DETERMINATION OF FACTORS AFFECTED THE FOUNDATION STUDENTS IN WITHDRAWAL

Zuwaina Rashid Salim AL-Riyami[1] and Aiman Moayid[2]
[1]Department of Information System, University of Nizwa, Nizwa, 44464642@uofn.edu.om
[2]Dr., Department of Information System, University of Nizwa, Nizwa,
aiman.moyaid@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 research investigate the factors that affect the students in withdraw from colleges of technology in the first year. The study target the students in colleges of technology in foundation level. Data mining techniques will be used to achieve the research goals.

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 be 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 [4]. As well as Dheeraj Raju & Randall schumacker they build a model to explore the important characteristic associated.

Data mining techniques are 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's retention at higher education institutes. Predictive model was developing to predict student at risk [7] [8]. Farshid Marbouti states in his study that itis possible to identify at risk student early and both student and instructors can be notifying early. In this paper the researcher compares different comparative methods to identify at risk student in course using grading standard. 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 assessment may have great advantage in prediction model. Among the seven prediction methods used to show the accuracy and usability of course. The best method was found is naïve Bayes classifier, which identify 86.2% of student who failed in the course. Chi-squared automatic iterative detection (CHAID) decision tree model is another predictive method was study by [8]. He develops predictive model for predicting short term student attrition. With 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 to be treated. The predictive model ware built using five years of historical data from 2009 to 2013.

Using 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 [4]. The researcher was able to identify the most factors that impact the student success as well as the input impact the student retention. In this study neural network algorithm and genetic algorithm was 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 neural network, and in order to improve the accuracy of the model genetic algorithm was used. It found that high school preparation has great impact on student success in college. In addition, student academic and social adjustment to college significantly reflect student performance and retention in 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 same procedure and models with larger dataset and more participants to provide better reflection of URM STEM student first year retention behaviors. Also, this study could be extended by adding more precollege inputs to the model like family background, high school grade and financial aid background. This study was having many limitations. The first limitation is sample of quantitative part which was limited to first time first year students. Limited number of URM student (N=498) was include in this study that affect the network accuracy.

A huge data is stored in educational database that contain useful data that can help college management in extracting right decision. Course performance is one of the data which is stored in data warehouse that need 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 [6]. They found that large volume of records, documents and image is store in educational database. Data mining techniques used to extract pattern from this information that help the college management to predict the student performance. With the help of wake software, ID3 decision tree used to get the measure of information gain. It generate set of rules from 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 which needed special attention to reduce failing ration and taking appropriate action at right time. Qasem A. Al-Radaideh and Emad Al-Shawakfa was state in their study how to apply decision tree in mining student data [9]. Data mining process was used to enhancing the quality of the higher educational system by evaluating student data to study the main attributes that may affect the student performance in courses. This study built to help the students to predict the final grade in a course. Questionnaire was used in this study to collect the data from the undergraduate students. They used Weka toolkit to rank the 12 attribute they found it based on questionnaire. CRISP-DM (Cross Industry Standard Process for data mining) methodology was used to build classification model. They found the higher managements can use such classification model to enhance the courses outcome 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 previous 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 data mining technique. The methodology will be done in four stages, which is 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 accurate and correct result. It found that the dataset has missing data which need to be clean. In this stage some of the record was deleted because most of the field data is missing. Some of the missing values was filed manually by predicting the value based on the previous field. The data are analysis using decision tree to extract the rules and calculate the accuracy of the classes.

4. FINDING

The objective of this study is to predict withdrew in foundation level and the most factors affect the students to leave. The history data is collecting from CIMS system for the period of [2009 -2018]. The variable of this research is 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

Decision tree is used in this study based on the recommendation of the previous study [6] and as its help to extract the rules in easiest way. In addition to that it helps the decision maker to see the logical view of the dataset. CART algorithm is used in this research. This algorithm can handle missing data and can create generalized models. In decision tree each node represent the attribute and each branch represent a decision rules finally the leaf node represent a single outcome of the class. In this study the entropy function will used. The dataset is split into two set train and test as 70% in the training set and testing is 30%. The accuracy of this model is 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 high score in applying withdraw in the first year. From the rules if the student is male and getting 90 as an allowance and his placement teat 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 test less than 50 and from low income category are expect to leave the college. Female students are also leaving the college if they are from low income category and their placement test mark is less than 50 with the same reason undesired to study.

5. CONCLUSION

Data mining was used in many different area health, commercial and education area. Many studies were done in education area to predict the performance of student in some courses, predicting student retention and success. Different kind of methods and algorithms were used to solve most of the issue in education area. In order to identify the accurate result, the data must be collected for long term as itis recommended by many researchers. This research was investigating the use of decision tree in predicting the student

withdraw and find the most factors affect the student to leave in early years. Set of factors found 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. 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).

5. 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.

6. 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

7. 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.

8. 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

9. 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.