Analysis of the Timeliness of Graduation of FMIPA College KIP Students at Bengkulu University Using Binary Logistic Regression

Riki Crisdianto(1), Muhammad Fathi Abdillah Rakafalih(2), Alya Saputri(3), Laga Sopiansyah(4), Athaya Fairuzindah(5), Dyah Setyo Rini(6)

[1,2,3,4,5,6] S1-Statistics Study Program, Bengkulu University

Jl. WR. Supratman, Kandang Limun, Kec. Muara Bangka Hulu, Sumatera, Bengkulu 38371
e-mail: rikic628@gmail.com*(1), FathiFalih017@gmail.com(2), alyasaputri9a@gmail.com(3), laga9621@gmail.com(4), athayazindah@gmail.com(5), dyah.setyorini@unib.ac.id(6)

ABSTRAK

Penelitian-penelitian terdahulu yang membahas mengenai ketepatan waktu lulus mahasiswa pada umumnya lebih difokuskan pada populasi mahasiswa secara keseluruhan, tanpa mempertimbangkan pengaruh dari program beasiswa secara spesifik seperti KIP Kuliah. Penelitian ini bertujuan untuk mendapatkan model ketepatan waktu lulus mahasiswa dan mendapatkan faktor-faktor yang mempengaruhi ketepatan waktu lulus mahasiswa KIP kuliah FMIPA Universitas Bengkulu menggunakan metode Regresi Logistik Biner. Hasil dari penelitian ini diharapkan dapat memberikan informasi yang berharga bagi pengelola KIP kuliah Universitas Bengkulu dalam meningkatkan efektivitas program beasiswa KIP Kuliah dan membantu mahasiswa KIP kuliah FMIPA Universitas Bengkulu agar bisa lulus tepat waktu. Dalam penelitian ini digunakan metode Regresi Logistik Biner karena variable respon memiliki skala biner nominal. Berdasarkan hasil pembahasan penelitian yang dilakukan diperoleh model Regresi Logistik Biner dari ketepatan waktu lulus mahasiswa KIP kuliah FMIPA Universitas Bengkulu dengan faktor-faktor yang memiliki pengaruh signifikan terhadap ketepatan waktu lulus mahasiswa KIP kuliah FMIPA Universitas Bengkulu adalah Asal prodi S1-Fisika, S1-Biologi dan S1-Statistika dan IPK. Kemudian Hasil klasifikasi menggunakan Regresi Logistik Biner memiliki tingkat ketepatan klasifikasi sebesar 77.97%. Sehingga dapat disimpulkan bahwa dalam klasifikasi ketepatan waktu lulus mahasiswa FMIPA Universitas Bengkulu pada model Regresi Logistik Biner cukup baik.
Kata kunci: KIP Kuliah; Regresi Logistik Biner; *Odds Ratio*, Klasifikasi

ABSTRACT

Previous studies that discuss the timeliness of graduating students are generally more focused on the student population as a whole, without considering the influence of specific scholarship programs such as KIP Kuliah. This study aims to obtain a model of the timeliness of student graduation and obtain factors that affect the timeliness of graduating KIP Kuliah students of FMIPA Bengkulu University using the Binary Logistic Regression method. The results of this study are expected to provide valuable information for KIP Kuliah managers of Bengkulu University in improving the effectiveness of the KIP Kuliah scholarship program and helping KIP Kuliah students of FMIPA Bengkulu University to graduate on time. In this study, Binary Logistic Regression method is used because the response variable has a nominal binary scale. Based on the results of the discussion of the research conducted, a Binary Logistic Regression model of the timeliness of graduating KIP Kuliah students of FMIPA Bengkulu University is obtained with factors that have a significant influence on the timeliness of graduating KIP Kuliah students of FMIPA Bengkulu University is the origin of the S1-Physics, S1-Biology and S1-Statistics study programs and GPA. Then the classification results using Binary Logistic Regression have a classification accuracy rate of 77.97%. So it can be concluded that the classification of the timeliness of graduating students of FMIPA Bengkulu University in the Binary Logistic Regression model is good enough.
Keywords: *College KIP; Binary Logistic Regression; Odds Ratio, Classification*

