

Analysis of Students Performance: Input to Program Enhancement of Students in Computing

Leah G. Rodriguez, Christopher A. Rodriguez, Jasmin D. Niguidula, and Dawn Iris Calibo

Abstract—The increase of students attending tertiary levels creates challenges for educational institutions to keep track on university status through student's academic performance. This study is a prospective investigation of the academic predictors of academic performance of Bachelor of Science in Information Technology students of selected university in the Philippines. It aims to analyze the student's performance using predictive data mining techniques, specifically, on classification. The results of the study shows that students' failure or success in passing their enrolled professional course has nothing to do with their gender. Moreover, consideration in the number of units in designing the curriculum should also be considered since data shows that a very high passing rate for summer class was evident, the period when the students have less course load.

Index Terms—Data mining, classification techniques, non-technical skills.

I. INTRODUCTION

The advent of information technology in different fields needs to lead the vast volumes of information storing in different arrangements like records, files, documents, pictures, sound, recordings, logical information and numerous new information groups. The information gathered from various applications requires appropriate technique for separating learning from substantial storehouses for better decision making [1].

This information like instructive information are accessible for examination these days as it keeps on developing quickly. The need to concentrate the enormous measures of information produced from the worldwide instructive biological system has brought forth the field of educational data mining (EDM) [2]. A case of data stored in databases kept up by educational institutions incorporates enlistment information, students' performances, instructors' assessments, gender differences, and various different measures. EDM can help universities better arrangement for the foreseen number of students selecting their projects, anticipate the dropout proportion, effectively recognize feeble students, and improve utilization of accessible assets going from the number of resources to the use of different assets [3]. Also, higher education assesses its students' progress to enhance its scholarly projects.

Today's education situation is quickly changing and demanding. The system requires greater levels of

communication between the university, student, and faculty members to have ideal utilization of assets.

Different information mining procedures can be connected to instructive report and data, and classification is one of them. Classification is a directed learning system that fabricates a model to characterize an information thing as per a predefined class mark. The aim of classification is to anticipate future yield in view of accessible information. Classification can be utilized to foresee students' performances in the most basic courses, for example, programming courses in information technology [4].

This paper aims to analyze the student's performance utilizing predictive data mining techniques, particularly, on classification. This prompts the correlation of influencing components of student performance in connection to their gender and year level.

II. RELATED LITERATURE

Students' scholastic performance execution are affected by various contributing component that incorporates, for example, sex, age, showing workforce, understudies mentoring, social monetary status, area of understudies, the medium of bearings in schools, instructive cost slant, step by step contemplate hour and convenience as day researcher [5].

As of late there has been an expanded enthusiasm for utilizing information digging for instructive purposes. Information mining speaks to promising regions of study in training, and it has addressed promising locales of investigates in direction, and it has specific requirements which diverse fields require. One of the educational issues that are comprehended with data mining is the figure of particular necessities which different fields need. One of the instructive issues that are fathomed with information mining is the expectation of understudies' scholastic exhibitions, whose goal is to predict an obscure variable that depicts understudies [6]. The estimation of understudies' exhibitions joins checking and controlling understudies through the showing procedure and appraisal. Evaluations, as the fundamental methodology for the estimation of considering results, demonstrate the level of understudies' execution, which is communicated subjectively and quantitatively [7].

Data mining ought to work an indistinguishable route from a human mind. It utilizes authentic data to learn. Be that as it may, all together for information mining innovation to get data out of the database, the user must "tell it" what the data resembles. It utilizes the depiction of that data to search for similar example in a database and utilizes these pieces of information from the past to build up a predictive model of what will occur later on [8].

Manuscript received December 8, 2018; revised April 5, 2019.

The authors are with the Pangasinan State University-San Carlos City Campus, Philippines (e-mail: mam_leah@yahoo.com.ph, thopsar7@yahoo.com, jasminniguidula@gmail.com, dawniris_19@yahoo.com).

