International Journal on Advanced Science Engineering Information Technology

Vol.14 (2024) No. 6
ISSN: 2088-5334

A Review of Graduate on Time Prediction

Theng-Jia Law [a], Choo-Yee Ting [a,*], Hu Ng [a], Hui-Ngo Goh [a], Quek Albert [a]

[a] Multimedia University, Cyberjaya, Selangor, Malaysia
Corresponding author: [*]cyting@mmu.edu.my

***Abstract*— In education, predicting students who can graduate on time is difficult. Identifying the significant variables is challenging to predict on-time graduation because human intervention in variable selection is required and time-consuming. It is essential to allow educational institutions to improve student learning experiences by focusing on the significant variables. Researchers have applied various methods, such as Artificial Intelligence, to predict graduation on time. This review has attempted to (i) summarize and compare the diverse methods used by researchers in predicting students who are likely to graduate on time, (ii) identify the gaps and central issues in the existing literature related to predicting on-time graduation, and (iii) establish future potential directions for research in predicting students who are likely to graduate on time, contributing to the ongoing discourse on enhancing educational outcomes. Drawing from an extensive literature analysis across diverse conferences and journals, a notable gap is underscored: the limited focus on predicting graduating on time among postgraduate students. The review addresses issues in learning from small amounts of data and variables, although the researchers demonstrated various techniques for predicting timely graduation, including their strengths and limitations. Future research direction is to consider additional features and improve the performance of predictive models by conducting a comparative analysis of different class treatment methods. As the educational landscape evolves, these considerations are paramount to developing more effective strategies and interventions to ensure timely graduation and foster positive educational outcomes.**

***Keywords*— Graduate on time; artificial intelligence; feature selection; machine learning.**

Manuscript received 22 Dec. 2023; revised 9 Apr. 2024; accepted 28 Sep. 2024. Date of publication 31 Dec. 2024.

I. INTRODUCTION

Graduate on Time (GOT) refers to those who completed their studies timely within the time frame specified by the university [1], [2]. The significance of student graduation transcends individual accomplishments and serves as a vital metric for evaluating educational success. A surge in the number of students unable to meet timely graduation targets adversely impacts the graduation rates, indicating a decline in academic quality and intuitional performance [1]. The challenges leading to delayed graduation often stem from students' struggles to fully commit to their studies, resulting in the retention of unsuccessful courses into subsequent semesters until successful completion is attained [3]. While alternative measures such as modifying assessment methods or lowering passing scores may appear to boost passing and graduation rates, [4] argued that such approaches compromise the quality of university graduates.

Moreover, students encounter various difficulties in completing their studies due to the complex and multidimensional nature of academic success [5]. The journey toward on-time graduation is notably complex, with more than half of attrition cases occurring in the first year of university [5]. The journey toward on-time graduation involves considerations during the university years and factors existing before a student's enrollment. As students navigate the intricacies of academic life, various personal, psychological, and environmental factors come into play, shaping their educational experiences and influencing the likelihood of timely completion [2], [3], [6]–[10]. While researchers have explored myriad factors, analyzing educational data containing these multifaceted factors manually proves challenging for educational institutions [6], [11]. Thus, exploring more efficient approaches to variable selection and prioritizing these factors is essential to foster academic success throughout the academic journey.

In addition to identifying the important study factors, predicting the likelihood of a student graduating on time becomes paramount for educational institutions in recognizing at-risk students and providing early interventions to enhance their academic performance [6], [12]. Researchers have delved into various methods, integrating Artificial

Intelligence (AI) as a notable approach. However, predicting GOT remains a formidable task due to the high dimensionality and the prevalence of missing data in educational data [4]. Researchers have acknowledged the difficulty of selecting the best algorithm from a pool of existing algorithms, mainly when performance evaluation and algorithm enhancement are integral to the process.

The objectives of this review are:

a. To summarize and compare the diverse methods used by researchers in predicting students who are more likely to graduate on time.
b. To identify the gaps and central issues in the existing literature related to predicting on-time graduation.
c. To establish future potential directions for research in predicting students more likely to graduate on time, contributing to the ongoing discourse on enhancing educational outcomes.

II. Materials and Method

In this study, the literature was considered from different conferences and journals in online databases such as Institute of Electrical and Electronics Engineers (IEEE) [13]–[17], Science Direct [18]–[23], Springer [24, 25], Atlantis Press [26], [27], and Google Scholar [3], [4], [6], [11], [12], [28]–[60]. To search this literature from the online databases, keywords or similar terms are considered such as “graduate on time”, “predicting graduate on time using machine learning”, “predicting graduate on time among postgraduate students using machine learning”, and “predicting graduate on time among undergraduate using machine learning”. Searches were restricted to the articles published from the year 2019 to 2023.

A. Challenges and Problems in the Context of Graduate on Time

Table 1 indicates the challenges and problems researchers address in the context of GOT. The literature on GOT reveals a range of challenges targeting diverse outcomes. These encompass the (i) development of a student monitoring system, (ii) prediction among undergraduate and postgraduate students during and after their university studies, (iii) identification of crucial variables contributing to GOT, and (iv) comparison and optimization of predictive models. Existing research focused on identifying the important variables contributing to GOT and determining the likelihood of students graduating on time. Assessing these challenges and solutions presented by researchers illuminates various gaps and opportunities for improvement in existing studies.

Notably, a predominant focus of existing research lies in predicting the likelihood of undergraduate students graduating on time, with limited attention given to postgraduate students [30], [39], [40], [53], [61], [62]. For example, researchers noted a need for more research on the graduation of postgraduate students using historical data in Nigeria [39]. This limitation may arise because models developed for use in one country may perform less effectively in others, given the variability in educational systems and factors influencing postgraduate studies. Attributes may also change as students transition from undergraduate to postgraduate studies, and differences among students from various departments within high-level educational institutions further complicate the predictive modelling landscape. Despite these challenges, researchers underscored the critical importance of addressing delayed graduation in accreditation processes for both undergraduate and postgraduate students [40].

