RESEARCH ARTICLE | MARCH 04 2026

Evaluation of learning effectiveness during and after COVID-19 pandemic using decision tree and Naïve Bayes

Maria Fatima Dineri De Jesus; Chatarina Enny Murwaningtyas

AIP Conf. Proc. 3326, 080034 (2026)
https://doi.org/10.1063/5.0308700

Evaluation of Learning Effectiveness During and After COVID-19 Pandemic Using Decision Tree and Naïve Bayes

Maria Fatima Dineri De Jesus[1,a)] and Chatarina Enny Murwaningtyas[1,b)]

[1]Department of Mathematics Education, Sanata Dharma University, Affandi Street, Santren, Caturtunggal, Depok, Sleman, Yogyakarta, 55281, Indonesia

[a)] Corresponding author: mariafddejesus@gmail.com
[b)]enny@usd.ac.id

Abstract. This study aims to evaluate the effectiveness of learning during and after the COVID-19 pandemic using data mining techniques by applying the Decision Tree algorithm that can handle data with complex features and the Naive Bayes algorithm applied to small data. Both algorithms were applied to Moodle data from the 2020 and 2023 cohorts, data obtained from the Moodle platform. The data include activity logs, attendance records, assignment grades, midterm exam grades, and final exam grades, with the final grade converted into graduation status serving as the target variable. For the 2020 cohort, the Decision Tree model has 100% accuracy, while for the 2023 cohort, the accuracy of the Naive Bayes model performs better with 92% accuracy. Feature importance analysis revealed that assignments, midterm exams, and final exams were most important for predicting student outcomes in the 2020 cohort, while in 2023 it was assignments, midterm exams, final exams, and attendance. Notably, Moodle activity data did not significantly affect the prediction, indicating that in-person assessments and attendance were more important for academic success. Insights from this study guide that academic assessments such as assignments, midterms, and final exams have the greatest impact on student graduation, both in the 2020 and 2023 cohorts. The presence of attendance indicates that students' active engagement in each lecture session plays a significant role in determining graduation. Therefore, instructors need to pay more attention and emphasize factors such as assignments, exams, and attendance in their teaching strategies to improve student learning outcomes.

INTRODUCTION

The emergence of Coronavirus Disease 2019, known by the abbreviation Covid-19, in Indonesia in early 2020, shifted traditional face-to-face classroom learning to online learning. Online learning became the primary solution to maintain the continuity of education, although it introduced many unforeseen challenges [1]. While online learning was not entirely new in education, as E-learning had previously been integrated with face-to-face learning[2], the pandemic prompted educational institutions to develop and modify existing online learning methods within higher education. Even in the 21st century, the use of technology has become a consumption or is considered a primary need by the global community. This reliance on technology offers several advantages in the context of online learning, such as providing a broader and more engaging platform for interaction, enchancing flexibility in teaching and learning, and facilitating the improvement and delivery of educational materials [3].

One of the most widely used online learning methods is E-learning, which leverages information and communication technology to facilitate and support the learning process [4]. According to Al-Smadi et al. [5], E-learning is considered a modern learning method because educational institutions transfer learning information to students through technology. As a result, almost all educational institutions in Indonesia have transitioned from conventional learning methods to digital ones.

The implementation of Learning Management Systems (LMS) supports modern digital-based learning methods. LMS platforms are software applications used for administration, documentation, reporting, and technology-based educational learning or training programs[6]. LMS platforms manage the learning process and offer a web-based integrative learning system with various easily accessible applications, whether open-source or commercial [7,8].

Despite the end of the Covid-19 pandemic, educational institutions continue to use LMS in their learning processes. According to Istiqomah et al. [9], learning has not reverted entirely to conventional methods but rather combines online and traditional approaches. LMS-based learning media are being further developed into blended learning models in the New Normal Era, with LMS-based e-learning media proving feasible for future learning.

One prominent open-source LMS application is Moodle, a web application that facilitates online learning through web pages accessible via computers or mobile devices [10]. Moodle supports various learning process features, such

Proceedings of the International Conference on Mathematics, Computational Sciences, and Statistics 2024
AIP Conf. Proc. 3326, 080034-1–080034-8; https://doi.org/10.1063/5.0308700
Published under an exclusive license by AIP Publishing. 978-0-7354-5299-2/$30.00

as learning modules, online assignments, chat features, and quizzes[11]. It is a valuable platform for recording extensive data generated in education [12].

