

Evaluating Pangasinan State University Faculty Performance Using Associative Rule Analysis

Frederick F. Patacsil, Paulo V. Cenas, Bobby F. Roaring, Jennifer M. Parrone, and Daniel Bezalel A. Garcia

Abstract—Evaluating faculty members' performance is a very complex area to study. In addition, predicting the performance of these faculty members is a very difficult and challenging task. However, the core of education is teaching and learning, and teaching-learning works to its fullest when there are effective teachers. Measuring the effectiveness of faculty members is done based on the student evaluation of faculty. This research aims to develop a model to predict the performance of the faculty members using associative rule based on the existing evaluation form used by PSU to evaluate faculty members. The model is designed to utilize the knowledge of text analytics rule capabilities that will provide great support for the decision-making of Pangasinan State University in the Philippines. The result reveals that the term good is still the top one terms occurred for all campuses followed by teaching. The results indicated that teacher/faculty members on all campuses are good teachers. Associating words reveal that "teaching good subject/topic," "explains simply" and other meaningful associated words can be utilized to evaluate the performance of the teacher. The results exposed not only the quantitative values of faculty evaluation it also exposed the qualitative opinion of the students in the performance of their faculty members. This study reveals important aspects of the faculty member's teaching performance in terms of words/association of words that will describe their teaching performance. The results can be utilized in coaching and mentoring faculty members to cope with their weaknesses. The proposed model can be utilized by Pangasinan State University to evaluate the faculty members in terms of their teaching performance by utilizing the comments/opinions of the students.

Index Terms—Association rule, data analytics, faculty performance.

I. INTRODUCTION

The core of education is teaching and learning, and teaching-learning works to its fullest when there are effective teachers. The effectiveness of teaching can be measured through the use of student evaluation of faculty, a survey that students complete each academic semester. It is used to determine the perception and knowledge of faculty members on the different basic and fundamentals of teaching, research, and service using the available resources [1]. This evaluative procedure is commonly referred to as student ratings of teaching (SRT), teacher rating form (TRF), student evaluation of teaching (SET), or student evaluation of faculty

(SEF).

In the Philippines, teacher evaluation is part of the educational institution's practice. The purpose of student evaluation of faculty is to facilitate faculty growth and development. Likewise, student evaluation is used to play a support tenure, promotion, transfer, and termination decisions as well as salary increases and faculty awards. In Vietnam, improving the quality of teaching performance, professional enhancement, and making positive improvements among teachers is the main goal of teacher evaluation [2], [3]. The main objective of educational institutions is to provide a quality education through regular teaching assessment to maintain quality of education [4]. According to Magno [5], this criterion implies that the educational institution measures the quality of the services through the delivery of instruction to their students. Teaching is a development process that improves faculty performance or a reviewing process that assists administrators in making personnel and strategic decisions [6]. The main objective of faculty evaluation is to identify the strengths and work on the weak areas to improve the performance of the teachers, thus achieve optimum process quality. According to Sanders & Horn [7], Once teachers are given feedback about classroom-level instructional outcomes, they start to modify their instruction to address their weak areas.

Evaluating teachers' performance is a difficult task, but it is made for years through pedagogical surveys. The most commonly used tool to evaluate a teacher's performance is through surveying students' responses. The concern about student evaluations is based mainly on 1) students not having enough experience and maturity to evaluate the course and the teachers, 2) students' evaluation is affected by the popularity of the course and/or teacher, grades are given by the teacher and course being compulsory: or elective. Furthermore, the teacher may have the best performance on research to find out the solution for community problems, have good performance in community service, social life inside and outside the university [8]. To evaluate teacher's performance, students individually fill a pedagogical survey, and the quantitative mean of the answers corresponds to the teacher evaluation. However, despite the spread of such tools, there is no study on automatically predicting teachers' performance using qualitative surveys that will directly identify the strength and weaknesses of a teacher [3].

Nowadays, data mining is becoming a more common platform in understanding and solving educational and administrative problems in higher education. Data mining techniques are applied in higher education more and more to give insights into educational and administrative problems to

Manuscript received June 9, 2021; revised July 28, 2021.

The authors are with Pangasinan State University, Philippines (e-mail: frederick_patacsil@yahoo.co.uk, arctangent2008@yahoo.com, bob_roaring@yahoo.com, plukjenniferparrone@gmail.com, dagarcia@up.edu.ph).

increase managerial effectiveness. For example, it is used to investigate factors related with the assessment of teaching performance of educators [9]. However, most of the researches carried out on educational mining focus mainly on modeling student's performance, but there is very a small number of research model on faculty performance [10]-[13].

Data mining is a powerful technology for analyzing important information from data warehouses. It is a data analysis methodology used to identify hidden patterns in a large set of data [2]. With data mining techniques, such knowledge can be extracted and accessed in transforming the database tasks from storing and retrieval of learning and extracting knowledge. It consists of a set of techniques that are used to extract relevant and interesting knowledge from data. Data mining has several tasks such as association rule mining, classification and prediction, and clustering. Classification techniques are supervised learning techniques that classify the data item into the predefined class label. It is made of the most useful techniques in data mining to build classification models from an input data set. It is the process of finding a set of models that describe and differentiate data classes and concepts, to predict the class whose label is unknown [9]. The users commonly build models used to predict future data trends. There are several algorithms for data classification such as decision tree (j48), Support Vector Machines (SVM), and Naive Bayes classifiers. With classification, the generated model would be able to predict a class for giving data depending on previously learned information from *historical* data [4].