Academic performance was hypothesized to be determined by a large group of variables, which included individual and family attributes, for example, student capacity, inspiration, the nature of secondary education obtained and additionally the gender of the student. It was declared that youth training and experience, contrasts in attitudes, parental and teacher expectations and behavior, differences in courses pursued and biological differences are all instrumental in offering rise to gender differences in accomplishment [9].

Measurement figures that have been seen to be fundamental effects on scholarly execution are age, sexual orientation, business commitments, and understudy workload. Thinks about have different results when measuring the association among age and scholarly accomplishment. A couple studies demonstrate a noteworthy number of relationships among age and academic achievement, in those school complete higher outcomes and will probably persist at school than create age understudies. Assorted reviews have found that create understudies, having a clearer calling and lower compromise needs, will in all likelihood satisfy higher insightful outcomes [10].

In present day's educational framework, a student's performance in any universities is controlled by the combination of internal evaluation and external check. An internal evaluation is completed by the instructors upon the student's performance in different assessment techniques, for example, tests, assignments, and workshops, participation, and extension activities. An external check is the one that is scored by the student in semester examination. Every student needs to get the base pass mark in internal and in addition to an external examination.

Moreover, unique methodologies and strategies of information mining were connected with the estimate of understudies' prosperity, applying the data assembled from understudies' database. Information mining strategies are used to deal with far reaching volumes of data to discover covered cases and associations obliging in basic decision making [11].

Data is a gathering of important data to model or record for noise. Selection is the recovering of information significant to the investigation errand from the database. Transformation is the combination of information suitable for mining by performing synopsis or collection operations. Data Mining is a use of intelligent methods used to extract data patterns. Evaluation is the assessment of information gathered [12]. The sequences of steps identified in extracting knowledge from data and Knowledge Discovery and Data Mining (KDD) are shown in Figure 1.

Fig. 1 shows the KDD framework used in combination of information suitable for mining by performing synopsis or collection operations.

Knowledge Discovery and Data Mining (KDD) is an interdisciplinary territory centering upon procedures for separating helpful learning from information [13].

Tasks appropriate to information mining incorporate the accompanying entities such as prediction-deciding the estimation of one variable in view of examples found in others; classification-isolating the information into predefined classes in view of their traits;

clustering-discovering likenesses and contrasts in an informational collection's ascribes with a specific end goal to distinguish an arrangement of the group to portray the information. The group might be fundamentally unrelated and comprehensive or comprise of covering classifications; and description - putting a given information example or relationship into human-interpretable frame [14].

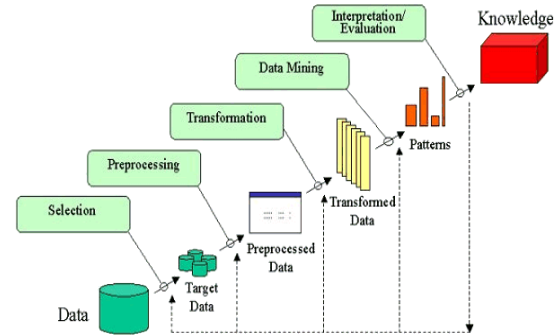


Fig. 1. Data mining: A KDD process.

III. METHODOLOGY

A. Data Set

The researcher used the collection of tertiary students' records from an educational institution for the period of two years. Specifically, the initial size of data set is 4,625 students' records.

TABLE I: STUDENT'S RECORDS ON DIFFERENT ATTRIBUTES, ITS DESCRIPTION AND POSSIBLE VALUES

Attribute	Description	Possible Values
Subject Description	The Subject enrolled by the students	Programming 1 Object Oriented Programming IT Elective I OOP Advance Web Programming
Year Level	The year level of student	I, II, III, IV
Gender	The gender of students	Male, Female
Grade	The final grades of the student	1.00, 1.25, 1.50, 1.75, 2.00, 2.25, 2.50, 2.75, 3.00, 4.00, 5.00, 0.00

As demonstrated in Table I, it described the attributes of the data and their possible values. The different attributes identified are: subject description, year level, gender and grade. These are corresponded by different descriptions and possible values for each entity.

