International Journal of Development Mathematics Vol 1 Issue 2 Page No. 160 -168

INTERNATIONAL JOURNAL OF DEVELOPMENT MATHEMATICS

ISSN: 3026-8656 (Print) | 3026-8699 (Online)

journal homepage: https://ijdm.org.ng/index.php/Journals

Survival Analysis of Students' Dropout in a Nigerian University System

Abubakar N. Ahmadu[a*], Adesupo A. Akinrefon[b], Emmanuel Torsen[a], Nasiru Yakubu[a]

[a]Department of Statistics, Faculty of Physical Sciences, Modibbo Adama University, Yola, Adamawa State, Nigeria.

ARTICLE INFO

Article history:
Received 02 March 2014
Received in revised form 17 May 2024
Accepted 27 Mayy 2024

Keywords:
Survival Analysis, Attrition, Dropouts, Event, Censored.

MSC 2020 Subject classification:
62N01, 62NO2, 62N99, 62P99.

ABSTRACT

Formal education remains the vehicle for socio-economic development and social mobilization in any society. Academic institutions are increasingly interested in monitoring the performance of their students. This interest has given rise to the need for research, collation, analysis, and interpretation of data to formulate evidence-based academic policies that aim to improve student performance and reduce dropout rates. Data obtained from the university comprises of Gender, Age at Entry, Marital Status, Sponsorship, Mode of Entry, Program of Study and Cumulative Grade Point Average (CGPA). This study investigated how these factors influence students' dropout at Modibbo Adama University, Yola Adamawa State Nigeria. Survival analysis was carried out using Log-logistic, Lognormal, Weibull and Cox Proportional Hazard models. The lognormal survival model was judged as the best model due to its lowest AIC value of 1199.773. The Gender, CGPA and students' program of study have significant effect on student dropout. Also, it was observed that 41.42% of male students dropped out when compared to 38.94% female students dropped out before their graduation year. Of all the programs of study considered, Physics has the highest number of dropped out before graduation. It accounted for 68.75%. In terms of academic performance, majority (85.05%) students who had CGPA of third class at the beginning of their studies dropped out before expected years of graduation. The study recommends that government and university management provide financial support to assist the male students who are on self-sponsorship, as financial pressure may contribute to their dropout rates.

1. Introduction

One of the long-term goals of any university in Nigeria and around the world is to reduce the student attrition rate. It is reported that about a quarter of the students dropped out of university after their first year and it increased to 50% by the end of the fourth semester (Tinto, 2006). The benefits of improving student retention is self-evident including a higher chance of having a better career and a higher standard of life (Thomas, 2002). On the other hand, the higher the student retention rate, the more likely that the university is positioned higher in rankings, secure more government funds, attract more intake, and will have an easier path to program accreditations. Because of these reasons, directors and administrators in universities are increasingly feeling the pressure to outline and implement strategies to decrease student attrition (Tinto, 2006).

Traditional methods such as regression and logistic regression have been used to identify dropout students for decades (Lin, Imbrie, and Reid, 2009). Recently, the student retention problem has drawn a lot of attention from researchers in data mining and machine learning communities (Lakkaraju *et al*, 2015). However, student attrition is not an abrupt event, but rather a lengthy process that completely depends on time (Tinto, 2006). Therefore, it would be appropriate to formalize it as a longitudinal problem and use sophisticated longitudinal data analysis techniques for modeling the problem. One of the important characteristics of student data is that it can be incomplete, due to the inability to continuously track the student, often referred to as censoring. This incompleteness in events or information is different from missing data encountered in routine data mining problems and not all modeling techniques can handle them (Vinzamuri and Reddy, 2013).