TABLE I
CHALLENGES AND PROBLEMS SOLVED BY RESEARCHERS IN THE CONTEXT OF GRADUATE ON TIME

Author	Student Monitoring Program	Predicting Graduate on Time among Postgraduate Students	Predicting Graduate on Time among Undergraduate Students	Predicting Graduate on Time during First Semester	Predicting Graduate on Time during each Semesters	Predicting Graduate on Time at Final Semester	Identifying Important Variables Contributing to Graduate on Time	Predictive Models Comparison	Predictive Models Optimization
[3]			✓					✓	
[4]			✓						✓
[12]			✓					✓	
[13]			✓					✓	✓
[14]			✓			✓		✓	
[15]			✓		✓				
[16]			✓			✓			
[17]			✓			✓	✓		✓
[18]			✓			✓			✓
[19]			✓				✓		
[20]			✓				✓		
[21]			✓			✓	✓		
[22]			✓			✓	✓		
[23]			✓				✓		
[24]							✓		
[25]			✓			✓	✓	✓	✓
[26]			✓						
[27]			✓					✓	
[28]	✓		✓						
[29]	✓		✓						
[30]	✓	✓							
[31]			✓	✓			✓		
[32]			✓				✓	✓	
[33]			✓						✓
[34]			✓		✓		✓	✓	
[35]			✓		✓		✓	✓	
[36]			✓	✓				✓	
[37]			✓	✓				✓	
[38]			✓				✓	✓	✓
[39]		✓						✓	
[40]		✓		✓				✓	
[41]			✓			✓		✓	
[42]			✓					✓	
[43]			✓						✓
[44]			✓		✓			✓	
[45]			✓					✓	✓
[46]			✓			✓			
[47]			✓			✓		✓	
[48]			✓						
[49]			✓			✓	✓		
[50]			✓		✓				
[51]	✓		✓			✓			
[52]			✓			✓	✓		
[53]		✓				✓		✓	✓
[54]			✓				✓		
[55]			✓						✓
[56]			✓			✓			
[57]			✓	✓			✓		
[58]	✓		✓	✓			✓		
[59]			✓						
[60]			✓			✓		✓	

Moreover, researchers recommended exploring opportunities for different performance focuses, such as predicting academic performance and student admission for postgraduate programs like Master's and Ph.D, as only limited research has targeted postgraduate students, with most studies concentrating on the undergraduate-level [62]. This inclination towards undergraduate studies suggests an avenue for further exploration and research in postgraduate education, identifying potential disparities and unique challenges that may impact timely graduation for this specific student demographic.

In the landscape of identifying the timely graduating on time, most researchers have concentrated on determining the timely graduating on time at the final semester of studies. The researchers determine the students with a higher probability of graduating on time by examining the Grade Point Average (GPA) for each semester, focusing particularly on the final semester [6], [16] – [18], [21], [41], [46], [47], [51], [52], [57], [60]. Although most research is conducted at the final semester, minority research is conducted at the phase before and during the semesters. One of the studies suggested that the students are more predisposed to complete their studies when there was a progression from the first year to the second year of university studies [8]. This is supported by other studies indicating a clear correlation between success in university mathematics and A-level mathematics among first year undergraduate students with pre-university study factors [63]. Nevertheless, the researchers did not consider the final degree due to the limited number of graduate students in the dataset collected.

B. Identifying the Important Variables Contributing to Graduate on Time

1) Feature selection techniques in identifying the critical variables contributing to graduating on time: Understanding the variables contributed to GOT provides benefits to educational institutions in managing and developing policies for those students who are facing challenges in graduating on time [21], [38]. To identify the critical variables contributing to graduating on time, researchers employed feature selection techniques using a high-dimensional dataset collected, as indicated in Table 2.

Pearson correlation is implemented in the study by [34] select the top 5 highly correlated variables among transactional course registration data and academic performance from 2006 to 2015. Furthermore, relevant courses are chosen to prevent overfitting and improve prediction performance using Pearson correlation based on the behavioral information and academic performance of undergraduate students at Beijing University [25]. On the other hand, IG is compared with Pearson correlation in identifying the critical variables to predict GOT when variables such as gender and GPA for each year were used [20].

Additionally, a study by [49] employed IG to compute the entropy in measuring the impurity in each variable and split it into each internal node of the decision tree. The feature selection is performed after cleaning and eliminating missing values in variables due to the numerous values in blood type, parent city, parent province, school hometown, and school province from year 2010 to 2016 collected. Other researchers implemented a similar approach but replaced the missing values based on the previous value on each variable [52]. The researchers selected the essential variables based on the student academic data from 2016 to 2019 collected from the Center for Information Technology and Database (PTIPD) of Universitas Isam Negeri Sultan Syarif Kasim Riau. These selected variables will be used as input features in predictive models to provide meaning and helpful information in predicting GOT.

TABLE II
FEATURE SELECTION TECHNIQUES USED BY RESEARCHERS IN IDENTIFYING THE ESSENTIAL VARIABLES TOWARDS GRADUATE ON TIME

Author	Chi-squared	Pearson Correlation	Information Gain	C4.5	Extreme Gradient Boosting	Classification and Regression Tree	Logistic Regression	Univariate Selection	Forward Selection	Ant Colony Optimization	Genetic Algorithm	Random Survival Forest
[6]								✓				
[17]									✓			
[19]												✓
[20]		✓	✓									
[21]							✓					
[22]				✓								
[23]	✓											
[24]							✓					
[25]		✓										
[31]							✓					
[32]				✓								
[34]		✓										
[35]					✓							
[38]	✓		✓									
[49]			✓									
[52]			✓						✓			
[53]										✓	✓	
[54]						✓						
[55]									✓			
[57]	✓											
[58]							✓					

Other than these two methods, studies by [23] and [57] used the Chi-square method to select the important variables contributing to GOT. The researchers performed the Chi-square method based on external and internal factors such as gender, religion, faculty, and GPA for first-year collected in the Universitas Advent Indonesia (UNAI) database [57]. Similarly, a study by [23] conducted a Chi-square method based on several variables collected to study the effect of demographic and high-school performance, such as the GPA on GOT. For further analysis, another study by [38] compared the IG and Chi-square methods to rank the 29 variables in ascending order based on the dataset collected with five years of academic information from a university database. Their study collects pre-college and post-admission data, such as admission test scores, demographic information, and GPA for the first semester. Before this feature selection technique is implemented and compared, data preprocessing is performed to remove the missing and duplicate records, normalize data, and detect outliers.

In addition to filter-based feature selection, a study [32] implemented the C4.5 algorithm to select the essential variables using the student information from 2017 and 2018 who completed the Final Project. Based on their study, the critical variable is chosen based on the variables in the root with the highest gain value. After collecting the student records from 2012 to 2014, the researchers performed a feature selection method using the C4.5 algorithm to select the essential variables from several variables collected, such as gender and GPA. Other than using Extreme Gradient Boosting (XGBoost) [35], Classification and Regression Tree (CART) [54] and Logistic Regression [21], [24], [31], [58], the researchers implemented Univariate Feature Selection to identify significant variables to predict GOT among several variables in the dataset collected [6]. The dataset includes variables such as name, gender, student status, enrolment year, parents' education level, parents' income, educational track, student's internal and external activities, Grade Point average from the first semester until the seventh semester, and GPA.

To improve the performance of Naïve Bayes, Forward Selection is performed to select five selected features from 12 variables, such as date of birth and GPA [17]. A study by [55] performed Forward Selection to identify significant variables based on the course scores of 4 semesters for engineering students from the year 2013 to 2015 batches. On the other hand, Ant Colony Optimization (ACO) is implemented to select the most significant variables based on 398 PhD postgraduate students and 22 variables such as the grade at degree and master level, gender, and type of course extracted from a local university [53]. Based on their study, the most significant variables are selected based on the highest fitness value obtained by each variable. Additionally, a study by [19] using gender, parents' occupation, and GPA as the variables in their study, they compared the Cox Proportional Hazards Regression and the Random Survival Forest method to select significant factors affecting GOT among undergraduate students in 2017.