Moodle LMS stores extensive data from educational processes, which can be analyzed using data mining techniques. Data mining, utilized across various fields, uncovers hidden patterns and makes predictions from existing data[13]. It involves large-scale data exploration techniques aimed at identifying recurring patterns, trends, or rules that explain data characteristics within specific contexts[14].

Research by Fakhri et al. [15] highlights the positive impact of active Moodle use on student learning outcomes, such as active participation in synchronous learning via Zoom. Similarly, research by Simbolon et al.[16] found a positive effect of Moodle use on student grades. Sokout and Usagawa [17] conducted a study to predict student success and failure in two courses using Moodle features like quizzes and assignments, identifying students at risk of failure. Hence, evaluating Moodle's impact on student learning outcomes is crucial for improving learning and positively influencing student performance.

The Mathematics Education Study Program at Sanata Dharma University continues to utilize Moodle for learning. However, evaluating learning outcomes using Moodle in this program has never been conducted. This research focuses on classifying student learning outcomes based on Moodle usage. The research includes three courses: Logic and Set Theory, Algebra and Trigonometry, and Plane Geometry for the 2020 and 2023 cohorts. The 2020 cohort experienced learning during the Covid-19 pandemic, while the 2023 cohort experienced learning post-pandemic.

Classification is the role of data mining that uses a predictive approach, which means that the data is divided into two, namely train data and test data, where train data serves to train the model, then evaluated using test data [18]. The results of the study Mariati et al. [19], assessing student satisfaction with facilities and learning with the Decision Tree c4.5 algorithm classification showed 88% accuracy and 98% precision. The result of the study Anugrah Putra & Kamayani. [20], using Naive Bayes and model evaluation using three different types of K-fold Cross Validation to predict student timeliness in graduation, obtained one model with the most accurate K-fold Cross Validation with an accuracy value of 80.19%, recall of 80.26%, precision of 92.75%, and F-Measure of 86.05%.

This study utilized two classification algorithm models: Decision Tree and Naive Bayes. These algorithms were applied to predict student learning outcomes across two different cohorts: the 2020 cohort during the COVID-19 pandemic and the 2023 cohort after the pandemic. The metric evaluation results of the two algorithm models, along with the feature importance results of the Decision Tree algorithm, were compared to determine the effectiveness of learning during and after the COVID-19 pandemic. The research involved data from 90 students, including activity logs, attendance records, assignment scores, midterm exam scores, and final semester exam scores. The target variable was the final grade, which was transformed into graduation status. The classification results of this study provide valuable insights for teachers in making informed decisions about learning methods within the Mathematics Education Study Program.

RESEARCH METHODOLOGY

This study focuses on three core courses: Algebra and Trigonometry, Plane Geometry, and Logic and Set Theory. The data used in this study are Moodle data from Mathematics Education students of the 2020 and 2023 cohorts obtained from the University's Moodle platform. Data including grades, activity logs, and attendance records were obtained from the University's Moodle system and arranged in an Excel file. To analyze this data used data mining techniques by applying the Decision Tree and Naive Bayes algorithm models. Several methods used in this data mining technique, that is Pipeline with steps in the Robust Scaler pipeline, Yeo-johnson, Random Over sampler, and algorithm models. The following methods are Stratified K-Fold, evaluation metrics, confusion matrix, and feature importance. The research steps are shown in Figure 1.

The initial stage of the research involves data pre-processing to prepare the data for the mining process, this stage needs to be carried out to obtain good quality data [21,22]. Preprocessing in this study includes data cleaning, namely deleting or removing empty row and column data that can cause missing values or columns that are not relevant to the study according to researchers such as name, NIM, date and time of access. Furthermore, the data transformation stage is carried out, the stage is the combination of several columns such as several task values that are combined to get an average, which is useful because it can get more concise data, reduce data dimensions, and can improve data quality. In the data transformation stage, coding is also carried out on courses for easier analysis: Algebra and Trigonometry are coded as 'A', Plane Geometry as 'B', and Logic and Set Theory as 'C' for both cohorts. After performing data transformation and obtaining data according to the needs for modeling, the data is then split into training and testing sets, with 70% allocated for training and 30% for testing. This split is conducted separately for the 2020 and 2023 cohorts to maintain the integrity of the cohort comparisons.

