Association Rule Analysis of Student Satisfaction Surveys for Teaching Quality Evaluation

Manuel J. Jiménez-Navarro, Belén Vega-Márquez, José María Luna-Romera, Manuel Carranza-García, and María Martínez-Ballesteros(✉)

Department of Computer Languages and Systems, University of Seville, 41012 Seville, Spain
{mjimenez,bvega,jmluna,mcarranzag,mariamartinez}@us.es

Abstract. The quality of university teaching is essential for the success of students and the academic excellence of an educational institution. The purpose of this work is to provide a methodology based on the Association Rule technique using the Apriori algorithm to analyze the results obtained from the student evaluation process regarding their satisfaction with the teaching received. This methodology has been applied in programming courses of students of several courses both in the Computer Engineering and Health Engineering degrees at University of Seville, Spain. The proposed methodology can serve as a starting point for a self-improvement process that clearly identifies strengths and weaknesses.

Keywords: Quality · teacher evaluation · machine learning · association rules

1 Introduction

The quality of university teaching is essential to the success of students and the academic excellence of an educational institution. Teachers must have effective teaching skills, stay up-to-date in their field of knowledge, create an inclusive and respectful learning environment, and be committed to the success of their students [4,11].

There are various mechanisms to evaluate teaching quality, which can include evaluations by students and colleagues, self-assessments, performance indicators, and data analysis [9]. Each mechanism can provide a different perspective on teaching quality and it is important to use a variety of approaches to obtain a complete picture of teaching effectiveness.

Among the usual indicators in these models, one of the most commonly used mechanisms to evaluate teaching quality is the implementation of student surveys to determine the level of satisfaction of students with the teaching activity of the faculty [3].

https://doi.org/10.1007/978-3-031-42519-6_30

Despite the data provided by these surveys, which would allow us to see if our average rating is better or worse compared to that of the rest of the teaching staff in the same area, degree or university, it would be necessary to evaluate these surveys in more detail to extract useful knowledge that helps us determine whether our teaching activity is of high or poor quality and which aspects should be improved. Another important aspect is to determine whether the scores provided are associated with the teachers or with the subject. Additionally, it would be interesting to analyze these questionnaires to better understand students and the environment in which they learn [5,7,10].

In this context, data mining techniques can be useful for analyzing the results of quality education surveys. In particular, the Association Rule (AR) technique is used to discover relationships and patterns between variables in a large dataset [1]. The Apriori algorithm is one of the most widely used algorithms for discovering AR due to its ability to efficiently find frequent patterns in large datasets and generate easily interpretable results [2].

Therefore, this work aims to provide a methodology based on the AR technique using the Apriori algorithm to analyze the results obtained from the student evaluation process regarding their satisfaction with the received teaching. The proposed methodology has been applied in programming courses of students of several courses both in the Computer Engineering and Health Engineering degrees at the University of Seville, Spain (US).

The remainder of this paper is structured as follows. Section 2 provides the main concepts of AR mining. Section 3 describes the data and the methodology used to obtain the survey rules. Section 4 presents and discusses the rules obtained after applying the methodology. Section 5 summarizes the main findings of this study.

2 Association Rules

Acquisition of knowledge through the use of ARs is a widely recognized and popular technique in the field of data mining to uncover interesting relationships among variables in large databases [2]. Rather than predicting the class of new data, AR aims to identify patterns that explain or summarize the data and are used to explore the properties of the data [1].

These rules are typically presented in the form of "if-then" statements, where the antecedent (if) represents the condition or item set, and the consequent (then) represents the outcome or item set that tends to co-occur with the antecedent.

The literature proposes several probability-based measures to assess the quality of ARs. In this work, we have selected the measures *support* (Eq. 1), *confidence* (Eq. 2), and *lift* (Eq. 3) to evaluate and analyze the AR obtained in order to assess the generality, reliability, coverage, and interest of the rules, respectively [8].