* Corresponding author. Tel.: +2348065745482
E-mail address: aa.norndigi@mau.edu.ng (Abubakar N. A.)
https://doi.org/10.62054/ijdm/0102.12

Although survival model has more flexibility to handle the student retention problem, predict the expected duration of the occurrence of the event (dropout), and identify the most effective factor(s) that lead to student attrition, it is rarely applied in student academic performance. Due to these reasons, this paper aims to model students' dropout using some survival analysis model. To the best of our knowledge, this has not been done anywhere else in the past. Survival analysis is defined as a set of methods for analyzing data where the outcome variable is time until an event of interest occurs. According to Sainani (2013), survival analysis is a statistical method for analyzing longitudinal data on the occurrence of events. The event is usually something that you do not want to happen such as failure, dropout and death. However, it might be a positive thing such as 'recovery' or healing or a specific treatment state such as remission.

Cox (1972) proficiently developed the Cox model survival analysis, which derives robust, consistent, and efficient estimates of the covariate effect. As contended by Cox (1972), the Cox model emerged as the most widely used method in survival analysis. The Cox model assumes the assumption of proportional hazard (PH) rate to be constant over time. In some literature, it was reported that most of the findings on survival analysis used Cox regression model which is a semi-parametric model also emerged as the most widely used method in survival analysis (Cox, 1972). The results from this model might lead to severe bias, wrong inferences, or lower power of a test if the PH assumption (the subjects have a constant hazard rate over time) in the Cox model is not met (Abrahamamowicz, Mackenzie and Esdaile, 1996). Due to this limitations of the Cox model, we decide to look for an alternative model with better precision, consistency, and less or zero bias. However, some competing parametric models emerged as the best alternative and can address the weaknesses of the Cox regression Model. These parametric models include the Log-logistic, Lognormal, and Weibull survival models. A study by Johnson (2006), using logistic regression revealed that students who were more likely to earn a CGPA above 2.00 were less likely to leave the university for academic reasons. Johnson further suggested that teaching methods, the support of the structure of the school, family background, and the type of secondary school attended were the significant factors that attributed to student dropout.

Akinrefon and Adejumo (2012) carried out a research to assess and investigate undergraduate students' academic performance in their first and final year to know which of them gives the student's true ability in his/her chosen course of study and also to find out the effect of some factors on their final grade. The conditional symmetry model and multinomial logit regression were used for the analysis. The result shows that, in all the various cross-classified data, that their first year performance rated the students' performance better than their final year performance, while the multinomial logit shows that the estimated odds of the student making a first/second class (upper division) as against a second class (lower division) is higher in their first year than in the final year. Sodipo and Adepoju (2015) examine students' academic performance at the University of Ibadan. Data for the study was collected from the faculty of Science University of Ibadan and Analysis of Variance (ANOVA) and Chi-square was used for the analyses. Their study revealed that a certain number of students who were admitted into the University, for the attainment of a degree, were unable to meet the requirements to be able to graduate on time. A situation where some students have to study longer than the four-year term is undesirable and should be checked. It is this kind of unwanted situation that prompted me to study this similar scenario but using survival analysis. Balogun *et al* (2016) used Cohen's Kappa to examine the performance of students in the Faculty of Science, University of Ilorin. The data was collected from eight departments in the faculty and it covers the performance of students measured by their Grade Point Average (GPA) and Cumulative Grade Average (CGPA) in both their first year and final year between the 2000-2006 academic sessions. The study determined the proportion of students who improved on their performance, dropped from their class of grade point which they started with, and maintained their performance using the psychometrics approach. Also, the strength of agreement that exists between the first and final year was examined.

3. Materials and Methods

The study was conducted on 649 students from the Faculty of Physical Sciences Modibbo Adama University, Yola Nigeria. The covariates considered for the study included Student CGPA, Age at Entry, Mode of

Entry, Sponsorship, Program of Study, Gender, and Marital Status, while the outcome or dependent variable is students' dropout before their graduation year.

