MINING PERFORMANCE OF STUDENTS IN COMPUTER FUNDAMENTALS COURSE USING CLASSIFICATION ALGORITHM: A MULTIDISCIPLINARY CASE STUDY

[1]TERESITA R. TOLENTINO, [2]JASMIN D. NIGUIDULA

[1]Cavite State University (CvSU) [2]Technological Institute of the Philippines (T.I.P.) - Manila
Email: [1]trtolentino@gmail.com, [2] jasniguidula@yahoo.com

Abstract - Data mining in educational institutions is currently utilized to discover information on the data relevant to the students and it is called Educational Data Mining. This paper applied the educational data mining to investigate the academic performance of the multidisciplinary students on Computer Fundamentals course considering the different parameters. In our case study, the record of the students was extracted from a database and used as the data set. After preprocessing the data, initial analysis leads the researcher to use the classification algorithm wherein the IF-THEN rule was derived. The result shows that students from non-computing courses can perform excellently over the others, however with almost the same passing percentage. Also, the gathered result identified that the gender of the students has nothing to do with their performance with regard to the average rating result.

Keywords - Educational data mining, kdd, classification algorithm, rule based classification, academic performance

INTRODUCTION

Universities today are operating in a very competitive environment. Like business organizations, they gather and collect data with reference to their students in an electronic form in order to transform it into useful information. The information gathered can be used an input to policy formulation to make the existing process more efficient or better yet contract new policy to accommodate the unanticipated events.
Nowadays, the common technique used to achieve this goal is through data mining. Data mining refers to extracting or mining knowledge from a large amount of data [1].

The process of mining knowledge stored in the database is known as Knowledge Discovery in Database or KDD. KDD refers to extracting or "mining" knowledge from a large amount of data. Discover patterns and relationships in the large volume of data can be performed using techniques in data mining [2]. There are several steps within the KDD model, these are the selection of data; preprocessing of data; the transformation of data; data mining; understanding the results and lastly the reporting. Many people use data mining as a synonym for KDD since it is a crucial and significant part of the KDD process [1]. A lot of developments happened in data mining field which contributes to the possibilities to mine the educational data and discover valuable information for innovation.

The implementation of data mining in an educational institution is called educational data mining (EDM). EDM is defined as the extent of systematic investigation associated with the development of methods for making discoveries to better understand students and in their dwellings [3]. The knowledge discovered can be used to do constructive recommendations to the academic planners in higher education institutions to improve their decision-making process, to improve student's academic performance and lessen the failure rate, to understand better the student's behavior, to assist instructors to improve learning style and other benefits [4].

This paper will investigate the student's data from a select university in the Philippines. This paper aimed to determine the academic performance of students in the Computer Fundamentals course according to program with respect to the ranking of computing courses; to compare dominant rating of students per program over seven-year period in Computer Fundamentals course; and to determine the academic performance of students in the Computer Fundamentals course according to gender.

II. RELATED WORKS

Data mining in educational institutions found valuable. In a survey between 1995 and 2005 on educational data mining conducted by Romero and Ventura, it was concluded that educational data mining is a promising area of research and it has specific requirements not presented in other fields [5].

There is a number of works at the university level predicting student's performance applying different data mining algorithms. The study of Golding, P. et al. finds that computer science students performance on the first year courses are strong predictors for overall academic performance by using the statistical method like regression and find the correlation [6]. Another study was conducted to predict student's academic performance in an engineering dynamic course using four mathematical models namely multiple linear regression, multilayer observation networks, radial basis functions and support vector

