

ID: 2016-ISFT-438

The Predictions of the Termination of Undergraduate Students of Factor Analysis with a Logistic Regression and Factor Analysis with Fuzzy Logic

Anutita Lekpech¹, Chanida Kaewphet², Supakij Puaprasert³

^{1,2,3}Faculty of Business Administration and Information Technology, Rajamangala University of Technology Suvarnbhumi, 450 Moo.6 Suphanburi-Chainat Street, yanyao, Samchuk, Suphanburi, 72130, Thailand ¹anutita.1@rmutsb.ac.th

Abstract: This paper was to forecast 199 undergraduates dismissal during academic year 2551-2554, the faculty of information technology, Rajamangala University of Technology Suvarnabumi, Suphanburi Campus with 31 variables using personal, family, financial and learning data and analyzed factors influencing their dismissal. Correlation analysis with 31 variables had showed that there were 17 influencing variables including personal, family, financial and learning data. We, then, analyzed and classified their components into five factors including learning data from the former school, learning data at the new school, rank in the family and number of sibling, father's education background, and gender. These 5 factors were used to forecast student's dismissal with logistic regression analysis compared with the fuzzy logic. Testing showed that logistic regression analysis could forecast correctly at 92.46 percent and fuzzy logic could forecast correctly at 91.46 percent.

Keywords: Correlation Analysis, Factor Analysis, Logistic Regression, Fuzzy Logic

1. INTRODUCTION

Department of business information technology, the faculty of business administration and information technology, Rajamangala University of Technology Suvarnabumi, Suphanburi Campus had been providing undergraduate degree teaching and learning since 2005 up to present time, and there have been 836 graduated students. In each academic year, the department admitted 90 freshmen and found many dismissed students each year. It was about 40.91 percent of students dismissed during academic year 2005 to 2011 caused by different reasons like dismissal further to measurement or financial regulation, resign and death. Students' dismissal tended to increase continuously and this affected to the department teaching and learning planning, determining numbers of instructors and budget provided for educational management. It also influenced graduate production in term of administration and management of the department which interested the researcher to investigate what factor influencing their dismissal in order to find the solution. The research aimed to study about students' personal, family, financial and learning data as shown in table 1 to determine influencing factors towards students' dismissal and develop its forecast model by using 199 undergraduates' data during academic year 2008-2011. Among these students, 66 students were dismissed and the numbers tended to increase continuously. Classified influencing factors showed dismissal and status maintenance which made data classification easier and were applied for correlation and factor analysis to determine influencing factors towards students' dismissal. Logistic regression analysis compared with fuzzy logic was applied to develop students' dismissal forecast model [1-8].

TABLE 1: Show Independentvariable is used to study 31 variables.

List	Variable				
Profile	Sex, Size of old school, Education Level graduate, Major graduated				
Family Information	Number of brothers and sisters, Number of brother and sister is studying, Birth order, Staying with, Housing characteristics, Status of Parents, Education of father, income/year of father, job of father, Education of Mother, income/year of Mother, job of mother				
Financial Information	The average cost / month, Source subsidized education				
Grades Information	Thai, math, science, religion and culture, health and physical, art, career and technology, Foreign language, GPA old school, Computer mathematic, Operating System, Computer Programming 1, GPS Term 1				

2. MATERIALS AND METHODS

Compare forecasting termination of undergraduates from technical Factor Analysis with logistic regression And Factor Analysis with fuzzy logic. The research process is as follows:



Fig. 1. A method to forecast the undergraduate students.

2.1 CORRELATION ANALYSIS

The results are shown in Figure 3.

Table 1 taken from the variable Profile, family information and financial information 18 variables were analyzed correlations **Spearman Rank Correlation** is a variable scale Ordinal scale.

The results are shown in Fig. 2.

	sex	Count sister	order born	Wutti Study Father	status student
sex	1.000				
Count_sister	-0.053	1.000			
order born	0.008	0.676**	1.000		
Wutti Study Father	0.030	0.090	0.009	1.000	
status student	-0.176*	0.179 *	-0.143*	0.182**	1.000

Fig. 2. The results of Spearman Rank Correlation

Table 1 taken from the variable Grades information 13 variables were analyzed correlations **Point Biserial Correlation** is a variable scale the Interval scale or Ratio Scale.