3.1 Cox Proportion Hazard Model

The most commonly used regression model in survival analysis is the Cox proportional hazard model. With this model, the distribution of the baseline hazard function is not specified and that is why it is called a semi-parametric model. The Cox-proportional hazard model is often used in modelling the hazard and survival function because it does not place distributional assumptions on the baseline hazard. The Cox model was introduced by Cox (1972). It has the form:

$$h(t|X) = h_0(t)\exp(X^T\beta) \tag{1}$$

$$h(t|X) = h_0(t)e^{(\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_4X_4+\beta_5X_5+\beta_6X_6+\beta_7X_7)} \tag{2}$$

Where the $_1, \beta_2, \beta_3, \dots, \beta_6$ are the estimated coefficient of the regression model

$h_0(t)$ is the baseline constant for the regression model for this study.

X_1 = Age (Continuous)
X_2 = Gender
X_3 = Marital Status
X_4 = CGPA (First year CGPA)
X_5 = Mode of Entry
X_6 = Sponsorship
X_7 = Program of Study and
t is time in years

The measure of the effect of the given covariates on survival time is given by the hazard ratio denoted as *HR*. Consider a categorical variable with two levels is stated as equation (3):

$$HR = \frac{h(t|X=1)}{h(t|X=0)} = \exp(\beta) \tag{3}$$

When $HR = 1$, it suggest that the individuals in the two categories are at the same risk of getting the event, when $HR > 1$, it suggest that the individuals in the first category (X=1) are at a high risk of getting the event (Abrahamamowicz, Mackenzie and Esdaile, 1996).

3.2 Log-logistic Model

In the log-logistic model, the hazard rate is stated as equation (4):

$$h(t,X) = \frac{\lambda^{\frac{1}{\gamma}} t^{\left[\left(\frac{1}{\gamma}\right)-1\right]}}{\gamma\left[1+(\lambda t)^{\left(\frac{1}{\gamma}\right)}\right]} \tag{4}$$

where

$$\lambda_i = e^{-(X_i\beta)} \tag{5}$$

As with the Weibull model, the log-logistic model has two parameters, λ is the location parameter and Υ is the shape parameter. Where the shape parameter works in the following way:

a. If $\hat{\gamma} < 1$then the conditional hazard first rises, then falls.
b. If $\hat{\gamma} \geq 1$then the hazard is declining.

However, it is worth noting that the hazard can never be monotonically rising in the log-logistic model.
Thus, the survivor function for the log-logistic is stated as in equation (6):

$$S(t) = \frac{1}{1+(\lambda t)^{\left(\frac{1}{\Upsilon}\right)}} \tag{6}$$

and the density function is stated as in equation (7):

$$f(t) = \frac{\lambda^{\frac{1}{\Upsilon}} t^{\left[\left(\frac{1}{\Upsilon}\right)-1\right]}}{\left\{\Upsilon\left[1+(\lambda t)^{\left(\frac{1}{\Upsilon}\right)}\right]\right\}^2} \tag{7}$$

and the integrated hazard function is stated as in equation (8):

$$H(t) = 1 + (\lambda t)^{\left(\frac{1}{\Upsilon}\right)} \tag{8}$$

(Abrahamamowicz, Mackenzie and Esdaile, 1996)

3.3 Lognormal Model

In the lognormal model, the survivor function is:

$$S(t) = 1 - \Phi\left\{\frac{\ln(t) - \mu}{\sigma}\right\} \tag{9}$$

where Φ is the standard normal cumulative density function and $\mu = X\beta$. The density function is:

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} exp\left[\frac{-1}{2\sigma^2}\{\ln(t) - \mu\}^2\right] \tag{10}$$

and the hazard rate is:

$$h(t) = \frac{\frac{1}{t\sigma\sqrt{2\pi}} exp\left[\frac{-1}{2\sigma^2}\{\ln(t)-\mu\}^2\right]}{1-\Phi\left\{\frac{\ln(t)-\mu}{\sigma}\right\}} \tag{11}$$

