ISSN 2278-3091
Volume 10, No.3, May - June 2021
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse461032021.pdf
https://doi.org/10.30534/ijatcse/2021/461032021

Impact of Postgraduate Students Dropout and Delay in University: Analysis Using Machine Learning Algorithms

Muhammad Nadeem[1], Sellapan Palaniappan[2],Waleej Haider[3]
[1] Department of Information Technology, School of Science and Engineering, Malaysia University of Science and Technology, Petaling Jaya 47810, Selangor, Malaysia, nadeem.muhammad@phd.must.edu.my
[2] Department of Information Technology, School of Science and Engineering, Malaysia University of Science and Technology, Petaling Jaya 47810, Selangor, Malaysia, sell@must.edu.my
[3]Department of Computer Science and Information Technology, Sir Syed University of Engineering and Technology, Karachi, Pakistan, directlyproportional2u@yahoo.com

ABSTRACT
Cost of education and economic background are some factors that influence student dropout from postgraduate studies. However, high dropouts do not affect the students only, but also impact university revenue. This research analyzes various literature on machine learning algorithms and applies suitable algorithm to produce a prediction model. This study indicates that decision tree and Random Forest algorithms have better accuracy, class recall, and class precision than Naïve Bayes. Therefore, the prediction model uses the Decision Tree algorithm to provide various approaches to maximize revenue in universities. The findings indicate high dropout rates negatively impact university revenue, while low rates influence revenue positively. Other aspects like grants received by students, the number of research publications, and degree level also positively or negatively impact revenue if the dropout rate is medium. A complete understanding of this prediction model can identify and minimize the risk of early withdrawal or delayed graduation and improve revenue generation by universities.

Key words- algorithm, dropout, delay, graduation, prediction model, decision tree, revenue, impact, university, postgraduate, students, random forest, naïve Bayes

1. INTRODUCTION

Education is a critical aspect of one's future because its effects are long-term. However, postgraduate students often find themselves in awkward situations that may cut short their education or make them delay their studies. According to [1], most institutions have a major concern about dropout and delay at the undergraduate level more than postgraduate level. Nevertheless, postgraduate students also face challenges leading to drop out or delay at the university that has remained unexplored for a longer time. Fortunately, dropout data have become widely documented across universities despite the lack of a substantial predictive model [2]. Besides, little is known about students' determination to terminate their studies, albeit the increasing enrollment for postgraduate studies [1], [3], and [4]. According to [3], the universities themselves may be unknowingly contributing towards the dropouts or delays. Several researchers are increasingly becoming concerned about the number of postgraduate students who fail to graduate as dropping out has diverse negative consequences for the affected [5], [6]. Other times, the rising dropout may significantly impact the institutions due to reputational damage and loss of financial gains. Notably, no student will enroll at an institution with a bad record of university dropout, making such instructions unprofitable.

Understandably, better methods have been unveiled to produce models for predicting learning outcomes based on student historical data. For instance, machine learning or artificial intelligence (AI), computational modelling, data visualization, and data mining are techniques used to predict learning outcomes and better understand students. Some of the known algorithms with greater accuracies tested and verified by other researchers include the Decision Tree algorithm, Neural Network and Logistic Regression [7], [8], [9]. Therefore, this research evaluates existing machine learning algorithms to develop a model for predicting and explaining the impact of postgraduate dropout delay on the University.

2. LITERATURE REVIEW

The number of students who graduate successfully from a university on time reflects on their operational cost. Therefore, the success rates will indirectly minimize a university's operational cost [10]. Suhaimi et al. [10] developed a predictive

model that would benefit university management in devising strategies for improving weak students' performance and increasing the graduation rate. The research applied five different machine learning algorithms (classifiers), including Decision Tree, Naïve Bayes, Random Forest, Support Vector Machine (KBFKernel), and Support Vector Machine (PolyKernel) [10]. Notably, their prediction model focused on student performance to develop a prediction model to help institutions understand and improve learning outcomes. The study outcome revealed SVM (PolyKernel) outperformed other classifiers, although the others also performed better. For instance, Random Forest and Decision tree provided consistent results across the test subjects, meaning they are reliable prediction algorithms. Naïve Bayes produced the lowest Accuracy, Recall, and F Score. Nevertheless, other studies indicate Naïve Bayes is a reliable algorithm that can predict loss of academic status at the University [11].