This research aims to develop a model to predict the performance of the faculty using associative rule based on the existing evaluation form use by PSU to evaluate faculty members.

Specifically, the study aims to:

- 1) Determine the most frequent word/words used to describe the performance of a faculty member.
- 2) Determine the performance of faculty members of every campus of Pangasinan State University using the associative approach text analytics technique.

The proposed model is designed to utilize the knowledge of associative text analytics capabilities that will provide great support for the decision-making of Pangasinan State University in the Philippines.

II. MATERIALS AND METHODS

A. Data Collection

The first step in this paper is to gather all the necessary data. Comments from the evaluation rating of the teaching staff of Pangasinan State University comprise the data for this study. Data were collected from the PSU Online Portal. Questionnaires were sent to each students' portal so that they could evaluate the performance of their instructors. Name of a faculty member, subject taught that semester (First Semester, A.Y. 2019 - 2020), and comments were gathered. Most of the comments are in the English language. Comments that are written in Filipino or those comments with two or more languages or dialects used were also

included. Refer to Table I for the sample comments derived from the actual data set. The estimated number of data gathered is 15,548.

TABLE I: SAMPLE COMMENTS FROM ACTUAL DATA SET WITH NUMBER OF LANGUAGE OR DIALECTS USED

Number of Language Used	Comments
One (English)	She is good at teaching. she do everything just to understand all the topics that she discuss. Sometimes she is late because of first subject she handle but we understand that because she wants us/them learned more about the topics she gave
One (Filipino, specifically Tagalog)	Tamad magturo, madalang.
Two (Filipino - English)	Matalino si Sir. Time is gold para sa kanya, ayaw niya ng plagiarism at higit sa lahat may sariling pinaniniwalaan Nag e-alien talk di ko na maintindihan minsan sa sobrang bilis o lalim ng mga words na ginagamit niya

Data translation was then employed after the data were compiled. Comments that are written in Filipino or Tagalog-English were translated to the English language. Filling in the missing values from the comments such as missing letter/s were next employed. Care was taken to make sure that the translation is consistent throughout the dataset. For example, comments like "instructing being not good in teaching" is translated as "bad in teaching." Lastly, inconsistencies such as misspelled words or grammatical errors were then corrected. The aforementioned processes were accomplished by the data processor.

B. Data Pre-processing

Data pre-processing was immediately employed after cleaning the data gathered. Data pre-processing operators include tokenization, filtering stop words, stemming, case transformation, and feature selection or n-gram. Translation, tokenization, case transformation, and filtering stop words are the only data preprocessing steps that were employed. The Fig. 1 shows the data preprocessing techniques used in this paper.



Fig. 1. The preprocessing of data in RapidMiner9.7.

Tokenization is the first process employed under the data preprocessing procedure. This operator breaks up the text of a document into pieces of words, keywords, symbols, and other similar things called tokens. Several options can be used in breaking up the text, but the default setting which is the non-letters was employed in this paper.

Transform cases were then employed after the data set was tokenized. This operation transformed the cases of letters from lower case to upper case or vice versa. In this paper, all letters in the data were transformed into the lower case for convenience and at the same time avoiding mix-ups.

The final preprocessing step employed is the Filter Stopwords (Dictionary) operator. This is done to remove all the tokens equal to the stopwords from the given text file and can be accomplished by listing down words that are assumed

to be unnecessary. Another option is to download a text file or document file containing pre-identified English stopwords via the Internet. The Fig. 2 shows the sample saved text file used in the paper to filter stopwords. If the document is not in text file format, it is converted into that format because it is the must-be-done requirement of RapidMiner9.7. The latter option was employed; the content was verified just in case that there are words that must be included or removed. Further, adjectives and some pronouns (such as *he* and *she*) were not included in the list of stopwords.

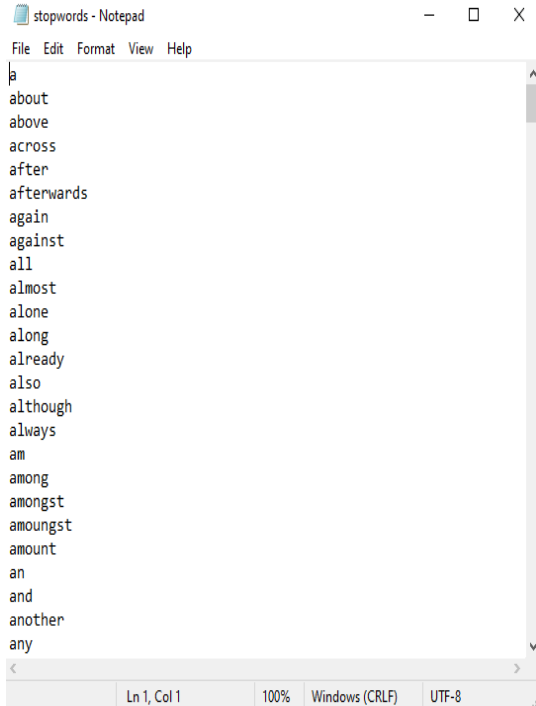


Fig. 2. Saved text file of stopwords.