				Social			Work		GPA school		Operating	Computer		Status
	Thai	Math	Science	and	Physical	Art	and	Language	Old	Commath	System	Programming	GPS Term 1	student
				Culture			Technology				System	riogramming		Judent
Thai	1.000													
Math	0.448**	1.000												
Science	0.609**	0.655**	1.000											
Social and Culture	0.577**	0.529**	0.768**	1.000										
Physical	0.513**	0.427**	0.678**	0.818**	1.000									
Art	0.436**	0.366**	0.568**	0.784**	0.898**	1.000								
Work and Technology	0.456**	0.420**	0.782**	0.699**	0.635**	0.550**	1.000							
Language	0.509**	0.450**	0.686**	0.837**	0.807**	0.792**	0.674**	1.000	0					
GPA school Old	0.703**	0.660**	0.884**	0.905**	0.863**	0.799**	0.801**	0.858*	* 1.000					
Commath	0.203**	0.137	0.268**	0.102	0.022	-0.048	0.281**	-0.00	9 0.159*	1.000				
Operating System	0.234**	0.231**	0.291**	0.194**	0.136	0.091	0.323**	0.206*	* 0.267**	0.196**	1.00	0		
Computer Programming	0.259**	0.276**	0.373**	0.270**	0.236**	0.137	0.350**	0.219**	* 0.331**	0.504**	0.568*	** 1.000		
GPS Term 1	0.341**	0.321**	0.476**	0.308**	0.221**	0.118	0.460**	0.208*	* 0.387**	0.685**	0.689*	* 0.792**	1.000	
Status student	-0.391**	-0.462**	-0.729**	-0.458**	-0.426**	-0.252**	-0.777**	-0.369*	* -0.610**	-0.422**	-0.360*	•• -0.447**	-0.577**	1.000

Fig. 3. The results of Point Biserial Correlation.

Results from the analysis of correlation Figure 2-3. Making variables were associated with the termination of the students have only 17 variables.

2.2 FACTOR ANALYSIS

Factor Analysis of 17 variable until only 5 factors. To include variables that are correlated in the same group.

The results are shown in Figure 4

Rotated Component Matrix"								
	Component							
	1	2	3	4	5			
Zscore: GPA_school_old	.965							
Zscore: Social and Culture	.921							
Zscore: Physical	.913							
Zscore: Language	.907							
Zscore: Art	.880							
Zscore: Science	.809	.330		.212				
Zscore: Work and Technology	.738	.340			254			
Zscore: Thai	.633	.216		.301				
Zscore: Math	.590	.235		.213				
Zscore: GPS_term1	.208	.934						
Zscore: Computer Programing1		.843						
Zscore: Operating System		.734		308				
Zscore: Commath		.705		.371	238			
Zscore: order_born			.932					
Zscore: count_sister			.922					
Zscore: sex				.859				
Zscore: wutti_study_father					.933			

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. ^a

a. Rotation converged in 5 iterations.

Fig. 4. The results of Factor Analysis.

Component 1 = Factor 1 Information about the results of the original study. {GPA old school, religion and culture, health

and physical, foreign language, art, science, career and technology, Thai, math}

Component 2 = Factor 2 Information about the results of the new study. {GPS term1, Computer Programming 1, Operating System, Computer Mathematic}

Component 3 = Factor 3 birth order and number of siblings. {Birth order, Number of brothers and sisters}

Component $4 = Factor 4 sex. \{Sex\}$

Component 5 = Factor 5 Education of father. {Education of father}

2.3 FORECASTING TERMINATION OF UNDERGRADUATE STUDENTS THE LOGISTIC REGRESSION METHOD

Factors affecting the student termination 5 factors predictions of the end of undergraduates using Binary Logistic Regression, because the dependent variable (status student) are two of the termination and untermination. Using selected variables into the equation, the **Enter** detailed below. The results are shown in Table 2

FABLE 2: Results	from logistic	regression	analysis 5 Factor
-------------------------	---------------	------------	-------------------

Factor	В	S.E.	Wald	df	Exp(B)	р		
Fac1	-3.377	.740	20.797	1	.034	.000		
Fac2	-3.240	.588	30.409	1	.039	.000		
Fac3	.058	.267	.047	1	1.060	.828		
Fac4	769	.311	6.100	1	.464	.014		
Fac5	1.021	.333	9.413	1	2.776	.002		
Chi-square (Omnibus test	Chi-square (Omnibus test of Model Coefficients) Model = 171.362, df = 5, Sig = .000							
Chi-square (Hosmer and Lemeshow Test) = 7.586, df = 8, Sig = .475 Constant = -2.693								
2 Log likelihood = 81.506	2 Log likelihood = 81.506, Cox & Snell R^2 = .577, Psecudo R^2 (Nagelkerke R^2) = .803							

Method Enter Correct 92.46%

Model Logistic Regression Analysis (Sawadchai, 2548)

Prob (event) =
$$\frac{e^{\beta_0 + \beta_1 X}}{1 + e^{-(\beta_0 + \beta_1 X)}}$$

or Prob(event) =
$$\frac{1}{1+e^{-(\beta_0+\beta_1X)}}$$

When β_0 and β_1 the coefficient is estimated from the data.

