KOMTEKINFO JOURNAL
LPPM Universitas Putra Indonesia YPTK Padang
Jl. Raya Lubuk Begalung, Padang, West Sumatera, Indonesia, Zip Code: 25221
Volume: 13, Issue: 1, Page: 18 - 26, March 30th, 2026, e-ISSN: 2502-8758
Available online at website: https://jkomtekinfo.org/ojs/index.php/komtekinfo

Implementation of K-Means Algorithm and C4.5 Classification in Analysis of Determinants of Student Timely Graduation

Rahma Yanti[1]✉, Musli Yanto[2], Syafri Arlis[3]
[1] Universitas Islam Negeri Sjech M. Djamil Djambek Bukittinggi, 26181, Indonesia
[2] Universitas Putra Indonesia YPTK Padang, 25145, Indonesia
rahmayanti210602@gmail.com

Abstract

This study was motivated by the importance of timely graduation as a key parameter affecting program accreditation. The timely graduation rate reflects the effectiveness of academic management and serves as an indicator of program quality. The purpose of this study was to apply the concept of data mining using the K-means and Decision Tree C4.5 methods to analyze the timely graduation of students in the Information Technology and Computer Education Study Program at UIN Bukittinggi. The research methods used are the K-Means and Decision Tree C4.5 methods. The K-Means algorithm is used to cluster student graduation data, which will then be processed in the next method. The Decision Tree C4.5 algorithm is used to classify student graduation data. The research data was sourced from the 2017 batch of the Information Technology and Computer Education Study Program at UIN Bukittinggi, with a total of 158 data points. The results of this study produced a model that was able to achieve an accuracy rate of 96% in the validation process. The accuracy results were relatively high, so the model produced can be used by the study program to improve academic quality. Based on the results of this study, it contributes as a basis for evaluating student academic performance, monitoring the risk of study delays, and supporting academic decision-making. In addition, this information contributes to maintaining and improving academic quality and supports the achievement and maintenance of the accreditation status of the PTIK UIN Sjech M. Djamil Djambek Bukittinggi Study Program.

Keywords: timely graduation, students, K-Means Clustering, Decision Tree C4.5, modle

1. Introduction

The rapid development of information technology has brought significant changes to various sectors of life, including education[1]. Technology, which was once only used as a tool, has now become an integral part of the learning process and information management in educational institutions [2]. The use of technology not only speeds up access to information, but also opens up opportunities to apply new methods and techniques in data analysis that can help make more accurate and evidence-based decisions [3]. The presence of this technology ultimately encourages universities to continue to adapt and improve the quality of their education in order to produce superior human resources that are relevant to the needs of the times [4]. Higher education plays a very important role in producing educated generations who are able to adapt to the changing times [5]. Higher education institutions are not only places for transferring knowledge, but also serve as vehicles for skills development, research, and community service [5]. Through a structured educational process, students are prepared to have academic competence, professional expertise, and character in line with global demands. The quality of higher education determines the competitiveness of human resources, both nationally and internationally, making efforts to improve the quality of education an unavoidable necessity [4].

A number of previous studies reinforce the role of higher education in addressing challenges of quality and technological adaptation. Previous studies explain that curriculum transformation is important to maintain the relevance of education to social, economic, and technological changes. Through curriculum adjustments, universities can increase the competitiveness of graduates and respond to the needs of society in the modern era [6]. Other studies show the importance of technology in the implementation of the Merdeka Curriculum. The results of the study show that the use of technology not only supports the effectiveness of learning but is also a key instrument in strengthening education policy in the digital era [7].

Timely graduation is an important indicator in measuring student success and the effectiveness of the education system in higher education [8]. This concept refers to the ability of students to complete all of their coursework within the time frame specified in the curriculum, for example, four years for a bachelor's degree [9]. A high on-time graduation rate indicates that the learning process, academic services, and support systems in higher education are functioning

Submitted: xx-xx-20xx | Revised: xx-xx-20xx | Accepted: xx-xx-20xx | Published: xx-xx-20xx

well [10]. Conversely, a low on-time graduation rate can be problematic because it indicates obstacles in the study process, whether they originate from internal factors related to students or external factors such as the curriculum, learning methods, or academic management systems [10].