TABLE II: THE BEFORE AND AFTER USING TRANSLATION AND DATA PREPROCESSING

Process	Before	After
Translation	She has good teaching skills. Nauunawaan ko ang mga tinuturo niya. Minsan busy siya kaya hindi kami nakakapag meet paero ok lang naman.	She has good teaching skills. I understand what she teaches. Sometimes busy that's why we can't meet but it's okay.
Transform Cases	She has good teaching skills. I understand what she teaches. Sometimes busy that's why we can't meet but it's okay.	she has good teaching skills. i understand what she teaches. sometimes busy that's why we can't meet but it's okay.
Tokenization	she has good teaching skills. i understand what she teaches. sometimes busy that's why we can't meet but it's okay.	she has good teaching skills i understand what she teaches sometimes busy that's why we can't meet but its okay
Stopwords (Dictionary)	she has good teaching skills i understand what she teaches sometimes busy that's why we cant meet but its okay	she has good teaching skills understand she teaches busy can't meet okay

The tokenization, case transformation, and filtering stopwords were done using the RapidMiner. The results after applying data preprocessing techniques are shown in Table

II.

C. Numeric to Binomial

The Numeric to Binomial operator was utilized after the data preprocessing. This operator transforms the type of any numeric attribute into a binary, whether true or false. This process is necessary because the output after utilizing this operator is then connected to FP-Growth to determine frequent itemset. Values for minimum and maximum parameters must be specified first. If the value of an attribute is between the specified values for the minimum and maximum parameters, then it will be mapped to false. Otherwise, the attribute will be mapped to true.

In this research paper, the default values for the minimum and maximum parameters were used, that is, both are set to 0. This means that if the value of an attribute is equal to 0, then it will be mapped to false. All values, except for 0, are mapped to true.

D. Vector Creation

The schema used in vector creation was TF-IDF or Term Frequency – Inverse Document Frequency. This schema is a numerical statistic that shows the importance of a word is to a document in a corpus and considered the most popular weight schema. In this schema, the TF-IDF value is directly proportional to the incidence of documents in the corpus that contain the word or term. Hence, as the word or term occurs most often, the value increases.

The TF-IDF will choose those words that weighted greater importance based on the dataset. The TF-IDF value increases proportionally to the number of times a term appears in the document but is offset by the frequency of the word in the corpus. It helps to control or adjust because some terms are generally more common than others.

E. Pattern Recognition

The method that was applied in determining the pattern recognition of comments was the Association Rule Mining. In Association Rule Mining, all itemsets must meet the set value for the minimum threshold for support and confidence to arrive at a strong relationship between or among items. The formula for computing the support and confidence is given below:

$$\text{SUPPORT} = \text{occurrences of } [A, B] / \text{total number of transactions} \quad (1)$$

$$\text{CONFIDENCE } [A,B] = \text{occurrences of } [A,B] / \text{total number of } [A] \quad (2)$$

Frequent itemsets are quantified by *support* which is defined as the ratio between the number of instances where [A, B] appeared together in a single transaction and the total number of transactions. On the other hand, *confidence* is the probability of finding [item A, item B] together.

This paper used a two-step process in analyzing the comments. The first step includes conducting comment word analysis using FP – Growth in terms of frequent patterns of itemsets. The value set for the minimum frequency for overall data (i.e., compiled data of every campus) is different for the per-campus data. The value set for the overall data is 67 and per campus differs depending on the data set. The second step includes the examination and determination of the strong relationship between a pair of words using the

Create Association Rules. The criterion used is confidence and the threshold value set for confidence is 0.90.

F. Flow Chart

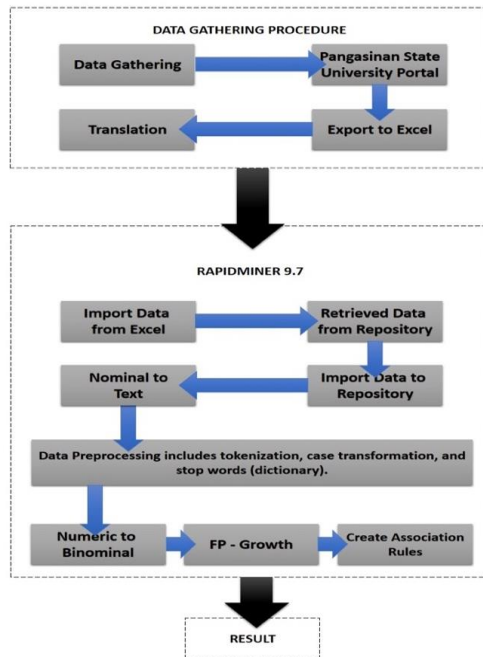


Fig. 3. The complete pattern recognition process.

Fig. 3 shows the process employed from data gathering up until the use of RapidMiner to arrive at the desired results.

III. RESULTS AND DISCUSSION

A. Descriptive Results

This subsection includes the terms with the highest frequency under the overall data set. Refer to Table III for the top 10 most occurred terms from the overall data set.

TABLE III: MOST OCCURRED TERMS FOR THE OVERALL DATA SET

Word	Rank
good	1
teaching	2
students	3
topic	4
understand	5
class	6
time	7
teach	8
topics	9
teacher	10

Based on the overall data, the term that occurred the most is *good*. Identifying the rank of the top 10 terms provided above concerning per campus provides a deeper understanding of the data set.

It can be gleaned from the table that the term *good* is still the top one terms occurred for all campuses. The San Carlos City Campus almost reflected the actual rankings for the overall data set except for the term *topics*. Table IV gives us an idea of the terms mostly used by the students per campus when giving comments on the performance of the faculty members. Say, for example, the term *teacher* under the Bayambang Campus is ranked number three, which has a

relatively higher rank compared to the other remaining campuses, tells us that most of the students on this campus give greater importance to the usage of this term.