machines. Their predictor variables were the students' cumulative GPA, grades earned in four pre-requisite courses, and scores on three dynamics midterm examinations. This work shows that previous marks can predict the grade of a course with high accuracy [7]. Meanwhile, a combination of different data mining algorithm was employed by Kabakchieva, D. et al. to predict student's university performance using student's personal and pre-university characteristics. They applied decision tree C4.5, Naïve Bayes, Bayesian networks, K-nearest neighbors (KNN) and rule learner's algorithms to classify the students into 5 classes i.e. Excellent, Very Good, Good, Average or Bad [8]. Classification rules together with association rules were used in El-Halees study to analyze student's learning behavior. His objective is to show how useful data mining in higher education can be to improve student's performance. In the end, he was able to present the ways how to be benefited from the discovered knowledge [9]. Moreover, Al-Radaideh et al. applied also the classification method to help in improving the quality of the higher education system by evaluating the student data to study the main attributes that may affect the student performance in the courses [10]. Classification is one of the data mining technique which is useful for predicting group membership for data instances. There are four classification methods in data mining, the Decision Tree induction, the Rule-Based classification, the Classification by Backpropagation and the Lazy learners [11]. This paper employed the Rule-Based classification. Moreover, the aim of the classification is to predict the future output based on the available data. Universities or educational institutions are using data mining techniques to predict the future output of the students based on the available data. Hence, classification is one of the techniques best suited for educational analysis [12].

III. METHODOLOGY

Data Collection

Student's record from a select university was used with a period of six years. Specifically, the record consists of 5,526 instances. Table 1 presents the attributes and their description that exist in the data set as taken from the source database.

Attribute	Description	Possible values
Course description	The courses enrolled by the students	COSC21 ITEC21 ITEC21A ITEC1
Program	Program from different discipline	ACT, BSCS, BSIT, BSE,BEE, BSBM, BSHRM
Gender	The gender of students	Female, Male
Grade	Rating of the final grades of the students	1.00, 1.25, 1.50, 1.75, 2.00, 2.25, 2.50, 2.75, 3.00, 4.00, 5.00, 6.00, 8.00

Table 1. Student's Records on Different Attributes, Its Description and Possible Values

The variables used for analyzing the performance of the students as shown in table 1 are the course description, the program or degree of the students, the gender and lastly the final grade of the students. The university offered the Computer Fundamentals across seven different programs on their first year level.

Data Processing

Using the process of knowledge data discovery model (Figure 1), data set was trimmed and summarized as shown in Table 1. From data warehouse, data integration and selection was performed to gather the attributes necessary in the study.

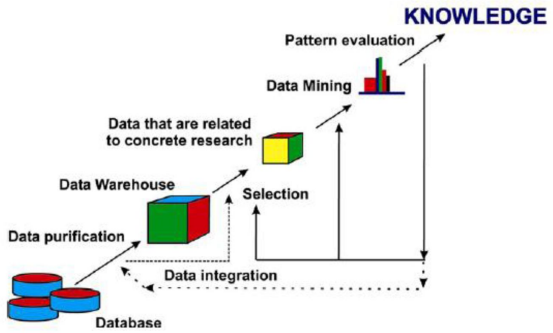


Figure 1 Process of Knowledge Data Discovery, Jiwei, H., et al. 2001

The selected data set was uploaded in Tableau software. Tableau is used by business industries to help them see and understand their data. This tool is for data discovery and visualization, for creating dashboards reports, self-service business intelligence and lastly it is for simple statistical analytics like trend and forecasting [13].

The initial visualization for the final grades of the students was classified into two classes only, the PASSED and FAILED rating. To further enhance the information retrieved, the classification algorithm was applied, specifically the IF-THEN rules classification. Rules are a good way of representing information or bits of knowledge. A set of IF-THEN rules is used as a classifier [1].

Figure 2 shows the derived classification model using the IF-THEN rules founded on the observation. It consists of converting constant attributes into distinct attributes that can be treated as categorical attributes. It uses four intervals and labels as Excellent, Very Good, Good, and Not Good.

Applying the derived IF-THEN rules, figure 2 illustrates the student's categories. Percent of total student count broken down by School Year alongside with Remarks. The data is filtered on Program, which has multiple members selected. The visualization is filtered on Remarks and School Year. The Remarks filter keeps Excellent, Very Good, Good and Not Good.