(Abrahamamowicz, Mackenzie and Esdaile, 1996)

3.4 Weibull Model

In the Weibull model, the hazard rate is characterized as:

$$h(t, X) = \lambda p(\lambda t)^{\rho-1} \tag{12}$$

where

$$\lambda_i = e^{X_i\beta} \tag{13}$$

The Weibull model is more general and flexible than the exponential model and allows for hazard rates that are non-constant but monotonic. It is a two parameter model (λ and ρ), where λ is the location parameter and ρ is the shape parameter because it determines whether the hazard is increasing, decreasing, or constant over time. The shape parameter works in the following ways:

a. If $\hat{\rho} < 1$, then the hazard is monotonically decreasing with time.
b. If $\hat{\rho} > 1$, then the hazard is monotonically increasing with time.
c. If $\hat{\rho} = 1$, then the hazard is flat and we have the exponential model i.e. the Weibull model nests the exponential model. This means that we can use the Weibull model to test to see if the exponential model is appropriate.

The survival function for the Weibull is:

$$S(t) = e^{(-\lambda t)^{\rho}} \tag{14}$$

And the density function is:

$$f(t) = \lambda\rho(\lambda t)^{\rho-1} e^{(-\lambda t)^{\rho}} \tag{15}$$

The expected duration from Weibull model is:

$$E(T) = \left(\frac{1}{\lambda}\right)^{\frac{1}{\rho}} \Gamma\left(1 + \frac{1}{\rho}\right) \tag{16}$$

Where denotes the gamma function (Abrahamamowicz, Mackenzie, and Esdaile, 1996).

3.5 Akaike Information Criterion (AIC)

Akaike information criterion (AIC) was used to compare all these survival models, which is defined as:

$$AIC = -2l + 2(k + c) \tag{17}$$

where l is the log-likelihood, k is the number of covariates in the model and c is the number of model-specific auxiliary parameters. The addition of 2(k +c) can be thought of as a penalty if non-predictive parameters are added to the model. Lower values of the AIC suggest a better model (Akaike, 1974).

4.0 Results

4.1 Descriptive Analysis of the Dropout Students

Table 1 shows the descriptive statistics of the 649 students considered for this study. It was observed that about 82.59% of the students are male while females accounted for the remaining 17.41%. Out of the total males, 536 (41.42%) dropped out of university due to poor performance while the remaining 58.58% continued their programs until they successfully graduated. Additionally, 38.94% of the 113 female students considered dropped out, while the remaining 61.06% of female students considered in this study successfully graduated.

Furthermore, students' first-year CGPA was also captured as one of the most significant variable that influence students' dropout. Thus, the results reveal that of the 649 students considered for the study, 2.77% had a First Class, 19.57% a Second Class Upper, 47.77% had a Second Class Lower and 29.89% had Third Class and below. However, it was apparent to note that none of the First Class students experienced the events throughout the study period. Similarly, of the 127 Second Class Upper candidates 97.64% successfully graduated and the remaining 2.36% experienced the event (dropped out). More so, of the 310 Second Class Lower, 68.39% succeeded in completing their programs within the expected year of graduation while the remaining 31.61% experienced the event. Additionally, of the 194 students who started their programs with Third Class and below majority (85.05%) dropped out before graduation.