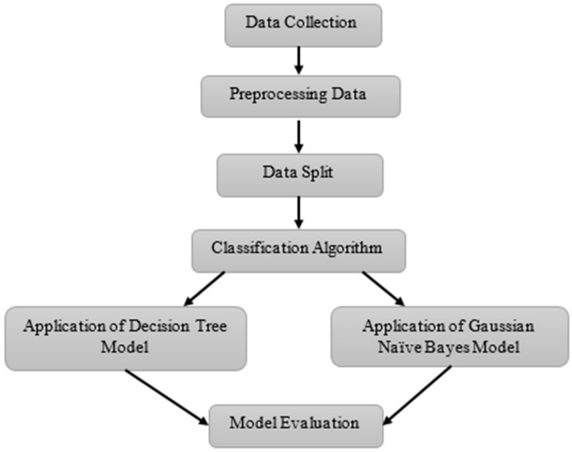


Figure 1. Methodology

The separated data is then handled by outliers and data normalization, although handling outliers and normalization is part of the transformation, the handling is done in the pipeline. Outliers in this research data are overcome by using Robust Scaler, where the use of this method is to reduce the impact of outliers without deleting or eliminating outliers that may have important or relevant information to the research by using statistical estimates in the form of medians and quartiles for scaling [23]. While normalization uses the Yeo-johnson method which changes abnormal data to approach normal data even though the data has a negative value [24]. The pipeline also performs Random Over Sampler after Robust Scaler and Yeo-Johnson are performed. This Random Over sampler is carried out with the aim of overcoming unbalanced data in the target class by adding synthetic samples to the minority class, so that the data becomes balanced for all classes [25]. At the end of the pipeline, modeling is carried out.

The GridSearchCV used to determine the best combination of parameters for the model to have optimal performance is Stratified K-Fold with data divided into 2 folds. This method is suitable for unbalanced data in the classification algorithm model and is good for evaluating model success [26]. The hyperparameters used for the Decision Tree model are maximum depth, minimum samples split, and minimum samples leaf. While the Naïve Bayes model uses the var-smoothing hyperparameter.

In analyzing data, especially in predictive models, it is necessary to know which features are most influential in predicting students' academic learning outcomes. Thus, information can be obtained that helps in determining future learning strategies. In Decision Tree, the most important features are determined based on how much they help in separating the data into the correct classes at each node. The features that provide the most useful information in distinguishing between different classes will be higher in the tree structure, indicating that they have a significant influence on the model's predictions. In Gaussian Naïve Bayes, feature importance is not measured in the same way as in Decision Tree, but the average variation across classes to identify features that have a significant influence on the classification results [27]. Feature importance is crucial for educators because it helps them design more effective teaching strategies. Identifying the key factors that affect learning outcomes allows educators to focus on the main elements that influence students' academic success.

After the pipeline and hyperparameters are carried out to obtain a model with optimal accuracy. then calculating the accuracy, precision, recall, F1-score, and confusion matrix values for the 2020 and 2023 cohorts using the following equations [28,29].

$$Accuracy = \frac{TP+TN}{TP+FP+FN+T} \tag{1}$$

$$Precision = \frac{TP}{TP+F} \tag{2}$$

$$Recall = \frac{TP}{TP+F} \tag{3}$$

$$F1\,Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

TABLE 1. Confusion Matrix

	Positive Prediction	Negative Prediction
Actual positive	TP	FN
Actual negative	FP	TN

The performance of these models is then evaluated using the test data. Key performance metrics such as accuracy, precision, recall, F1-score, and the confusion matrix (as shown in Table 1), are calculated to provide a comprehensive assessment of model performance for each cohort. These metrics are essential for determining the model's ability to correctly classify student outcomes and highlight the model's strengths and weaknesses.

