 4, 7, 3, 9, and 10, question 2 was found to be a necessary for the finding of outstanding instructors.

5. Discussion

The researchers found obvious disadvantages of Association Rules due to the complication of the question scores as the algorithm was characterized by a decimal number. K-Means was found to be the solution for grouping information. The finding is consistent with the research of Suvarnno, N. and Singeiam, A. (2011) stating that using data discretization to set entrance exam scores to a certain range, such as 0 to 20 or 21 to 40. In addition, in this research, the researchers only focus on the finding of the relationship of being the outstanding instructors where a confidence value is equal to 1.

As for the Rough Set, it was found that all the selected outstanding instructors had no ambiguous scores. All the 281 instructors were absolutely outstanding instructors, as they did not have duplicate scores as the instructors who were not outstanding instructors did.

6. Conclusion

The use of Rough Set, K-Means, and Association Rules to find the association of questions affected the selection of outstanding instructors. It was found that Q12 with a score range of 4.55 to 4.66 was considered to be the most important factor, followed by Q1 with a score range of 4.58 to 4.70 and Q13 with a score range of 4.55 to 4.66. On the other hand, Q5 and Q8 were not probably be important factors. Moreover, Q2 with a score range of 4.56 to 4.68 showed an additional relationship that seemed to be one of the important factors in finding outstanding instructors whereas other score ranges were not probably considered as important factors to evaluate outstanding instructors.

Compared to the previous time, to become an outstanding instructor was determined by an average teaching evaluation score of at least 4.50. Nowadays, no evidence or articles about studying outstanding instructors by teaching evaluation were found, so new algorithms were suggested to be included and applied to the prediction of results in future research.

7. Acknowledgements

We would like to express our special thanks to Somchai Lekcharoen who gave us an opportunity to conduct this studyand helped us do a lot of research and learn useful knowledge. Finally, we would also like to thank our parents and friends who helped us a lot in completing this project within the limited time frame.

8. References

Jamesmanoharan, J. & Ganesh, S.H & Felciah, M.L.P & Shafreenbanu, A.K. (2014). Discovering Students' Academic Performance Based on GPA Using K-Means Clustering Algorithm" Computing and Communication Technologies (WCCCT), 2014 World Congress on Feb. 27 2014-March 1 2014, pp. 200 – 202.

Jetpenpas, V. (2012). Pattern of Disbursed Supplies Using Association Rules. (Master's Thesis). King Mongkut's University of Technology North Bangkok, Bangkok, Thailand.

Lekcharoen, S. & Pankham, S. (2016). A rough set control token leaky bucket in policing mechanism schemes over high speed network. Rangsit Journals of Art and Sciences. 6(2), pp. 109-116.

Office of the Higher Education Commission. (n.d.). Thai Qualifications Framework for Higher Education. Retrieved December 15, 2016, from http://www.mua.go.th/users/tqf-hed/news/news8.php

Poomithes, C. (2012). Decision Support Systems Development Stock Medicine by using Association Rules Technique Case Study : Srithanya Hospital. (Master's Project). King Mongkut's University of Technology North Bangkok, Bangkok, Thailand.

Rangsit University. (2014). Teaching evaluation scores by students in the 2014 academic year. Pathumthani, Thailand.

Suvarnno, N. & Singeiam, A. (2011). The use of Assocation rules to find the factors effected to the risk of low proficiency students Case study: Prince of Songkla University.Journal of management sciences, 28(1), 65-79.