The support of the rule $X \implies Y$ is the percentage of transactions in the dataset that contain the items in X (antecedent) and Y (consequent) simultaneously. Note that $frq(X, Y)$ is the number of instances that satisfy the conditions for the antecedent X and Y in the dataset simultaneously. N is the total number of instances in the dataset. The support values range in the interval [0, 1].

$$Support(X \Longrightarrow Y) = \frac{frq(X,Y)}{N} \tag{1}$$

The confidence is the probability that instances containing X, also contain Y. The confidence values range in the interval [0, 1].

$$Confidence(X \Longrightarrow Y) = \frac{frq(X,Y)}{frq(X)} \tag{2}$$

Lift is another quality measure that captures the correlation by indicating whether the antecedent positively (lift $>$ 1) or negatively (lift $<$ 1) influences the consequent.

$$Lift(X \Longrightarrow Y) = \frac{Support(X \Longrightarrow Y)}{Support(X) \cdot Support(Y)} \tag{3}$$

The Apriori algorithm proposed in [2] is the most well-known and cited algorithm in the literature to find AR and is the basis for most existing algorithms. The main improvement over the previous algorithms lies in the way candidate sets are generated, as it enforces the property of frequent sets: Any subset of a frequent set must also be a frequent set. This property ensures that many of the frequent sets required by other algorithms are not constructed unnecessarily.

This algorithm is based on prior knowledge or "a priori" of frequent sets, and uses a breadth-first search strategy. Note that this algorithm obtains AR by discretizing continuous values in the database prior to processing.

Conceptually, the Apriori algorithm has the following steps for generating frequent itemsets:

- Generation of all itemsets with one item. Use these itemsets to generate those with two items, and so on.
- Calculation of the support of the itemsets, in order to obtain the resulting set by eliminating those subsets that do not exceed the minimum support.

The algorithm addresses the problem by reducing the number of sets considered so that the user defines minimum support and Apriori generates all sets that meet the condition of having support greater than or equal to that threshold. Any rules that do not satisfy the restrictions imposed by the user, such as the minimum confidence, are discarded, and the rules that do satisfy them are retained.

3 Methodology

This section details the context (Sect. 3.1), the main items analyzed (Sect. 3.2), and preparation from student satisfaction surveys (Sect. 3.3).

3.1 Student Satisfaction Surveys

The procedure for obtaining completed questionnaires from students regarding the teaching activity of their faculty can be carried out using a self-managed or online system. Each teacher can freely choose the system they consider most appropriate. However, the online system is only available to faculty included in the main activity of the subject, meaning that it is not available to lecturers who only teach practical classes.

The items in the Student Satisfaction Survey Questionnaire with the US Teaching Activity are organized into 18 questions, as shown in Table 1. It can be observed that the questions are organized into four categories: educational planning and organization (Q1, Q2, Q5, Q6, Q7, and Q8), student support (Q3, Q4, Q9, Q10, Q11, Q12, Q13, Q14, and Q15), evaluation (Q17) and general satisfaction (Q18).

Table 1. The items of the questionnaire of the student satisfaction survey with the teaching activity using in US.

Questions
Q1: It has given me the orientation to learn about the teaching project of the subject
Q2: Its teaching is in accordance with the planning foreseen in the teaching project
Q3: I adequately attended tutorials
Q4: The tutoring schedule is adequate
Q5: The bibliography and other recommended teaching materials are proving to be useful for me to follow the course
Q6: The teaching is well organized
Q7: The means you use to teach are adequate for my learning
Q8: The bibliography and other recommended teaching materials are available to students
Q9: Explains clearly
Q10: Is interested in the degree of comprehension of their explanations
Q11: Provides examples to put into practice the contents of the course
Q12: Resolves the doubts that arise
Q13: Promotes a work and participation climate
Q14: Motivates students to take an interest in the subject matter
Q15: Treats students with respect
Q16: The teaching is helping me to achieve the objectives of the course
Q17: The evaluation criteria and systems are adequate to assess my learning
Q18: In general, I am satisfied with the teaching performance of this professor

3.2 Main Items Analyzed

In this section, the main elements assessed from the student satisfaction surveys are presented. The various components, along with their descriptions and potential values, are as follows:

– Question: refers to the question evaluated in the survey presented. The values have been shortened to Qn, where n is the question number. There are 18 questions listed in Table 1.