Budiman et al. [12] researched with C4.5 and ID3 algorithms to predict student dropout. Notably, [13] conducted a similar study using the C4.5 algorithm using student sample data from Simón Bolívar University. The C4.5, derived from the algorithm Iterative Dichotomiser (ID3), is a type of the Decision Tree algorithm. The study result indicated C4.5 had better accuracy in evaluating student academic performance. Therefore, it can be used as an alternative prediction model to evaluate student performance. A research by Gil et al. [21] also used two classification algorithms- C4.5 and Naïve Bayes to predict dropout indicators in private schools based on academic and demographic details. The classification algorithms discovered a clear and understandable models of student dropout prediction, with 97.89% and 98.94% accuracy for Naïve Bayes and C4.5, respectively [21].

Another research by [8] examined IoT systems for school dropouts using machine learning algorithms based on socioeconomic data. Notably, the authors propose the automation of the prediction process that contributes to the accurate prediction of student learning outcomes without analyzing the bulk of data. The result of this research indicated Decision Tree is a viable option to predict university graduation outcomes with 99.34% accuracy, 99.34%F1 score, 100%recall, and 98.69%precision [8]. According to [8], it outperformed others like Naïve Bayes, Logistic Regression, Support Vector Machines, K-Nearest Neighbors, and Multi-Layer Perceptron. Other studies like [9], [14], [15], [16] and [17] have achieved better accuracies with other algorithms, such as educational data mining (EDM), Neural Network, Support Vector Machines, and Machine Learning, that would be applicable in predicting university dropout.

3. METHODOLOGY

The first stage in the analysis is understanding the data and transforming it into a format that can be analyzed using a data mining tool. The technique applied in gathering data was through documents and records, by examining student historical data from the university database. The cohort selected were students enrolled from 2011 to 2018 at three distinct universities. Next, the data were examined to identify primary attributes (features) and their range values, as shown in Tables 1 and 2.

Table1. Features selection and description

Features	Value	Description
Year	2011- 2018	Year
Prog_desc	Science and Engineering programme	Programme taken by the students
Degree	MS, Phd	Name of postgraduate degree
St_Drop_Rate	High, Medium and Low	Student dropout rate
Int_Grants	High, Medium and Low	International Research & Training Grants taken by the student
Nat_Grants	High, Medium and Low	National Research & Training Grants taken by the student
Res_Pub	High, Medium and Low	Research publication(s)
Enroll_Rate	High, Medium and Low	Rate of enrollment in university
Revenue	High, Medium and Low	Impact in University in term of Revenue (Target Variable)

Table 2. Features values and ratings

Value	Rate in Percentage
High	Above 60%
Medium	Below 50% to 60%
Low	Below 50%

Several attributes were selected from the raw data set based on relevance to the goals of this study, as shown in table 1. Further cleaning was done to exclude attributes that did not give meaning to achieve the research objectives. Notably, the study involved analyzing data for students in the Software Engineering, Management Science, Computer Science, Computer Engineering, and Software Engineering fields- totaling 1728 data instances. The modeling phase involved a search for useful patterns

in the data. Modeling algorithms were narrowed down to three classifiers that matched research objectives and the types of variables involved. Therefore, the data was trained on the three classifiers, namely Decision Tree, Random Forest, and Naïve Bayes. The performance of the classifiers was evaluated on the value of accuracy score, precision, recall, and F-measure. Understandably, the highest-scoring classifier was selected for the research. Notably, the predictive models are developed using RapidMiner Tool.

4. ANALYSIS AND FINDINGS

The study applied three classifiers, namely Decision Tree, Naïve Bayes, and Random Forests, to determine the best algorithm for creating the prediction model. Notably, the Decision tree had a 100% score for all the metrics (accuracy, class recall, and class precision). Similar results were achieved using Random Forest, as shown in fig. 1 and 3, respectively. However, Naïve Bayes scored slightly lower with 94.4% accuracy. The class recall and class precision were also lower than Decision Tree and Random Forest (see fig 2). Although Random Forest and Decision Tree recorded similar results, the Decision Tree algorithm was selected for the prediction model. Notably, both generated a similar tree, but the Decision Tree was chosen for convenience and suitability to this research.