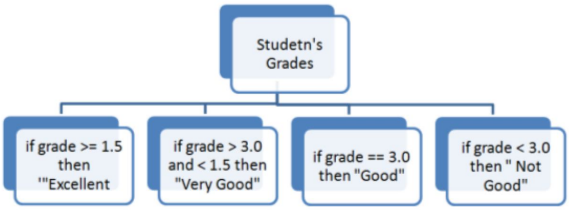


Figure 2 Derived Classification Model

Remarks	2010-2011	2011-2012	2012-2013	2013-2014	2014-2015	2015-2016	2016-2017
Excellent	69	109	59	51	56	55	11
VeryGood	120	108	87	82	79	82	28
Good	478	432	408	517	427	865	276
Not Good	44	26	37	62	31	89	48

Table 2. Number of Students Classified Using the Derive Model

To validate the figures illustrated in table 2, the population of students per program and per gender must be identified. Table 3 shows the number of students enrolled per program per academic year. Thus, the following ranking of programs was derived.

Program	2011	2012	2013	2014	2015	2016	2017
ACT	136	114	111	38	42	39	1
BEE	34	36	46	36	46	34	10
BSBM	10	89	7	201	145	258	97
BSCS	52	41	47	38	36	29	7
BSE	63	48	52	50	50	66	28
BSHRM	137	117	64	110	91	166	68
BSIT	278	228	263	239	183	499	150

Table 3. Number of Students per Program

Another parameter was considered in analyzing the student's performance. The average of student's grade per program was computed based on the final grade of each student under each program per academic year.

Program	Population Share
BSIT	33.45%
BSHRM	15.86%
BSBM	15.53%
BSE	13.05%
BEE	8.83%
ACT	8.72%
BSCS	4.56%

Table 4. Student's Population Share per Program

Table 4 shows the population share of the programs involved in the study. On the lead was Bachelor of Science in Information Technology (BSIT) with 33.45.% followed by Bachelor of Science in Hotel and Restaurant Management (BSHRM) with 15.86%. The third is Bachelor of Science in Business Management (BSBM) with 15.53% followed by Bachelor of Science in Education (BSE) with 13.05%. on the fifth spot is the Bachelor of Science in Elementary Education (BSEE) with 8.83%. on the sixth is the Associate in Computer Technology with 8.72%, and lastly with 4.56% share is the Bachelor of Science in Computer Science. The average of student's grade for six consecutive years was computed. For illustration purposes, Figure 3 shows the computed average grade of students for two academic years.

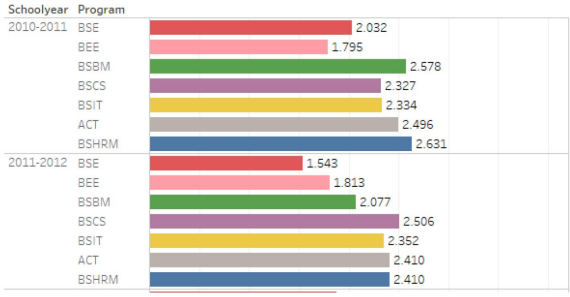


Figure 3 Average of Students Per Program

III. RESULTS AND DISCUSSION

The result shows that majority of the students are female. Figure 4 shows the total number of students enrolled in the course per academic year. The color shows the details about gender and the size shows the number of records.

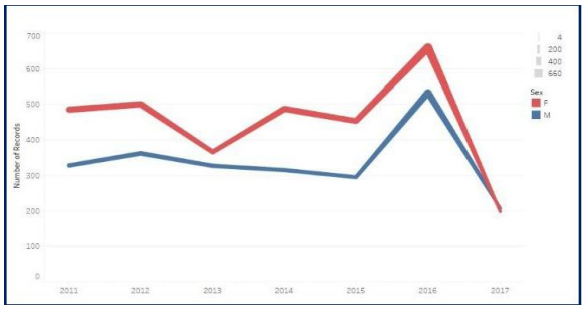


Figure 1 Number of Female and Male Students per Academic Year

When it comes to the gender of the students, it shows a slight effect on their performance. Figure 5 shows that female students across the seven programs performed well compared with male students. Although, there was a particular academic year wherein the male students excel over the female. Hence, on the forecast, it shows that for next school year female students will still outsmart the male students.