INTRODUCTION

Higher education has an important role in developing quality and highly competitive human resources. The percentage of students graduating on time is one of the indicators of assessing the success and feasibility of a study program in the implementation of higher education [1]. The timeliness of student graduation has different criteria for each program available at the tertiary level. D3 (Diploma) program students are said to graduate on time if they can complete their studies in less than or equal to three years. Students of the S1 (Bachelor) program are said to graduate on time if they can complete their studies in less than or equal to four years. Students of the S2 (Master) program are said to graduate on time if they can complete a study of less or equal to two years and the S3 (Doctoral) program if they can complete a study of less or equal to three years [2]. The longer students complete their studies in college, the more they need a lot of money. One of the government programs that helps students in terms of costs to continue their education in higher education is the education fee for poor students with achievements (Bidikmisi) or what is now known as the Indonesia Smart College Card (KIP-K).

KIP-K is higher education assistance in the form of cash, expanding access, and learning opportunities from the government given to students who come from poor/vulnerable families to finance education [3]. The period of KIP-K provision for Diploma 3 programs is a maximum of 6 semesters, Bachelor programs are a maximum of 8 semesters and Doctoral programs are a maximum of 4 semesters. It is feared that if KIP-K students do not graduate on time or study beyond the KIP-K granting period, they will not be able to continue their studies due to financial constraints. So the timeliness of graduating KIP-K students is important and needs to be considered by study programs and universities in the implementation of higher education. This research is important to understand the factors that specifically affect the timeliness of graduating KIP-K students and can provide valuable information to related parties and also the students themselves, to improve the effectiveness of the KIP-K scholarship program.

Previous research conducted by Agwil, Fransiska, and Hidayati in 2020 discussed the timeliness of student graduation [2]. The research focused on the student population as a whole, without considering the influence of specific scholarship programs such as KIP-K. One of the public universities in Bengkulu province that has the trust to manage KIP-K is Universitas Bengkulu (UNIB). There are still many KIP-K UNIB students who do not graduate on time, especially in FMIPA UNIB. The method used to analyze the timeliness of graduating KIP-K students at FMIPA UNIB in this study is Binary Logistic Regression.

Binary Logistic Regression is one of the effective analysis methods in predicting binary variables, such as student graduation on time. So in this study, the Binary Logistic Regression method is applied in modeling the timeliness of student graduation where category 1 represents students who graduate on time (success) and category 0 represents students who do not graduate on time (failure). In addition, based on previous research related to the Binary Logistic Regression method, namely research by Nikie Ramsi Tamnge on the effect of service quality on student satisfaction at the Faculty of Teacher Training and Education, Muhammadiyah University of Surabaya which has a classification accuracy of 86% included in the Good Classification category [4].

LITERATURE REVIEW

Binary Logistic Regression

Logistic regression is a regression analysis technique used to explore the relationship between dichotomous response variables (have two categories) or polycotomous (have more than two categories) with a group of predictor variables that are continuous or categorical [5]. According to Hosmer & Lemeshow (2000) [6], Binary Logistic Regression is a statistical method used to find the relationship between the response variable (y) which has a nominal data scale (two categories or binary) with predictor variables (x) which are categorical or continuous.

The logistic regression model with p predictor variables is as follows:

$$\pi(x) = \frac{e^{(\beta_O+\beta_1 X_1+\cdots+\beta_p X_p)}}{1+e^{(\beta_O+\beta_1 X_1+\cdots+\beta_p X_p)}} \quad (1)$$

Description:

$\pi(x)$: probability of "success" or chance of a successful event

$X_1, \ldots, X_p$: Independent variable

β_O: Constant of the model

$\beta_O, \ldots, \beta_p$: Regression coefficient parameter.

The logit transformation model of $\pi(x)$ from the equation above can be written as follows:

$$g(x) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_O + \beta_1 X_1 + \cdots + \beta_p X_p \quad (2)$$

with g(X): the logit transformation of π(x).