Several studies have examined the prediction of student graduation on time, including the first study which showed that the C4.5 algorithm was able to produce a fairly high accuracy rate, around 90%, in predicting the likelihood of a student completing their studies within the specified period. This finding provides an important contribution to universities in designing more targeted academic assistance strategies, especially for students who are potentially late in graduating [11]. Further research states that Random Forest is the best model with an accuracy of 85% and an AUC value of 0.875 [12].

Data mining is a solution for uncovering hidden patterns in big data that can provide important insights in various fields, including education [13]. Through the application of statistical techniques and machine learning, data mining enables the processing of student data to identify factors that influence academic outcomes, such as timely graduation [14]. With this approach, universities can analyze learning patterns, time management, and other supporting factors related to student academic success[15]. The results of this data mining analysis can be used to provide early intervention and optimize the academic support system to improve the effectiveness of the educational process[16].

The C4.5 Decision Tree algorithm is a widely used classification method in data mining to produce prediction models in the form of decision trees [17]. This method works by dividing data into branches based on specific attributes until a clear final decision is reached [17]. The main advantage of C4.5 is its ability to handle attributes with continuous and discrete values and generate classification rules that are easy to understand [18]. In educational research, Decision Tree C4.5 is often applied to analyze student data, such as academic performance, attendance, and learning achievements, so that it can help predict timely graduation in a more structured manner [19].

This is reinforced by research conducted by researchers, which shows that the K-Means algorithm is capable of dividing students into three main groups based on their academic characteristics. The first group consists of high-achieving students, characterized by a GPA above 3.5, a number of credits in line with the curriculum, and a consistent attendance rate. The second group consists of students with moderate achievements, who generally have a GPA ranging from 2.75 to 3.49, a number of credits that are relatively in line with the schedule, but an unstable attendance rate [20]. Research by Gusmelia Testiana et al. (2024) used the C4.5 algorithm to classify student graduation predictions at UIN Raden Fatah Palembang based on GPA and timely study requirements. The data was divided using the k-fold cross-validation technique (k=3), which resulted in an accuracy of 83.31%, classified as good. A total of 242 rules were formed, with GPA as the main attribute influencing graduation status [21]. Further research shows that the combination of the K-Means and C4.5 algorithms can identify the main factors that influence the risk of dropping out among vocational high school students [22].

The process of analyzing academic problems related to student graduation requires the selection of appropriate research concepts and methods so that the results can provide an accurate picture and relevant solutions. The selection of methods is not only based on theoretical considerations but also takes into account their effectiveness in handling the complexity of diverse educational data. On that basis, this study uses a data mining approach because it has the ability to explore hidden patterns in student academic data and present useful information to support the decision-making process. The selected method can answer the research questions and make a real contribution to improving the academic evaluation system, particularly in the Information Technology and Computer Education Study Program (PTIK) at UIN Sjech M. Djamil Djambek Bukittinggi. The purpose of this study is to apply the concept of data mining using the K-Means and Decision Tree C4.5 methods in analyzing the timely graduation of students in the PTIK Study Program at UIN Sjech M. Djamil Djambek Bukittinggi. The contribution of this research is to provide an overview of the application of the K-Means and Decision Tree C4.5 methods in analyzing the timely graduation of students in the PTIK Study Program at UIN Sjech M. Djamil Djambek Bukittinggi.

2. Methods

This research method will explain in detail the stages of work used in this study. The research method is a series of stages that include steps for processing and managing data, with the aim of producing high-quality and relevant research. It consists of various stages that will be explained sequentially, starting from the needs analysis process, data collection, to data processing. Each stage will be presented in the form of a framework diagram for easy visual understanding, so that the entire research process can be understood well from start to finish. The research framework includes the stages that will be carried out by the author in completing this research. These stages will be carried out in a structured manner by the researcher, from the initial stage to the obtaining of research results. The stages of this research can be seen in Figure 1:

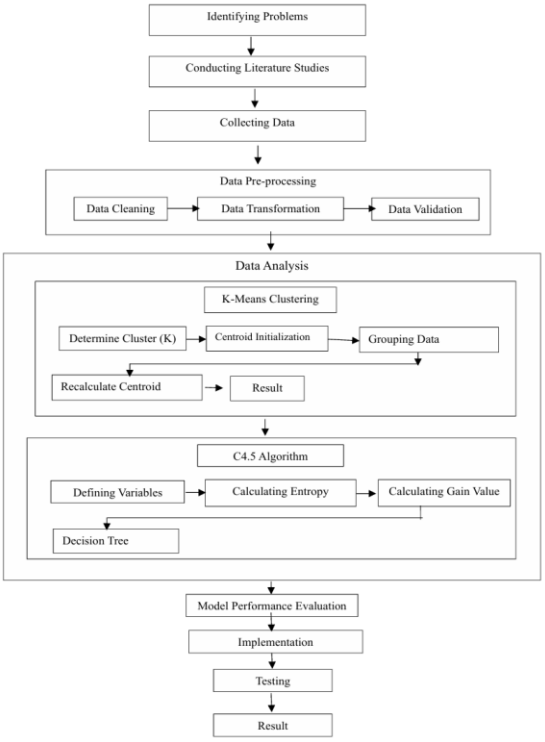

Figure 1. Research Framework

Figure 1 shows the research framework consisting of several main stages, namely identifying the problems to be used in the research, collecting literature in line with the research, collecting research data, performing pre-processing data analysis, grouping process using the K-Means method, and classification analysis using the C4.5 Decision Tree. These stages are arranged systematically to ensure that the data used is ready for processing, then grouped according to its characteristics, and finally analyzed further through a classification model to produce accurate prediction patterns. The output results are presented in the form of prediction information and a knowledge base in the form of a decision tree from student data.

2.1 Research Data Set

The research data set used in this study is data on students majoring in Information Technology and Computer Education, class of 2017, at UIN Sjech M. Djamil Djambek Bukittinggi. The data is presented in Table 1:

Table. 1. PTIK Student Data (initial data)

Data	sks	study_period	ipk	deposit amount
1	153	Study period: 4 years, 2 months, 11 days	3.47549	Already
2	153	Study period: 4 years, 2 months, 19 days.	3.614379	Already
3	153	Study period: 4 years, 2 months, 19 days	3.55719	Already
….	….	….	….	….
154	153	Study period: 4 years, 6 months, 10 days	3.560458	Already
155	153	Study period: 4 years, 3 months, 16 days	3.663399	Already
156	153	Study period: 4 years, 6 months, 14 days	3.454248	Already
157	22	Study period: 1 year, 10 months, 0 days.	0	Not yet
158	155	Study period: 4 years, 11 months, 30 days	3.475806	Already

Based on Table 1 above, it can be seen that the initial data of PTIK students consists of several important attributes, namely the number of credits taken, study period, GPA, and deposit status. The total number of data is 158.

2.2 K-Means Algorithm

The K-Means algorithm is an unsupervised learning technique used to group data into several clusters based on the degree of data similarity[23]. This technique works by finding the cluster center (centroid) and grouping data based on the closest distance to that center[24]. This algorithm is very popular because of its ability to process large amounts of data with relatively short computation time. K-Means can only be used for numerical attributes and is very effective in clustering data that has a clear distribution[15].

1. Determining the Number of k
   The first step in the K-Means method is to determine the value of k, which is the number of clusters to be formed. Determining this value is the basis for the data clustering process[25].
2. Determining the Initial Cluster Centers (k-Centroid)
   The next step is to determine the initial cluster centers or centroids. The selection of these starting points is very important because it will affect the final results of the clustering process[25].
3. Calculating the Distance of Each Data Point to Each Centroid
   After the cluster centers are determined, the distance of each data point to the centroid is calculated. This calculation uses Euclidean Distance, which is a formula for measuring the distance between objects and the cluster center, as shown in Equation 2.1[25].

$$D_{ij} = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

   Equation 1 is used to calculate the Euclidean distance between data X and centroid Y by measuring the difference between each variable, then summing them and taking the square root to obtain the distance between data points.

4. Data Clustering

Based on the distance calculation results, each data point is then grouped into clusters according to the closest distance between the data and the centroid [25].

5. Determining the New Centroid
   The new centroid is calculated based on the average value of the data in the same cluster. This process is repeated until there are no more changes in the cluster positions. The new centroid can be determined using Equation 2.2 [25].

$$D_{ij} = \frac{x1+x2+x3+\cdots+xn}{\sum x}$$

This formula is used to calculate the relative average value of a data point by comparing the sum of all records (x1 to xn) to the total data (∑x).