TABLE IV: TOP 10 MOST OCCURRED TERMS FOR THE OVERALL DATA SET COMPARED TO PER CAMPUS DATA SET

Word	Alaminos		Bayambang		Binnmaley		Lingayen		San Carlos		Sta Maria		Urdaneta	
	rank	freq.	rank	freq.	rank	freq.	rank	freq.	rank	freq.	rank	freq.	rank	freq.
good	1	547	1	344	1	421	1	940	1	841	1	391	1	1052
teaching	2	389	2	215	2	363	2	932	2	679	2	295	2	1007
students	3	312	4	101	3	154	3	504	3	323	3	192	3	783
topic	4	240	5	90	5	133	5	347	4	227	4	127	4	570
understand	11	131	10	47	8	104	4	391	5	200	5	110	8	440
class	6	146	8	52	4	135	7	337	6	182	10	83	11	369
time	21	76	11	45	17	67	19	196	7	168	21	47	7	462
teach	13	109	7	109	22	62	8	286	8	165	11	79	20	268
topics	12	111	17	32	24	52	12	253	16	122	13	57	10	388
teacher	8	138	3	122	9	98	16	215	10	161	7	98	35	150

B. Pattern Recognition Using FP-Growth

This research paper applied *FP - Growth* to determine the frequent patterns in the data set. This operator efficiently calculates all frequent itemsets from the given *ExampleSet* using the *FP-tree* data structure. All attributes of the input *ExampleSet* must be binominal.

FP - Growth has two basic working modes in identifying the most frequent itemset/s: 1) Looking for the least specified number of itemsets with the highest support without considering the 'minimum support', and 2) looking for every itemset with support larger than the specified minimum support. This approach uses the *FP-Tree algorithm* which encodes the data set into a tree and then extracts the frequent itemsets from this tree. Frequent itemsets are groups of items that often appear together in the data.

Using the RapidMiner 9.7, the dataset was generated and calculated based on *FP - Growth method*. The most frequent items and itemsets were displayed in the Table V:

TABLE V: RESULTS AFTER USING FP-GROWTH FOR THE OVERALL DATA SET

Size	Support	Item1	Item2	Item3	Item 4
3	0.031	good	teaching	she	
3	0.029	good	teaching	he	
3	0.012	good	teaching	students	
3	0.009	good	teaching	topic	
3	0.012	good	teaching	understand	
4	0.005	good	teaching	she	understand
4	0.004	good	teaching	he	students
4	0.004	explain	way	things	difficult

(min freq = 67, min items = 3, max items =5).

The most frequent itemsets occurred are the following: itemset *good* and *teaching* for size 2 with support equal to 0.141; itemset *good*, *teaching*, and *she* for size 3 with support equal to 0.031; and itemset *good*, *teaching*, *she*, and *understand* for size 4 with support equal to 0.005. For example, the terms *good* and *teaching* are the most common terms used in the comments on the performance of the faculty members, that is, this item set occurs 1,410 times for every 10,000 comments. Another, the itemset containing terms *explain*, *way*, *things*, and *difficult* means that it occurs 40 times for every 10,000 comments. The results in this section

are useful in considering the rules that are worthy of further analysis. The FP -Growth results per campus are provided in the Table VI – XII. Also, Fig. 4 exhibits the comparison in the support count of the most occurred terms among the campuses of Pangasinan State University.

TABLE VI: RESULTS AFTER USING FP-GROWTH FOR THE ALAMINOS DATA SET

Size	Support	Item1	Item2	Item3
3	0.025	he	good	teaching
3	0.005	he	good	students
3	0.008	he	good	topic
3	0.021	good	she	teaching
3	0.007	good	she	students
3	0.01	good	she	topic
3	0.006	topic	can	understand
3	0.006	can	explain	topics

(min freq = 10, min items = 3, max items =5).

TABLE VII: RESULTS AFTER USING FP-GROWTH FOR THE BAYAMBANG DATA SET

Size	Support	Item1	Item2	Item3
3	0.02	good	teaching	she
3	0.017	good	teaching	he
3	0.01	good	she	teacher
3	0.011	good	he	instructor
3	0.014	communication	skills	social

(min freq = 10, min items = 3, max items =5).

TABLE VIII: RESULTS AFTER USING FP-GROWTH FOR THE BINMALEY DATA SET

Size	Support	Item1	Item2	Item3	Item 4
3	0.046	good	teaching	she	
3	0.029	good	teaching	he	
4	0.009	good	teaching	she	understand
4	0.01	good	teaching	dont	like
4	0.009	good	topic	giving	examples
4	0.01	teaching	she	dont	like
4	0.009	she	dont	noisy	like

(min freq = 10, min items = 3, max items =5).

TABLE IX: RESULTS AFTER USING FP-GROWTH FOR THE LINGAYEN DATA SET

Size	Support	Item1	Item2	Item3	Item 4
3	0.033	she	good	teaching	
3	0.007	she	good	students	
3	0.012	she	good	understand	
3	0.008	she	good	topic	
4	0.008	she	good	teaching	understand
4	0.009	she	good	teaching	shes
4	0.006	she	teaching	understand	shes
4	0.007	good	teaching	he	students
4	0.006	good	teaching	voice	low

(min freq = 20, min items = 3, max items =5).

TABLE X: RESULTS AFTER USING FP-GROWTH FOR THE SAN CARLOS CITY DATA SET

Size	Support	Item1	Item2	Item3	Item 4
3	0.022	good	teaching	she	
3	0.023	good	teaching	he	
3	0.014	good	teaching	students	
3	0.021	good	teaching	understand	
3	0.015	good	teaching	class	
3	0.016	she	topic	discuss	
3	0.014	explain	way	difficult	
3	0.015	explain	way	simple	
3	0.014	explain	difficult	simple	
3	0.014	way	difficult	simple	
4	0.014	explain	way	difficult	simple

(min freq = 32, min items = 3, max items =5).