Additionally, the study conducts a detailed comparison of the metric evaluation results and feature importance between the 2020 and 2023 cohorts. This comparison aims to reveal insights into the effectiveness of learning during and after the COVID-19 pandemic. Understanding these differences is crucial for identifying how learning processes and outcomes have evolved over time and under varying circumstances.

Result and Discussion

This study uses 90 student data from two different cohorts: the 2020 cohort that experienced the COVID-19 pandemic and the 2023 cohort that has passed the COVID-19 pandemic period. The data used is Moodle data in the form of activity, attendance, assignment score, midterm exam scores, and final semester exam scores, where these data are independent variables. Final grade data as a dependent variable which has been transformed to serve as the target class. Then the algorithm model is applied to each second cohort of data and obtained quite different model metric evaluation result.

The evaluation result for the 2020 cohort, shown in Table 2 for the Decision Tree model and Table 3 for the Naïve Bayes model. Evaluation result for the 2023 cohort can be seen in Table 4 for the Decision Tree model and Table 5 for the Naïve Bayes model.

TABLE 2. Evaluation Metric of the Decision Tree Model for the 2020 Cohort

Class	Accuracy	Precision	Recall	F1-Score
Not passing	1.00	1.00	1.00	1.00
Passed	1.00	1.00	1.00	1.00

TABLE 3. Evaluation Metric of the Naïve Bayes Model for the 2020 Cohort

Class	Accuracy	Precision	Recall	F1-Score
Not passing	0.87	0.67	1.00	0.80
Passed	0.87	1.00	0.82	0.90

Based on Table 2 and Table 3, obtained Evaluation metric of the Decision Tree model for the cohort of the 2020 shows that the data is balanced for class not passing and class passed, it can be seen from the accuracy of 100%, he precision value of 100% which means the model correctly predicts for the 'Passed' class, recall 100% indicates that the model correctly finds all passed samples in the 'Passed' class, and the F-1 value of 100% also indicates that the model has good performance in the precision and recall aspects of the 'Passed' class. Similar to the model's performance in predicting the 'Passed' class which shows 100%, in the 'Not passing' class the model also works well which produces a value of 100% for each prediction aspect. This shows that the Decision Tree model works better using the 2020 cohort data and shows that the model is able to classify all examples in this cohort perfectly. The opposite is true for the Naïve Bayes model evaluation metrics, indicating that the accuracy of the Decision Tree model is poor, with an accuracy of only 83%. This is because the precision in the 'Not passing' class is only 67%, meaning that there is a model error in predicting failure in this class, although in the 'Passed' class the model predicts very accurately at 100%. And the recall of the 'Passed' class of 82% indicates that the model did not find all samples that passed in this 'Passed' class. The F1-Score in both classes shows that the model has an unbalanced performance in terms of precision and recall. This means that the accuracy value is influenced by the precision and recall values, where if the precision and recall values are high in both classes, the accuracy will also be high. Conversely, accuracy will decrease if the precision and recall are small.

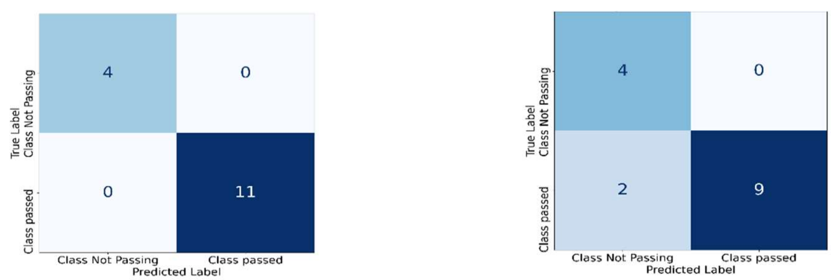
TABLE 4. Evaluation Metric of the Decision Tree Model for the 2023 Cohort

Class	Accuracy	Precision	Recall	F1-Score
Not passing	0.83	0.50	0.50	0.50
Passed	0.83	0.90	0.90	0.90

TABLE 5. Evaluation Metric of the Naïve Bayes Model for the 2023 Cohort

Class	Accuracy	Precision	Recall	F1-Score
Not passing	0.92	1.00	0.50	0.67
Passed	0.92	0.91	1.00	0.95