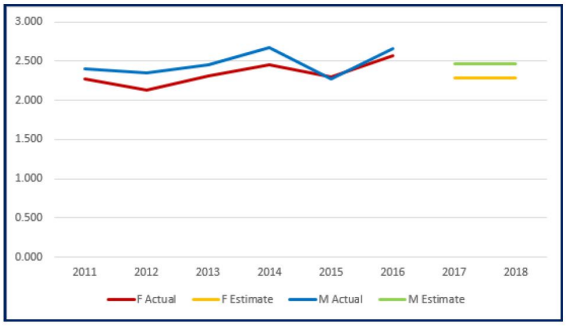


Figure 5 Average Trend of Female and Male Students with Forecast Across the Different Programs

When it comes to batch and program performance, student's average trend varies. Figure 6 illustrates the average grade of students per program for six consecutive academic years. It shows that BSE got the highest average. But on the analysis per academic year, students from different program earned the highest average on a certain academic year. Like for instance, for 2011 and 2015, BEE got the highest average rating and for 2016, ACT leads the ranking. It was discovered as well that the computing courses was outsmarted with non-computing courses. BSHRM, on the other hand, is the least performing program in computer fundamentals course. While BSIT and ACT rank as well in the bottom three.

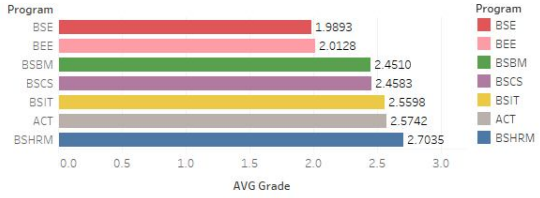


Figure 2 Average Rating of Students per Program

On the other hand, to classify the students based on the derived rules, Figure 7 to 13 illustrates the different results. Each figure presents the percentage of students belong to each class per program in an academic year.

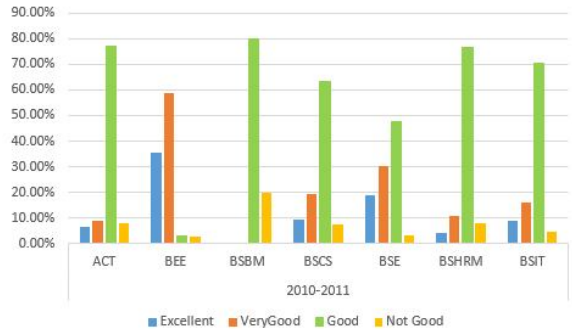


Figure 3 Students Classification for AY 2010-2011

For Academic Year 2010-2011, BEE got the highest percentage of excellent and very good students. Almost 94% of BEE students got high ratings.

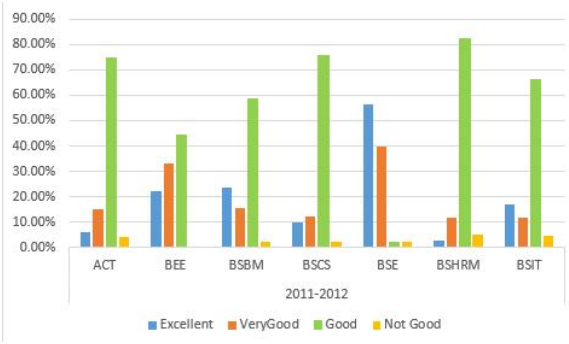


Figure 8 Students Classification for AY 2011-2012

In figure 8, it was observed that BSE is the leading program when it comes to the number of students classified as excellent and very good students. A total of 96% of students had an excellent and very good rating.

