

DETERMINATION OF FACTORS AFFECTED THE FOUNDATION STUDENTS IN WITHDRAWAL

Zuwaina Rashid Salim AL-Riyami¹ and Aiman Moayid²

¹Department of Information System, University of Nizwa, Nizwa, 44464642@uofn.edu.om

²Dr., Department of Information System, University of Nizwa, Nizwa, aiman.moayid@unizwa.edu.om

Abstract

In 21 century the education become more important like the health and the economic sector. In Oman, the students join the higher educational institutes after they complete twelve years in school. During their studies the students affected by many factors that can lead them to incomplete their studies and get the higher certificate. This study investigate the factors that affect the students in withdraw from colleges of technology in the first year using data mining techniques. This study target the students in colleges of technology in foundation level. Using decision tree list of the rules extracted that may affect the students in withdrawing from the college. It found that most of students leave the college with reason undesired to study, which affected the college performance.

Keywords: Data mining, withdraw, higher education, Colleges of Technology, factors.

1. INTRODUCTION

Collaboration between higher education institutes and secondary schools are highly needed to prepare the students for higher education level. Based on many research done on higher education institutes it found that the students need to well oriented about the higher education rules and regulations. This will help the higher education in improve the student's retention. Colleges of technology is one of the higher education institutes in Oman. It come under Ministry of manpower. A number of students join colleges of technology each year divided into two batch September batch and January batch. The purpose of this study is to apply data mining techniques to investigate the factors that affect foundation students in withdraw from the college. Many researchers conducted studies to develop models using data mining techniques to predict student retention in higher education institutes and highlight the important factors affect the students in withdraw. Most of researches were carried out in US institutes. Ruba Alkhasawneh developed a hybrid model to predict the student retention in the first year . As well as Dheeraj Raju & Randall schumacker they build a model to explore the important characteristic associated.

Data mining techniques used to predict the student's retention as they provide more accuracy and itis easy to use and analysis the data. Many reasons were found by previous study that lead students to leave the institute before they graduate as in [1] [2] [3]. Some of those reasons that lead the student to leave the college in the first year are academic performance, challenges in their financial and social situations,

availability of vacancies for those students who have post-secondary diploma certificate and health problem. Management of the colleges of technology need to find a way to reduce the number of students who are leaving the college as this will affect the quality of the college.

2. BACKGROUND AND LITERATURE REVIEW

Over past years many studies were done to improve student retention at higher education institutes. The predictive model was developing to predict student at risk [4] [5]. Farshid Marbouti states in his study that it is possible to identify at-risk students early and both students and instructors can be notified early. In this paper, the researcher compares different comparative methods to identify at-risk students in the course using grading standards. Logistic regression, support vector machine, decision tree, multi-layer perceptron (MLP), naïve Bayes classifier (NBC) and k-nearest neighbor (KNN) were used to find best predictive method. It was identified that using learning objective scores and score based graded assessments may have a great advantage in the prediction model. Among the seven prediction methods used to show the accuracy and usability of course. The best method was found in the naïve Bayes classifier, which identifies 86.2% of students who failed in the course. Chi-squared automatic iterative detection (CHAID) the decision tree model is another predictive method that was studied by [5]. He develops a predictive model for predicting short term student attrition. With the help of CHAID tool he was able to divide the population into subset according to student characteristics. In addition to that this tool can handle missing data which is needing be treated. The predictive model were built using five years of historical data from 2009 to 2013.

Using a different kind of method in data mining help to identify the factors that lead the student to leave the college before they complete their study. Ruba Alkhasawneh was using data mining in the study to develop a hybrid model to predict student first-year retention [6]. The researcher was able to identify the most factors that impact student success as well as the input impact student retention. In this study, the neural network algorithm and the genetic algorithm were used. Qualitative design was applies focus group in order to collect more details about student success and student retention as it is an effective way in term of analyzing non-quantitative data. With the help of the neural networks, and in order to improve the accuracy of the model genetic algorithm was used. It found that high school preparation has a great impact on student success in college. In addition, student academic and social adjustment to college significantly reflects student performance and retention in the first year. The network accuracy was improved using an optimum set of student inputs for the three-group selected. The recommendation of this study was the researcher can use the same procedure and models with a larger dataset and more participants to provide a better reflection of URM STEM student first-year retention behaviors. Also, this study could be extended by adding more precollege inputs to a model like family background, high school grade, and financial aid background. This study was having many limitations. The first limitation is the sample of the quantitative part which limited to first-time first-year students. The limited number of URM students (N=498) was included in this study that affects network accuracy.

A huge data is stored in an educational database that contains useful data that can help college management in extracting the right decision. Course performance is one of the data which is stored in a data warehouse that needs to be used in order to track student performance in their course. Abeer Badr El Din Ahmed and Ibrahim Sayed Elaraby predict student performance using classification methods [7]. They found that large volume of records, documents and images is store in an educational database. Data mining techniques used to extract a pattern from this information that helps the college management to predict student performance. With the help of wake software, ID3 decision tree used to get the measure of information gain. It generates a set of rules from the decision tree. An example of this rule is (IF Midterm='Excellent' AND LG='Good' AND SP='No' AND HW='No' AND SEM='Good' Dep='Scientific Mathematics' THEN FG='Very Good'). This study will help the college to improve the student's performance, to identify those students who needed special attention to reducing failing ration and taking appropriate action at the right time. Qasem A. Al-Radaideh and Emad Al-Shawakfa stated in their study how to apply decision trees in mining student data [8]. Data mining process used to enhance the quality of the higher education system by evaluating student data to study the main attributes that may affect student performance in courses. This study built to help the students to predict the final grade in a course. The questionnaire was used in this study to collect the data from undergraduate students. They used the Weka toolkit to rank the 12 attributes they found it based on the questionnaire. CRISP-DM (Cross Industry Standard Process for data mining) methodology used to build the classification model. They found the higher management can use such a classification model to enhance the outcome of the course according to the extracted knowledge. Such knowledge can be used to give a deeper understanding of student's enrollment pattern in the course under study.