The result of the metric valuation metric in the 2023 cohort shows a different result from the result of the 2020 cohort metric evaluation. Where evaluation metric of the Decision Tree model experienced a decrease in performance with an accuracy of 83%. This is due to the precision and recall values in the 'Not passing' class which are only 50%, this indicates that the model is not accurate in predicting the class and determining that all samples do not pass in this class. The F1-score in this class shows that the model does not work very well as seen from the precision and recall. While in the 'Passed' class, the precision and recall of only 90% indicate that the model is less accurate in predicting and finding all samples in this 'Passed' class. From the F1-score of the 'Passed' class, it can be seen that the model is not working well. In contrast, the Naïve Bayes model showed better performance for the 2023 cohort with an overall accuracy of 92%. It can be seen from the precision value of the 'Not passing' class in this model that there has been an increase in value to 100%, although the recall value of this model is not different from the Decision Tree model, the Naïve Bayes model has a slightly higher performance in this class as seen from the F1-score of 67%. In the 'Passed' class, it can be seen that the Naïve Bayes model is slightly better at predicting the 'Passed' class and recall shows that this model is good at finding all samples in this 'Passed' class. The F1-score shows 95%, this means that the model has good performance in predicting the 'Passed' class as seen from the precision and recall. Thus, it means that when the precision and recall values increase, the accuracy value will also be better. Although the application of the two models to the 2023 cohort data shows that the data is unbalanced, the Naïve Bayes model still works better on small and unbalanced data, resulting in better accuracy than the Decision Tree model. With these accuracy results, it shows that the Naïve Bayes model works better for the 2023 cohort with smaller data.

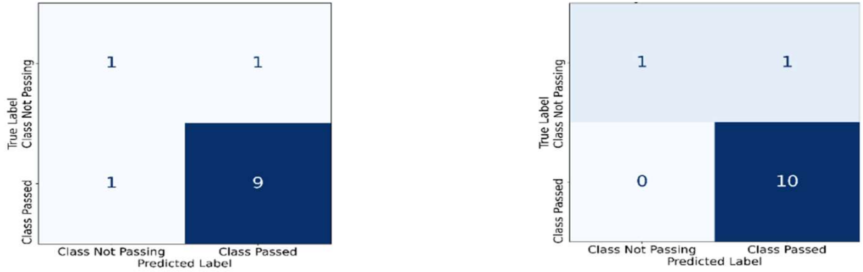


(a) (b)

Figure 2. Confusion Matrix of the 2020 Cohort, (a) Decision Tree models and (b) Naive Bayes models

Figure 2(a) shows the confusion matrix of the Decision Tree model for the 2020 cohort, where the model classified all instances perfectly. It accurately identified 4 'Not Pass' and 11 'Pass' students without any errors, achieving 100% accuracy, precision, and gain. Figure 2(b) displays the confusion matrix of the Naive Bayes model for the 2020 cohort, highlighting a notable difference in performance. The Naive Bayes model for this cohort correctly classified 4 'Not Pass' students and 9 'Pass' students but misclassified 2 'Pass' students as 'Not Pass,' resulting in an accuracy of approximately 87%.