- Subject: indicates the subject in which the survey was conducted, such as Algorithm Theory, Data Analysis, or Web Development. Subjects have been abbreviated and may be represented by one of the following: ADDA (Analysis and Design of Data and Algorithms), DT (Design and Testing), DSA (Data Structures and Algorithms), MSIT (Management of Services and Information Technologies), ISEIS (Introduction to Software Engineering and Information Systems), OS (Operating Systems), or PF (Programming Fundamentals).
- Degree: represents the academic program in which the survey was conducted, all of which fall under the computer science discipline. The values may be one of the following: SE (Software Engineering), IT (Information Technology), CE (Computer Engineering), or ITM (Information Technologies and Mathematics).
- Course: refers to the period of years during which the survey was collected, spanning from 2017-18 to 2021-22.
- ClassType: distinguishes between two types of classes: theoretical, in which the teacher provides a theoretical explanation of the subject; and practical, in which the class focuses mainly on student work, in addition to teacher explanation.
- Covid: indicates whether surveys were conducted before 2020 (Precovid), during 2020 (Covid), or after 2020 (Postcovid) [6].
- QType: differentiates between two types of questions: direct, which are related to some aspect of the teacher or the teaching approach, and indirect, which are generally based on some aspect more related to the subject than to the teacher. Indirect questions include Q5, Q6, Q7, Q8, and Q17.
- Score: refers to the target of the analysis, which is the value of the student's answer in the range [1, 5] or Do not know /Do not answer (Dk/Da).

3.3 Analysis Process

To analyze the survey results, individual surveys are collected for each interested teacher. The surveys include a table with each row representing a question and columns displaying the different scores. The values in the table represent the number of students who evaluated the question with a particular score. Additional information is included in the metadata, as detailed in Sect. 3.2.

Once the surveys have been collected, the tables are transformed to build an itemset. For each score in each question evaluated by a student, a transaction is built that includes all the metadata from the survey, as well as the specific question and score obtained.

Finally, the Apriori algorithm described in Sect. 2 is applied to the itemset. The algorithm uses minimum support and confidence to filter out irrelevant features, which must be established based on the quantity and quality of the target rules obtained. The output of the algorithm includes the rules and associated metrics. To eliminate redundant information, any rule with a subset of its antecedents having the same or higher confidence is filtered out.

4 Results and Discussion

In this section, we present an analysis of 1673 surveys collected between 2017 and 2021, supported by the use of AR, which items are described in Sect. 3.2. Our analysis focuses on rules with support greater than 1%, confidence greater than 50%, lift greater than 1, and the score item as the consequent of the rule. We divide this section into five subsections that describe the analysis, focusing on the different items from the survey except for the Covid and QType as these items usually appear just in combination with other items.

4.1 Overview

This section shows an overview of the main rules obtained by the study using Fig. 1.

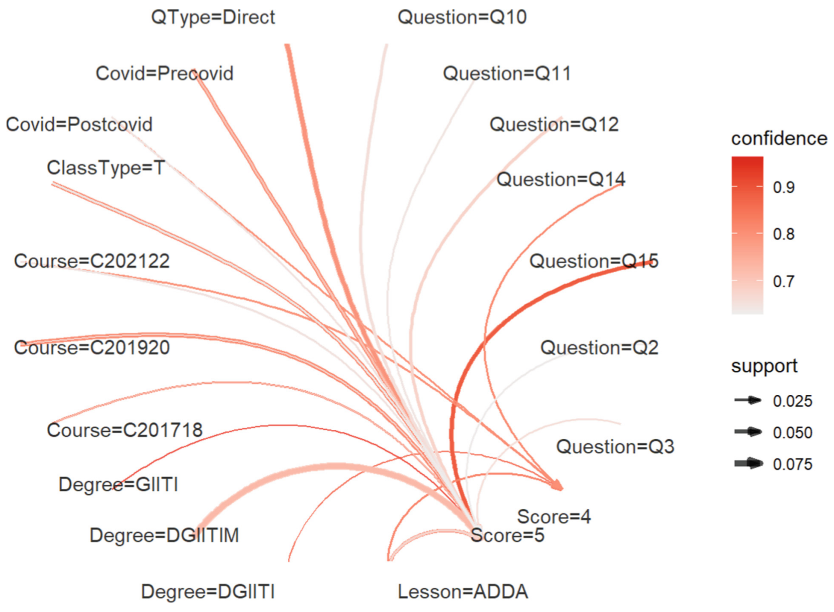

Fig. 1. Summary of the main rules obtained.

As observed, the items that appear as antecedents with a higher support and confidence score of 5 in the consequence are Q15, direct questions, precovid surveys, and the DGIITIM degree. For a score of 4, Q14, surveys collected during postcovid, and the DGIITI degree seem to be evaluated with greater confidence and support.