2.3. Algorithm Decision Tree C4.5

The C4.5 Decision Tree is a machine learning algorithm that falls under the category of supervised learning and is often used to solve classification problems[26]. This algorithm was developed by Ross Quinlan as an extension of the ID3 (Iterative Dichotomiser 3) algorithm with several improvements, particularly in its ability to handle more complex and diverse data[27]. C4.5 has the ability to build decision trees based on input data and automatically determine which variables are most influential in making decisions[28]. The steps are as follows:

1. Determining Variables
   Determining relevant variables is an important step in the classification process. This step is performed after the dataset has undergone preprocessing to ensure that the data is clean and ready for use. Selecting the right variables will improve the accuracy and effectiveness of the classification model being built [29].
2. Calculating Entropy
   Calculating entropy is done to measure the level of instability or uncertainty of data in each attribute. This process helps determine how well an attribute can distinguish data into existing classes. The formula for calculating entropy can be explained as follows [29]:
   Formula for Calculating Entropy:

$$\boldsymbol{Entropy(S)} = \sum_{i=1}^{n} - \boldsymbol{pi} * \log_2 \boldsymbol{pi} \quad (3)$$

   The entropy formula is used to measure the level of uncertainty or diversity of data in a set of cases (S) based on the number of partitions (n) and the proportion of each partition (pi) to the overall data.
3. Calculating Gain Values
   Gain information is used to determine the most relevant and influential attributes in the data division process in a classification model. The information gain value indicates the extent of the decrease in uncertainty or entropy after the data is divided based on a particular attribute. Thus, the higher the gain value of an attribute, the better the attribute is used to separate data, and the formula for calculating gain can be explained as follows [29].
   Formula for Calculating Gain:

$$\boldsymbol{Gain(S,A)} = \boldsymbol{Entropy(S)} - \sum_{i=1}^{n} \frac{|\,\boldsymbol{Si}\,|}{|\,\boldsymbol{S}\,|} * \boldsymbol{Entropy\,(Si)} \quad (4) \quad (2)$$

   This formula is used to calculate the division of cases in a set S based on attribute A, taking into account the number of attribute partitions and the proportion of cases in each partition relative to the entire data set.

4. Decision Tree Construction
   After the attributes are selected and the branches are created, the algorithm continues to repeat this process on each new branch until the data on that branch is fully classified or reaches the final result. [29]

3. Results and Discussions

Based on the research framework described earlier, this study consists of several stages, namely identifying problems, determining research objectives, conducting literature studies, collecting data, preprocessing data, applying the K-Means algorithm for clustering, applying the C4.5 Decision Tree algorithm for classification, and testing the results using RapidMiner. To facilitate understanding of the analysis and system design process, the following flowchart illustrates the research flow from the initial stage to the evaluation of results.

3.1 Pre-processing

Data pre-processing is the initial stage that serves to prepare data so that it can be processed optimally in the subsequent analysis process. The data used is data on students majoring in PTIK at UIN Bukittinggi in the 2017 cohort. This stage is important to ensure that the data used is of good quality, free from errors, and consistent. The pre-processing process includes three main steps, namely data cleaning, data transformation, and data validation. The results of pre-processing are presented in Table 2:

Based on Table 2 above, all PTIK student data consisting of attributes such as credits, study period, GPA, and tuition fees have met the criteria for analysis. The data shows that there are no missing values or anomalies, making it suitable for use in the next modeling stage. Thus, this validated dataset is ready to be processed using the K-Means algorithm to group students based on their academic and non-academic characteristics.