TABLE XI: RESULTS AFTER USING FP-GROWTH FOR THE STA. MARIA DATA SET

Size	Support	Item1	Item2	Item3	Item 4
3	0.082	good	he	teaching	
3	0.027	good	he	students	
3	0.023	good	he	topic	
3	0.075	good	teaching	she	
3	0.032	good	teaching	students	
3	0.027	good	teaching	student	
3	0.031	good	teaching	kind	
3	0.027	good	teaching	attention	
3	0.023	good	teaching	abuse	
3	0.023	good	she	students	
3	0.023	good	kind	abuse	
3	0.025	good	attention	catch	
3	0.023	he	teaching	kind	
3	0.027	he	topic	understand	
3	0.023	teaching	kind	abuse	
4	0.023	good	teaching	kind	abuse

(min freq = 22, min items = 3, max items =5).

TABLE XII: RESULTS AFTER USING FP-GROWTH FOR THE URDANETA CITY DATA SET

Size	Support	Item1	Item2	Item3	Item 4
3	0.021	he	teaching	good	
3	0.009	he	teaching	students	
3	0.006	he	teaching	topic	
3	0.006	he	teaching	subject	
3	0.018	teaching	good	she	
3	0.009	teaching	good	students	
3	0.006	teaching	good	topic	
3	0.004	teaching	good	subject	
4	0.003	he	teaching	good	students
4	0.004	he	topic	can	explain
4	0.004	teaching	good	she	understand
4	0.004	she	students	topic	understand
4	0.003	students	topic	subject	understand

(min freq = 15, min items = 3, max items =5).

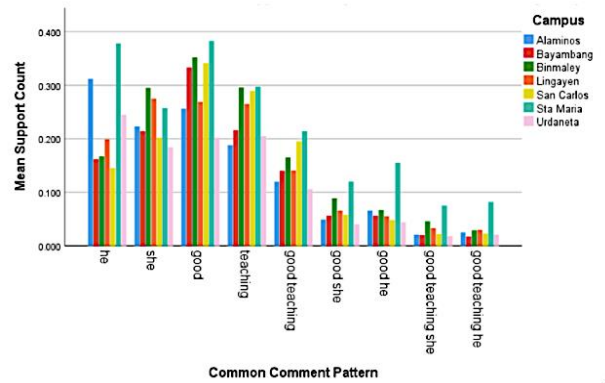


Fig. 4. Comparison of support count among the campuses of Pangasinan State University.

C. Pattern Recognition by Creating Association Rules

Create Association Rules that are based on market basket analysis was used for pattern recognition. This operator generates a set of association rules from the given set of frequent itemsets. This technique is often applied in the supermarket and other shopping industries to identify the trending items that are purchased together; thus, contributing to the possible increase in profit of the said businesses by determining the items that are subject to optimization or mass production. When this method is applied to text mining, each

term is considered as an item and the text field is the transactions. Hence, the goal is to further identify the terms that occur together that will give us an overview of the perspective of the students towards their instructors. Applying the *Create Association Rules* operator gives a clear view and understanding of the terms that have a strong relationship with each other and are commonly used together by the students in giving comments or evaluations to the faculty members. The complete result for the overall data set is shown in the below:

TABLE XIII: RESULTS AFTER CREATING ASSOCIATION RULES FOR THE OVERALL DATA SET

Premises	Conclusion	Support	Confidence
explain, way, simple	things	0.004309513	0.917808219
explain, way, simple	difficult	0.004309513	0.917808219
explain, way, things	simple	0.004309513	0.930555556
way, difficult	explain, simple	0.004309513	0.943661972
explain, simple	way	0.00469544	0.948051948
way, difficult	explain, things	0.004373834	0.957746479
explain, things, difficult	way	0.004373834	0.971428571
things, difficult	explain	0.004502476	0.972222222
way, difficult, simple	explain	0.004309513	1.000000000

(min freq = 67, min items = 1, max items =5, min confidence = 0.9).

A total of 9 rules were derived wherein each numerical value for confidence exceeds the set threshold value (minimum confidence is 0.90) after applying the Create Association Rules operator under the RapidMiner. This implies that the terms included in Table XIII have a strong relationship with each other. For example, the terms *way* and *difficult* are strongly related with the terms *explain* and *simple* with support equal to 0.004309513 and confidence equal to 0.943661972. Also, if *way* and *difficult* (premise) were included in the comment, then there is a high possibility that the terms *explain* and *simple* (conclusion) will also appear. It is not appropriate to say that *explain* and *simple* (conclusion) occurred in the comment because of the existence of *way* and *difficult* in the same comment. Take note that Association Rules are not *consequential*; instead, they are *co-existential*, that is, when the premise occurred, then the conclusion will also occur at the same time. This is also because the result has a low support value but with a high confidence value. Further, this implies that the rules might not stand out in the data set but they occurred frequently due to continued analysis. The same explanation goes for the rest of the rules included in the list.

Conducting the same Association Rule process per campus shown in Table XIV gives us the idea about the usual characteristics of its faculty members based on how the students perceived them. In Alaminos Campus, if *lessons* and *explaining* (premise) were included in the comment, then there is a high possibility that the term *good* (conclusion) will also appear, hence most of the students perceived their instructors as good in explaining the lessons and critiqued them based on their teaching strategies. Also, if the term *loud* was in the comment, then there is a high possibility that the term *voice* will follow. This implies that students thought that their instructors are loud-voiced that may indicate whether

resounding and clear (positive) or piercing (negative).

