

Determinants of Students' Retention in Higher Education

Myelinda A. Baldeovar

Surigao del Sur State University Tagbina Campus, Tagbina, Surigao del Sur

Abstract: Student retention has become one of the most important measures of success for higher education institutions. This study discusses about educational data mining on the students' retention in Bachelor of Science in Computer Science ladderized program of Surigao del Sur State University-Tagbina Campus. It has been observed that after earning an associate diploma, a drastic number of students opted not to continue attaining the BSCS degree. This study aims to predict students' retention rate, the relationship between high school GPA in finishing the BSCS degree and the impact of student performance in computer programming subjects in predicting student retention. Results of the study revealed that retention rate of females are higher than males. In addition, sex and HS GPA is not a significant predictor of the retention rate of BSCS students. Correlational analyses showed that average grade in programming 1 and 2 has a strong positive relationship with students' retention.

Keywords: Correlation, data mining, predictive modelling, retention

1. Introduction

High school graduates enrol in colleges to earn a college degree; however, some students do not finish the degree. An institution fails to retain its student if the student does not finish the degree from where they started. Seidman (2005) defines student retention as the "ability of a particular college or university to successfully graduate the students that initially enrol at that institution". The Surigao del Sur State University – Tagbina Campus offers a ladderized program in Bachelor of Science in Computer Science where students will obtain an Associate in Computer Technology diploma after finishing the first two years of study. After obtaining the associate diploma, the students may opt to continue for another two years leading to BSCS degree. It has been observed that after obtaining the associate diploma, more than 50% of the students did not return to continue the BSCS degree. The key to better understand this phenomenon is to apply application of data mining and predictive modelling. The issues that are addressed in the study are the following: a) predict students' retention rate, b) the relationship between high school GPA in finishing the BSCS degree c) the impact of student performance in computer programming subjects in predicting student retention

Data mining was used in different areas of education. Educational Data Mining is concerned with developing methods that extract knowledge from data come from the educational context. Uncovering hidden information is the fundamental goal of data mining. The data can be collected from historical and operational data reside in the databases of educational institutes. The student data can be personal or academic. (Romero et. al, 2008). Delavari et al. (2005) proposed a data mining analysis model to be used in higher educational system, which identified various research areas in higher education that could use data mining. Luan and Serban (2002) listed some of the applications of data mining to higher education, and provided some case studies to showcase the application of data mining to the student retention problem.

Studies in the past shows that student retention problems have been researched extensively to realize individual and institutional factors over the past 30 years that impact student retention and graduation rates but little research on students retention in a ladderized program and the significance of high school GPA and student performance in predicting student retention in higher education has been done.

Predictive modelling is important to identify factors to improve students' retention while reducing the cost of education processes. Improving student success and increasing student retention rates yields a higher financial benefit. It is more cost effective to keep students who are already enrolled than to invest in recruitment efforts to drive up enrolment numbers.

2. Literature Survey

Data mining can be applied to a number of different applications, such as data summarization, learning classification rules, finding associations, analyzing changes, and detecting anomalies (Han et al., 2006, Westphal et al., 1998). Sometimes, data mining has to deal with unstructured or semi-structured data. For universities, data mining techniques could help provide more personalized education, maximize educational system efficiency, and reduce the cost of education processes. It may guide us to increase student's retention rate, increase educational improvement ratio, and increase student's learning outcome.

The transition from high school to college has received much attention (Astin, 1993; Atkinson and Geiser, 2009; Conley, 2005; Fetter, 1997; Willingham et al., 1985). According to Hagedorn (2005) retention rate means first-time Freshmen students who graduate within six years of their original enrolment date. Druzel and Glymour (1999) define "student retention rate" as the percent of entering Freshmen who eventually graduate from the university where they enrolled as a Freshmen. Kramer (2007) suggested an uncomplicated definition of retention as an

“individual who enrolls in college and remains enrolled until the completion of a degree.”

Herrera (2006) concluded that many variables vary in their success at predicting persistence, depending on the academic level. In other words variables that affect persistence at one academic level won't necessarily affect persistence at a different academic level. This means that different models which differentiate between dropout and persistent student should be constructed for each academic level. The same results could be expected at the course levels. That would mean that we would get different probabilities of leaving or staying on the course even for the same student depending upon the course.

Three indicators are of particular importance when predicting future success during that transitional period: GPA obtained at the secondary school, intelligence quotient (IQ), and self-efficacy (Preckel and Frey, 2004). A metaanalysis by Trapmann et al. (2007) measures correlations (mean corrected validities) between high school GPA and undergraduate grades in the range of 0.26 to 0.53. In addition, Atkinson and Geiser (2009) emphasize that high school grades are the best known predictors of student readiness for undergraduate studies, regardless of the quality and type of high school attended, while a standardized admissions test provides useful supplementary information.

3. Methodology

3.1 Data Preparations

The dataset analyzed in this research was obtained manually from the University Registrar Office of Surigao del Sur State University-Tagbina Campus. The dataset included only full-time first year Bachelor of Science in Computer Science students of academic year 2013-2014 and their grades from academic year 2013-2014 and 2014-2015. The data included high school GPA and grades in every computer programming subjects in the first two years of study in the university. Data stored in different tables was joined in a single table. After the joining process, errors were removed.

This research used Rapidminer data mining software to investigate the most important variables that are associated with retention and for all the analyses to answer specific research questions. Rapidminer facilitated data mining analyses to generate very high accurate predictive models based on all the data gathered from the university registrar office. In mining the data, the target variable was graduation and all other variables will be used as predictors.