Biner Parameter Estimation of Binary Logistic Regression Model

The Maximum Likelihood Estimation (MLE) method is used to estimate parameters in logistic regression. The MLE method was chosen because it has several advantages compared to other methods, including that it can be used to form non-linear models such as logistic regression, and the estimation results are unbiased [6]. Mathematically, the likelihood function $x_i, y_i)$ can be expressed [7]:

$$f(x_i) = \pi(x_i)^{y_i}[1-\pi(x_i)]^{1-y_i} \quad (3)$$

If each observation has been assumed to be an independent variable, the likelihood function is also a multiplication of each likelihood function, which is as follows:

$$L(\beta) = \Pi_{i=1}^{n} f(x_i) \quad (4)$$

By using the Maximum Likelihood method to estimate the Binary Logistic Regression parameters, the estimator is obtained $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k)$ by solving the system of equations $\Sigma_i y_i x_{ia} - \Sigma_i y_i x_{ia}\pi(x_i) = 0, a = 1, 2, \ldots, k$ with $\pi(x_i) = \frac{\exp(x_i\beta)}{1+\exp(x_i\beta)}$ using numerical methods.

Parameter Estimation Testing

The parameter estimation test aims to determine whether there is a significant relationship between the independent variables (parameters) and the dependent variable.

Simultaneous Test

The simultaneous test is used to test parameter estimates simultaneously (together). The test hypothesis is as follows:

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$

H_1 : There is at least one $\beta_j \neq 0; j = 1,2, \ldots, p$

Test statistics:

$$G = -2ln \frac{\left(\frac{n_1}{n}\right)^{n_1}\left(\frac{n_o}{n}\right)^{n_o}}{\sum_{i=1}^{n} \hat{\pi}_i^{y_i}(1-\hat{\pi}_i)^{1-y_i}} \quad (5)$$

Reject H_0 if $G > x^2_{(a,p)}$ with p number of predictor variables in the model [6].

Partial Test

Partial tests are used to test parameter estimates partially (separately). The test hypothesis is as follows:

$H_0 : \beta_j = 0$

$H_1 : \beta_j \neq 0; j = 1,2, \ldots, p$

Test statistics:

$$W^2 = \left[\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}\right]^2 \quad (6)$$

Reject H_0 if $W^2 > \chi^2\ (a, p)$ with p number of predictor variables in the model [6].

Fit Test

The Goodness of fit test is used to evaluate whether the model obtained is suitable or not with the observed data The test hypothesis is as follows [8]:

H_0 : Model is fit

H_1 : Model does not fit

Test statistics:

$$\hat{C} = \sum_{j=1}^{k} \frac{(o_j - n_j{}'\overline{\pi}_j)^2}{n_j{}'\overline{\pi}_j(1-\overline{\pi}_j)} \quad (7)$$

Reject H_0 if $\hat{C} > \chi^2\ (p-2)$ or $P - value < \alpha$.

Parameter Coefficient Interpretation

The parameter coefficient can be interpreted using the *odds ratio* (ψ) value. The definition of odds ratio for $x = 1$ dan $x = 0$ is as follows:

$$OR = \frac{\pi(1)/[1-\pi(1)]}{\pi(0)/[1-\pi(0)]} = \frac{\exp\ (\beta_O + \beta_1)}{\exp\ (\beta_O)} = \exp\ (\beta) \quad (8)$$

The odds ratio is a measure of association that can be interpreted broadly, especially in epidemiology. From the equation above, the odds ratio is the average of the tendency of the response variable to have a certain value if $x = 1$ compared to $x = 0$.[6].

Classification Accuracy

Classification accuracy of the model is used to determine whether the data is classified correctly or not [9]. The evaluation metrics used are APER and Accuracy. The APER value expresses the proportion of samples incorrectly classified by the classification function [10]. The Accuracy value expresses the proportion of samples that are correctly classified by the classification function. Calculating the APER value can use the confusion matrix as follows [11]:

Table 1. Contingency Table of Classification Accuracy

Actual Class	Prediction Class		Total
	Yes	No	
Yes	TP	FN	P
No	FP	TN	N
Total	P'	N'	P+N

$$APER = \frac{FP+FN}{P+N} \quad (9)$$

$$Accuracy = \frac{TP+TN}{P+N} = 100\% - APER \quad (10)$$

Smart Indonesia Card College (KIP-K)

According to [3], KIP-K is higher education assistance in the form of cash, expansion of access, and learning opportunities from the government provided to students from poor/vulnerable families to finance education. KIP-K aims to increase the economic potential and social mobility for students from poor/vulnerable families to attend college. KIP-K is an expansion of the Bidikmisi scholarship.