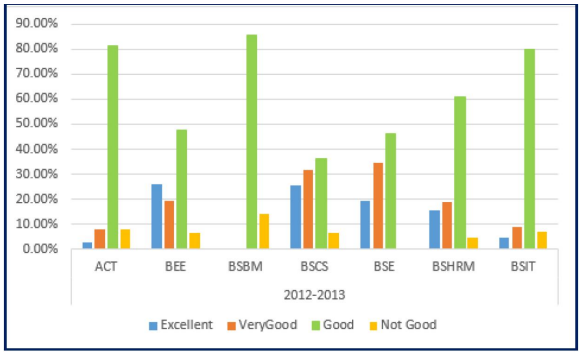


Figure 4 Students Classification for AY 2012-2013

For the academic year 2012-2013, again BEE soared high in the rating, it has 26% excellent students and 20% very good students, while BSCS ranks second.

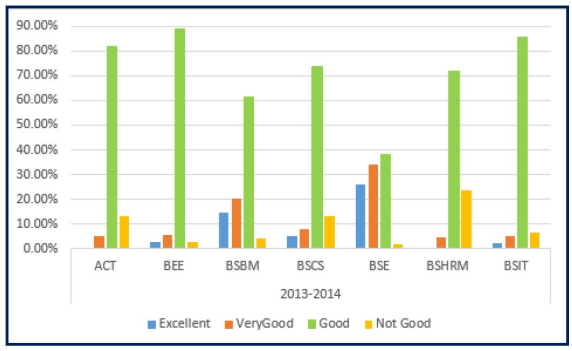


Figure 5 Students Classification for AY 2013-2014

BSE once again ranked top as shown in figure 10 for the academic year 2013-2014. It is followed by BSBM students. But it was noticeable that BSHRM does not have a record of the excellent rating.

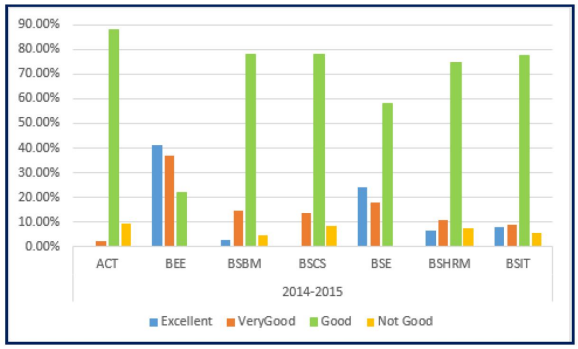


Figure 6 Students Classification for AY 2014-2015

For the academic year 2014-2015, no ACT students classified as excellent, while BEE students for the third time, rank first again when it comes to the percentage of students belong to excellent and very good classes.

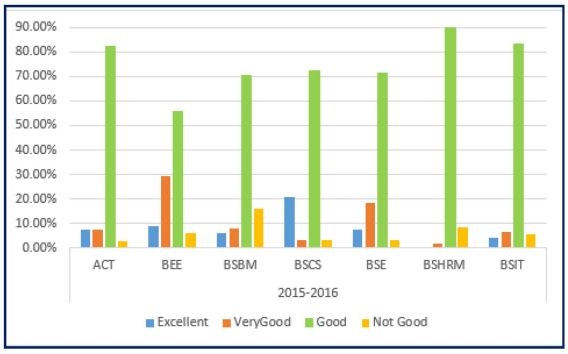


Figure 7 Students Classification for AY 2015-2016

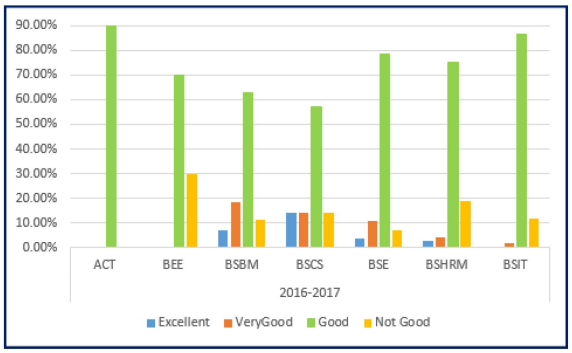


Figure 8 Students Classification for AY 2016-2017

For two consecutive academic year, BSCS lead in the ranking as shown in figure 12 and figure 13. For the academic year 2015-2016, there was a remarkable decrease in the percentage of BEE and BSE students with an excellent and very good rating. BSCS acquires 21% with excellent rating compared with 9% from BEE and 8% from BSE. While for the academic year 2016-2017, BSCS got 14% excellent rating, BEE does not have excellent rating while BSE got 4%. It was noticed as well that BSIT like BEE has no excellent rating for the academic year 2016-2017. On the other hand, BSBM rank second for this academic year.