4.2 Question

Table 2 displays the principal rules that contain the question item in its antecedent sorted by confidence. To enhance visual clarity, the rules have been condensed, highlighting the most illustrative ones. Some questions are not included in the table, as the rules evaluated in our surveys did not produce significant findings regarding support and confidence.

Table 2. Principal rules related to questions. The rules are summarized and sorted by confidence.

Antecedents	Score	Support	Confidence
{Q15,ADDA,2019-20}	5	0.016	0.963
{Q14,DIT}	4	0.012	0.909
{Q15}	5	0.054	0.882
{Q12,ADDA,Postcovid}	5	0.013	0.759
{Q12}	5	0.041	0.673

The top rule obtained contains Q15 item obtaining a maximum score of 88.5% confidence overall. The rule with best confidence establishes that during the 2019-20 course, in the ADDA subject, the Q15 score was maximum with more confidence. This appears to indicate that this group of students was particularly satisfied with the treatment they received from their teacher, with an increase in confidence of 7.8%.

Q14 exhibits a higher degree of variability; as a consequence, there is no rule with this item in the antecedent. However, certain combinations with other items yield useful information. All the rules discovered achieved a score of 4, with confidence being particularly high in the DIT degree at 90.9%.

Q12 achieved a maximum confidence of 67.3% overall. This query achieved the highest confidence score in ADDA subject and prior to COVID at a 75.9% confidence level. This suggests that the scores in other subjects are more varied and do not provide adequate support or confidence to draw any conclusions.

4.3 Subject

The rules in Table 3 show the antecedents related to the subject that have the strongest correlation with scores of 4 or 5.

The first two rules ({DSA, 2021-22} and {DSA, Postcovid} antecedents) suggest that students who took the DSA course during the 2021-22 academic year or after the COVID pandemic rated the subject quality higher. These results may indicate that changes in teaching methods or adjustments made in response to the pandemic may have positively impacted subject quality. However, since the support values are relatively low, we should be cautious when making strong conclusions based on these rules alone.

The next two rules ({ADDA, Practical, Direct} and {ADDA, Direct} antecedents) suggest that students who took the ADDA course and had direct interactions with the professor rated the quality of the subject higher. These findings align with common educational practices that emphasize teacher-student engagement, which has been shown to positively impact learning outcomes. The second rule with higher support value suggests that direct interactions with the professor may be more important than other factors, such as practical sessions, in determining subject quality for ADDA students.

Table 3. Principal rules related with the subject. The rules are summarized and sorted by confidence.

Antecedent	Score	Support	Confidence
{DSA, 2021-22}	4	0.034	0.559
{DSA, Postcovid}	4	0.034	0.559
{ADDA, Practical, Direct}	5	0.069	0.542
{ADDA, Direct}	5	0.250	0.520

4.4 Degree

Table 4 shows the principal ARs related to the degree of the students enrolled in the courses. The rule {ITM, 2019-20, Direct} $\rightarrow$ 5 indicates that when a student enrolled in ITM in the 2019-20 academic year gives a high score, it is a high probability that they also responded positively to direct questions that evaluate the professor. Similarly, the rule ITM, Theory $\rightarrow$ 5 indicates that when a student enrolled in ITM takes a theoretical subject, there is a high probability that they also give a high score.

Table 4. Principal rules related with the degree. The rules are summarized and sorted by confidence.

Antecedent	Score	Support	Confidence
{ITM, 2019-20, Direct}	5	0.044	0.881
{ITM,Theory}	5	0.056	0.797
{ITM}	5	0.097	0.706
{SE,2021-22}	4	0.034	0.559
{CE, Direct}	5	0.045	0.524

These results suggest that professors should pay special attention to their performance when teaching theoretical subjects to ITM students, as this seems to be a key factor in obtaining high scores. Additionally, professors may want to encourage students to answer direct questions in the survey, as these questions are highly associated with high scores, indicating that they provide valuable feedback for professors.

4.5 Course

Table 5 shows the main ARs related to the courses enrolled. Following the same nomenclature as in the previous sections, we can see that {2017-18, Theory, Direct} and {2018-18, Direct} $\rightarrow$ 5 indicate that the direct answers for the 2017-18 course get a 5 with a confidence of 52.4% and 51.8%, respectively. We can see a similar rule ({2019-20, Theory, Direct}$\rightarrow$ 5) for the 2019-20 course with similar confidence (51.4%). On the other hand, there exists the rule {2021-22, Practice, Direct}$\rightarrow$ 5 that indicates with confidence of 50.5% that the questions for the group of practices of the 2021-22 course on direct questions about the teacher get a 5. Furthermore, we find {2021-22, Direct}, which with a confidence of 50.1% obtained a score of 5.