TABLE XIV: RESULTS AFTER CREATING ASSOCIATION RULES FOR PER CAMPUS DATA SET

Campus	Premises	Conclusion	Support	Confidence
Alaminos	chance	he	0.005867971	1.000000000
	loud	voice	0.006845966	1.000000000
	strategies	teaching	0.007334963	0.937500000
Bayambang	lessons, explaining	good	0.005867971	0.923076923
	social	communication	0.013527575	1.000000000
	social	skills	0.013527575	1.000000000
Binmaley	communication, social	skills	0.013527575	1.000000000
	topic, discussing	good	0.00969163	1.000000000
	discussing, giving	good	0.00969163	1.000000000
Lingayen	students, situation	understand	0.00969163	1.000000000
	noisy, like	dont	0.014977974	1.000000000
	like, messy	dont	0.008810573	1.000000000
San Carlos	she, noisy, like	dont	0.008810573	1.000000000
	low	voice	0.014096916	0.941176471
	attend	class	0.020264317	0.920000000
Sta Maria	good, teaching, low	voice	0.0059988	1.000000000
	good, low	voice	0.00989802	0.970588235
	soft	voice	0.0074985	0.961538462
Alaminos	loud	voice	0.016196761	0.947368421
	she, low	voice	0.00869826	0.935483871
	way, difficult	explain	0.013864818	1.000000000
Bayambang	difficult, simple	explain	0.014298094	1.000000000
	explain, way, difficult	simple	0.013864818	1.000000000
	explain, simple	way	0.014731369	0.971428571
Binmaley	explain, difficult	simple	0.014298094	0.970588235
	explain, difficult, simple	way	0.013864818	0.969696970
	simple	way	0.01559792	0.947368421
Lingayen	way, simple	explain	0.014731369	0.944444444
	explain, simple	difficult	0.014298094	0.942857143
	explain, way, simple	difficult	0.013864818	0.941176471
San Carlos	simple	explain	0.015164645	0.921052632
	explain, simple	way, difficult	0.013864818	0.914285714
	abuse	good	0.023329799	1.000000000
Sta Maria	abuse	teaching	0.023329799	1.000000000
	abuse	kind	0.023329799	1.000000000
	teaching, attention	good	0.026511135	1.000000000
Alaminos	abuse	good, teaching	0.023329799	1.000000000
	good, abuse	teaching	0.023329799	1.000000000
	teaching, abuse	good	0.023329799	1.000000000
Bayambang	abuse	good, kind	0.023329799	1.000000000
	good, abuse	kind	0.023329799	1.000000000
	kind, abuse	good	0.023329799	1.000000000
Binmaley	abuse	teaching, kind	0.023329799	1.000000000
	teaching, abuse	kind	0.023329799	1.000000000
	kind, abuse	teaching	0.023329799	1.000000000
Lingayen	abuse	good, teaching, kind	0.023329799	1.000000000
	good, abuse	teaching, kind	0.023329799	1.000000000
	teaching, abuse	good, kind	0.023329799	1.000000000
San Carlos	good, teaching, abuse	kind	0.023329799	1.000000000
	abuse	good, teaching	0.023329799	1.000000000
	kind, abuse	good, teaching	0.023329799	1.000000000
Sta Maria	good, kind, abuse	teaching	0.023329799	1.000000000
	teaching, kind, abuse	good	0.023329799	1.000000000
	teaching,	good	0.026511135	0.961538462

student				
	audible	voice	0.004361371	1.000000000
	inaudible	voice	0.004361371	1.000000000
	faint	voice	0.004153686	1.000000000
	aids	visual	0.007476636	1.000000000
	teaching, matter	subject	0.003530633	1.000000000
	depth, matter	subject	0.003946002	1.000000000
	late, comes	class	0.003115265	1.000000000
Urdaneta	life, relates	real	0.003322949	1.000000000
	life, situations	real	0.003738318	1.000000000
	real, situations	life	0.003738318	1.000000000
	style	teaching	0.003738318	0.947368421
	he, reading	materials	0.003738318	0.947368421
	soft, hear	voice	0.003322949	0.941176471
	explains, matter	subject	0.005607477	0.931034483
	loud	voice	0.017860852	0.914893617

(min items = 1, max items =5, min confidence = 0.9).

In Bayambang Campus, the majority of the students gave critic to their instructors through their *communication* and *social skills* because these terms have a strong relationship with each other. Bayambang Campus is best known for its teacher education program because of having been granted by The Accrediting Agency of Chartered Colleges and University in the Philippines (AACCU) the Level IV status for Bachelor of Secondary Education and Bachelor of Elementary Education until 2021; hence, good communication and social skills are necessary for the instructors to mold future educators and to maintain the excellence and quality teaching.

As for Binmaley Campus, the students perceived their instructors as good at discussing the topic. This implies that students are satisfied with the teaching methods employed, and the knowledge and pieces of information delivered by their instructors. Also, students perceived their instructors as understanding especially when it comes to the situation of their students. The only downfall is that students perceived their instructors as having low voice quality. Hence, it explains why they do not like noise particularly for female instructors due to the relationship between the terms *she*, *noisy*, *like* to *don't*.