Table View Plot View

accuracy: 100.00%

	true High	true Low	true Medium	class precision
pred. High	209	0	0	100.00%
pred. Low	0	277	0	100.00%
pred. Medium	0	0	32	100.00%
class recall	100.00%	100.00%	100.00%	

Figure 1: Performance of the Decision Tree algorithm

Table View Plot View

accuracy: 94.40%

	true High	true Low	true Medium	class precision
pred. High	192	0	0	100.00%
pred. Low	0	268	3	98.89%
pred. Medium	17	9	29	52.73%
class recall	91.87%	96.75%	90.62%	

Figure 2: Performance of the Naive Bayes algorithm

Table View Plot View

accuracy: 100.00%

	true High	true Low	true Medium	class precision
pred. High	209	0	0	100.00%
pred. Low	0	277	0	100.00%
pred. Medium	0	0	32	100.00%
class recall	100.00%	100.00%	100.00%	

Figure 3: Performance of the Random Forest algorithm

4.1Analysis of the Prediction Model

Dropping out of university has serious financial ramifications for students and can impact them over the short and long term [18], [19]. However, universities also share the burden as high dropout means loss of funds to run the institution. Higher dropout rates can have other consequences, such as reputational damage that can discourage students from joining the university. The prediction model provided in fig. 4a and 4b will help universities evaluate the impact of dropout on their revenue and devise proactive measures to curb the issue. The decision tree (fig. 4a and 4b) is divided into two parts for visibility.

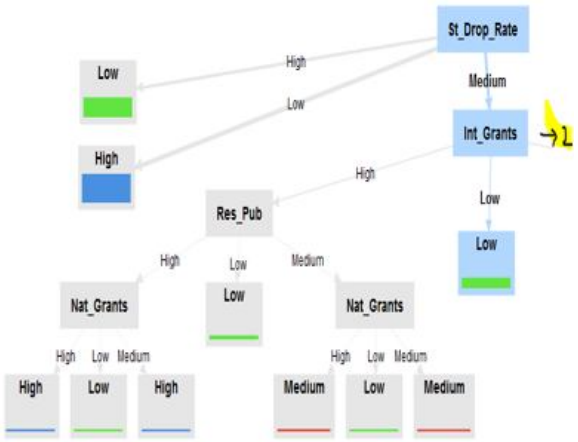

Figure 4a: Decision Tree prediction model for Impact of Student Dropout Rate on University Revenue

Student dropout rate (St_Drop_Rate) is the root of the tree, indicating high (above 60%) dropout lowers the university revenue, while the opposite increases revenue. A low rate means students can pay tuition and sustain themselves for the period of stay at the university. They also utilize other facilities and services, such as the cafeteria, that generate additional revenue for the institution. Therefore, it is apparent universities should strive to minimize dropout or delay to maximize revenue. However, there are several dynamics whenever the dropout rate is medium (between 50% and 60%). Understandably, if the university received low International Research and Training Grants (Int_Grants), its revenue will reduce significantly. Postgraduate studies can be expensive for students to cater for tuition by themselves. Similarly, the higher the university fees are, the more expensive the delayed university becomes for the students [20]. Therefore, the university revenue can decline if they get minimal grants to support students.

However, if the university gets high international research grants, university revenue will depend on research publications (Res_Pub). Both high and medium research publications mean the National Research and Training Grants (Nat_Grant) will influence the revenue. Nevertheless, the university revenue can reduce significantly if they have fewer research publications, meaning low research grants. For universities with high research publications, a high or medium Nat_Grant generates high revenue, while low grants impact revenue negatively. On the other side, medium Res_Pub implies the university will receive at most medium national grants (Nat_Grant). Otherwise, their revenue is minimum if the Nat_Grant is low. Overall, international grants, research publications and national grants have significant impact on university revenue based on the rating scale (High, Medium, or Low).

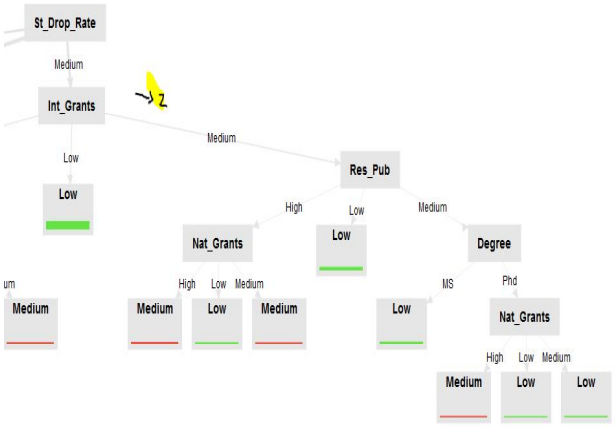

Figure 4b: Decision Tree prediction model for Impact of Student Dropout Rate on University Revenue

Furthermore, revenue impact on universities with medium Student dropout and International Research Grants received by students depends on other factors, as shown in fig. 4b. These factors include the Degree and National Research and Training Grant taken by the student. Understandably, universities in this category (tree branch) with low national grants will record low revenue despite having high research publications (Res_Pub). However, medium or high national grants (Nat_Grants) will improved university revenue, at least to medium if the university has high research publications. Low research publication will impact revenue negatively despite having medium dropout and international grants.