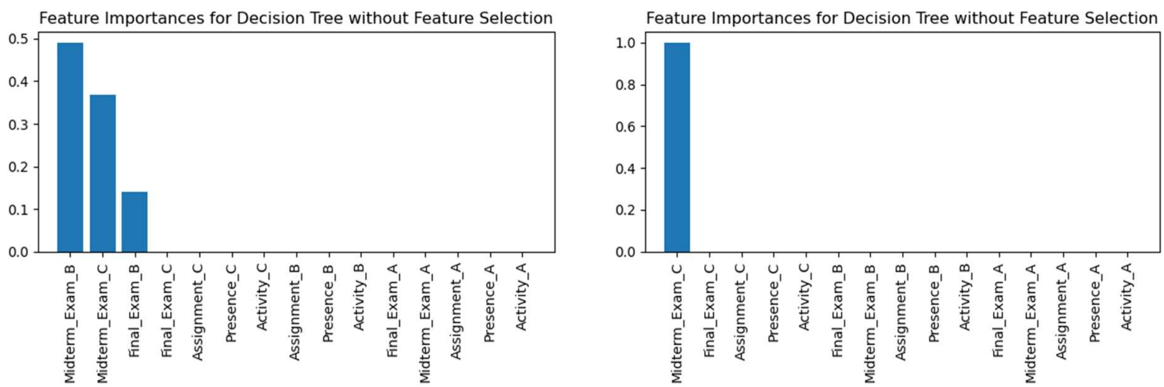
For the 2023 cohort confusion matrix shown in Figure 3(a), where the Decision Tree model correctly classified 1 'Not Passed' student and 9 'Passed' students but misclassified 1 'Passed' student as 'Not Passed' and 1 'Not Passed' student as 'Passed', resulting in a lower score and thus an overall accuracy of only 83%. However, as shown in Figure 3(b), the Naive Bayes model demonstrated a more balanced performance by correctly classifying 1 'Not Passed' and 10 'Passed' students, but it also misclassified 1 'Not Passed' student as 'Passed'. This shows the robustness of the model in handling imbalanced data quite well, obtaining 92% accuracy.



(a) (b)

Figure 3. Confusion Matrix of the 2023 Cohort, (a) Decision Tree models and (b) Naive Bayes models

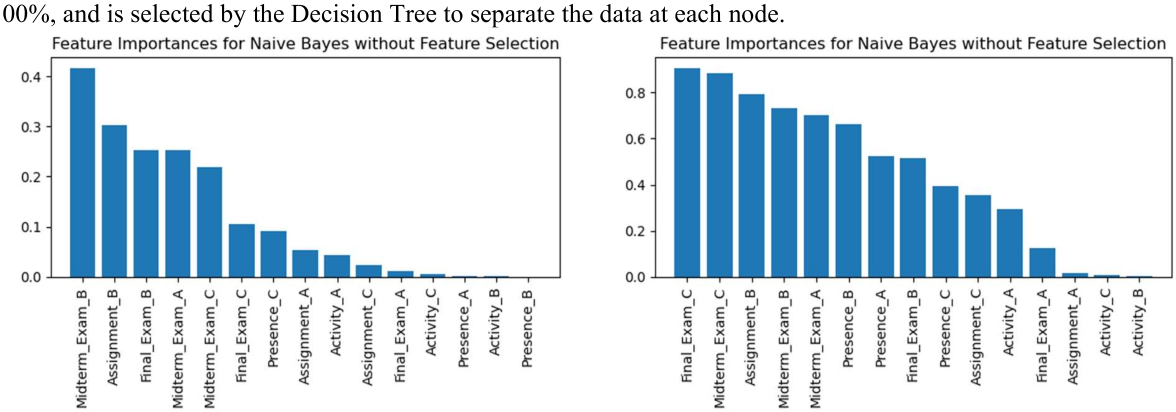
Furthermore, to determine the variables that have the most important role in the accuracy results of the Decision Tree model applied to 2020 and 2023 cohort data. The results of the feature importance level and feature importance level can be seen in Figure 4 for the 2020 and 2023 cohorts.



(a) (b)

Figure 4. Feature Importance of the Decision Tree Model, (a) 2020 and (b) 2023 Cohorts

Based on Figure 4(a), shows the importance of features in the Decision Tree model for the 2020 cohort. The Decision Tree model identifies only three main features: midterm exam B, midterm exam C, and final exam B. These three features play a significant role in decision-making in the Decision Tree, indicating that the graduation decision in the 2020 cohort is strongly influenced by these three aspects. In contrast, Figure 4(b), illustrates the importance of features in the Decision Tree model for the 2023 cohort. There is a significant difference in the importance of features in the 2023 cohort compared to the 2020 cohort. In the 2023 cohort, there is only one feature that affects the accuracy of the Decision Tree, which is the midterm exam C. This feature becomes the most dominant, with an influence of 100%, and is selected by the Decision Tree to separate the data at each node.