Table 1 Summary Statistics of Data Collected

Variables	Event (Dropout in %)	Survived (Graduated in %)	Total in %
Gender			
Male	222 (41.42%)	314 (58.58%)	536 (82.59%)
Female	44 (38.94%)	69 (61.06%)	113 (17.41%)
Marital Status			
Single	257 (40.92%)	371 (59.08%)	628 (96.76%)
Married	9 (42.86%)	12 (57.14%)	21 (3.24%)
CGPA			
4.5-5.00	0 (0.00%)	18 (100%)	18 (2.77%)
3.50-4.49	3 (2.36%)	124 (97.64%)	127 (19.57%)
2.40-3.49	98 (31.61%)	212 (68.39%)	310 (47.77%)
0.00-2.39	165 (85.05%)	29 (14.95%)	194 (29.89%)
Mode of Entry			
UTME	187 (45.06%)	228 (54.94%)	415 (63.94%)
DE	79 (33.76%)	155 (66.24%)	234 (36.06%)
Program			
CHM	18 (36.00%)	32 (64.00%)	50 (7.70%)
CSC	62 (33.33%)	124 (66.67%)	186 (28.66%)
GLY	19 (39.58%)	29 (60.42%)	48 (7.40%)
ICH	16 (26.23%)	45 (73.77%)	61 (9.40%)
MEC	16 (53.33%)	14 (46.67%)	30 (4.62%)
MTH	33 (49.25%)	34 (50.75%)	67 (10.32%)
OPR	45 (46.39%)	52 (53.61%)	97 (14.95%)

PHY	22 (68.75%)	10 (31.25%)	32 (4.93%)
STA	35 (44.87%)	43 (55.13%)	78 (12.02%)
Sponsorship			
Parent	189 (40.38%)	279 (59.62%)	468 (72.11%)
Guardians	35 (34.65%)	66 (65.35%)	101 (15.56%)
Self	42 (52.50%)	38 (47.50%)	80 (12.33%)

4.2 Kaplan-Meier (KM) Plot

Figure 1 depicts the KM plots of students drop outs by CGPA. However, all the 18 students that began their respective programs with a first-class CGPA successfully graduated within the prescribed timeline of their programs. Similarly, of the 127 students that started their program with a Second Class Upper, 2 students dropped out after their third year and the survival probability dropped from 100% to 98.4%. Similarly, of the remaining 125 students, only 1 student dropped out after their fifth year, and survival probability reduced from 98.4% to 97.2% (this is the right censored probability for this category of students). Subsequently, of the 310 students that began their program with a Second Class Lower, 47 students dropped out after their second year and the survival probability reduced from 100% to 84.8%. Also, of the remaining 263 students, 42 dropped out after their third year and the survival probability reduced from 84.8% to 71.2%. More so, of the remaining 221 students, 6 dropped out after their fourth year and the survival probability reduced from 71.2% to 69.3%. Furthermore, of the remaining 215 students, 3 students dropped out after their fifth and the survival probability reduced from 69.3% to 67.9%. Additionally, of the 194 students that started their program with a Third Class and below, 35 students dropped out after their first year, and the survival probability reduced from 100% to 82.0%. Similarly, of the remaining 159 students, 112 students dropped out after their second year, and the survival probability was reduced from 82.0% to 24.2%. Subsequently, of the remaining 47 students, 14 students experienced the event after their third year and the survival probability reduced from 24.2% to 17.0%. Furthermore, of the remaining 33 students, 4 students experienced the event after their fourth year and the survival probability reduced from 17.0% to 14.9%. However, the log-rank test was also conducted and the calculated Chi-square value is 347 at 3 degree of freedom with a P-value of 0.0000 this implies that there is significant difference between the students who dropped out by CGPA.