B. Data Exploration

In this study, the data set is processed statistically in order to In order to visualize using graphical plots and diagrams. In mining data, this phase becomes an important activity because it allows the researcher and the readers to comprehend the information that is derived from the data before creating a conclusion that applies to a more complex mining tasks and algorithms.

Table II reveals the ranges of data as treated using frequency based on the different entities such as subject description, year-level, gender and grade.

Table III demonstrates the statistical summary of student's record based on different attributes as it is classified from the most or least type of data which is later treated using

frequency.

TABLE II: STATISTICAL RANGE OF DATA ON STUDENT’S RECORDS

Attribute	Range
Subject Description	Computer Organization (268), Accounting Principles (235), Logic Design (234), Software Engineering (227), Database Management System 1(226), Electronic Communication System 1 (221), Web Development (220) Network Management (217), Operating System Applications (217), Database Management System 2 (216), Application System Development 2 (207), Discrete Structures (200), ICT Elective 1 (OOP) 194, System Analysis and Design (193), Object Oriented Programming (189), Professional Ethics (187), Technopreneurship (154), Free Elective 1 (119), Free Elective 2 (104), Free Elective 3 (104), Integrated Application Software & Productivity Tools (101), Programming 1 (90), IT Fundamentals (86), Application Systems Development 1(80), ICT Elective 4 (Review on IT Certification) (80), Management Information Systems (77), Multimedia Systems (77), ICT Elective 2 (Advanced Web Programming) (45), Internship (at least 486 hrs) (29), ICT Elective 3 (Open Source Advanced Web Programming) (28)
Year Level	I (511), II(1748), III (1688), IV (678)
Gender	Female (2157), Male(2468)
Grade	1.00 (1), 1.25 (142), 1.50 (313), 1.75 (831), 2.00 (1631), 2.25(571), 2.50 (554), 2.75 (309), 3.00 (228), 4.00 (0), 5.00 (0), 0.00 (45)

TABLE III: STATISTICAL RANGE OF DATA ON STUDENT’S RECORDS BASED ON DIFFERENT ATTRIBUTES

Attribute	Most	Least
Subject Description	Computer Organization (268)	IT Elective 3 (28)
Year Level	II (1748)	I (511)
Gender	Male (2468)	Female (2157)
Grade	2.00 (1631)	1.00 (1)

TABLE IV: CLASS-WISE ACCURACY FOR FINAL GRADE PREDICTION

CLASS	True Positive (TP)	False Positive (FP)	Correct Precision (%)
0	1.000	0.000	100%
1	1.000	0.000	100%
1.25	1.000	0.000	100%
1.5	1.000	0.000	100%
1.75	1.000	0.000	100%
2	1.000	0.000	100%
2.25	1.000	0.000	100%
2.5	1.000	0.000	100%
2.75	0.997	0.000	99%
3	1.000	0.000	100%
Weighted Average	1.000	0.000	100%

TABLE V: CONFUSION MATRIX OF THE STUDENT’S CLASSIFICATION OF FINAL GRADE

a	Class										Classification Value
	b	c	d	e	f	g	h	i	j		
45	0	0	0	0	0	0	0	0	0	0	a=0
0	1	0	0	0	0	0	0	0	0	0	b=1
0	0	142	0	0	0	0	0	0	0	0	c=1.25
0	0	0	313	0	0	0	0	0	0	0	d=1.5
0	0	0	0	831	0	0	0	0	0	0	e=1.75
0	0	0	0	0	1631	0	0	0	0	0	f=2
0	0	0	0	0	0	571	0	0	0	0	g=2.25
0	0	0	0	0	0	0	554	0	0	0	h=2.5
0	0	0	0	0	0	0	0	308	0	0	i=2.75
0	0	0	0	0	0	0	0	0	228	0	j=3

Table IV presented the class-wise accuracy, which compared the true positive and false positive classifications. In addition, the correct precision for the final grade class outcome categories was presented.