 \boldsymbol{X} The independent variables

e Natural Logarithms) Estimated 2.718

From the above equation. We can write a new equation.

$$Prob(event) = \frac{e^z}{1+e^z}$$

(1)

or Prob(event) = $\frac{e^z}{1+e^{-z}}$		2)
By $z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$ The prosp	pect of non-event is estimated from the equation.	
Prob (no event) = 1-Prob (event)	(3)
From the Table 2 can be represented by Equation 3.		
	e ^z	

Prob (The termination of student) = $\Pr{ob}($ การพ้นสภาพนักศึกษา) = $\frac{e}{1+e^{-z}}$

Z = -2.693 + 1.021 (Fac5) - 0.769 (Fac4) + 0.058 (Fac3) - 3.240 (Fac2) - 3.377 (Fac1)

2.4 FORECASTING TERMINATION OF UNDERGRADUATE STUDENTS THE FUZZY LOGIC METHOD

Create fuzzy systems by using of the Membership Function Mamdani is Trapesoidal and Triangular.





Fig. 5. Show Scatter Plot Compare 5 Factor

TABLE 4: Statistical data shows the value of Class 0 and Class 1

Factor			aduation		Class 1 Undergraduate					
racioi	N	mean	stdev	minimum	maximum	Ν	mean	stdev	minimum	maximum
F1	133	0.33666	1.05541	-1.33660	2.80995	66	-0.67842	0.29127	-1.16324	0.52950
F2	133	0.36144	0.93559	-1.29356	2.70979	66	-0.72836	0.68014	-2.79690	1.05636
F3	133	-0.01158	1.01342	-1.78728	3.02041	66	0.02334	0.97964	-1.25264	5.75558
F4	133	0.13164	0.97679	-2.15667	1.82409	66	-0.26527	1.00082	-2.38620	1.40094
F5	133	-0.14757	0.91877	-2.10536	2.42834	66	0.29737	1.09433	-1.72208	2.63495

Scatter Plot from Figure 5 and the statistics in Table 4.The researchers designed a fuzzy set (Linguistic Terms) in each dimension follows.

The Predictions of the Termination of Undergraduate Students of Factor Analysis with a Logistic Regression and Factor Analysis with Fuzzy Logic



Fig. 6. The design of fuzzy sets of dimensions 6 variable output status students.

Dimension 1-5 Factor1-5 divided into two fuzzy sets = {Low, High} Types of Membership function = {trapmf, trapmf} Dimension 1 Factor 1 Low Parameter = [-1.336 -1.336 -0.678 0.337] High Parameter = $[-0.678 \ 0.337 \ 2.81 \ 2.81]$ Dimension 2 Factor 2 Low Parameter = [-2.796 -2.796 -0.728 0.361] High Parameter = [-0.728 0.361 2.71 2.71] Dimension 3 Factor 3 Low Parameter = [-1.786 -1.786 -0.012 0.023] High Parameter = [-0.012 0.023 5.756 5.756] Dimension 4 Factor 4 High Parameter = [-0.265 0.132 1.825 1.825] Low Parameter = [-2.386 -2.386 -0.265 0.132] **Dimension 5 Factor 5** Low Parameter = [-2.105 -2.105 -0.148 0.297] High Parameter = [-0.148 0.297 2.635 2.635] Dimension 6 Output divided into two fuzzy sets = {PASS, FAIL} Types of Membership function = {Triangular, Triangular} PASS = [-1 0 1] $FAIL = [0 \ 1 \ 2]$ Rules to be used for prediction by Fuzzy Logic 44 rules.