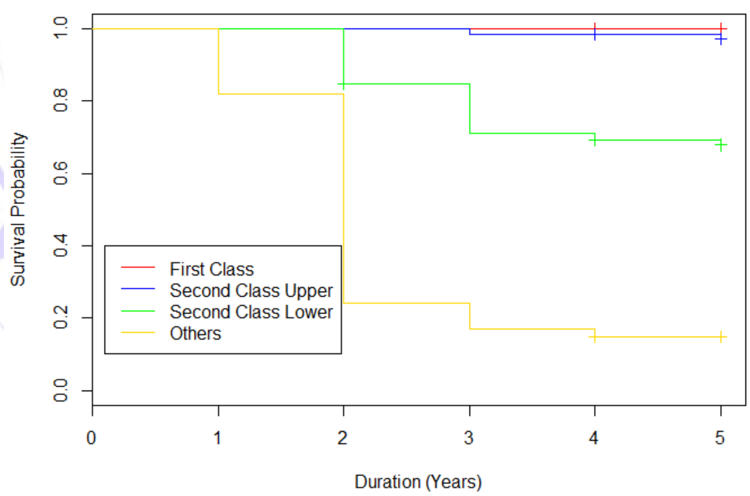

Figure 1 Survival Curve of Students Dropout by CGPA

Table 2 shows the performance of the models fitted on the survival data. Based on AIC values, it was found that Lognormal appears to be the best model since it has the minimum AIC value of 1199.773.

Table 2 Performance of the Models Based on AIC

Models	**Log-likelihood**	**AIC**	**Rank**
Cox PH	-1476.0	2966.043	4
Lognormal	-590.9	**1199.773**	1
Log logistic	-594.5	1207.046	2

Weibull	-608.4	1234.722	3

Since the lognormal model outperformed the other models then we find the parameter estimates as shown in table 3, based on the output, it was found that the following covariates: Gender, Program of study, and CGPA are statistically significant at 0.05 because their P-values are less than 0.05. Moreover, based on gender, male students had a higher hazard rate (29.12%) than the female students. Also, students offering Physics had a higher hazard rate of about 44.08% but the hazard rate of students in the other 7 programs is insignificant. Furthermore, students who began their respective program with a Third Class CGPA had a very high hazard rate of about 73.061% as compared to those in the other category of CGPAs.

Table 3 Results from Fitting Lognormal Model

Covariates	Estimate	Std. Error	Z	P-value	Hazard Ratio
Intercept	5.45449	0.2699	20.20907	8.15e-91	0.004277058
Mode of Entry, UTME	**Reference**				
DE	0.09464	0.0707	1.33834	1.81e-01	0.909700368
Gender, Female	**Reference**				
Male	-0.25560	0.0876	-2.91652	**3.54e-03**	1.29123613
Marital Status, Single	**Reference**				
Married	-0.15887	0.1705	-0.93201	3.51e-01	1.172185553
Sponsorship, Parent	**Reference**				
Guardians	-0.00625	0.0868	-0.07208	9.43e-01	1.006269572
Self	-0.05646	0.0891	-0.63347	5.26e-01	1.058084291
Program, CHM	**Reference**				
CSC	-0.03854	0.1260	-0.30581	7.60e-01	1.039292299
GLY	0.16468	0.1542	1.06799	2.86e-01	0.848165073
ICH	-0.06391	0.1515	-0.42189	6.73e-01	1.065996455
MEC	0.00083	0.1671	0.00497	9.96e-01	0.999170344
MTH	-0.02296	0.1391	-0.16512	8.69e-01	1.023225610
OPR	0.10664	0.1320	0.80783	4.19e-01	0.898849200
PHY	-0.36518	0.1626	-2.24649	**2.47e-02**	1.440773324
STA	-0.10025	0.1392	-0.72005	4.71e-01	1.105447245
Age at Entry	-0.00943	0.0120	-0.78756	4.31e-01	1.009474603
CGPA, 4.50-5.00	**Reference**				
3.50-4.49	-2.24574	0.1674	-13.41803	**4.74e-41**	9.447404049
2.40-3.49	-3.29258	0.0668	-49.28157	**0.00e+00**	26.91220766
0.00-2.39	-4.29129	0.0000	-Inf.	**0.00e+00**	73.06065598