In summary, most researchers focused on using filter-based feature selection methods such as Information Gain (IG), Chi-square, and Pearson Correlation in determining students who were more disposed to graduate on time as these methods execute faster than wrappers [20], [23], [25], [38], [49], [52], [57]. By focusing on the filter-based feature selection methods, researchers expedite the identification of key variables contributing to GOT, thus informing targeted interventions and policies to support student success. While wrapper methods offer alternative approaches to feature selection, filter-based techniques remain favored for their speed and scalability, enabling researchers to navigate the complexities of educational datasets.

2) Important variable contributing to graduating on time (See Table 3 for the literature supported): Grade Point Average (GPA) serves as a pivotal variable in identifying students who are apt to graduate on time, as highlighted by various filter-based feature selection methods [38], [49]. For instance, research by [49] demonstrated that the GPA for each semester was selected as one of the influential variables using IG. Similarly, a research by [57] conducted a Chi-square analysis, concluding GPA is paramount in determining whether a student graduates on time, irrespective of gender. Besides, another research [23] indicated that the high school degree and grades could be strongly associated with the student graduation rates. Conversely, according to [34], Pearson correlation analysis revealed that the number of credit hours taken was one of the highly correlated variables with graduation time.

TABLE III
LITERATURE OF ESSENTIAL VARIABLES CONTRIBUTING TO GRADUATING ON TIME

Author	**Grade Point Average**	**Gender**	**Number of credit hours**	**Age**	**High school grades**	**Parent' s income**	**Cumulative Grade Point Average**
[6]	✓					✓	
[17]	✓						
[19]		✓					
[20]		✓					
[21]	✓	✓		✓			
[22]	✓						
[23]					✓		
[24]					✓		
[25]	✓						
[31]		✓					
[32]	✓					✓	
[34]			✓				
[35]	✓				✓		
[38]	✓						
[49]	✓						
[52]	✓						
[53]							✓
[54]	✓	✓			✓		
[55]	✓						
[57]	✓						
[58]		✓					

In alignment with the filter-based feature selection methods findings, a study by [35] stated that a higher high school GPA significantly correlates with timely graduation, with factors such as gender and income exerting less influence when the researchers analyzed the feature importance of XGBoost. This finding was supported by another study conducted by [54]. In their study, the high school GPA substantially influences GOT utilizing CART because it helps convey how well students perform during the preliminary year. Additionally, the researchers mentioned that female students perform better in the first semester of university, and the students who attended university sooner were significantly more inclined to complete their studies promptly compared to those who enrolled more than a year after completion. Furthermore, another study by [21] examined the student demographics factors such as age at graduation and gender as well as the GPA for each year to determine the on-time graduation in the pharmacy program. In their study, the researchers concluded that students with higher GPAs throughout the initial one to two years of pharmacy school were more inclined to graduate on time. Based on other findings by [24], the researchers found that students with high school exam grades were less likely to drop out of Science, Technology, Engineering, and Mathematics (STEM) higher education in the first year.

Nevertheless, there is little association between high school exam grades and success in studying. Besides that, the researchers illustrated that GPA is the most influential factor in predicting student graduation using C4.5 [22], [32]. In addition to that, the forward selection method selected semester GPA as the essential variable [52].

Although most studies indicated that GPA was the most influential variable, other studies concluded that gender could be one of the most influential variables in identifying GOT. Using the Random Survival Forest method, gender was selected as one of the essential factors influencing the study periods for undergraduate students [19]. Studies by [31] and [58] revealed that gender was one of the significant variables in predicting those liable to complete their program within the specified time frame using Logistic Regression. Another study supported this by asserting gender as the essential feature in predicting students who graduate on time using IG, with average female students statistically outperforming their male counterparts [20]. However, a study by [24] presented a contrasting perspective, suggesting that female students perform comparably to male students regarding STEM graduation within ten years despite female students being less likely to graduate than their male counterparts.

Other than GPA and gender, another study [32] illustrated that the parent's income could be related to the student's graduation, achieving the highest correlation using univariate selection. They mentioned that additional expenses in extending the studies could motivate students with low parental income to graduate on time. Furthermore, the number of current semesters and the Cumulative Grade Point Average (CGPA) of the Bachelor level were selected as the influential variables in determining whether a student graduates on time rather than the CGPA of the intake using a metaheuristic algorithm, Ant Colony Optimization (ACO) [53].

The plethora of variables identified by researchers using various feature selection methods underscores the complexity of factors influencing GOT. While demographic factors such as gender, parent income, and high school performance consistently emerge as influential, nuanced variations exist across universities and programs. Thus, the researchers advocate for the inclusion of additional features to analyze their impact on GOT comprehensively [15], [21], [25], [34], [35], [37], [39], [42], [[44], [49], [57], [59]. By analyzing these findings, it becomes evident that the multifaceted nature of factors contributing to timely graduation requires a nuanced understanding and tailored strategies to address the specific challenges faced by students in different contexts.

C. Graduate on Time Prediction

1) Artificial Intelligence Techniques in predicting graduates on time: In addition to identifying the crucial variables influencing GOT, researchers have directed their attention towards applying AI techniques to detect students who are more likely to graduate on time, as outlined in Table 4. The tabulated data highlights techniques such as Naïve Bayes (NB), C4.5, and Decision Trees, which have garnered substantial attention in the research landscape. Conversely, less common approaches, such as CART and XGBoost, have been relatively underutilized in detecting GOT, showcasing a potential avenue for exploration and diversification in applying AI methodologies.

Decision Tree (DT), Random Forest (RF), NB, Support Vector Machine (SVM) with PolyKernel and RBFKernel were trained and compared based on the 74670 historical student information with 31 variables collected from Universiti Teknologi MARA (UiTM) [3]. Their study analyzes the predictive models through cross-validation (CV) with various partition sizes, evaluating performance based on accuracy, precision, recall, and F1-score. On the other hand, the researchers performed a comparison between RF, NB, SVM, K-Nearest Neighbor (KNN), Adaptive Boosting (AdaBoost), and LR after duplicate and inconsistent values were cleaned [12].