METHOD

Research Type and Data Source

The type of research used is quantitative. This method must also use quantitative tools in the form of the R program in processing the data. The data used is data on undergraduate students receiving Bidikmisi scholarships 2017-2019 FMIPA Bengkulu University. The type of data used is secondary data. Secondary data is data obtained from a previously available source. The data source comes from the Academic section of FMIPA Bengkulu University.

Research Variables

The variables used in this study are as follows:

Table 2. Research Variables

Label	Variable	Category	Scale
Y	Timeliness of graduation	0 = Not on time (> 8 semester) 1 = on time (≤ 8 semester)	Binary
X_1	Study Program	1 = S1 Mathematics 2 = S1 Chemistry 3 = S1 Physics 4 = S1 Biology 5 = S1 Statistics 6 = S1 Geophysics 7 = S1 Pharmacy	Nominal
X_2	Gender	0 = Female 1 = male	Nominal

X_3	District of Origin	0 = Outside Bengkulu City 1 = Bengkulu City	Nominal
X_4	School of Origin	1 = SMA 2 = SMK 3 = MAN 4 = Lainnya	Nominal
X_5	Entrance Path	0 = SBMPTN 1 = SNMPTN	Nominal
X_6	GPA Score	-	Numerical

Research Stages

Research activities outline starting from designing, collecting references and data, making analysis, making models, model testing, and interpretation. In detail, the stages of this research include activities:

1. Literature review and methodology exploration
2. Data collection.
3. Data exploration and categorization
4. Creating an initial model
5. Perform parameter significance testing
6. Create a new model with significant variables
7. Performing model fit test
8. Calculating and interpreting the odds ratio
9. Assessment of classification accuracy.

RESULT AND DISCUSSION

Initial Model of Binary Logistic Regression

The initial Binary Logistic Regression model obtained from the timeliness of graduating KIP students of FMIPA Bengkulu University is as follows:

$$\hat{g}(x) = -11.60552 + 0.24629x_{12} - 2.05500x_{13} - 1.65856x_{14} - 2.56339x_{15} - 1.09088x_{16} - 0.58078x_{17} - 0.83477x_{21} + 0.23606x_{31} - 0.41358x_{42} + 1.04611x_{43} + 0.04034x_{44} - 0.77481x_{51} + 3.75888x_6 \quad (11)$$

Parameter Estimation Testing

The parameter estimation test aims to determine whether there is a significant relationship between the independent variables (parameters) and the dependent variable.

Test Partial

The simultaneous test is used to test parameter estimates simultaneously (together). The test is as follows:

1) Hypothesis

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_8 = 0$

H_1 : There are at least $\beta_j \neq 0; j = 1,2, \ldots ,6$

2) Required quantities
   n = 168
   α = 5%
3) Test statistics

Table 3. G Test(simultaneously)

G	df	$x^2_{(a,p)}$
66.0853579	6	12.59159

4) Rejection criteria
   Reject H_0 if $p - value <$ α or $G > x^2_{(a,p)}$
   Accept H_0 if $p - value >$ α or $G < x^2_{(a,p)}$
5) Conclusion
   Based on the results of the G test in Table 3, it can be seen that the value of the G test on the model obtained is 66.0853579 with a free degree of 6. Because the value of $G = 66.0853579 > x^2_{(0,05,6)} = 12.59159$ then H_0 is rejected. This means that there is at least one predictor variable that is simultaneously significant to the graduation of KIP College students at FMIPA Bengkulu University.

Partial Test

Partial tests are used to test parameter estimates partially (separately). The test is as follows:

1) Hypothesis
   $H_0 : \beta_j = 0$
   $H_1 : \beta_j \neq 0; j = 1,2, \dots, p$
2) Required quantities
   n = 168
   α = 5%
3) Test statistics

Table 4. Wald Test (Partial)

Parameters	***Coef***	***S.E.***	***Wald Z***	***Pr(>|Z|)***
Intercept	-11.606	3.7045	-3.13	0.0017
X_{12}	0.2463	0.5952	0.41	0.679
X_{13}	-2.055	0.638	-3.22	0.0013
X_{14}	-1.6586	0.6551	-2.53	0.0114
X_{15}	-2.5634	0.8669	-2.96	0.0031
X_{16}	-1.0909	0.9151	-1.19	0.2332
X_{17}	-0.5808	1.338	-0.43	0.6642
X_{21}	-0.8348	0.5852	-1.43	0.1537
X_{31}	0.2361	0.4115	0.57	0.5662
X_{42}	-0.4136	1.278	-0.32	0.7462
X_{43}	1.0461	0.7402	1.41	0.1575