In Lingayen Campus, if the terms *good*, *teaching*, and *low* appeared in the comment, then the term *voice* will appear at the same time. This implies that the majority of students recognized their instructors as good in teaching but with a low voice, especially when the instructor is female because of the strong relationship between the terms *she*, *low*, and *voice*. Also, the terms *loud* and *voice* that appeared on the result asserts the existence of instructors that possess audible voices. Just like in Alaminos Campus, this may indicate whether resounding and clear (positive) or piercing (negative). The ambiguity of the usage of *loud* and *voices* is probably due to the insufficiency of objects and complements in the comments of the students, thus, the majority of the comments are deemed as incomplete sentences.

In San Carlos Campus, most of the students perceived that their instructors can explain difficult topics, lessons, and other similar things thoroughly since the terms *simple* and *way* have a strong relationship, also the itemset *simple*, *way*, *difficult*, and *explain* has a strong relationship. This is a good skill for any educator in facilitating the understanding and acquisition of knowledge among the students. Furthermore,

this implies that the instructors have a profound understanding of the topics and lessons they are teaching, because their students deemed that they can thoroughly explain things.

In Sta Maria Campus, the students perceived their instructors as good in teaching just like the other campuses but abused them for being kind. This indicates that the students recognized the kindness and potential of their instructors in teaching, but their kindness made them vulnerable to abuse which imposes a serious concern. One of the things that an educator must possess is the ability to assert authority inside the classroom. It is believed and through experiences that too much authority inside the classroom affects the learning of the students and impedes innovation and feedback on the part of the students that is why instructors need to set boundaries on when to show kindness and when to impose strict authority. Thus, instructors in this campus should exercise balance between imposing authority and showing kindness inside the classroom.

Lastly, in Urdaneta Campus, most of the students perceived their instructors as good in relating topics on real-life situations to the subject matter. This might indicate that the instructors are good at using an analogy in discussions; hence, this further indicates their thorough understanding of the subject matter they are teaching. Also, students recognized their instructors whether loud-voiced or soft-voiced akin to Lingayen Campus. Also, instructors are deemed by students as late in class or tardy. Tardiness in work implies irresponsibility and unreliability. Tardiness of instructors breeds indifference, truancy or absenteeism in the students. Instructors must take their attendance and punctuality seriously so as to foster a good classroom environment and to be role models to the students.

It is meaningful and just to introduce another parameter since the support value is low with a high confidence value to measure the effectiveness of the association rules - which is the lift. The lift is the ratio between the support and the product of the probabilities of the premise and the conclusion (or the left and right-hand side) occurring assuming that there is no association between them or independent from each other. The lift measures the strength of the association between the premise and conclusion; hence, the larger the value of the lift the stronger the association between them. A lift value equal to 1 suggests that there is no meaningful association between the premise and conclusion. Notice that every lift value of each of the following derived rules is greater than 1; thus, there is a strong association between the premise and conclusion. Table XV exhibits the top three rules with the strongest association per campus highlighted in yellow.

TABLE XV: LIFT VALUE OF RULES PER CAMPUS

Campus	Premises	Conclusion	Lift
Alaminos	chance	he	3.205329154
	loud	voice	49.87804878
	strategies	teaching	4.992675781
	lessons, explaining	good	3.609354317
Bayambang	social	communication	41.7826087
	social	skills	41.7826087
	communication, social	skills	41.7826087

Binmaley	topic, discussing	good	2.844611529
	discussing, giving	good	2.844611529
	students, situation	understand	11.58163265
	noisy, like	dont	12.20430108
	like, messy	dont	12.20430108
	she, noisy, like	dont	12.20430108
	low	voice	22.72841051
Lingayen	attend	class	8.420967742
	good, teaching, low	voice	12.97276265
	good, low	voice	12.5912108
	soft	voice	12.47381024
	loud	voice	12.28998566
San Carlos	she, low	voice	12.13581022
	way, difficult	explain	19.23333333
	difficult, simple	explain	19.23333333
	explain, way, difficult	simple	60.73684211
	explain, simple	way	22.42057143
	explain, difficult	simple	58.9504644
	explain, difficult, simple	way	22.38060606
	simple	way	21.86526316
	way, simple	explain	18.16481481
	explain, simple	difficult	57.26616541
	explain, way, simple	difficult	57.16408669
	simple	explain	17.71491228
	explain, simple	way, difficult	65.94285714
	teaching, attention	good	17.46296296
	abuse	good, teaching	2.61944444
	good, abuse	teaching	4.668316832
	teaching, abuse	good	3.367857143
	abuse	good, kind	2.61944444
	good, abuse	kind	27.73529412
	kind, abuse	good	17.46296296
	abuse	teaching, kind	2.61944444
	teaching, abuse	kind	29.46875
	kind, abuse	teaching	17.46296296
	abuse	good, teaching, kind	3.367857143
	good, abuse	teaching, kind	32.51724138
	teaching, abuse	good, kind	29.46875
	good, teaching, abuse	kind	27.73529412
kind, abuse	good, teaching	17.46296296	
good, kind, abuse	teaching	4.668316832	
teaching, kind, abuse	good	3.367857143	
teaching, student	good	2.61944444	
Urdaneta	audible	voice	12.57180157
	inaudible	voice	12.57180157
	faint	voice	12.57180157
	aids	visual	102.4468085
	teaching, matter	subject	11.12009238
	depth, matter	subject	11.12009238
	late, comes	class	13.67897727
	life, relates	real	63.35526316
	life, situations	real	63.35526316
	real, situations	life	54.1011236
	style	teaching	4.612314406
	he, reading	materials	52.43194192
	soft, hear	voice	11.83228383
	explains, matter	subject	10.35318946
	loud	voice	11.50186101

IV. CONCLUSION

This research aims to develop a model to predict the performance of the faculty using associative rule based on the existing evaluation form use by PSU to evaluate faculty

members. Further, this study aims to determine the most frequent word/words to describe the performance of a faculty member and to determine the performance of faculty members of every campus of Pangasinan State University using the associative approach text analytics technique. RapidMiner9.7. was used to evaluate the results in this study. All members of this paper evaluated the results manually to ensure validity.