TABLE IV
ARTIFICIAL INTELLIGENCE TECHNIQUES USED BY RESEARCHERS IN PREDICTING GRADUATE ON TIME

Author	**C4.5**	**Classification and Regression Tree**	**Naïve Bayes**	**Logistic Regression**	**Decision Trees**	**Random Forest**	**Support Vector Machine**	**Adaptive Boosting**	**Extreme Gradient Boosting**	**K-Nearest Neighbors**
[3]			✓		✓	✓	✓			
[4]							✓			
[6]				✓	✓		✓	✓		
[11]	✓				✓					
[12]			✓	✓		✓	✓	✓		✓
[13]			✓	✓	✓	✓	✓			
[14]			✓	✓						
[15]										✓
[16]	✓				✓					
[17]			✓							
[18]					✓					
[20]						✓				
[22]	✓				✓					
[25]			✓							
[26]			✓							
[27]			✓							✓
[28]	✓				✓					
[29]		✓								
[31]			✓							
[32]	✓		✓		✓					
[33]			✓							
[34]				✓	✓	✓				
[35]				✓					✓	
[36]			✓	✓	✓					
[37]			✓	✓	✓	✓	✓			
[38]				✓						
[39]					✓		✓			
[40]			✓	✓	✓	✓	✓			✓
[41]	✓				✓					✓
[42]	✓		✓		✓	✓				✓
[43]							✓			
[44]					✓	✓	✓			
[45]			✓			✓	✓			✓
[46]	✓				✓					
[47]							✓			
[49]			✓							
[50]			✓							
[51]						✓				
[52]			✓							
[53]							✓			
[54]		✓								
[56]			✓							
[57]	✓				✓					
[58]					✓					
[59]	✓				✓					
[60]	✓				✓			✓		

Furthermore, a similar approach has been constructed when NB, LR, DT, RF, and SVM were constructed [13], [37], [40]. Nevertheless, research by [40] focused on predicting GOT among postgraduate students by collecting a sample of 1257 engineering students from the last five years who have completed their master's program credits and registered in the final project process based on student, academic, and study program information. The researchers mentioned that most studies have focused on student dropout at different levels but not on the graduation stage, especially postgraduate students.

Notably, NB has been used individually in predicting the likelihood of students graduating on time widely due to its simplicity [17], [25] – [27], [31], [33], [49], [50], [52], [56]. NB is a probabilistic method based on Bayes' Theorem, operating that attributes are conditionally independent when considering the class label [17], [31], [33]. The researchers constructed NB based on the student demographic data, entrance examination, and grades in the year's first semester [31]. In their study, conditional probabilities for each variable are computed using WEKA based on the occurrence of each category item and the total number of evidence per class label. In addition to that, [33] addressed the superiority of NB in terms of speed and accuracy when a huge dataset is applied. The researchers computed the probability of sample conditions in the dataset and evaluated NB based on the accuracy, precision, and recall, determining the GOT for Information Systems (IT) students at the Universitas Dirgantara Marsekal Suryadarma. [26] constructed and evaluated NB based on accuracy, precision, recall, and error rate when only 114 training and 127 testing data were used. To study the gap between the students in 2015 and 2016 graduating on time, the researchers computed the average difference between those IT students who graduated on time and late.

Boosting algorithms have been implemented in the prediction of GOT, such as AdaBoost [6], [12], [60] and XGBoost [35]. AdaBoost enhances performance by combining multiple weak learners and focusing on complex cases through weight adjustments [60], while XGBoost is an advanced, scalable gradient-boosting algorithm that incorporates features such as regularization, parallel processing, and tree pruning [35]. The researchers implemented the AdaBoost algorithm to improve the C4.5 and C5.0 algorithms in determining the GOT when 140 samples of graduated students and ten variables, including the GPA of each semester and gender, were used and collected from the UNTAN Statistics Study Program Period I of the 2017/2018 to Period II of the 2022/2023 Academic Year [60]. In their study, C4.5 and C5.0 algorithms are boosted by updating and normalizing the data weights until the error value is less than 0.5 and the maximum iteration. Besides that, [35] compared the performance of LR and a gradient boosting machine, XGBoost, to determine the GOT based on the dataset collected from registration information of large enrollment in American research universities such as gender, race, GPA, total number of credit hours taken, and income. Due to the missing records in the dataset, the researchers imputed the missing values with the means during the rescaling step when LR was used. In contrast, the missing values are not imputed prior to fitting the XGBoost model since it has a built-in imputation engine.

2) Model Evaluation Metrics: After the predictive models are constructed, the researchers evaluated the predictive models based on accuracy, precision, recall, and F1-score as described in Table 5. Besides that, the Area Under the Curve (AUC) is computed to indicate the trade-off between correctly predicted positive classes and incorrectly predicted negative courses. In the context of evaluation metrics, four crucial measurements are employed: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). TP represents the total number of positive classes correctly identified, while TN refers to the accurate identification of negative classes. If the model incorrectly predicts the negative class as positive, it results in FP. Conversely, FN denotes the total instances where positive classes are mistakenly predicted as negative. These metrics offer an in-depth insight into the models' effectiveness, revealing their ability to classify positive and negative cases accurately.

TABLE V
PERFORMANCE METRICS IN EVALUATING PREDICTIVE MODELS

Metric	Formula	Description
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	Evaluates the number of correct predictions from a model.
Precision	$\frac{TP}{TP+FP}$	Evaluates the percentage of positive predicted cases that are positive.
Recall	$\frac{TP}{TP+FN}$	Measures the total number of the positive cases that are captured by the positive predictions.
F1-score	$2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$	Represents the harmonic mean of precision and recall.

3) Model Performance: By performing an analysis of students from the year 2008 until 2014, a study by [11] achieved 54.76% average recall and 57.77% F1-score using the C4.5 algorithm, although the researchers achieved 90% average accuracy. To achieve the best result of performance metrics, the researchers recommended feature selection techniques. On the other hand, the Synthetic Minority Oversampling Technique (SMOTE) is recommended to resolve the class imbalance issue in the data used, although the researchers achieved the highest accuracy (93.10%) using C4.5 [28]. By applying SMOTE, [57] improved around 2% accuracy of C4.5. The researchers suggested using different algorithms to address the class imbalance issue in further improving the performance. Other than feature selection and class imbalance treatment techniques are suggested, studies by [16] and [57] stated that a more significant size of training data could improve the accuracy of C4.5 in determining the likelihood of students towards GOT. In addition to that, NB, KNN, and SVM are suggested to be used as classification algorithms in identifying the GOT other than C4.5 [22].

NB has emerged as a widely implemented algorithm in the context of predicting GOT, as evidenced by several researchers [3], [12] – [14], [17], [25], [26], [31], [33], [36], [37], [40], [42], [45], [49], [50], [52], [56]. Notably, [32] it demonstrated the superiority of NB over C4.5, achieving higher accuracy rates with 34% and 25% testing data. This performance difference has prompted researchers to advocate for developing and comparing various classification methods to determine the likelihood of students graduating on time. However, it is essential to recognize that NB exhibited slightly lower performance metrics in direct comparison with other models. A research by [3] provided a notable example wherein SVM outperformed NB. The study illustrated that NB faced challenges in predicting students' graduation status, generating the lowest average performance among the classifiers assessed.

