University Dropout Prevention through the Application of Big Data

Yeajou Shiau
Associate Professor
Zhaoqing University, China
+86 0758-2716325
shiau.yj@qq.com

ABSTRACT

This study explores the reasons for the suspension and dropout of full-time university students at a university in Taiwan and suggests better timing and strategies for student counseling. In this study, the narrative statistical analysis is used to analyze and discuss the sample objects, and then use data mining technology to find characteristic phenomena and classification conditions of the students who are suspended or dropout. Other studies related to dropouts rarely use leading indicators to predict the student dropout probability in real-time, most likely because of the timeliness and availability of student data. Therefore, this study proposes to use daily changes in absence indicators as predictive variables. Through the use of discriminant analysis to construct discriminant functions, the coefficient value of each student's withdrawal from university and early warning threshold for determining withdrawal from university can be presented in real-time in order to effectively provide the student with immediate counseling.

CCS Concepts

• Information systems→Information systems applications→Data mining→Association rules.

Keywords

Data mining, Decision tree model, Bayesian probability classification table, Counseling decision.

1. RESEARCH BACKGROUND AND MOTIVATION

According to the Taiwan Ministry of Education's April 2015 report on the first-year university freshmen forecast from 2016 to 2031 [1], the average number of first-year university students will decrease by 5,000 per year in the next 16 years or approximately 2.19% per year. Except for the growth in 2018, 2029 and 2030, the rest of the years experience a downward trend. The decrease in 2016, 2020 and 2028 was larger than average. They reduce by 15,000, 27,000 and 13,000 respectively compared with the previous school years. The report also pointed out that although higher education is becoming more and more popular, the corresponding school-age population is declining. In the past five years, colleges and universities (including the four-year and two-year specialist for the full-time students and vocational schools) have a shortage of 46,000 people. Under the influence of the phenomena of a declining birthrate, it is for everyone's attention to how to cope with the problem of the decrease in the number of students in colleges and universities in the future is really a concern for all schools.

According to the Taiwan Appraisal Association [2], Taiwan's higher education has implemented the idea of using big data or empirical information for professional management of university affairs under the challenges of international competition, declining birthrate impact, and globalization. At the same time, the coming crisis of lower enrollment also drives the trend of institutional research in Taiwan colleges and universities.

In fact, it is as important to avoid students abandoning their studies and dropping out of school as much as recruiting new students. Lin Dongqing pointed out that the word-of-mouth recommendation statistics [3], a satisfied customer will be recommended to five friends on average, and the recommendation has a high turnover rate. Therefore, satisfied former students are equal to the recruiters who are employed by the school, and they can promote to and serve as role models for the high school students. In many business management studies, it is pointed out that the cost of developing a new customer is five times that of retaining an old customer. An unsatisfied customer will tell his dissatisfaction to 8 to 10 people on average. Especially in the current age of internet information and network social community activities, word-of-mouth influence is more likely to cause indelible harm to schools, which in turn affects enrollment operations.

Therefore, it is very important for colleges and universities to build a dropout early warning and counseling mechanism to reduce suspension and withdrawal [4]. Due to the fact that the university students in the case universities have full-time and night divisions, the proportion of students in the night division accounts for less than 5% of the students, and most of them leave school due to work and military service issues. In order to explore the reasons for the suspension and withdrawal of the four-year university full-time students, and to find out the best counseling opportunities and target student strategies, the direction of this research was formulated.

1.1 Existing Counseling Mechanism for This Case University

This study focuses on student data from a technology university in Taiwan. This sample university has two campuses, namely, Wenshan Campus in Taipei and Hukou Campus in Hsinchu County. At present, there are three colleges, namely the "School of Planning and Design", "School of Management" and "School

IMMS 2020, August 7–9, 2020, London, United Kingdom

ACM ISBN 978-1-4503-7546-7/20/08…$15.00
DOI: https://doi.org/10.1145/3416028.3416029

of Information", with a total of 18 departments, 6 master programs, and a "General Education Center". Established in 1965 as a national junior college for the cultivation of youth, it is the first municipal vocational school in Taiwan. In order to cope with the changing times of the future, it has undergone periods such as municipal administration, business colleges, and technical colleges. Due to its excellent academic performance, it was restructured from the technical college to the university of technology in 2005, becoming the first private university of technology in Taipei.