Similarly, universities with medium dropouts, international grants, and research publications depend on the Degree level. Master's (MS) degree and research publications are medium guarantees the institution will record low revenue. However, Ph.D. programs are further dependent on national grants, where high grants mean medium revenue. Low or medium national grants for Ph.D. programs can reduce the university revenue to the minimum.

4.2Findings

The cost of education has a cross-cutting impact on all aspects of human life and learning institutions. The findings of this research are drawn from the analysis of the prediction model generated by the decision tree algorithm. Some of the key findings from the prediction model include the following:

- High dropout impacts the revenue of a university negatively, while low dropout improves its financial capacity.
- International grants, research publications, and national grants play a crucial role in financing postgraduate studies. Therefore, universities should maximize these areas to retain more students and boost revenue.
- Although the programs undertaken by students and enrolment rates may influence revenue, they have inconsequential impacts on the overall revenue compared to other aspects like degree level, research publication, and grants. Nonetheless, sustaining more students who complete their studies on time has a significant impact on the revenue.
- Universities that strive to maintain dropout at medium or high have a better revenue generation.

5. CONCLUSION AND RECOMMENDATION

In summary, the research analyzes empirical literature on machine learning algorithms to produce a prediction medal for evaluating revenue impact on universities. Based on the algorithm evaluation results, the Decision Tree and Random Forest performed best, scoring 100% on the accuracy, class recall, and class precision. Naïve Bayes recorded 94.4% accuracy that was lower than the other two. Therefore, the prediction model generated by the decision tree algorithm was used in this study.

Overall, on-time graduation indirectly influences the operation cost per student. High dropout rates have a greater impact on management and consequently the institution's revenue. A university that keeps the dropout rate low manages to generate more revenue due to constant cash flow from grants, research publications, and degree level. A medium student dropout rate also impacts revenue, but this depends on degree level, number of research publications, and grants received by students. Understandably, research and training grants and research publications positively influence the revenue of an institution. Other aspects like education programs and enrolment rate may positively influence revenue generation, but they are overshadowed by the others included in the prediction model. Therefore, this research identifies the impacts of student dropout on university revenue to help learning institutions make educated choices when enrolling students in their programs.

5.1Recommendations

Although postgraduate studies are becoming more expensive, universities can utilize this prediction model to leverage financial resources and advice to their students. Here are some recommendations that will help universities minimize student dropouts and improve cash flow:

- Universities should help students apply for and source funds, such as grants to finance their studies. The approach benefits the student in settling tuition fees and completing education on time, while the university gains financially from fees paid in the form of grants.
- Universities should intensify research to produce more publications that will generate more funds to minimize research spending.
- Universities should focus on improving MS degree programs since they impact the overall revenue negatively compared to Ph.D. programs.

REFERENCES

[1] N. Rotem, G. Yair and E. Shustak, "Dropping out of master's degrees: objective predictors and subjective reasons," *European Journal of Higher Education Research & Development, DOI: 10.1080/07294360.2020.1799951,* pp. 1-8, 2020.

[2] T.-E. Patterson, "Why Do So Many Graduate Students Quit?," The Atlantic, 6 July 2016. [Online]. Available: https://www.theatlantic.com/education/archive/2016/07/why-do-so-many-graduate-students-quit/490094/. [Accessed 12 May 2021].

[3] A. Styger, G. W. van Vuuren and A. Heymans, "Case Study Of Postgraduate Student Dropout Rate At South African Universities," *International Business & Economics Research Journal,* vol. 14, no. 1, pp. 1-14, 2015.

[4] R. Wollast, G. Boudrenghien, N. V. der Linden, B. Galand, N. Roland, C. Devos, M. D. Clercq, O. Klein, A. Azzi and M. Frenay, "Who Are the Doctoral Students Who Drop Out? Factors Associated with the Rate of Doctoral Degree Completion in Universities," *International Journal of Higher Education,* vol. 7, no. 4, pp. 143-155, 2018.