No	Rule
Rule 1	IF Factor1 is High and Factor2 is High and Factor3 is low and Factor4 is low and Factor5 is High Then Class is PASS (1)
Rule 2	IF Factor1 is low and Factor2 is High and Factor3 is low and Factor4 is High and Factor5 is High Then Class is PASS (1)
Rule 3	IF Factor1 is low and Factor2 is High and Factor3 is High and Factor4 is low and Factor5 is low Then Class is PASS (1)
Rule 4	IF Factor1 is low and Factor2 is High and Factor3 is low and Factor4 is low and Factor5 is High Then Class is PASS (1)
Rule 5	IF Factor1 is High and Factor2 is High and Factor3 is High and Factor4 is low and Factor5 is High Then Class is PASS (1)
Rule 6	IF Factor1 is High and Factor2 is High and Factor3 is low and Factor4 is High and Factor5 is low Then Class is PASS (1)
Rule 7	IF Factor1 is High and Factor2 is low and Factor3 is low and Factor4 is High and Factor5 is low Then Class is PASS (1)
Rule 8	IF Factor1 is High and Factor2 is High and Factor3 is low and Factor4 is High and Factor5 is High Then Class is PASS (1)
Rule 9	IF Factor1 is High and Factor2 is High and Factor3 is High and Factor4 is High and Factor5 is low Then Class is PASS (1)
Rule 10	IF Factor1 is High and Factor2 is High and Factor3 is low and Factor4 is low and Factor5 is low Then Class is PASS (1)
Rule 11	IF Factor1 is High and Factor2 is High and Factor3 is High and Factor4 is low and Factor5 is low Then Class is PASS (1)
Rule 12	IF Factor1 is High and Factor2 is low and Factor3 is High and Factor4 is High and Factor5 is low Then Class is PASS (1)
Rule 13	IF Factor1 is High and Factor2 is low and Factor3 is low and Factor4 is low and Factor5 is low Then Class is PASS (1)
Rule 14	IF Factor1 is High and Factor2 is low and Factor3 is High and Factor4 is low and Factor5 is low Then Class is PASS (1)
Rule 15	IF Factor1 is High and Factor2 is low and Factor3 is High and Factor4 is low and Factor5 is High Then Class is PASS (1)
Rule 16	IF Factor1 is low and Factor2 is low and Factor3 is low and Factor4 is low and Factor5 is low Then Class is FAIL (1)
Rule 17	IF Factor1 is low and Factor2 is low and Factor3 is low and Factor4 is High and Factor5 is High Then Class is FAIL (1)
Rule 18	IF Factor1 is low and Factor2 is low and Factor3 is low and Factor4 is low and Factor5 is High Then Class is FAIL (1)
Rule 19	IF Factor1 is High and Factor2 is low and Factor3 is High and Factor4 is High and Factor5 is High Then Class is FAIL (0.33)
Rule 20	IF Factor1 is High and Factor2 is low and Factor3 is High and Factor4 is High and Factor5 is High Then Class is PASS (0.67)
Rule 21	IF Factor1 is low and Factor2 is High and Factor3 is low and Factor4 is High and Factor5 is low Then Class is PASS (0.67)
Rule 22	IF Factor1 is low and Factor2 is High and Factor3 is low and Factor4 is High and Factor5 is low Then Class is FAIL (0.33)
Rule 23	IF Factor1 is low and Factor2 is High and Factor3 is High and Factor4 is High and Factor5 is High Then Class is FAIL (0.14)
Rule 24	IF Factor1 is low and Factor2 is High and Factor3 is High and Factor4 is High and Factor5 is High

TABLE 5: Show rules apply to forecast the termination of the student.

The Predictions of the Termination of Undergraduate Students of Factor Analysis with a Logistic Regression and Factor Analysis with Fuzzy Logic