Conversely, researchers have highlighted the efficacy of RF over other models, such as SVM, and NB [12], [37], [40], [42], [44], [45]. Despite the accurate predictions achieved by NB, SVM, and RF, specific models demonstrated more incredible promise than others [13]. To further improve the performance in predicting the GOT, the research by [13] implemented an ensemble method using SVM, RF, NB, and LR, achieving superior accuracy exceeding 90%, surpassing all other classifiers. However, the increased predictive accuracy achieved through integrating diverse models within the ensemble method comes at the cost of substantial computational resources and time, which can hinder educational institutions from seeking timely predictions for GOT.

To navigate the challenges, simpler models such as LR emerge as a viable alternative for detecting GOT. In determining the students who have poor academic performance in the computer science subject offered, LR achieved 74.53% accuracy, 71.91% recall, and 74.87% F1-score, obtaining higher performance compared to NB [36]. Similarly, a study by [14] LR achieved higher accuracy than NB in predicting the student's academic performance. However, researchers mentioned that gradient-boosting algorithms predicted GOT more effectively than existing statistical algorithms [35]. XGBoost outperformed LR in predicting the likelihood of students graduating on time during their study semesters [35]. Beyond XGBoost, [6] concluded that AdaBoost and DT obtained the highest 82% F1-score, showcasing superior performance compared to other models like LR.

While researchers have showcased promising results in identifying students likely to graduate on time, it remains a non-trivial task due to the complexity of student behavior and the difficulty of handling missing data. The researchers addressed the challenges in the complexity of student behavior when numerous factors come into play in different datasets. A prevailing recommendation among them is the necessity for further data collection, even alongside ongoing model refinement efforts employing various algorithms [21] – [23], [37], [43], [45], [54], [57]. The researchers mentioned that additional data underscores the complexity and multifaceted nature of factors influencing graduation outcomes, suggesting that a more comprehensive dataset could yield more profound insights and more robust predictions. Besides that, predicting GOT was hindered by the biases of data due to the missing records when the missing records correlate with the target variable [35]. Different missing imputation methods, such as mean, median, and mode, could contribute to different biases, leading to misrepresentation in predicting GOT.

In the context of class imbalance, SMOTE is recommended in their studies for future research to create more data by generating synthetic samples [43, 45, 57], aiming to address the class imbalance in the dataset and enhance the model's performance in predicting GOT. This collective call for expanded data collection and utilization of advanced techniques underscores the commitment to refining predictive models and improving their efficacy in supporting student success.

III. Result and Discussion

While researchers have successfully identified the important variables using various feature selection methods, there is still room for improvement in this research area. Most studies have applied feature selection methods on a dataset with limited features, typically focusing on the demographic and course achievements, such as GPA for each semester. In such cases, the variables selected by these feature selection methods might need to be more significant in other datasets with more features. Moreover, these studies lacked the investigation of other critical variables, such as financial, environmental, and psychological factors that could contribute to GOT. Therefore, it is crucial to include additional features, such as student location information, during data collection to encompass a wide range of potential significant factors related to GOT.

In addition to broadening the scope of features, it would be intriguing to study the impact of different subsets of features on GOT and compare the computational complexity of various feature selection methods, especially when dealing with datasets containing a vast number of features. The efficiency of feature selection methods can vary significantly based on the dataset's complexity and the subset of features under consideration.

Additionally, many researchers have implemented feature selection methods on the dataset with limited records, primarily focusing on specific universities. This approach may lead to an inaccurate representation of the impact of selected features on GOT, as the chosen university needs to accurately reflect the diversity of universities across the country. Thus, collecting more extensive data encompassing different universities, study years, and programs is strongly recommended to ensure a more comprehensive understanding of the factors influencing timely graduation on a broader scale. This approach would enhance the generalizability and applicability of the findings to a broader educational context.

While existing research has proved the superiority of predictive models such as LR, DT, and boosting algorithms to ascertain students with a higher probability of finishing their studies, specific issues, such as the challenge of learning from small amounts of data, needed to be addressed. Many studies have focused on identifying GOT among undergraduate or postgraduate students with limited information and records. Although class imbalance treatment such as SMOTE could be applied to the imbalanced data, insignificant data sizes remain one of the challenges to the learning performance of predictive models. Applying SMOTE to address the class imbalance and insufficient data can be challenging, mainly when dealing with smaller datasets. The generation of synthetic data may lead to overlapping regions, resulting in less useful information being learned and captured by predictive models. This is particularly true when noise and overlapping areas of the data distribution significantly influence the learning performance in finding an accurate decision boundary [64] – [67]. Thus, it is essential to explore and implement alternative class imbalance treatment methods that can effectively address these challenges and enhance the performance of predictive models in identifying students who are more likely to graduate on time. Comparative analysis of different imbalance treatment methods can contribute to determining the most effective approach for improved model performance in scenarios involving data sizes.

IV. Conclusion

In this review, the significance of timely graduation as a key metric has been highlighted to evaluate educational success, emphasizing its impact on graduation rates and institutional performance. Researchers have successfully identified and addressed various challenges, including developing student monitoring systems, prediction among undergraduate and postgraduate students, identification of important variables contributing to GOT, and model comparison and optimization. Despite these advancements, this review has identified several gaps and opportunities for improvement in the existing body of research. Notably, a predominant focus on predicting the likelihood of GOT among undergraduate students has limited the exploration of postgraduate student scenarios.

Moreover, there is room for enhancement in considering a broader array of financial, environmental, and psychological features to provide a more holistic understanding of the factors influencing the GOT. While SMOTE has been a common approach in addressing class imbalance, exploring alternative class imbalance treatment methods could further improve model performance when dealing with a smaller dataset. Further research involves incorporating additional features, exploring different class imbalance treatments, and improving the performance of predictive models. As the educational landscape evolves, addressing these considerations will aid in creating more efficient strategies and interventions, ensuring a higher likelihood of a student achieving timely graduation across diverse academic settings.

Acknowledgment

The researchers thank for the financial support provided by the TM Research and Development Grant (TM R&D), MMUE/220028.

References

[1] K. Anwar, H. Hanafiah, and A. Ebun, "Predicting Student Graduation Using Artificial Neural Network: A Preliminary study of Diploma in Accountancy Program at UiTM Sabah," 2020.

[2] G. Sidhu, S. Kannan, A. S. Samsul Kamil, and R. Du, "Sustaining Students' Quality Learning Environment by Reviewing Factors to Graduate-on-Time: A case study," *Environment-Behaviour Proceedings Journal*, vol. 8, pp. 127–133, 2023, doi:10.21834/ebpj.v8i24.4649.

[3] N. Mohammad Suhaimi, S. Abdul-Rahman, S. Mutalib, N. H. Abdul Hamid, and A. Md Ab Malik, "Predictive model of graduate-on-time using machine learning algorithms," in *Soft Computing in Data Science: 5th International Conference, SCDS 2019, Iizuka, Japan, August 28–29, 2019, Proceedings 5*, 2019, pp. 130–141. doi:10.1007/978-981-15-0399-3_11.