Table 2. PTIK Student Data

Data	sks	study_period	ipk	deposit
1	153	4.2	3.47549	1
2	153	4.2	3.614379	1
3	153	4.2	3.55719	1
….	….	….	….	….
154	153	4.6	3.560458	1
155	153	4.3	3.663399	1
156	153	4.6	3.454248	1
157	22	1.1	0	0
158	155	4.11	3.475806	1

3.2 K-Means Method Analysis

After the data has been validated and is ready for use, the next step is to apply the K-Means algorithm for the data grouping process. This method plays an important role in identifying patterns and grouping students based on similarities in academic characteristics. The K-Means process will begin with determining the number of clusters, calculating the initial centroid, and grouping the data into clusters that have the closest distance until a stable group is formed. The results of the student graduation data clustering from the K-Means algorithm are presented in Table 3 :

Table 3. PTIK Student Data (Cluster)

Data	sks	study_period	ipk	deposit amoun	Cluster
1	153	4.2	3.47549	1	Graduated on Time
2	153	4.2	3.614379	1	Graduated on Time
3	153	4.2	3.55719	1	Graduated on Time
….	….	….	….	….	….
154	153	4.6	3.560458	1	Graduated on Time
155	153	4.3	3.663399	1	Graduated on Time
156	153	4.6	3.454248	1	Graduated on Time
157	22	1.1	0	0	Discontinued Study
158	155	4.11	3.475806	1	Graduated on Time

Based on Table 3 above, the results of data clustering with the following distribution were obtained: Cluster 0 (C0) contained 131 data points, Cluster 1 (C1) contained 14 data points, Cluster 2 (C2) contained 4 data points, and Cluster 3 (C3) contained 9 data points. The K-Means Clustering algorithm successfully grouped the data into four clusters, namely:

1. Cluster 0 (C0) contains 131 data points with students who graduated on time
2. Cluster 1 (C1) contains 14 data points with students who discontinued study
3. Cluster 2 (C2) contains 4 data points with students who dropped out
4. Cluster 3 (C3) contains 9 data points with students who graduated late

As the final stage of the K-Means Clustering analysis process, the results of the student data grouping show the distribution in each cluster. These results illustrate the pattern of similarities in characteristics among students based on the research attributes used. An overview of the overall cluster results is presented in Table 4.13, which shows student data based on the clusters that have been formed.

3.3 C4.5 Method Analysis

The C4.5 Decision Tree method is used to form a decision tree from data that has been grouped using the K-Means method. After the clustering process is complete, the results of each cluster will be used as input data to determine patterns or decision rules through the C4.5 algorithm. The stages carried out in this method are determining variables, calculating entropy, calculating gain values, and constructing a decision tree. The classification of student graduation data from the C4.5 Decision Tree algorithm is presented in the figure below :

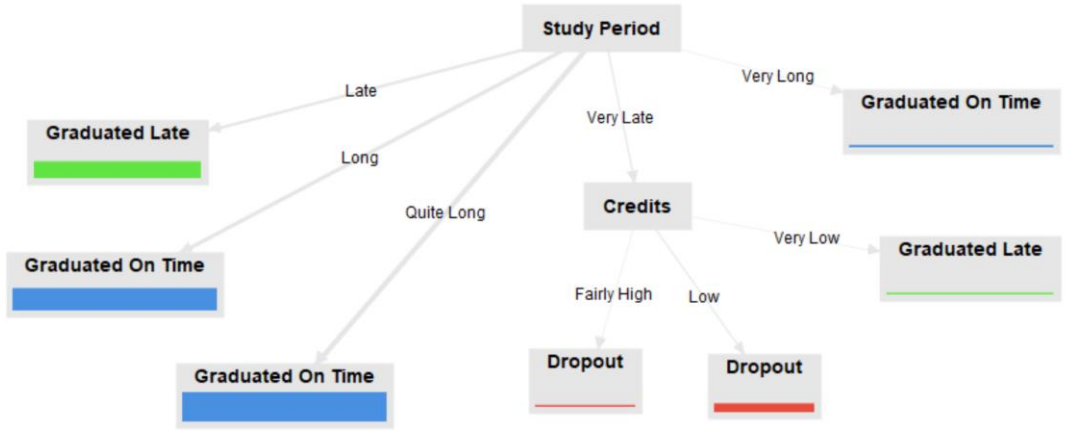

Figure 2. Decision Tree

Based on Figure 2, which is a decision tree diagram for predicting student graduation status with the main root "study_period," it can be seen that students with very short, moderate, or long study periods tend to graduate on time, while students with late study periods will be further evaluated based on their GPA. if the GPA is high (A or A-), there is still a chance of graduating on time, but if the GPA is low (B or B+) accompanied by low or very low credit acquisition ability, then the student is at high risk of dropping out, while a very late study period almost always results in late graduation or dropping out despite other supporting factors.