In fact, the current university method of counseling students who have the possibility of being suspended or withdrawing from school has achieved remarkable results. In recent years, based on the analysis of key factors of suspension and withdrawal, the following measures have been taken to improve counseling measures:

1.1.1 Class academic advisor assistance

In the second week after the mid-term exam, online grading system provides the mid-term examination results for the academic advisor to inquire and provides a list of poorly-performing students as early warnings. The academic advisor assists in understanding the student's life or learning problems, and actively contacts the parents as needed to help solve the student's learning difficulties. In the 2018 school year, the scope of early warning for students was expanded. The threshold of failing 66% or more of their courses in the original mid-term exam was expanded to students failing 50% or more of their courses being included in the early warning list to strengthen counseling. In addition, the academic advisors also use social media platforms such as Facebook to build a student community to meet the interactive needs of the new generation of network social communities and to guide students to care about their work and learning through their peers. Since the 2011 school year, the university has cooperated with Far EasTone Telecommunications Co. to handle the "Academic Advisor Internet" providing a mobile phone contact system, and to strengthen the academic advisor contact function. This platform offers a schedule for the full-time teacher office hours which provides students with online appointments and remedial teaching after class.

1.1.2 Course teacher guidance

The remedial teaching methods and available times are clearly specified in the teaching standards and provide students who have fallen behind with remedial opportunities.

1.1.3 Parental care system

Developed in the 2008 school year, it provides a method of regularly interacting with parents, so that they can understand the learning dynamics of their students.

1.1.4 Student information system

Students can take the initiative to understand the dynamic information, such as: personal absence records, grades, and other information. Starting from the 2010 school year, the language and information self-study area is opened to provide students with self-study after class. After the mid-term examination, students with early warnings start remedial classwork, and the student participation rate has been over 80%.

The early warning counseling mechanism and results of students with poor academic performance in this sample university are detailed in Figure 1 and Table 1.

According to the selection of high-risk groups for dropping out of school and being suspended from school based on key factors, this university has turned passive remedial teaching into active preventive counseling to solve students' learning dilemmas. Counseling to improve learning outcomes was 81.4% effective in the 2014 school year and has increased to 87.4% in the 2015 school year. Obviously, the counseling program has begun to take effect. As they continue to identify the key factors of dropping out of school and being suspended from school, they will build early warning models, find students who need assistance, and provide counseling and remedial teaching to improve students' willingness to learn and improve their effectiveness.

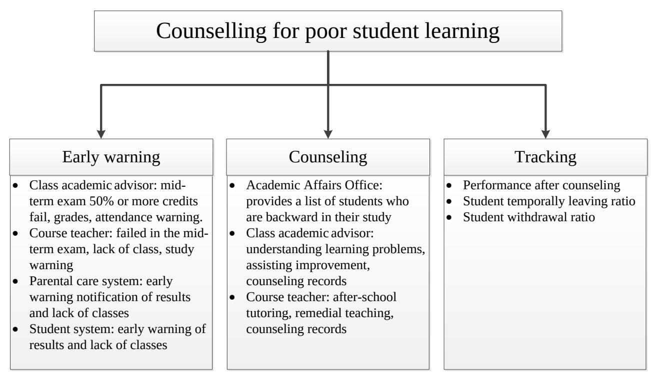

Figure 1. Counseling mechanism for poor learning

Table 1. Early warning counseling situation for students with poor learning outcomes in the 2012-2015 school year

School year	Number of students (Bachelor's degree, including the further education school) **(A)**	Number of students who are warned **(B)**	Number of students receiving early warning and counseling **(C)**	Ratio of students receiving counseling **(C/B)%**	Number of students who improved their learning after counseling **(D)**	Improvement ratio **(D/B)%**	Number of students who have withdrawn **(E)**	Withdraw ratio **(E/A)%**
2012	11,006	1,118	933	83.5	907	81.1	287	2.6
2013	10,974	1,235	1,030	83.4	1,016	82.3	145	1.3
2014	11,163	1,270	1,070	84.3	1,034	81.4	218	2.0
2015	11,507	1,306	1,218	93.3	1,141	87.4	283	2.5