[4] K. T. Chui, D. C. L. Fung, M. D. Lytras, and T. M. Lam, "Predicting at-risk university students in a virtual learning environment via a machine learning algorithm," *Comput Human Behav*, vol. 107, p. 105584, 2020, doi: 10.1016/j.chb.2018.06.032.

[5] R. Garcia-Ros, F. Pérez-González, F. Cavas-Mart\'\inez, and J. M. Tomás, "Effects of pre-college variables and first-year engineering students' experiences on academic achievement and retention: a structural model," *International Journal of Technology and Design Education*, vol. 29, pp. 915–928, 2019. doi: 10.1007/s10798-018-9466-z.

[6] A. Desfiandi and B. Soewito, "Student Graduation Time Prediction using Logistic Regression, Decision Tree, Support Vector Machine, and AdaBoost Ensemble Learning," *International Journal of Information System and Computer Science*, vol. 7, no. 3, pp. 195–199, 2023. doi: 10.56327/ijiscs.v7i2.1579.

[7] T. S. Hoon, G. Narayanan, and G. K. Sidhu, "Motivation to graduate on time: A case study in Malaysia," *Pertanika Journal of Social Sciences and Humanities*, vol. 27, no. 4, pp. 2417–2439, 2019.

[8] J. G. Lisciandro, "First-year university retention and academic performance of non-traditional students entering via an Australian pre-university enabling program," *Australian Journal of Adult Learning*, vol. 62, no. 2, pp. 167–201, 2022.

[9] P. Muthukrishnan, K. Gurnam, T. Hoon, N. Geethanjali, and Y. F. Chan, "Key Factors Influencing Graduation on Time Among Postgraduate Students: A PLS-SEM Approach," *Asian Journal of University Education*, vol. 18, p. 51, 2022, doi:10.24191/ajue.v18i1.17169.

[10] E. N. Okwuduba, K. C. Nwosu, E. C. Okigbo, N. N. Samuel, and C. Achugbu, "Impact of intrapersonal and interpersonal emotional intelligence and self-directed learning on academic performance among pre-university science students," *Heliyon*, vol. 7, no. 3, 2021. doi: 10.1016/j.heliyon.2021.e06611.

[11] H. Yuliansyah, R. A. P. Imaniati, A. Wirasto, and M. Wibowo, "Predicting students graduate on time using C4. 5 algorithm," *Journal of Information Systems Engineering and Business Intelligence*, vol. 7, no. 1, pp. 67–73, 2021. doi: 10.20473/jisebi.7.1.67-73.

[12] F. T. Anggraeny, A. K. Darmawan, A. Anekawati, I. Yudhisari, and others, "Early Prediction for Graduation of Private High School Students with Machine Learning Approach," 2023. doi:10.47577/technium.v16i.9971.

[13] Z. Bitar and A. Al-Mousa, "Prediction of Graduate Admission using Multiple Supervised Machine Learning Models," in *2020 SoutheastCon*, 2020, pp. 1–6. doi:10.1109/SoutheastCon44009.2020.9249747.

[14] M. N. Razali, H. Zakariah, R. Hanapi, and E. A. Rahim, "Predictive Model of Undergraduate Student Grading Using Machine Learning for Learning Analytics," in *2022 4th International Conference on Computer Science and Technologies in Education (CSTE)*, 2022, pp. 260–264. doi: 10.1109/CSTE55932.2022.00055.

[15] A. P. Salim, K. A. Laksitowening, and I. Asror, "Time Series Prediction on College Graduation Using KNN Algorithm," in *2020 8th International Conference on Information and Communication Technology (ICoICT)*, 2020, pp. 1–4. doi:10.1109/ICoICT49345.2020.9166238.

[16] E. Haerani, F. Syafria, F. Lestari, N. Novriyanto, and I. Marzuki, "Classification Academic Data using Machine Learning for Decision Making Process," *Journal of Applied Engineering and Technological Science (JAETS)*, vol. 4, no. 2, pp. 955–968, 2023, doi:10.37385/jaets.v4i2.1983.

[17] F. Nuraeni, Y. H. Agustin, S. Rahayu, D. Kurniadi, Y. Septiana, and S. M. Lestari, "Student Study Timeline Prediction Model Using Naïve Bayes Based Forward Selection Feature," in *2021 International Conference on ICT for Smart Society (ICISS)*, 2021, pp. 1–5. doi:10.1109/ICISS53185.2021.9532502.

[18] A. Maulana, "Prediction of student graduation accuracy using decision tree with application of genetic algorithms," *Institute of Physics Conference Series: Materials Science and Engineering*, vol. 1073, no. 1, p. 12055, 2021, doi: 10.1088/1757-899X/1073/1/012055.

[19] M. A. A. Arib, "Survival Analysis of Students Not Graduated on Time Using Cox Proportional Hazard Regression Method and Random Survival Forest Method," *Journal of Statistics and Data Science*, pp. 13–21, 2023. doi: 10.33369/jsds.v2i1.24312.

[20] S. Noviaristanti, G. Ramantoko, A. T. Hadi, and A. Inayati, "Predictive Model of Student Academic Performance in Private Higher Education Institution (Case in Undergraduate Management Program)," in *2022 International Conference on Data Science and Its Applications (ICoDSA)*, 2022, pp. 262–267. doi:10.1109/ICoDSA55874.2022.9862822.

[21] C. A. Spivey, M. A. Chisholm-Burns, and J. L. Johnson, "Factors Associated with Student Pharmacists' Academic Progression and Performance on the National Licensure Examination," *American Journal of Pharmaceutical Education*, vol. 84, no. 2, p. 7561, 2020, doi: 10.5688/ajpe7561.

[22] G. Hanes, and C. Catherine, "Information Systems Students' Study Performance Prediction Using Data Mining Approach," in *2019 Fourth International Conference on Informatics and Computing*, 2019, pp. 1–8. doi: 10.1109/ICIC47613.2019.8985718.

[23] A. Tamimi *et al.*, "Admission criteria and academic performance in medical school," *BMC Medical Education*, vol. 23, no. 1, p. 273, 2023. doi: 10.1186/s12909-023-04251-y.

[24] M. Vooren, C. Haelermans, W. Groot, and H. M. van den Brink, "Comparing success of female students to their male counterparts in the STEM fields: an empirical analysis from enrollment until graduation using longitudinal register data," *International Journal of STEM Education*, vol. 9, no. 1, pp. 1–17, 2022. doi: 10.1186/s40594-021-00318-8.