3.4 Model Performance Evaluation

Model classification performance evaluation in this study was conducted to determine how well the model predicts student graduation status based on the attributes of credits, study period, GPA, and deposit_verse. In this study, the model was tested against 158 student data with the label Validation as actual data, and the prediction results were obtained through the applied method. The data to be evaluated is shown in Table 4.

Table 4. Validation of Student Graduation Predictions

Data	sks	study period	ipk	deposit amount	Prediction	Validation	
						Valid	Invalid
1	Very High	Quite a long time	A	Already	Graduated on Time	√	
2	Very High	Quite a long time	A+	Already	Graduated on Time	√	
3	Very High	Quite a long time	A	Already	Graduated on Time	√	
….	….	….	….	….	….	….	
154	Very High	Quite a long time	A	Already	Graduated on Time	√	
155	Lowest	Quite a long time	A+	Already	Graduated on Time		√
156	Very High	Quite a long time	A	Already	Graduated on Time	√	
157	Lowest	Very Fast	C	Not yet	Discontinued Study	√	

158	Very High	Quite a long time	A	Already	Graduated on Time	√

Based on Table 4, the student graduation prediction data was validated directly by the research subjects. There were 151 valid data points according to the prediction and 7 invalid data points. To calculate the confusion matrix, it was then used as the basis for calculating the model performance evaluation metrics. The calculation was performed by calculating the accuracy. Accuracy is one of the main metrics for evaluating the performance of the classification model in this study. This metric describes the proportion of correct predictions compared to the total data tested. In other words, accuracy shows the model's success rate in predicting the overall graduation status of students. The calculation is performed using Equation 5.

$$Accuracy = \frac{\text{Number of correct predictions}}{\text{Amount Of Data}} \times 100\% \quad (5)$$

$$Accuracy = \frac{151}{158} \times 100\%$$

$$= 95{,}57\%$$

Based on calculations, the model successfully predicted 151 out of 158 student data correctly. When calculated using the accuracy formula, the value obtained was 95.57%, which can be rounded up to 96%. This value indicates that the model has a high level of accuracy in predicting student graduation status.

4. Conclusions

Based on research conducted in analyzing the determinants of timely graduation of students in the PTIK UIN Bukittinggi study program, it can be concluded that the K-Means and C4.5 algorithms have succeeded in obtaining fairly precise and accurate results. These results were obtained based on an evaluation with an accuracy rate of 96%. The results of this study can be used as an alternative solution in improving the academic performance of the study program and maintaining the accreditation status of the study program. For future researchers, the author suggests using other variables and developing them into other models.

References

[1] Hasanah, "Pemanfaatan Teknologi Informasi dalam Penguatan Karakter Pancasila di Kalangan Pelajar," *J. Agama dan Sos.*, vol. 1, no. 3, pp. 158–169, 2024.

[2] V. E. Picauly, "Transformasi Pendidikan di Era Digital," *Indones. Res. J. Educ.*, vol. 4, no. 3, pp. 1528–1535, 2024.

[3] V. Listy and I. Ilham, "Revolusi Sistem Informasi Manajemen di Era AI dan Big Data Mengubah Cara Bisnis Bekerja," *Simpatik J. Sist. Inf. dan Inform.*, vol. 5, no. 1, pp. 27–36, 2025, doi: 10.31294/simpatik.v5i1.7621.

[4] F. Abdillah, "Peran Perguruan Tinggi Dalam Meningkatkan Sumber Daya Manusia," *JMM - J. Masy. Merdeka*, vol. 3, no. 1, pp. 13–24, 2023, doi: 10.51213/jmm.v3i1.46.

[5] Ahmad Faizal, Edy Soesanto, and Mohamad Aren Gracia Shandy, "Membangun Generasi Cerdas dan Berkarakter Yang Kompeten Serta Pendewasaan Diri Sebagai Seorang Mahasiswa Perguruan Tinggi," *Konsensus J. Ilmu Pertahanan, Huk. dan Ilmu Komun.*, vol. 2, no. 1, pp. 33–43, 2025.