1.2 Research Purposes

This study explores the reasons for the suspension and withdrawal of the four-year undergraduate students of this university through two steps, as well as the strategies for better counseling opportunities and better targeting of at-risk students. Firstly, it presents a descriptive statistical analysis and discussion of the sample objects and the characteristics of the at-risk students; such as, suspension, withdrawal, and classification conditions of students are identified through data mining techniques. Secondly, many studies on withdrawals rarely use leading indicators to predict the probability of dropping out of school in real-time. The reason for this is mostly the timeliness and availability of data acquisition. For example, if the student fails 50% or 67% of their midterm exams the semester has nearly ended. A failure of 50% at the midterm often leads to a failure of 67% of courses by semester end and results in student withdrawal. The data in this instance is already a backward indicator and cannot be an early warning to help students before dropping out of school. Therefore, this study proposes to use the daily variation of the absence index as a predictor variable. Through the use of differential analysis to construct a discriminant function, the coefficient value of each student's withdrawal probability and the threshold for determining the withdrawal are presented in real-time. In addition to cross-referencing with other current indicators, it is possible to effectively provide immediate advice to counselors. With limited manpower and material resources, the most counseling care can be placed on students with a higher possibility of dropping out of school in order to reduce the loss of students.

Finally, this study provides a comparison of the preventative student dropout counseling program to the current counseling strategies and discriminant functions.

Based on the above, the research objectives of this study are divided into the following five points:

1. Provide a reference for decision-making conditions for suspension from school and withdrawal from school through a decision matrix.
2. Use the decision tree model to construct rules for dropout conditions, to establish early warning module, and to fine-tune the early warning model.
3. A Bayesian probability classification table is used to provide the probability that students from different origins will be suspended in order to construct counseling and decision-making decisions for students of various types in different grades.
4. Form a discriminant function through the daily change of the leading indicator of attendance rate to make a calculation with discriminant analysis and provide instructors with real-time early warning and counseling for students who are at-risk of withdrawal.
5. Compare the differential analysis and prediction results with the current counseling strategies to find better counseling opportunities and student-targeting strategies to provide supervisors with a preventative student withdrawal counseling program.

2. LITERATURE REVIEW

Because almost 100% of high school students in Taiwan can enter university, many students enter who cannot adapt to university courses, exams, assignments, etc. Compared with other countries, the suspension and dropout rates of universities in Taiwan are very high, so it is very important to establish a more convenient early warning suspension or dropout mechanism.

An at-risk early warning system can bring potential withdrawal students back to successful path as early as possible, and guide them towards a graduation path [5]. Because the method is effective, the State of Wisconsin has developed a Dropout Warning System (DEWS) to predict students at-risk of school dropout [6], and the Department of Education and Early Childhood Development in Victoria, Australia has developed a Student Mapping Tool (SMT) to help schools identify students at-risk of dropping out [7].

Nandeshwar et al. found that the following factors are highly informative in predicting school dropout: family background, family social-economic status, and high school GPA and test scores [8]. Saqr et al. found that student commitment and consistency in the use of online resources is the most important factor in predicting dropout risk [9]. Tan and Shao selected student's personal characteristics and academic performance as input attributes and developed predictive models using ANN, Decision Trees (DT) and Bayesian Networks (BN) [10].

Wang Jianhua used decision tree to analyze absenteeism and academic performance, analyzed the behavioral characteristics of college students, and compared the analysis of personality tests to explore the relationship between absenteeism and academic performance of dropout students. His research found that students who dropped out were severely absent from school and had poor academic performance [11].

Zhao Ruizheng studied the issue of how to reduce the loss of students through the data mining technology [12]. Through the use of decision trees and neural network technologies, students' loss prediction models are constructed. His research found that the C5.0 decision tree algorithm performs best in predicting student withdrawal or suspension. He further found that unregistered students, late registrations, and work factors were the main factors for student withdrawal and suspension. In addition, Chen Tingbin and Gao Shuzhen used the ID3 algorithm to investigate the patterns for the suspension and withdrawal of students from a university in southern Taiwan, and finally obtained 231 prediction rules [13].