The result reveals that the term *good* (excluding the pronouns *he* and *she*) is the most occurring term on the students' evaluation using associative rule analysis. It can be deduced that the majority of the teaching staff of Pangasinan State University across every campus are good at teaching their subject or topic and can explain it simply so that students can easily comprehend it.

It is revealed that "teaching good subject/topic," "explain simple," and other meaningful associated words can be utilized to evaluate the performance of the teacher. The results reveal not only the quantitative values of faculty evaluation it also exposed the qualitative opinion of the students in the performance of their faculty members.

This study reveals important aspects of the teaching performance of the faculty members of Pangasinan State University. The results can be used for coaching and mentoring by university and campus heads to their faculty members in terms of their weaknesses. Moreover, the results can be utilized by Pangasinan State University to evaluate the teaching performance of their faculty members based on the comments or opinions of the students.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

F.F.P., P.V.C., B.F.R., J.M.P. conceived the presented idea of this paper; P.V.C. and J.M.P. gathered every necessary data; D.B.A.G. conducted the preprocessing of data; F.F.P. and P.V.C. performed the analysis; F.F.P., P.V.C., B.F.R., and J.M.P. supervised the findings of this work. All authors stated on this paper discussed the results and contributed to the completion of the final manuscript.

REFERENCES

- [1] J. P. Sampson, M. P. Driscoll, D. F. Foulk, and P. S. Carroll, "Successful faculty performance in teaching, research and original creative work, and service," 2010.
- [2] N. Temmerman, "Transforming higher education in Vietnam," University World News, 2019.
- [3] G.W. Fry, "Higher education in Vietnam," *The Political Economy of Educational Reforms and Capacity Development in Southeast Asia*, Education in the Asia-Pacific Region: Issues, Concerns, and Prospects, vol 13. Springer, Dordrecht. 2009.
- [4] A. K. Pal and S. Pal, "Evaluation of teacher's performance: A data mining approach," *International Journal of Computer Science and Mobile Computing*, vol. 2, issue 12, pp. 359-369, 2013.
- [5] C. Magno, *A Meta-Evaluation Study on the Assessment of Teacher Performance in an Assessment Center in the Philippines*, vol. 3, 2003.
- [6] R. I. Miller, *Evaluating Faculty for Promotion and Tenure. The Jossey Bass Higher Education Series*, Jossey-Bass Publishers, 350 Sansome Street, San Francisco, CA 94104, 1987.
- [7] W. L. Sanders and S. P. Horn, "Research findings from the Tennessee Value-Added Assessment System (TVAAS) database: Implications for educational evaluation and research," *Journal of Personnel Evaluation in Education*, vol. 12, pp. 247-256, 1998.

- [8] A. Mustafa, *Predicting Instructor Performance Using Data Mining Techniques in Higher Education*, 2016.
- [9] S. Badur and B. Badur, "Analyzing teaching performance of instructors using data mining techniques," 2011.
- [10] R. S. J. D. Baker and K. Yacef, "The state of educational data mining in 2009: A review and future vision," *J. Edu. Data Mining*, vol. 1, no. 1, pp. 3-17, 2009.
- [11] F. D. Kentli and Y. Sahin, "An SVM approach to predict student performance in manufacturing processes course," *Energy, Edu., Sci. Technol. B*, vol. 3, no. 4, pp. 535-544, 2011.
- [12] A. Pena-Ayala, "Review: Educational data mining: A survey and a data mining-based analysis of recent works," *Expert Syst. Appl.*, vol. 41, no. 4, pp. 1432_1462, 2014.
- [13] J. Zimmerman, K. H. Brodersen, H. R. Heinimann, and J. M. But-unarm, "A model-based approach to predicting graduate-level performance using indicators of undergraduate-level performance," *J. Edu. Data Mining*, vol. 7, no. 3, pp. 151-176, 2015.

Copyright © 2022 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).



Frederick F. Patacsil received his master of computer science and doctor in information technology degree from the University of the Cordilleras in 2003 and 2016 respectively. Since 1994, he has been with Pangasinan State University, where he is currently a faculty at the Department of Information Technology. His current research interests include data mining, software development, and machine learning.



Paulo V. Cenas received his M.A. applied statistics degree from Benguet State University and Ed.D. Mathematics degree from Pangasinan State University. Currently, he is the vice president for research, extension, and innovation at Pangasinan State University. His current research interests include system development, digital learning, and business management.



Bobby F. Roaring received his M.A. applied statistics degree from Benguet State University. He serves as the head of the Statistics Center at Pangasinan State University. Mr. Roaring's main research interests include machine learning and data mining.



Jennifer M. Parrone received her M.S. in mathematics at the University of the Philippines (UP) Baguio City. She serves as the chair of the Mathematics and Natural Sciences Department at Pangasinan State University – Urdaneta City Campus. Her main research interests include digital learning and game-based learning.



Daniel Bezael A. Garcia is an M.S. Mathematics student at the University of the Philippines (UP) Baguio City. He is a faculty member of the Mathematics and Natural Sciences Department at Pangasinan State University – Urdaneta City Campus. Mr. Garcia's main research interests include parameter estimation, digital learning, and data mining.