[6] D. Dendodi, N. Nurdiana, Y. D. Astuti, A. Aunurrahman, and W. Warneri, "Dampak dan tantangan terhadap Transformasi kurikulum di Satuan Pendidikan," *J. Educ. Res.*, vol. 5, no. 2, pp. 1071–1080, 2024, doi: 10.37985/jer.v5i2.960.

[7] Oemi Koelsoem and Kusmiyati Kusmiyati, "Peran Teknologi dalam Implementasi Kurikulum Merdeka," *Morfol. J. Ilmu Pendidikan, Bahasa, Sastra dan Budaya*, vol. 2, no. 6, pp. 297–303, 2024, doi: 10.61132/morfologi.v2i6.1191.

[8] T. H. Hasibuan and D. Mahdiana, "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma C4.5 Pada Uin Syarif Hidayatullah Jakarta," *Skanika*, vol. 6, no. 1, pp. 61–74, 2023.

[9] M. Zulkarnain, Ahmad Syukri, "TRANSFORMASI SEKOLAH TINGGI AGAMA ISLAM NEGERI (STAIN) MENJADI INSTITUT AGAMA ISLAM NEGERI (IAIN) DALAM MENINGKATKAN PELAYANAN AKADEMIK," *Abdau J. Pendidik. Madrasah Ibtidaiyah*, vol. 4, no. 1, pp. 167–186, 2024.

[10] R. P. Bendriyanti, "Manajemen Mutu Layanan Service Quality Management Competence in Improving Academic Graduate in Private Universities in the Province Bengkulu," *J. Pendidik. Tambusai*, vol. 6, no. 2, pp. 16252–16260, 2022.

[11] H. Yuliansyah, R. A. P. Imaniati, A. Wirasto, and M. Wibowo, "Predicting Students Graduate on Time Using C4.5 Algorithm," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 7, no. 1, p. 67, 2023, doi: 10.20473/jisebi.7.1.67-73.

[12] M. A. S. Pawitra, H.-C. Hung, and H. Jati, "A Machine Learning Approach to Predicting On-Time Graduation in Indonesian Higher Education," *Elinvo (Electronics, Informatics, Vocat. Educ.*, vol. 9, no. 2, pp. 294–308, 2024.

[13] E. Haryatmi and S. Pramita Hervianti, "Penerapan Algoritma Support Vector Machine Untuk Model Prediksi Kelulusan Mahasiswa Tepat Waktu," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 2, pp. 386–392, 2022, doi: 10.29207/resti.v5i2.3007.

[14] F. M. Almufqi and A. Voutama, "Perbandingan Metode Data Mining Untuk Memprediksi Prestasi Akademik Siswa," *J. Tek.*, vol. 15, no. 1, pp. 61–66, 2023.

[15] A. Nur Khormarudin *et al.*, "Teknik Data Mining: Algoritma K-Means Clustering," *J. Lebesgue J. Ilm. Pendidik. Mat. Mat. dan Stat.*, vol. 1, no. 2, pp. 116–123, 2022.

[16] H. Okprana, "Algoritma C4.5 Dalam Data Mining Untuk Menentukan Klasifikasi Penerimaan Calon Mahasiswa Baru," *Terap. Inform. Nusant.*, vol. 2, no. 5, pp. 358–364, 2023.

[17] B. A. R. P. Wahyu, A. F. Farozi, C. P. Mahendra, and R. K. Hapsari, "Klasifikasi Penderita Penyakit Diabetes Berdasarkan Decision Tree Menggunakan Algoritma C4.5," *INTEGER J. Inf. Technol.*, vol. 8, no. 1, pp. 80–89, 2023.

[18] M. H. Sukri and Y. Handrianto, "Penerapan Algoritma C4.5 Dalam Menentukan Prediksi Prestasi Siswa Pada SMPN 51 Jakarta," *Informatics Comput. Eng. J.*, vol. 4, no. 1, pp. 11–24, 2024.

[19] D. Cahyani, D. Rachmatullah, D. Akbar, H. Malik, K. A. Akhmallahudin, and M. Fansyuri, "Data Mining Menggunakan Algoritma Decision Tree untuk Menentukan Kelulusan Mahasiswa dengan RapidMiner," *J. Cendekia Indones.*, vol. 1, no. 1, pp. 113–117, 2025.