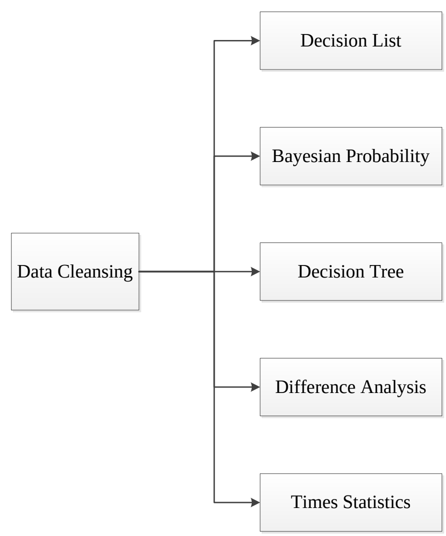

Figure 2. Data mining and statistical analysis methods used in this research

According to the research purpose, the data mining and statistical analysis methods used in this research are shown in Figure 2. This study processed relevant variables and used statistics through decision matrices, Bayesian probability classification tables, decision tree models, discriminant analysis methods, and reasons for suspension and withdrawal to analyze and provide an instant warning mechanism for reference by decision-making units.

3. RESEARCH OBJECT

The objects and related information of this study are as follows:

1. Research object:

The research object of this study is full-time, four-year students of the Taipei campus of this sample university.

Source: Case University Institutional Research Office.

Date of integration: May 20, 2016

Scope of data: Information on students at the Taipei Campus in the second semester of the 2014 academic year. In total, there were 5,375 students in the second semester, 5,157 students in the university, 130 students dropped out, and 88 students were suspended.

2. Study period:

This study was conducted from May 2016 to April 2017.

3. Data collection and research tools:

The data of the full-time students of the sample university after admission to the university are very detailed. This study divides these data into three types of impact variables, which are: categorical variables, numerical variables, and dates. The relevant aggregations are shown in Table 2.

4. This study has the following limitations:

(1) The scope of the study is limited to the full-time four-year students of the Taipei campus of the sample university. The students in the Hsinchu campus, evening students, and other education systems are not included in this study.

(2) If a university student chooses to suspend or retake the admission exam, he or she will not necessarily notify the university with the real reason on the suspension form, so it is not easy to know the real reason. This student is often classified as "overdue registration, suspended overdue, or not resumed", so the number of people meeting these conditions will be higher. It is not easy to apply corresponding counseling measures and strategies because of unreliable reasons.

Table 2. Consolidated table of variables affecting student suspension and withdrawal

<table>
<tr><th>Variable name</th><th>Nature</th><th>Update frequency</th><th>Usage analysis / application</th></tr>
<tr><td>Admission channels, household registration, gender</td><td rowspan="2">Categorical variable</td><td>fixed</td><td rowspan="4">1. Decision matrix / Provision of suspension and withdrawal rates when the conditions for suspension and withdrawal are met
2. Bayesian probability classification table / Probability statistics of students from different sources in different grades
3. Early warning module for decision tree model / Provide early warning module for decision tree rules, provide early warning list every semester
4. Discriminant analysis / Form a discriminant function through leading indicators of daily change in attendance, make calculation forms to provide instructors to pre-judge real-time early warning counseling students
5. Statistics on the number of reasons for suspension and withdrawal /Reasons for suspension and withdrawal to provide reference for decision-making units</td></tr>
<tr><td>• Year
• Whether tuition reduction student
• Whether to apply for a school loan
• Whether living for rent
• Status (at school, withdrawal, suspension)
• Reason for withdrawal or suspension</td><td>every semester</td></tr>
<tr><td>Average academic grades last semester
Scores earned last semester
Last semester credit</td><td rowspan="2">Numerical variable</td><td>every semester</td></tr>
<tr><td>Absence status (late, cut classes, official leave, personal leave, sick leave, funeral leave, marriage leave)</td><td>every day</td></tr>
<tr><td>Others (nationality, campus, division, college, department, identity, admission results, school year, birthday ...)</td><td>categorical variables, numerical variables, and dates</td><td></td><td>Reasons not used in this analysis
1. The student may transfer to a department or transfer exam or for other reasons
2. Too many missing data
3. Exclude insignificant impacts
4. Analysis methods or issues assumptions are inappropriate
5. No management implications or difficult to explain</td></tr>
</table>