[5] K. Levecque, F. Anseel, A. De Beuckelaer, J. Van der Heyden and L. Gisle, "Work organization and mental health problems in PhD students," *Research Policy,* vol. 46, no. 4, p. 868–879, 2017.

[6] J. C. Van Ours and G. Ridder, "Fast track or failure: A study of the graduation and dropout rates of PhD students in eonomics," *Economics of Education Review,* vol. 22, no. 2, pp. 157-166, 2003.

[7] S. A. Kumar and M. D. Vijayalakshmi, "fficiency of Decision Trees in Predicting Students' Academic Performance," *Cs & It-Cscp*, p. 335–343, 2011.

[8] F. Freitas, F. Vasconcelos, S. Peixoto, M. M. Hassan, A. A. Dewan, V. H. Albuquerque and P. R. Filho, "IoT System for School Dropout Prediction Using Machine Learning Techniques Based on Socioeconomic Data," *MDPI*, pp. 2-11, 2020.

[9] G. W. Dekker, M. Pechenizkiy and J. M. Vleeshouwers, "Predicting Students Drop Out: A Case Study," *Educational Data Mining,* pp. 1-13, 2009.

[10] N. M. Suhaimi, S. Abdul-Rahman, S. Mutalib, N. H. Abdul Hamid and A. Malik, "Predictive Model of Graduate-On-TimeUsing Machine Learning Algorithms," *Communications in Computer and Information Science,* pp. 131-141, 2019.

[11] E. Budiman, Haviluddin, N. Dengan, A. H. Kridalaksana, M. Wati and Purnawansyah, "Performance of Decision Tree C4.5 Algorithmin Student Academic Evaluation," *Lecture Notes in Electrical Engineering,* pp. 380-389, 2018.

[12] Y. Amaya, E. Barrientos and D. Heredia, "Student dropout predictive model using data miningtechnique," *IEEE Lat. Am. Trans,* vol. 13, no. 9, p. 3127–3134, 2015.

[13] C. E. Guarín, E. L. Guzmán and F. A. González, "A model to predict low academicperformance at a specific enrollment using data mining," *EEE Rev. Iberoam. Tecnol. Aprendiz,* vol. 10, no. 3, p. 119–125, 2015.

[14] B. Dushimimana, Y. Wambui, T. Lubega and P. E. McSharry, "Use of Machine Learning Techniques to Createa Credit Score Model for Airtime Loans," *J. Risk Financ. Manag,* vol. 13, no. 8, p. 180, 2020.

[15] S. Herzog, "Estimating student retention and degree-completion time: Decision trees and neural networks vis-a-vis regression," *New Directions for Institutional Research,* pp. 17-33, 2006.

[16] N. Guenther and M. Schonlau, "Support vector machines," *Stata J,* vol. 16, p. 917–937, 2016.

[17] R. Al-Shabandar, A. Hussain, A. Laws, R. Keight, J. Lunn and N. Radi, "Machine learning approaches topredict learning outcomes in Massive open online courses," in *In Proceedings of the 2017 International JointConference on Neural Networks (IJCNN),*, Anchorage, AK, USA, , 2017.

[18] R. Lake, "The Cost of Being a College Dropout," The balance, 22 April 2021. [Online]. Available: https://www.thebalance.com/the-cost-of-college-dropout-4174303. [Accessed 18 May 2021].

[19] A. Latif, A. I. Choudhary and A. A. Hammayun, "ECONOMIC EFFECTS OF STUDENTS' DROPOUTS: A COMPARATIVE STUDY OF THE CAUSES OF STUDENTS' DROPOUTS GLOBALLY," *International Journal of Economics, Commerce and Management,* vol. 3, no. 6, pp. 1511-1521, 2015.

[20] A. Carmen, B. Eliana, C. Giorgia and P. Francesco, "The economics of university dropouts and delayedgraduation: a survey, GLODiscussion Paper, No. 189," Global Labor Organization (GLO), Maastricht, 2018.

[21] J. S. Gil, A. J. Delima and R. N. Vilchez, "Predicting Students'Dropout Indicators in Public School using Data Mining Approaches," *International Journalof Advanced Trends in Computer Science and Engineering,* vol. 9, no. 1, pp. 774-778, 2020.