[20] M. A. M. A. Aqil and Z. Fatah, "Analisis Pengelompokan Prestasi Mahasiswa Universitas Malang Menggunakan Metode K-Means Clustering," *JITU J. Inform. Technol. Commun.*, vol. 8, no. 2, pp. 29–40, 2024.

[21] V. Fitriyanti, G. Testiana, and C. Eri Gunawan, "Klasifikasi Predikat Kelulusan Mahasiswa Menggunakan Algoritma C4.5," *J. SAINTEKOM*, vol. 14, no. 2, pp. 217–232, 2024, doi: 10.33020/saintekom.v14i2.626.

[22] S. Anwar, "Application of K-Means and C4.5 Algorithms for dropout risk prediction in vocational high schools," *Indones. J. Multidiscip. Sci.*, vol. 4, no. 5, pp. 428–435, 2025, doi: 10.55324/ijoms.v4i5.1102.

[23] A. Sulistiyawati and E. Supriyanto, "Implementasi Algoritma K-means Clustring dalam Penetuan Siswa Kelas Unggulan," *J. Tekno Kompak*, vol. 15, no. 2, p. 25, 2022.

[24] R. Rianti, R. Andarsyah, and R. M. Awangga, "Penerapan PCA dan Algoritma Clustering untuk Analisis Mutu Perguruan Tinggi di LLDIKTI Wilayah IV," *Nuansa Inform.*, vol. 18, no. 2, pp. 67–77, 2024.

[25] M. P. A. Ariawan, I. B. A. Peling, and G. B. Subiksa, "Prediksi Nilai Akhir Matakuliah Mahasiswa Menggunakan Metode K-Means Clustering (Studi Kasus : Matakuliah Pemrograman Dasar)," *J. Nas. Teknol. dan Sist. Inf.*, vol. 9, no. 2, pp. 122–131, 2023, doi: 10.25077/teknosi.v9i2.2023.122-131.

[26] A. Fatma Ayu Rahman, S. Wartulas, J. K. Raya Pagojengan, and P. Brebes, "Prediksi Kelulusan Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus Di Universitas Peradaban)," *IJIR*, vol. 1, no. 2, pp. 70–77, 2020.

[27] A. Pratiwi, "Decision Tree Prediction Model in Patient Mortality Rate based on Risk Factor," *J. Kesehat. Masy.*, vol. 8, no. 1, pp. 30–39, 2023.

[28] R. Triyandika, "Penerapan Metode Decision Tree Dengan Algoritma C4.5 Untuk Klasifikasi Penyakit Jantung," p. 1, 2022.

[29] D. Lase, "Metode Decision Tree untuk Mahasiswa Penerima Beasiswa," *J. Inform. Press*, vol. 1, pp. 20–23, 2024.

Biographies of Authors

Rahma Yanti Born in Tanjung Durian on June 21, 2002, is a Master's student in Informatics Engineering at UPI YPTK Padang with a concentration in Data Science. Currently, she works as an Editorial Board member of the UIN Bukittinggi e-journal and as a freelance project worker in the IT and scientific fields. She earned her bachelor's degree from UIN Bukittinggi in 2024, majoring in Informatics Engineering and Computer Education, and previously served as an ICT teacher at Pondok Pesantren Al-Makrif Bukittinggi. She can be contacted at: rahmayanti210602@gmail.com

Musli Yanto he is a lecturer at Universitas Putra Indonesia YPTK Padang.Educational background since 2014 has completed postgraduate studies in the field of informatics engineering. His skills include data science analysis, algorithms and programming, big data and artificial intelligence. He can be contacted at email: musli_yanto@upiyptk.ac.id.

Syafri Arlis Syafri Arlis was bornin Padang on October 23, 1986. His undergraduate study was completed in 2009 at Universitas Putra Indonesia YPTK. He completed his master'sdegree at Putra Indonesia University, YPTK Padang. He currently serves as a lecture in the Informatics Engineering study program at the Universitas Putra IndonesiamYPTK Padang. Teaching history that has been carried out starting from 2011 until now, such as databasesanddigital image processing. Published research history places more emphasis on the atificial neural network and digital image processing. He can be contacted at email: syafri_arlis@upiyptk.ac.id.