Table 5. Principal rules related to the course. The rules are summarized and sorted by confidence.

Antecedent	Score	Support	Confidence
{2017-18, Theory, Direct}	5	0.045	0.524
{2017-18, Direct}	5	0.094	0.518
{2019-20, Theory, Direct}	5	0.096	0.514
{2021-22, Practice, Direct}	5	0.088	0.505
{2021-22, Direct}	5	0.127	0.501

5 Conclusions

This work presents a simple methodology for analyzing teaching quality using association rules mining in student satisfaction surveys. This method is intuitive due to its "if-then" structure, which provides useful information about the strong points and weak points during teaching. Additionally, the method is generalizable to almost any survey with minimal changes. In our study, most of the rules obtained good confidence and support scores, indicating that students are generally satisfied with the treatment of teachers. However, some aspects still require reinforcement since rules related to important teaching skills do not appear with sufficient confidence. This methodology can serve as a starting point for a self-improvement process that clearly identifies strengths and weaknesses.

As future work, we intend to enhance our study by including surveys from a greater number of years and a wider range of subjects. This will help us to enrich our analysis and yield more comprehensive findings. Additionally, we're interested in exploring additional types of academic data to uncover patterns related to student dropouts within the university context.

Acknowledgements. The authors would like to thank the Spanish Ministry of Science and Innovation for the support under the projects PID2020-117954RB-C22 and TED2021-131311B-C21, and the Junta de Andalucía for the project PYC20 RE 078 USE.

References

1. Agrawal, R., Imielinski, T., Swami, A.: Mining association rules between sets of items in large databases. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, vol. 22, pp. 207–216, (1993)
2. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules in large databases. In: Proceedings of the International Conference on Very Large Databases, vol. 20, pp. 478–499 (1994)
3. Castro Morera, M., Navarro Asencio, E., Blanco Blanco, A.: The quality of teaching as perceived by students and university teachers: analysis of the dimensionality of a teacher evaluation questionnaire. Educacion XX1 **23**(2), 41–65 (2020)
4. García-Concepción, M.A., Jiménez-Ramírez, A., Martínez-Ballesteros, M., Gasca, R.M., Parody, L., Soria-Morillo, L.M.: Extensiones para el ciclo de mejora continua en la enseñanza e investigación de ingeniería informática. Revista de Enseñanza Universitaria **38**, 4–26 (2011)
5. Isla-Díaz, R., Hess-Medler, S., Marrero-Hernández, H.: The role of students in the docentia process after ten years of evaluation: evaluating the teacher or the subject? that is the question. Revista Espanola de Pedagogia **79**(280), 393–411 (2021)
6. Kumar, A., Sarkar, M., Davis, E., Morphet, J., Maloney, S., Ilic, D., Palermo, C.: Impact of the covid-19 pandemic on teaching and learning in health professional education: a mixed methods study protocol. BMC Med. Educ. **21**, 439 (2021)
7. García Martín, A., Teresa Montero Cases, Josefina García León, Vázquez Arenas, G.: Validez de las encuestas de satisfacción de los estudiantes para evaluar la calidad docente: el caso de la UPCT (cartagena). REDU: Revista de Docencia Universitaria (2020)
8. Martínez Ballesteros, M., Troncoso, A., Martínez-Álvarez, F., Riquelme, J.C.: Improving a multi-objective evolutionary algorithm to discover quantitative association rules. Knowl. Inf. Syst. **49**, 481–509 (2016)
9. Mas Torelló, O.: Las competencias del docente universitario: la percepción del alumno, de los expertos y del propio protagonista. Revista de Docencia Universitaria, REDU (2012)
10. Rubio-Escudero, C., Martínez-Álvarez, F., Atencia-Gil, E., Troncoso, A.: Implementation of an Internal Quality Assurance System at Pablo de Olavide University of Seville: Improving Computer Science Students Skills. In: Proceedings of International Conference on European Transnational Education, vol. 10, pp. 340–348 (2020)
11. Thijssen, M.W.P., Rege, M., Solheim, O.J.: Teacher relationship skills and student learning. Econ. Educ. Rev. **89**, 102251 (2022)