Table V shows the confusion matrix of the student’s classification of final grade. Specifically, the classes are categorized according to the equivalent value of each class. Further, generally shows the student’s performance as it reveals the number of students that has a grade based on the classification value.

IV. RESULTS AND DISCUSSION

The model in the Form of ID3 algorithm generated from studentmodel.arff is shown in Table IV and Table V. The accuracy of the predictive model obtained is 99% which mean that 4624 student out of 4625 are correctly classified.

It is obvious from the confusion matrix shown in Table IV that the model is ideal in the prediction of grade 1 because out of 4625 students it has been classified 100 % correctly.

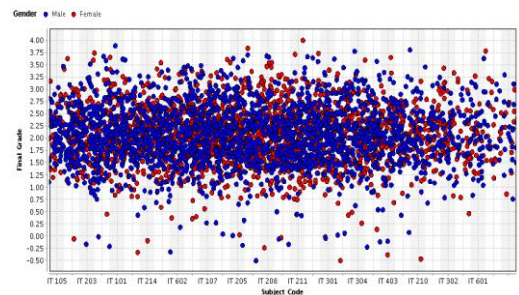


Fig. 2. Visualization of passing and failing rate of students according to course.

Fig. 2 shows the visualization of the courses taken by the students according to gender and grade result. It shows that the number male and female students in the select university has equal distribution same with their grades. Majority of the students taking computing professional courses are getting passing rating. However, there are some students who still fails on their enrolled courses. There are some predictors being considered why these students failed their classes.

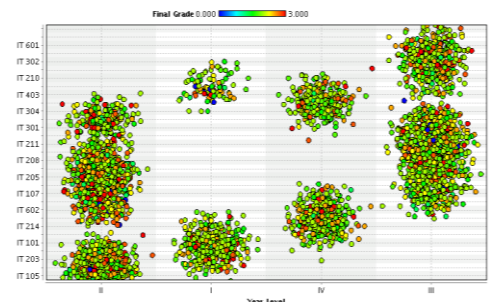


Fig. 3. Visualization of students grade according to course and year level.

Fig. 3 shows the distribution of students’ grade according to year level. It shows that the junior students have the least failing grades as compared to the other year level (the blue dot corresponds to the failing marks), while the freshmen and senior students have the most number of failing marks. . Some causes students to fail in college are as follows: lack of preparedness, poor attendance, study habits, and time management [15]. The result shows that freshmen students

are still in adjustment period while some sophomores haven't adjusted yet that is why they still have failing grades. The senior students are on their terminal year commonly working on their capstone project that is why some of their enrolled courses are being neglected

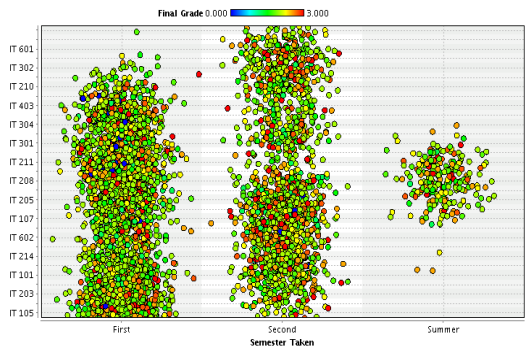


Fig. 4. Visualization of students grade according to semester.

Figure 4 shows the grades of the students for 1st semester, second semester, and summer. It shows that there was no failing grade on summer where professional courses were taken, but it also shows that there were very few enrollees during summer. In the Philippines, particularly where the data set was taken not all curriculum requires summer, some students enroll summer class because of their failing grades or they want to advance course. The maximum number of units that a student can enroll is 9 units only, this can be related to the passing rate of the students during summer because they are concentrating only on few courses but it does not apply that students will get passing rate if they enroll in summer.