X_{44}	0.0403	0.6131	0.07	0.9475
X_{51}	-0.7748	0.43	-1.8	0.0715
X_6	3.7589	1.1124	3.38	0.0007

4) Rejection criteria

   Reject H_0 if $p - value <$ α or $W^2 > \chi^2\ (a, p)$

   Accept H_0 if $p - value >$ α or $W^2 < \chi^2\ (a, p)$

5) Conclusion

   Based on the Wald test results in Table 4, it can be seen that in some variables the $p - value < a = 0{,}05$, then H_0 is rejected. student graduation. The variables that have a significant effect on the timeliness of graduating KIP students at FMIPA Bengkulu University are the variable origin of study programs in categories 3, 4, and 5 and the GPA variable.

Binary Logistic Regression Model

Based on testing the significance of parameters both simultaneously and partially, it shows that the variables of Study Program Origin and GPA have a significant effect so in the new model the variables of Study Program Origin and GPA are included. The Binary Logistic Regression model obtained is as follows:

$$\hat{g}(x) = -12.76537 + 0.09213x_{12} - 1.78109x_{13} - 1.26683x_{14} - 2.16772x_{15} - 1.23262x_{16} - 0.75135x_{17} + 3.97552x_6 \quad (12)$$

Model Fit Test

The goodness of fit test is used to evaluate whether the model obtained is suitable or not with the observed data. The test is as follows:

1) Hypothesis

   H_0 : Model is fit

   H_1 : Model does not fit

2) Required quantities

   n = 168

   α = 5%

3) Test statistics

Table 5. Model fit test

$\widehat{C}$	Df	$x^2_{(a,p)}$	$P - value$
6.9512	2	15.50731	0.5419

4) Rejection criteria

   Reject H_0 if $p - value <$ α or $\hat{C} > x^2_{(a,p)}$

   Accept H_0 if $p - value >$ α or $\hat{C} < x^2_{(a,p)}$

5) Conclusion

Based on the model fit test on the model, it can be seen that the $\hat{C}$ or chi-square value is 120 with a free degree of 4. Because the value of $\hat{C} = 6.9512 < x^2_{(a,p)} = 15.50731$ and $p - value = 0.5419 > a = 0{,}05$, then H_0 is accepted. This means that the model used is appropriate.

Parameter Coefficient Interpretation

The interpretation of the parameter coefficient is to determine the functional relationship between the response variable and the predictor variable and define any changes in the response variable caused by the predictor variable. The odds ratio results are as follows:

Table 6. *Odds ratio value*

Variable	***Coef***
Intercept	$2.858049e-06$
X_{12}	1.096509
X_{13}	0.1684552
X_{14}	0.2817225
X_{15}	0.1144383
X_{16}	0.2915274
X_{17}	0.4717298
X_6	53.27807

Based on Table 6 above, it can be seen that the odds ratio value is obtained from the exponential value (β) interpretation that the odds ratio value of variable X_{12} as a student who comes from the S1-Chemistry Study Program tends of 1.096509 times compared to students who come from other study programs at FMIPA Bengkulu University to graduate on time. The odds ratio value of variable X_{13} as a student who comes from the S1-Physics Study Program tends of 0.1684552 times compared to students who come from other study programs at FMIPA Bengkulu University to graduate on time. The odds ratio value of variable X_{14} as a student who comes from the S1-Biology Study Program tends of 0.2817225 times compared to students who come from other study programs at FMIPA Bengkulu University to graduate on time. The odds ratio value of variable X_{15} as a student who comes from the S1-Statistics Study Program tends of 0.1144383 times compared to students who come from other study programs at FMIPA Bengkulu University to graduate on time. The odds ratio value of variable X_{16} as a student who comes from the S1-Geophysics Study Program tends of 0.2915274 times compared to students who come from other study programs at FMIPA Bengkulu University to graduate on time. The odds ratio value of variable X_{17} as students who come from the S1-Pharmacy Study Program tends of 0.4717298 times compared to students who come from other study programs at FMIPA Bengkulu University to graduate on time. The odds ratio value of variable X_6 is GPA, for a one-unit change in the GPA value, it can be said that if the GPA value increases by one unit, the odds ratio or the student's chance to graduate on time will increase by 53.27807 times.