Moreover, to further enhance the knowledge discovered using the classification algorithm, all students from seven programs were classified as well using the derived rule. Figure 14 shows that majority of the students belong to GOOD classification. For the excellent classification, result in the academic year 2011-2012 shows an extreme increase compared with previous and subsequent results. It is also alarming to see the percentage of excellent students and very good students are decreasing.

CONCLUSION

This academic performance results presented in this paper for computer fundamentals course from a select university over a period of seven years showed an unexpected result. It confirms the potential of data mining technique to be used in the higher educational institution to further enhance the services they are providing. The classification model was derived from a given parameters. The academic performance of students in the Computer Fundamentals course according to gender, based on the result, gender of the students has nothing to do with their performance. There were slight differences but could be considered statistically equal. In terms of program ranking in computer fundamental course, top performing program were non-computing programs which are BSE, BEE, and BSBM. In comparing the dominant rating of students per program over seven-year period, it was observed that almost half of the BSE and BEE students could get an excellent rating compared just a quarter of the population of BSCS with the same rating, while above 10% of BSBM, BSIT, and BSHRM, and below 10% coming from ACT. On the other hand, 65% of students from BSCS, BSBM, BSIT, BSHRM, and ACT got a very good rating.

The university should look into the entrance requirements of the students, particularly in the computing programs. There can be grade requirement for education programs because they require board exam. An intervention programs like tutorials and buddy system, and the like should be institutionalized to help out enrolled students who made it to the university because their program doesn't have entrance grade requirement. This should be looked into in order to better improve the quality of the students in the university.

REFERENCES

[1] J. &. K. M. Han, Data Mining Concepts and Techniques, 2nd Edition, San Francisco: Morgan Kaufmann, 2006.

[2] B. e. a. Baradwaj, "Mining Educational Data to Analyze Student's Performance," International Journal of Advanced Computer Science and Applications, vol. 2, no. 6, pp. 63-69, 2011.

[3] R. S. Baker, "Data Mining for Education". International Encyclopedia of Education (3rd edition). Oxford, UK: Elsevier.

[4] V. &. C. A. Kumar, "An Empirical Study of the Applications of Data Mining Techniques in Higher Education," International Journal of Advanced Computer Science and Applications, vol. 2, no. 3, p. 80, 2011.

[5] Educational Data Mining: A Survey from 1995 to 2005, 2007.

[6] P. &. D. O. Golding, "Predicting Academic Performance," in 36th ASEE/IEEE Frontiers in Education Conference, 2006.

[7] S. &. F. N. Huang, "Predicting Student Academic Perofrmance in an Engineering Dynamic Course: A Comparison of Four Types of Predictive Mathematical Models," in Computer and Education, 2013, pp. 133-145.

[8] D. e. a. Kabakchieva, "Analyzing University Data for Determining Student Profiles and Predicting Performance," in 4th International Conference on Educational Data Mining, 2011.

A. El-Halees, "Mining Students Data to Analyze Learning Behavior: A Case Study," in International Arab Conference of Information Technology, Tunisia, 2008.

[9] Q. e. a. Al-Radaideh, "Mining Student Data Using Decision Trees," in International Arab Conference on Information Technology, 2006.

[10]R. e. a. Kumar, "A Modified Tree Classification in Data Mining," Global Journals Inc. , vol. 12, no. 12, pp. 58-63, 2012.

[11]M. &. A.-R. M. Al-Barrak, "Predicting Students Final GPA Using Decision Trees: A Case Study," International Journal of Information and Education Technology, vol. 6, no. 7, 2016.

[12]"Tableau," [Online]. Available: https://onlinehelp.tableau.com. [Accessed 1 August 2017].

★★★