4.3 Discussion

Among the factors that affect students' academic performance considered for this study, it was found that student's Gender, program of study and CGPA are the most significance with a p-value less than 0.05 (i.e. p-value<0.05).The finding revealed that students' Program of study is a significant factor that determines students' attrition rate in the university. The attrition or dropout rate in the Physics department is twice the attrition rate in the other departments in the Faculty of Physical Sciences MAU, Yola. This is attributed to the fact that about 70% of the students didn't apply for the course and coupled with the complexity of the course students' passion for the course is been demoralized. This is in line with the findings of Thaithanan *et al* (2021) that change of student's choice of course of study is a major factor that affects students' academic performance in tertiary institution. Students first year CGPA which is a tool for measuring student academic performance in Nigeria tertiary institutions was also found significant at 5% significance level. This is because students that began their program with a First Class or Second Class Upper CGPA tends to graduate at the due time without experiencing the event than students who began their program with another class of degree. This finding is in accordance with the findings of Ali *et al* (2013) which established that students with low CGPA are more at risk to dropping out of their respective programs than students with high CGPA.

5. Conclusions

Based on the findings, it was found that Gender, students' Programs of Study, and CGPA contributed significantly to students' dropout. The lognormal model outperformed other competitive models considered for the study. The hazard rate is higher among Physics students compared to students in other departments.

Acknowledgment

The first author would like to thank the Petroleum Technology Development Fund (PTDF) for their kind support and assistance. The authors would like to express their appreciation to the editor and the reviewers for the constructive comments that improved the quality of this article.

References

Abrahamowicz, M., Mackenzie. T. and Esdaile, J. M. (1996). Time-dependent Hazard Ratio: Modeling and Hypothesis Testing with Application in Lupus Nephritis. *Journal of American Statistical Association*, Vol. 91. No. 433 pp. 1432–1439.

Akaike H (1974). A new look at the statistical model identification. IEEE Trans Automatic Control.

Akinrefon, A. A. and Adejumo, A. O. (2012). A Study of Undergraduate Students Performance in Nigeria. *Bulgarian Journal of Science and Education Policy (BJSEP), Volume 6.*

Ali, S., Zubair, H. and Fahad, M. (2013). *Factors contributing to the students' academic performance: A case study of Islamia University sub-campus*, «American Journal of Educational Research», 1 (8), 283-289.

Balogun, O. S. and Elepo, T. A. (2016): Statistical Analysis on Students' Performance. *Covenant Journal of Informatics and Communication Technology (CJICT)* Vol. 4, No. 1

Cox, D.R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society*, Vol. 34. No. 2 pp.187-220.

Johnson, V. E. (2006). Grade Inflation: A crisis in College education *Springer Science and Business media.*

Lakkaraju, H., Aguiar, E., Shan, C., Miller, D., Bhanpuri, N., Ghani, R. and Addison, K. L. (2015). A machine learning framework to identify students at risk of adverse academic outcomes. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1909–1918. ACM, 2015.

Lin, J., Imbrie, P. and Reid, K. J. (2009). Student retention modelling: An evaluation of different methods and their impact on prediction results. *Research in Engineering Education Sysmposium*, pages 1–6.

Sodipo, E. O., Sodipo, A. A. and Adepoju, K. A. (2015). Statistical analysis of students' academic performance in Nigeria universities: a case study of the university of Ibadan, Nigeria. *International Journal of Recent Advances in Multidisciplinary Research* Vol. 02 Issue 02 pp 0187-0192

Thaithanan, J., Thinnukool, O., Chaichana, M. and Wanishsakpong, W. (2021). Investigating Undergraduate Student Dropout Rates using Survival Analysis in the College of Arts, Media and Technology, Chiang Mai Univesity. *Journal of Educational Psychology, 7(10).*

Thomas, L. (2002). Student retention in higher education: the role of institutional habitus. *Journal of Education Policy*, 17(4):423–442.

Tinto, V. (2006). Research and practice of student retention: what next? *Journal of College Student Retention:*

Research, Theory and Practice, 8(1):1–19, 2006.

Vinzamuri, B. and Reddy, C. K. (2013). Cox regression with correlation based regularization for electronic health records. *Proceedings - IEEE International Conference on Data Mining, ICDM*, pages 757–766.