So it can be concluded that the higher the GPA value, the greater the chance for students to graduate on time.

Classification Accuracy

The classification accuracy results of the model are as follows:

Table 7. Confusion Matrix

Actual Class	Prediction Class		Total
	1	0	
1	55	17	72
0	20	76	96
Total	75	93	168

$$APER = \frac{20 + 17}{168} = \frac{37}{168} = 0.220238$$

$$Accuracy = \frac{55 + 76}{168} = \frac{131}{168} = 0.7797619$$

The results of the APER calculation, it can be seen that the classification error value is 22.03%. From the accuracy calculation, it can be seen that the classification accuracy value is 77.97%. So it can be concluded that the Binary Logistic Regression model obtained in classifying the graduation time of college KIP FMIPA Bengkulu University students on time and not on time is good enough.

CONCLUSION

Based on the results of the discussion of the research conducted, a Binary Logistic Regression model of the timeliness of graduating KIP College students FMIPA Bengkulu University is obtained with variables that have a significant effect as follows:

$$\hat{g}(x) = -12.76537 + 0.09213x_{12} - 1.78109x_{13} - 1.26683x_{14} - 2.16772x_{15} - 1.23262x_{16} - 0.75135x_{17} + 3.97552x_{6}$$

Factors that have a significant influence on the timeliness of graduating Bidikmisi students of FMIPA Bengkulu University are the Origin of S1-Physics, S1-Biology, and S1-Statistics study programs and GPA. Then the classification results using Binary Logistic Regression have a classification accuracy rate of 77.97%. So it can be concluded that the Binary Logistic Regression model obtained in classifying the graduation time of college KIP FMIPA students at Bengkulu University on time and not on time is good enough.

ACKNOWLEDGMENTS

Thank you to the supervisors and the friends as well as relatives who have helped.

REFERENCE

[1] Badan Akreditasi Nasional Perguruan Tinggi, "Naskah Akademik IAPT 3.0.," BANPT, Jakarta, 2019.

[2] W. Agwil, H. Fransiska and N. Hidayati, "Analisis Ketepatan Waktu Lulus Mahasiswa Dengan Menggunakan Bagging CART," *Jurnal Pendidikan Matematika dan Matematika*, vol. 6, no. 2, pp. 155-166, 2020.

[3] Pusat Layanan Pembiayaan Pendidikan, "Pedoman Pendaftaran Kartu Indonesia Pintar Kuliah KIP Kuliah Merdeka," Puslapdik, Jakarta, 2023.

[4] N. R. Tamnge, "Regresi Logistik Biner Dalam Menentukan Pengaruh Kualitas Pelayanan Terhadap Kepuasan Mahasiswa Fakultas Keguruan dan Ilmu Pendidikan Universitas Muhammadiyah Surabaya," *Journal of Mathematics Education, Science and Technology*, vol. 1, no. 2, pp. 222-233, 2016.

[5] A. Agresti, Categorical Data Analysis, New York: John Wiley & Sons, Inc, 1990.

[6] D. W. Hosmer and S. Lemeshow, Applied Logistic Regression, Canada: John Wiley & Sons, Inc, 2000.

[7] W. Alwi, Ermawati and S. Husain, "Analisis Regresi Logistik Biner Untuk Memprediksi Kepuasan Pengunjung Pada Rumah Sakit Umum Daerah Majene," *Jurnal MSA*, vol. 6, no. 1, pp. 20-26, 2018.

[8] B. Peeters, R. Dewil and I. Y. Smets, "Improved Process Control of an Industrial Sludge Cemtrifuge-dryer Installation Through Binary Logistic Modelling of The Fouling Issues," *Journal of Process Control*, vol. 22, no. 7, pp. 1387-1396, 2012.

[9] A. Agresti, Categorical Data Analysis Second edition, New Jersey: John Wiley & Sons, 20002.

[10] R. Johnson and D. Wichern, Applied Multivariate Statistical Analysis, New Jersey: Prentice Hall, 1992.

[11] J. Han, M. Kamber and J. Pei, Data Mining Concepts and Techniques, USA: Elsevier, 2012.