No	Rule
	Then Class is PASS (0.86)
Rule 25	IF Factor1 is low and Factor2 is High and Factor3 is High and Factor4 is High and Factor5 is low Then Class is PASS (0.82)
Rule 26	IF Factor1 is low and Factor2 is High and Factor3 is High and Factor4 is High and Factor5 is low Then Class is FAIL (0.18)
Rule 27	IF Factor1 is low and Factor2 is low and Factor3 is High and Factor4 is low and Factor5 is low Then Class is PASS (0.13)
Rule 28	IF Factor1 is low and Factor2 is low and Factor3 is High and Factor4 is low and Factor5 is low Then Class is FAIL (0.88)
Rule 29	IF Factor1 is low and Factor2 is low and Factor3 is High and Factor4 is low and Factor5 is High Then Class is PASS (0.17)
Rule 30	IF Factor1 is low and Factor2 is low and Factor3 is High and Factor4 is low and Factor5 is High Then Class is FAIL (0.83)
Rule 31	IF Factor1 is low and Factor2 is low and Factor3 is low and Factor4 is High and Factor5 is low Then Class is FAIL (0.73)
Rule 32	IF Factor1 is low and Factor2 is low and Factor3 is low and Factor4 is High and Factor5 is low Then Class is PASS (0.27)
Rule 33	IF Factor1 is low and Factor2 is low and Factor3 is High and Factor4 is High and Factor5 is low Then Class is FAIL (0.6)
Rule 34	IF Factor1 is low and Factor2 is low and Factor3 is High and Factor4 is High and Factor5 is low Then Class is PASS (0.4)
Rule 35	IF Factor1 is High and Factor2 is High and Factor3 is High and Factor4 is High and Factor5 is High Then Class is PASS (0.92)
Rule 36	IF Factor1 is High and Factor2 is High and Factor3 is High and Factor4 is High and Factor5 is High Then Class is FAIL (0.08)
Rule 37	IF Factor1 is low and Factor2 is High and Factor3 is low and Factor4 is low and Factor5 is low Then Class is FAIL (0.17)
Rule 38	IF Factor1 is low and Factor2 is High and Factor3 is low and Factor4 is low and Factor5 is low Then Class is PASS (0.83)
Rule 39	IF Factor1 is High and Factor2 is low and Factor3 is low and Factor4 is High and Factor5 is High Then Class is FAIL (0.25)
Rule 40	IF Factor1 is High and Factor2 is low and Factor3 is low and Factor4 is High and Factor5 is High Then Class is PASS (0.75)
Rule 41	IF Factor1 is low and Factor2 is low and Factor3 is High and Factor4 is High and Factor5 is High Then Class is FAIL (0.83)
Rule 42	IF Factor1 is low and Factor2 is low and Factor3 is High and Factor4 is High and Factor5 is High Then Class is PASS (0.17)
Rule 43	IF Factor1 is low and Factor2 is High and Factor3 is High and Factor4 is low and Factor5 is High Then Class is FAIL (0.75)
Rule 44	IF Factor1 is low and Factor2 is High and Factor3 is High and Factor4 is low and Factor5 is High Then Class is PASS (0.25)

3. RESULTS

TABLE 6: Comparison of forecast analysis using logistic regression with fuzzy logic

The results of predictive	Logistic	Regression	Fuzzy Logic		
The results of predictive	Amount	Percent	Amount	Percent	
correct	184	92.46%	182	91.46%	
Mistake	15	7.54%	17	8.54%	
Total	199	100.00%	199	100.00%	

Vth International Symposium on "Fusion of Science & Technology", New Delhi, India, January 18-22, 2016

Regression analysis logistics can predict the termination of the students were required percentage 92.46 and use equations to predict only one equation in comparison to how fuzzy logic can predict the termination of the student's correct. 91.46 percent and the 44 Rules of prediction.

4. CONCLUSIONS

The results showed that factors affecting the termination of a student with 17 variables and apply Factor Analysis grouping variables that are related to the same group until only five factors is Information about the results of the original study, Information about the results of the new study, Birth order and number of siblings, sex, Education of father and factors were analyzed Regression analysis logistics can predict the termination of the students were required percentage 92.46 and use equations to predict only one equation in comparison to how fuzzy logic can predict the. Termination of the student's correct. 91.46 percent and the 44 Rules of prediction.

REFERENCES

[1] Hatchavanich, D. A Comparison of the Coefficient of Determination Measures for Binary Logistic Regression. Thai KMUTNB Journal 17 (2): 61-70.

- [2] Meesad, P. Fuzzy Systems and Neural Network. 1st edition. KMUTNB Book center, Bangkok
- [3] Thaiparnit, S. Panintorn Sueapae and Phayung Meesad. Apply regression analysis and fuzzy logic to forecast the student's status in curriculum of Business Information. PhD thesis. KMUTNB.
- [4] Sripanomtanakorn, S. Logistic regression analysis. source: http://www.thairenu.com/logistic.htm, October 15,
- Jiawiwatkul, U. Multiple regression analysis logistics source: http://rlc.nrct.go.th/download/Research_Zone/phase1 4_4.pdf, October 1,
- [6] Nguyen, H.A. First Course in Fuzzy logic. 3 rd. editon, CRC Press 2005.
- [7] Pongpatrakant P. An Application of Logistic Regression Analysis Technique in Predicting Undergraduate Student's Retirement. Computer Science, Loei Rajabhat University.
- [8] Pongpatrakant, P. Factors Analysis of Undergraduate Student's Retirement Using Filtering by Committee Machine. Computer Science, Loei Rajabhat University. Proceeding of NCCIT2010, 2010.