[25] K. and H. X. and L. N. and P. X. and G. P. and J. R. Jia Baoting and Niu, "Prediction for Student Academic Performance Using SMNaive Bayes Model," in *Advanced Data Mining and Applications*, S. and Q. S. and L. X. and W. S. Li Jianxin and Wang, Ed., Springer International Publishing, 2019, pp. 712–725. doi: 10.1007/978-3-030-35231-8_52.

[26] A. Meiriza, E. Lestari, P. Putra, A. Monaputri, and D. A. Lestari, "Prediction Graduate Student Use Naive Bayes Classifier," in *Proceedings of the Sriwijaya International Conference on Information Technology and Its Applications (SICONIAN 2019)*, Atlantis Press, 2020, pp. 370–375. doi: 10.2991/aisr.k.200424.056.

[27] M. Windarti and P. T. Prasetyaninrum, "Prediction Analysis Student Graduate Using Multilayer Perceptron," in *Proceedings of the International Conference on Online and Blended Learning 2019 (ICOBL 2019)*, Atlantis Press, 2020, pp. 53–57. doi:10.2991/assehr.k.200521.011.

[28] D. Kurniawan, A. Anggrawan, and H. Hairani, "Graduation Prediction System on Students Using C4.5 Algorithm," *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 19, no. 2, pp. 358–366, 2020, doi: 10.30812/matrik.v19i2.685.

[29] A. Maesya and T. Hendiyanti, "Forecasting Student Graduation with Classification and Regression Tree (CART) Algorithm," *Institute of Physics Conference Series: Materials Science and Engineering*, vol. 621, no. 1, p. 12005, 2019, doi: 10.1088/1757-899X/621/1/012005.

[30] S. Joseph, E. Dada, D. Mshelia, and A. Hamidu Alkali, "Students Graduation on Time Prediction Model Using Artificial Neural Network," vol. 21, pp. 28–35, 2019, doi: 10.9790/0661-2103012835.

[31] A. C. Lagman *et al.*, "Embedding Naive Bayes Algorithm Data Model in Predicting Student Graduation," in ICTCE '19. Association for Computing Machinery, 2020, pp. 51–56. doi:10.1145/3369555.3369570.

[32] D. Dikriani and A. T. I. Karim, "Comparison of C4. 5 and Naive Bayes Algorithm Methods in Prediction of Student Graduation on Time (Case Study: Information Systems Study Program)," *Journal of Dinda: Data Science, Information Technology, and Data Analytics*, vol. 3, no. 1, pp. 40–44, 2023. doi: 10.20895/dinda.v3i1.782.

[33] M. Muryan Awaludin, "Optimization of Naive Bayes Algorithm Parameters for Student Graduation Prediction at Universitas Dirgantara Marsekal Suryadarma," *Journal of Information System, informatic and Computing*, vol. 6, no. 1, pp. 91–106, 2022. doi:10.52362/jisicom.v6i1.785.

[34] H. S. Brdesee, W. Alsagaff, N. Aljohani, and S.-U. Hassan, "Predictive model using a machine learning approach for enhancing the retention rate of students at-risk," *International Journal on Semantic Web and Information Systems (IJSWIS)*, vol. 18, no. 1, pp. 1–21, 2022. doi: 10.4018/IJSWIS.299859.

[35] J. M. Aiken, R. De Bin, M. Hjorth-Jensen, and M. D. Caballero, "Predicting time to graduation at a large enrollment American university," *The Public Library of Science One*, vol. 15, no. 11, p. e0242334, 2020. doi: 10.1371/journal.pone.0242334.

[36] H. Altabrawee, O. Ali, and S. Qaisar, "Predicting Students' Performance Using Machine Learning Techniques," *Journal of University of Babylon for pure and applied sciences*, vol. 27, pp. 194–205, 2019, doi: 10.29196/jubpas.v27i1.2108.

[37] A. Sadqui, M. Ertel, H. Sadiki, and A. Said, "Evaluating Machine Learning Models for Predicting Graduation Timelines in Moroccan Universities," *International Journal of Advanced Computer Science and Applications*, vol. 14, 2023, doi:10.14569/ijacsa.2023.0140734.

[38] J. Febro, "Utilizing Feature Selection in Identifying Predicting Factors of Student Retention," *International Journal of Advanced Computer Science and Applications*, vol. 10, 2019, doi:10.14569/ijacsa.2019.0100934.

[39] A. F. U., G. B. S., & G. M. Bako H. S., "Predicting Timely Graduation of Postgraduate Students using Random Forests Ensemble Method," vol. 7, pp. 177–185, 2023, doi: 10.33003/fjs-2023-0703-1773.

[40] D. Ruete *et al.*, "Early Detection of Delayed Graduation in Master's Students," *ASEE Annual Conference and Exposition, Conference Proceedings*, 2021. doi: 10.18260/1-2—36999.

[41] S. Suwitno and A. Wibowo, "On-Time Graduation Prediction System Using Data Mining Classification Method," EAI, 2019. doi:10.4108/eai.20-1-2018.2281900.

[42] G. Gunawan, H. Hanes, and C. Catherine, "C4.5, K-Nearest Neighbor, Naïve Bayes, and Random Forest Algorithms Comparison to Predict Students' on TIME Graduation," *Indonesian Journal of Artificial Intelligence and Data Mining*, vol. 4, no. 2, pp. 62–71, 2021, doi:10.24014/ijaidm.v4i2.10833.

[43] A. Anggrawan, H. Hairani, and C. Satria, "Improving SVM Classification Performance on Unbalanced Student Graduation Time Data Using SMOTE," *International Journal of Information and Education Technology*, vol. 13, no. 2, pp. 289–295, 2023. doi:10.18178/ijiet.2023.13.2.1806.

[44] N. Suresh, V. Hashiyana, G. T. Nhinda, I. Stephanus, and P. Kautwima, "Graduates' Prediction System Using Artificial Intelligence," in *Proceedings of the International Conference on Data Science, Machine Learning and Artificial Intelligence*, in DSMLAI "21." New York, NY, USA: Association for Computing Machinery, 2022, pp. 317–327. doi: 10.1145/3484824.3484873.

[45] R. Bakri, N. P. Astuti, and A. S. Ahmar, "Machine Learning Algorithms with Parameter Tuning to Predict Students' Graduation-on-time: A Case Study in Higher Education," *Journal of Applied Science, Engineering, Technology, and Education*, vol. 4, no. 2, pp. 259–265, 2022, doi: 10.35877/454RI.asci1581.

[46] A. Pradipta, D. Hartama, A. Wanto, S. Saifullah, and J. Jalaluddin, "The Application of Data Mining in Determining Timely Graduation Using the C45 Algorithm," *International Journal of Information System and Technology*, vol. 3, no. 1, pp. 31–36, 2019. doi:10.30645/ijistech.v3i1.30.

[47] Y. Yennimar, M. R. Faturrahman, S. Nesen, M. A. Guci, and S. R. Pasaribu, "Implementation of artificial neural network and support vector machine algorithm on student graduation prediction model on time," *Jurnal Mantik*, vol. 7, no. 2, pp. 925–934, 2023. doi:10.35335/mantik.v7i2.3992.