4. RESEARCH RESULTS AND ANALYSIS

4.1 Analysis Results of Decision Matrix

Through the analysis of the decision matrix, we provide the probability of the suspension and withdrawal when the conditions for the suspension and withdrawal are established, and the results are as follows:

1. List of suspension decisions:

(1) When the academic average performance in the previous semester is ≦ 63.60, the probability of suspension of school this semester is 10.17%.

(2) When the number of absences in the previous semester is >28 class periods, the probability of suspension in the next semester is 3.68%.

2. List of withdrawal decisions:

(1) When the number of sick leave days in the previous semester is >21, and the average academic performance in the previous semester is ≦68.78, the probability of withdrawal is 16.78%

(2) When the average academic performance in the previous semester is ≦63.10, the dropout rate in the next semester is 11.33%.

4.2 Bayesian Probability Classification Table Analysis Results

Table 3. The conditional probability table of students' admission channels and suspension and withdrawal

Different admission channels / in school or suspension		Occurrence rate by year			
Admission channel	Status	1st year	2nd year	3rd year	4th year
Others	in school	0.11	0.53	0.32	0.04
	suspension	0.00	0.44	0.44	0.11
Excellent technical and artistic ability	in school	0.30	0.23	0.25	0.22
	suspension	0.44	0.22	0.22	0.11
Apply for admission	in school	0.28	0.24	0.27	0.21
	suspension	0.25	0.50	0.25	0.00
Joint selection	in school	0.34	0.29	0.20	0.17
	suspension	0.34	0.47	0.19	0.00
Joint registration distribution	in school	0.26	0.21	0.26	0.27
	suspension	0.44	0.35	0.12	0.09

According to the probability diagram of a Bayesian condition of suspension:

1. "Other" admission channel: the third year is the highest rate of suspension for accounting 44% of the number of suspensions for this admission channel.
2. "Excellent technical and artistic ability" admission channel: The first year is the highest rate of suspension for accounting 44% of the number of suspensions for this admission channel.
3. "Apply for admission" admission channel: The second year is the highest rate of suspension for accounting 50% of the number of suspensions for this admission channel.

The conditional probability table of students' admission channels and suspension and withdrawal is shown in Table 3.

4.3 Decision Tree Analysis Results

The decision tree model results show:

1. There are 185 students with an average academic performance in the previous semester of ≦53.33, of which 67 students dropped out, accounting for 36.24% of the number of students dropped out.
2. For seniors in years three and four with the average academic performance in the previous semester of ≦41.39, of which 73.18% will drop out.
3. Among the first and second year students, if the average academic performance in the previous semester is ≦45.6, and those students pay full tuition, 74.49% of students will drop out.

4.4 Discriminant Analysis Results

Using the discriminant analysis method, the discriminant function is formed through the daily change of the absentee rate- the leading indicator, and the calculation form is provided to the instructor to identify the real-time early warning of dropout in order to counsel the students.

The analysis results are displayed based on the analysis function of all students' differences:

F(withdrawal) Difference function = Late *-0.292+ Cut classes *0.434+ Official leave *-0.099+ Personal leave *0.353+ Sick leave *0.690+ Funeral leave *-0.023 ……………………………………………………..…..(1)

The central value of students is -0.057, the central value of dropout students is 2.276, and the hit rate is 88.58%.

Among them, the slope (impact) of sick leave is the highest, indicating that students who have dropped out of the university in practice will be shown through sick leave indicators, followed by cut classes; this part also reflects the influence of the university's daily supervisor of the student leave system, who records students who cut classes often into sick leave requests.