V. CONCLUSION

The accuracy rate of the data set used is 99%, the results of the study shows that students failure or success in passing their enrolled professional course has nothing to do with their gender. Lack of preparedness, poor attendance, study habits, and time management [15] can be considered in order to improve the grades of the students especially in the freshmen, and sophomore level. Technical skills of the students should not be the focus in improving the academic performance of the students but also the non-technical skills which could also affect their class performance. Consideration in the number of units in designing the curriculum should also be considered since data shows that a very high passing rate for summer class was evident, the period when the students have less course load.

ACKNOWLEDGMENT

We are grateful to all of those with whom we have had the pleasure to work during this study. Each of them has provided us extensive personal and professional guidance and taught us a great deal about both research and life in general. We would especially like to thank Pangasinan State University-San Carlos Campus, Technological Institute of the Philippines-Manila and Siquijor State College. They taught us more than we could ever give them credit for here. They have shown us, by their example, what a good person

should be.

REFERENCES

- [1] V. R. Basili *et al.*, "Experience factory," *Encyclopedia of Software Engineering*, 1994.
- [2] C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005," *Expert Systems with Applications*, vol. 33, no. 1, pp. 135-146, 2007.
- [3] R. S. Baker and K. Yacef, "The state of educational data mining in 2009: A review and future visions," *JEDM-Journal of Educational Data Mining*, vol. 1, no. 1, pp. 3-17, 2009.
- [4] D. M. Patil and R. N. Phursule, *Knowledge Transfer Using Cost Sensitive Online Learning Classification*.
- [5] S. L. Christenson, "The family-school partnership: An opportunity to promote the learning competence of all students," *School Psychology Review*, vol. 33, no. 1, p. 83, 2004.
- [6] E. Osmanbegović and M. Suljić, "Data mining approach for predicting student performance," *Economic Review*, vol. 10, no. 1, 2012.
- [7] J. Biggs, "What the student does: Teaching for enhanced learning," *Higher Education Research & Development*, vol. 18, no. 1, pp. 57-75, 1999.
- [8] C. Puschmann and J. Burgess, "Big data, big questions | metaphors of big data," *International Journal of Communication*, vol. 8, p. 20, 2014.
- [9] E. K. Nyhus and E. Pons, "The effects of personality on earnings," *Journal of Economic Psychology*, vol. 26, no. 3, pp. 363-384, 2005.
- [10] Y. K. R. Chan, "The relationship between gender, age, study mode, locus of control, extracurricular activities, learning approaches and academic achievement: the case of full-time and part-time Hong Kong Chinese sub-degree students," Doctoral dissertation, University of Leicester, 2010.
- [11] U. Fayyad *et al.*, "From data mining to knowledge discovery in databases," *AI Magazine*, vol. 17, 3, p. 37, 1996.
- [12] Q. Yang, "Data preparation for data mining," *Applied Artificial Intelligence*, vol. 17, no. 5-6, 2003, pp. 375-381.
- [13] G. Vishal and G. S. Lehal, "A survey of text mining techniques and applications," *Journal of Emerging Technologies in Web Intelligence*, vol. 1, no. 1, 2009, pp. 60-76.
- [14] U. Fayyad *et al.*, *AI Magazine*, 1996.
- [15] N. Kukemuller, "What causes students to fail courses in college?" 2017.



Leah G. Rodriguez was born in the Philippines, July 28, 1975. She is graduated of master in information technology in the Philippines and has units in doctor in information technology.

She is a professor of information technology and has made research contribution in the areas of system development and design, and project management.



Christopher A. Rodriguez is graduated of master of arts major in computer education in the Philippines and has units in Doctor of Philosophy. He is an instructor of information technology in Pangasinan State University – San Carlos Campus.



Jasmin D. Nigidula was born in the Philippines. She is a professor of information technology and has made research contribution in the areas of information theory, data mining, system development and design, project management and artificial intelligence.



Dawn Iris Calibo is graduated of Master in science teaching computer science in the Philippines and has units in Doctor in Information Technology. She is a professor of information technology and has made research contribution in the areas of information theory, data mining, system development and design, project management and artificial intelligence.