3.2 Data selection and transformation

In this step only those fields were selected which were required for data mining. A few derived variables were selected. While some of the information for the variables was extracted from the database. All the predictor and response variables which were derived from the database are given in Table 1 below.

Table 1: Predictor and Response variables

List of Variables	Type	Description
Sex	Binomial	Student Gender
Finished_BSCS	Binomial	Graduated BSCS or not
Finished ACS	Binomial	Graduated ACS or not
HS_GPA	Integer	High school GPA
Progg1and2average	Integer	Average grade in programming 1 and 2 subjects

3.3 Correlational Analysis

Correlation is a statistical technique that can show whether and how strongly pairs of attributes are related.

4. Results and Discussion

After applying data mining techniques in the data, results are shown below.

Table 2: BSCS Retention Rate

Sex	Male	Female
Total	77	92
Finished ACS	25	71
Did Not Finish ACS	52	21
Retention Rate	32.47%	77.17%
Finished BSCS	13	39
Did Not Finish BSCS	12	32
Retention Rate	52%	54.93%

Values in Table 1 above revealed that there are more female than male students enrolled in BSCS. It can be noted that there are more male students who did not complete the associate diploma than female students. Furthermore, retention rate in associate diploma of BSCS male students is lesser compared to female students. In addition, results showed that after completing the associate diploma only few students continued and finished the BSCS degree.

Table 3: Correlation Matrix of BSCS Students Retention

First Attribute	Second Attribute	Correlation
Sex	Finished BSCS	0.044
Sex	Finished ACS	0.173
Sex	HS GPA	0.008
Sex	Progg1and2average	0.153
Finished BSCS	Finished ACS	0.330
Finished BSCS	HS GPA	0.008
Finished BSCS	Progg1and2average	0.421
Finished ACS	HS GPA	0.008
Finished ACS	Progg1and2average	0.515
HS GPA	Progg1and2average	0.009

As shown in Table 3 above, the average grade in programming 1 and 2 subjects is highly correlated with students' retention in BSCS program. This implies that students who have a good performance in programming subjects tend to finish the BSCS degree. On the other hand, sex and high school GPA are not correlated with students' retention in the program. This means that students' sex and GPA in high school does not determine whether the students will finish the program or not.

5. Conclusions and Recommendations

The following are the conclusions that can be drawn from the analysis:

- There are more female than male students enrolled in BSCS program.
- Retention rate of females is higher than males.
- Sex and HS GPA is not significant in predicting the retention rate of BSCS students.
- Average grade in programming 1 and 2 has a strong positive relationship with students' retention.

Future research in this area should focus on the socio-economic and faculty factors that could directly relate to the individual student's decision to stay or leave.

References

- [1] Atkinson, R. C., and Geiser, S. 2009. Reflections on a century of college admissions tests. *Educational Researcher* 38, 9, 665-676.
- [2] Braxton, J. M., Hirschy, A. S., & McClendon, S. A. (2004). *Understanding and reducing college student departure*. NJ: Wiley Periodicals, Inc.
- [3] Druzel, M. J., & Glymour, C. (1999). Causal inferences from databases: why universities lose students. In C. Glymour, & G. F. Cooper (Eds.), *Computation, causation, and discovery* (pp. 521-539). Menlo Park, CA: AAAI Press.
- [4] Han, J. W., Kamber, M., 2006. *Data Mining: Concepts and Techniques*, 2nd Edition, *The Morgan Kaufmann Series in Data Management Systems*, Gray, J. Series Editor, Morgan Kaufmann Publishers.
- [5] Herrera, O. L. (2006). Investigation of the role of pre- and post-admission variables in undergraduate institutional persistence, using a Markov student flow model. PhD Dissertation, North Carolina State University, USA.
- [6] J. Luan and A. M. Serban. Data mining and its application in higher education. In *Knowledge Management: Building a Competitive Advantage in Higher Education: New Directions for Institutional Research*. Jossey-Bass, 2002.
- [7] N. Delavari, M.R. Beikzadeh, and S. Phon-Amnuaisuk. Application of enhanced analysis model for data mining processes in higher educational system. *ITHET 6th Annual International Conference*, pages 7{9, July 2005.
- [8] PRECKEL, D., AND FREY, K. 2004. Ein Überblick über Prädiktoren für Studienerfolg. Retrieved from: <http://www.ifvf.ethz.ch/news/index>.
- [9] Raju, Dheeraj(n.d) Predicting Student Graduation In Higher Education Using Data Mining Models: A Comparison
- [10] Romero, C., Ventura, S. and Garcia, E. (2008) 'Data mining in course management systems: Moodle case study and tutorial', *Computers & Education*, vol. 51, no. 1, pp. 368-384.
- [11] Romero, C., Ventura, S. and Garcia, E. (2008) 'Data mining in course management systems: Moodle case study and tutorial', *Computers & Education*, vol. 51, no. 1, pp. 368-384.

[12] Seidman, A. (2005). *College student retention: Formula for student success*. Westport, CT: Praeger.

[13] Trapmann, S., Hell, B., Weigand, S., And Schuler, H. 2007. Die Validität von Schulnoten zur Vorhersage des Studienerfolgs: Eine Metaanalyse :The validity of school grades

Author Profile



Myelinda A. Baldeovar is an Associate Professor at Surigao del Sur State University-Tagbina Campus located in Tagbina, Surigao del Sur, Philippines. Presently she has completed her academic requirements in Doctor of Information Technology at Saint Paul University Philippines located in Tuguegarao, Cagayan, Philippines.