However, due to the different status of compulsory, elective, and internship courses in each grade, this analysis further analyzes each years' students differently for the use of academic advisors of different years. The results are as follows (the values of the enrollment and withdrawal centers of each grade in parentheses):

F (1st year dropout) =Late*-0.307+ Cut classes *0.508+ Official leave *-0.108+ Personal leave *0.038+ Sick leave *0.749+ Funeral leave *-0.007 (in school=-0.062，out of school=2.480) ……………………………….……..…………(2)

F (2nd year dropout) = Late *-0.314+ Cut classes *0.373+ Official leave *-0.164+ Personal leave *0.488+ Sick leave *0.654+ Funeral leave *-0.028 (in school =-0.096，out of school=1.803)…….……………………………..…………..…..(3)

F (3rd year dropout) = Late *-0.124+ Cut classes *0.167+ Official leave *-0.074+ Personal leave *0.388+ Sick leave *0.852+ Funeral leave *-0.042 (in school =-0.035，out of school =3.281)…….………………………..…….……..………(4)

F (4th year dropout) = Late *-0.342+ Cut classes *0.701+ Official leave *0.045+ Personal leave *-0.032+ Sick leave *0.622+ Funeral leave *-0.094 (in school =-0.023 , out of school =3.180)……………………………………..…………..……...(5)

This research produces a form for academic advisors to determine the drop-out in advance and provides a real-time warning to assist students in the calculation form shown in Figure 3. Each input field is multiplied by the application coefficient and added up to get the total score of the discriminant function, and then the dropout threshold is compared to 2.2186. The higher the score, the higher the probability of dropout.

This research further uses the data sources of the 2014 academic year to verify the actual data of the 2015 academic year with the results of the discriminant function analysis. The description of the data is as follows: On November 29, 2016, a total of 5877 students were enrolled full-time at the Taipei campus. (Of which 2015-1 209 students dropped out, 2015-1 56 students suspended, 2015-2 104 students dropped out, 2015-2 37 students suspended). After data validation, there were 5,612 students enrolled in the second semester of 2015. A total of 5,525 students were put into the identification and prediction system.

The following results were obtained:

1. The system predicts that the difference function threshold is 2.2186, and there are 783 students who will dropout.
2. Less than the differential function threshold of 2.2186, the system predicts that there will be 4720 students who will not drop out.
3. The system predicts 734 students who are at risk without dropping out of school (wrong) and actually withdrawals of 49 students (correct).
4. The system predicts that 4715 students who are not at risk and no withdrawal (correct), and 5 students actually withdrawn (wrong).
5. Hit rate is (49 + 4737) / 5525 = 86.62% (88.58% for the 2014 academic year).
6. The accuracy rate is as high as 85% or more for two consecutive years.

Differential function menu for early warning students to drop out of school						
	2014-1 Late	2014-1 Cut classes	2014-1 Official leave	2014-1 Personal leave	2014-1 Sick leave	2014-1 Funeral leave
Normalization coefficient	-0.292	0.434	-0.099	0.353	0.69	-0.023
Structure matrix	-0.062	0.601	-0.093	0.427	0.813	0.001
Canonical coefficient	-0.169	0.04	-0.028	0.087	0.055	-0.014
Application coefficient	-0.169	0.04	-0.028	0.087	0.055	-0.014
Please enter the student information	5	15	5	10	20	0
					In school	-0.057
					Out of school	2.276
F(dropout) Discriminant function=	0.896					
The threshold for difference in out of school scores is	2.2186					
If the obtained value exceeds the threshold, please counsel the student immediately and fill in the counseling record						

Figure 3. Calculation form for real-time early warning counseling to identify dropouts

5. CONCLUSION AND SUGGESTIONS

5.1 Conclusion

According to the sample university data, the reasons for the dropout in the 2014 school year show: (1) For the reasons of 88 suspension students, the highest proportion of majors differ from interests is 28 (32%); the second is 18 students for part-time employment (20%), and the third is 10 students who are not adapted to life (11%). (2) The reasons for 130 students who dropped out were at most 2/3 of the credits failed, a total of 87 students accounted for 67%; followed by 34 transfers students, 5 unregistered students after the due date, over 3 students with majors differ from interests, and only 1 student unfit for life. This study draws the following conclusions:

1. According to the statistics from the Statistics Department of the Ministry of Education, in the 2014 academic year, the suspension and drop-outs of full-time students in universities indicate that the highest reasons for university students' suspension are other reasons, followed by academic interest, and the third is the part-time employment. The sample university in this study excludes "other reasons" that cannot be studied. Among the reasons for the suspension of full-time students in the Taipei campus in the 2014 academic year, the highest proportion of students with majors that differ from interests accounted for 32%; the second was part-time employment of 20%, and the results were quite consistent.
2. According to statistics from the Statistics Department of the Taiwan Ministry of Education on the suspension and dropout of full-time students in universities in the 2014 academic year [14], it is pointed out that the highest reason for university dropouts is that they have not registered after the due date or have not returned to university after suspension, and their majors differ from their interests (31%) and academic performance (14%) are 2nd and 3rd respectively. After investigation and analysis, the full-time students of the sample university found that the largest number of dropouts were 2/3 credits fail, accounting for 67%, followed by transfer students with 26%, and the third was overdue registration. It is suggested that the sample university can provide counseling on academic performance, and through the decision tree and threshold analysis of differential analysis obtained in this study, they can formulate counseling plan strategies to identify students at-risk of dropping out.

5.2 Suggestion

1. The number of students in these statistics only accounts for 50% of the number of students in the sample university. It is recommended that the study can be expanded in the future to students in other academic systems and Hsinchu campuses and the information can be exchanged with other universities to obtain more data.
2. The current practice of counseling students in the sample university with the at-risk of suspension and withdrawal has been very effective. Follow-up with this study provides various decision-making, early warning, and distinguishing function tables, which can be more effective in time and accuracy.

6. REFERENCES

[1] Lin Xiaoli 2016. Forecast report on the number of first-year students in colleges and universities (2016 - 2031 academic year). Department of Statistics, Taiwan Ministry of Education.

[2] Wu Shuyuan 2016. *School Research-Explore the implications behind the data*. Review Bimonthly, November 2016, 64.

[3] Lin Dongqing 2016. *Information management: the core competitiveness of e-businesses*. Zhisheng Publishing House, Taipei, August 2016.

[4] Balfanz, R.; Herzog, L.; Maciver, D.J. *Preventing Student Disengagement and Keeping Students on the Graduation Path in Urban Middle-Grades Schools: Early Identification and Effective Interventions*. Educ. Psychol. 2007, 42, 223–235.

[5] Dynarski, M.; Gleason, P. How Can We Help? What We Have Learned From Recent Federal Dropout Prevention Evaluations. J. Educ. Stud. Placed Risk (JESPAR) 2002, 7, 43–69.

[6] Knowles, J.E. Of needles and haystacks: Building an accurate statewide dropout early warning system in Wisconsin. JEDM J. Educ. Data Min. 2015, 7, 18–67.

[7] Lamb, S.; Rice, S. Effective Strategies to Increase School Completion Report: Report to the Victorian Department of Education and Early Childhood Development; Communications Division, Department of Education and Early Childhood Development: Melbourne, Australia, 2008.

[8] Nandeshwar, A.; Menzies, T.; Nelson, A. Learning patterns of university student retention. Expert Syst. Appl. 2011, 38, 14984–14996.

[9] Tan, M.; Shao, P. Prediction of student dropout in E-learning program through the use of machine learning method. Int. J. Emerg. Technol. Learn. 2015, 10, 11–17.

[10] Saqr, M.; Fors, U.; Tedre, M. How learning analytics can early predict under-achieving students in a blended medical education course. Med Teach. 2017, 39, 757–767.

[11] Wang Jianhua 2004. *Application of data mining technology to tutoring students who leave the school halfway through colleges*. Master Thesis, Institute of Defense Information, National Defense University, Taipei.

[12] Zhao Ruilin 2007. *Research on Reducing Student Attrition with Data Mining Technology*. Master Thesis. Institute of Information Management, Yunlin University of Science and Technology, Taiwan

[13] Chen Tingbin and Gao Shuzhen. Exploring the rules of student withdrawal from school using the ID3 algorithm—Taking a university as an example. "The 23rd International Symposium on Information Management", May 2012, Kaohsiung University, Taiwan

[14] Department of Statistics, Taiwan Ministry of Education 2014. *Overview of the suspension of college and university students in the 2014 academic year.* Retrieved from http://stats.moe.gov.tw/files/chart/103