3. METHODOLOGY

Based on the earlier studies and their recommendations, this study will investigate the factors that affect the students in withdraw from colleges of technology in the first year using the data mining technique. The method will be in four stages, which are collecting data, pre-processing the data, analyzing data and Train the data. From CIMS (Student Information Management system), 24,987 records were retrieved between 2009 and 2018. The quality of the dataset is important to get an accurate and correct result. It found that the dataset has missing data which need to be clean. In this stage, some of the records were deleted because most of the field data is missing. Some of the missing values were filled manually by predicting the value based on the previous field. After the preprocessing phase, the dataset was to train and test using the Decision tree. The decision tree is used in this study based on the recommendation of the previous study [7] and as its help to extract the rules in the easiest way. In addition to that, it helps the decision-maker to see the logical view of the dataset. The CART algorithm is used in this research. This algorithm can create generalized models. In decision tree each node represent the attribute and each branch represent a decision rules finally the leaf node represents a single outcome of the class. In this study, the entropy function will be used. Python tool was used to build a decision tree and analyzing the dataset.

4. FINDING

The aim of this study is to predict withdrew at foundation level and the most factors that affect the students to leave. The history data is collecting from the CIMS system for the period of [2009 -2018]. The variable of this research classified and analysis as in table (1.1):

Table (1.1), Variable classification & analysis.

| Variable Type | Variable | Description |
|----------------------|------------------|---|
| Demographic variable | Gender | Male / Female |
| | Region | Which region student from |
| | Student category | Low income / Social welfare /other |
| | Disability | Yes / No |
| College variable | Amount | Student allowance in the college 90/45 |
| | Study status | Withdraw / Completed |
| | College Name | Name of the college the student left from |
| | Placement test | English placement test score |

In decision tree dataset was split into two set train and test as 70% in the training set and testing is 30%. The accuracy of this model was 0.83 for the train data. There are set of rules were extracted from the Decision tree, some of this rule mention below:

Decision tree some of this rule mention below:

IF (Gender = Male, Allowance=90, PT<50, SG =other) = Students Withdraw

IF (Gender = Male, Allowance=45, PT<50, SG=LI) = Students withdraw

IF (Gender =Female, Allowance =90, PT<50, SG=other) = Students withdraw

IF (Gender =Female, Allowance =45, PT<50, SG=LI) = Students withdraw

From the analysis phase, it found that male students get a high score in applying to withdraw in the first year. From the rules, if the student is male and getting 90 as an allowance and his placement test mark is less than 50 the student are expecting to leave the college mostly with the reason of undesired to study. Moreover, male students with placement tests less than 50 and from low-income groups are expected to leave the college. Female students are also leaving the college if they are from low-income group and their placement test mark is less than 50 with the same reason undesired to study.

5. CONCLUSION

Data mining used in many different area health, commercial and education areas. Many studies were done in the education area to predict the performance of the student in some courses, predicting student retention and success. Different kinds of methods and algorithms were used to solve most of the issues in the education area. In order to identify the accurate result, the data must be collected for the long term as its recommended by many researchers. This research was investigating the use of a decision tree in predicting the student withdraw and find the most factors affect the student to leave in the early years. A set of factors found to affect the students in completing their studies like: undesired to study, financial issues since more students live in a hostel and social issues. The limitation of this study was the secondary school data was missing and the course performance in the first year also was missing. I believe that if this two-variable will help more in improving the student's retention.

REFERENCE LIST

1. Bylaska, W. P. (2015). *The impact of financial issues on the departure of first-year, first-time-in-college, full-time students from a public, four-year university* (Order No. 3689887). Available ProQuest Dissertations & Theses Global. (1678199469).
2. Christie, H., Munro, M., & Fisher, T. (2004). Leaving university early: Exploring the differences between continuing and non-continuing students. *Studies in Higher Education*, 29(5), 617-636..
3. Davies, R., & Elias, P. (2003). *Dropping out: A study of early leavers from higher education* London: Department for Education and Skills.
4. Marbouti, F., Diefes-Dux, H. A., & Madhavan, K. (2016). Models for early prediction of at-risk students in a course using standards-based grading. *Computers & Education*, 103, 1-15.
5. Seidel, E., & Kutieleh, S. (2017). Using predictive analytics to target and improve first year student attrition. *Australian Journal of Education*, 61(2), 200-218. doi: <http://dx.doi.org/10.1177/0004944117712310>
6. Alkhasawneh, R. (2011). *Developing a hybrid model to predict student first year retention and academic success in STEM disciplines using neural networks* (Order No. 3473939). Available from ProQuest Central; ProQuest Dissertations & Theses Global. (896615741).
7. Abeer Badr El Din Ahmed and Ibrahim Sayed Elaraby (2014) , Data Mining: A prediction for Student's Performance Using Classification Method, *World Journal of Computer Application and Technology* 2(2): 43-47
8. Al-Radaideh, Q. A., Al-Shawakfa, E. M., & Al-Najjar, M. I. (2006, December). Mining student data using decision trees. In *International Arab Conference on Information Technology (ACIT'2006)*, Yarmouk University, Jordan
9. Raju, D., & Schumacker, R. (2015). Exploring student characteristics of retention that lead to graduation in higher education using data mining models. *Journal of College Student Retention: Research, Theory & Practice*, 16(4), 563-591.