[48] S. P. Kristanto *et al.*, "Implementation of ML Rough Set in Determining Cases of Timely Graduation of Students," *Journal of Physics: Conference Series*, vol. 1933, no. 1, p. 12031, 2021, doi:10.1088/1742-6596/1933/1/012031.

[49] G. F. Halim, and Djoni, "Students' Timely Graduation Attributes Prediction Using Feature Selection Techniques, Case Study: Informatics Engineering Bachelor Study Program," in *2022 IEEE International Conference of Computer Science and Information Technology (ICOSNIKOM)*, 2022, pp. 1–7. doi:10.1109/icosnikom56551.2022.10034873.

[50] C. Kuncoro, "Analysis of UMN Student Graduation Timeliness Using Supervised Learning Method," *IJNMT (International Journal of New Media Technology)*, vol. 8, no. 2, pp. 89–95, 2022, doi:10.31937/ijnmt.v8i2.2366.

[51] D. A. Rachmawati, N. A. Ibadurrahman, J. Zeniarja, and N. Hendriyanto, "Implementation of the Random Forest Algorithm in Classifying the Accuracy of Graduation Time for Computer Engineering Students at Dian Nuswantoro University," *Jurnal Teknik Informatika (Jutif)*, vol. 4, no. 3, pp. 565–572, 2023, doi:10.52436/1.jutif.2023.4.3.920.

[52] D. Fahrudy and others, "Classification of Student Graduation using Naive Bayes by Comparing between Random Oversampling and Feature Selections of Information Gain and Forward Selection," *JOIV: International Journal on Informatics Visualization*, vol. 6, no. 4, pp. 798–808, 2022. doi: 10.30630/joiv.6.4.982.

[53] S. R. Kumaran, M. S. Othman, L. M. Yusuf, and A. Yunianta, "Educational Business Intelligence Framework Visualizing Significant Features using Metaheuristic Algorithm and Feature Selection," in *2019 International Conference on Advances in the Emerging Computing Technologies (AECT)*, 2020, pp. 1–6. doi:10.1109/AECT47998.2020.9194221.

[54] A. Moraga-Pumarino, S. Salvo-Garrido, and K. Polanco-Levicán, "Profiles of University Students Who Graduate on Time: A Cohort

Study from the Chilean Context," *Behavioral Sciences*, vol. 13, no. 7, p. 582, 2023, doi: 10.3390/bs13070582.

[55] M. Anwar, "Prediction of graduation rate of engineering education students using Artificial Neural Network Algorithms," *Education (Chula Vista)*, vol. 5, no. 1, pp. 15–23, 2021. doi:10.24036/00411za0002.

[56] A. D. Rachmatsyah, B. Wijaya, and others, "Data Mining Predicts the Graduation of Students of STMIK Atma Luhur Information System Using Neive Bayes Algorithm," *Jurnal Mantik*, vol. 4, no. 3, pp. 2100–2105, 2020. doi: 10.35335/mantik.Vol4.2020.1087.pp2100-2105.

[57] Y. T. Samuel, J. J. Hutapea, and B. Jonathan, "Predicting the Timeliness of Student Graduation Using Decision Tree C4.5 Algorithm in Universitas Advent Indonesia," in *2019 12th International Conference on Information & Communication Technology and System (ICTS)*, 2019, pp. 276–280. doi:10.1109/ICTS.2019.8850948.

[58] Ma. C. G. Fernando, A. C. Lagman, M. V. S. Solomo, J. H. J. Ortega, M. L. I. Goh, and J. P. Lalata, "Development of Predictive Decision Support System for Student Graduation Using Decision Tree Algorithm," *International journal of simulation: systems, science & technology*, 2019. doi: 10.5013/IJSSST.a.20.S2.27.

[59] A. T. I. Karim and S. Sudianto, "Dominant Requirements for Student Graduation in the Faculty of Informatics using the C4. 5 Algorithm," *Journal of Dinda: Data Science, Information Technology, and Data Analytics*, vol. 3, no. 2, pp. 50–58, 2023. doi:10.20895/dinda.v3i2.1040.

[60] Y. Crismayella, N. Satyahadewi, and H. Perdana, "Comparison of Adaboost Application to C4.5 and C5.0 Algorithms in Student Graduation Classification," *Pattimura International Journal of Mathematics (PIJMath)*, vol. 2, no. 1, pp. 7–16, 2023, doi:10.30598/pijmathvol2iss1pp07-16.

[61] Y. Baashar *et al.*, "Evaluation of postgraduate academic performance using artificial intelligence models," *Alexandria Engineering Journal*, vol. 61, no. 12, pp. 9867–9878, 2022, doi: 10.1016/j.aej.2022.03.021.

[62] S. Jeganathan, S. Parthasarathy, A. R. Lakshminarayanan, P. M. Ashok Kumar, and Md. K. A. Khan, "Predicting the Post Graduate Admissions using Classification Techniques," in *2021 International Conference on Emerging Smart Computing and Informatics (ESCI)*, 2021, pp. 346–350. doi: 10.1109/ESCI50559.2021.9396815.

[63] N. C. King and G. Hambrook, "Examining the impact of pre-university qualifications on success in interdisciplinary science," *Journal of Further and Higher Education*, vol. 45, no. 9, pp. 1192–1205, 2021, doi: 10.1080/0309877X.2020.1854695.

[64] G. A. Pradipta, R. Wardoyo, A. Musdholifah, I. N. H. Sanjaya, and M. Ismail, "SMOTE for Handling Imbalanced Data Problem: A Review," in *2021 Sixth International Conference on Informatics and Computing (ICIC)*, 2021, pp. 1–8. doi: 10.1109/ICIC54025.2021.9632912.

[65] N. A. Azhar, M. S. Mohd Pozi, A. Mohamed Din, and A. Jatowt, "An Investigation of SMOTE based Methods for Imbalanced Datasets with Data Complexity Analysis," *Institute of Electrical and Electronics Engineers Transactions on Knowledge and Data Engineering*, p. 1, 2022, doi: 10.1109/TKDE.2022.3179381.

[66] M. M. Hussain, S. Akbar, S. A. Hassan, M. W. Aziz, and F. Urooj, "Prediction of Student's Academic Performance through Data Mining Approach," *Journal of Informatics and Web Engineering*, vol. 3, no. 1, pp. 241–251, 2024. doi: 10.33093/jiwe.2024.3.1.16.

[67] T. A. Khan, R. Sadiq, Z. Shahid, M. M. Alam, and M. B. M. Su'ud, "Sentiment Analysis using Support Vector Machine and Random Forest," *Journal of Informatics and Web Engineering*, vol. 3, no. 1, pp. 67–75, 2024. doi: 10.33093/jiwe.2024.3.1.5.