(a) (b)

Figure 5. Feature Importance of the Naïve Bayes Model, (a) 2020 and (b) 2023 Cohorts

Figure 5(a), shows the four most influential features in the Naïve Bayes model for the 2020 cohort, this is midterm exam B, assignment B, final exam B, and midterm exam A. These four features have the greatest influence on the Naive Bayes modeling of the 2020 cohort compared to other features. Figure 5(b), shows the most influential features of the Naïve Bayes model for the 2023 cohort. Different from the 2020 cohort which only had the four most influential features, in the 2023 cohort there are six features of importance: final exam C, midterm exam C, assignment B, midterm B, midterm exam B, and presence B. These six features have the greatest influence on the modeling.

This difference indicates that the activity data from Moodle, such as login frequency and participation in discussion forums, were not identified as important features in either cohort. This suggests that while these activities are part of overall student engagement, they do not make a significant contribution to predicting final grades. In the 2020 cohort, graduation is influenced by a combination of three features, that is midterm exam, final exam, and assignment. In the 2023 cohort, four features were the determinants: midterm exam, final exam, assignment, and attendance. This slight shift suggests a possible change in educational dynamics and assessment strategies post-pandemic.

CONCLUSION

Based on the results of research on the application of the Decision Tree algorithm model and the Naïve Bayes algorithm in two different cohorts, it can be concluded that in the 2020 cohort, the Decision Tree algorithm has an accuracy of 100%, and the Naïve Bayes algorithm has an accuracy of 87%. In the 2023 cohort, the Decision Tree algorithm has an accuracy of 83% and the Naïve Bayes algorithm of 92%. Both algorithms applied to different cohort data experience changes in accuracy values due to data imbalance, this can be seen in the metric evaluation of each model, there is a difference between the metric evaluation values of classes that do not pass and classes that pass.

In the analysis of the most influential features, it is known that academic assessments have the greatest influence on predicting graduation in both cohorts. In the 2020 cohort, it was seen that midterm exams, assignments, and final semester exams had a major influence on graduation in the 2020 cohort. Meanwhile, in the 2023 cohort, it was found that the factors that influenced graduation were midterm exams, final exams, assignments, and attendance. This shows that these factors are the determinants for teachers in choosing teaching strategies to improve learning outcomes. In this study, it was found that Moodle activities did not affect students' final learning outcomes.

Direct assessments in academics such as assignments, midterm exams, final exams, and attendance in the 2023 cohort showed that teachers need to pay more attention to these four factors because these factors determine student graduation. Assignments, midterm exams, and final exams are measures of the level of understanding and application of the material by students directly, so they are the main benchmarks in academic evaluation. However, in the 2023 cohort, presence also emerged as a significant factor, indicating that active student involvement in each lecture session is considered important. This comparison between direct assessment and attendance underlines the change in the instructor's approach to assessing student completion. In this way, instructors can emphasize active student participation during the learning process, not just assessment based on final results.

For further research, deeper exploration can be carried out regarding the various types of online activities in Moodle, such as forum discussions, quizzes, and additional materials, which can be integrated more effectively into the curriculum. Further research can also explore student learning outcomes online, offline, and hybrid. It is also suggested that further research can also use a larger and more diverse data set, not only limited to certain courses but also covering all courses that take place in one semester. In addition, using more complex and powerful algorithms such as Random Forest, which is a development of a Decision Tree, can help produce more accurate predictions.

ACKNOWLEDGMENTS

This research was funded by the Institute for Research and Community Service of Sanata Dharma University through the Master Doctoral Research grant program. We express our deepest gratitude for the support and funding provided. Without this support, the research would not have been successfully conducted. We also extend our thanks to all parties who have contributed and supported this research, both directly and indirectly.

REFERENCES

[1] D. Y. Irawati, and Jonatan, Jurnal Rekayasa Sistem Industri, **9**(2), 135–144 (2020).
[2] S. L. Nasution, F. Windari, S. Z. Harahap, and Elvina, Ecobisma (Jurnal Ekonomi, Bisnis Dan Manajemen), **8**(1), 67–75 (2021).
[3] U. H. Salsabila, S. N. Utami, A. Zahra, F. Haikal, and A. Cyahono, Jurnal Ilmiah Wahana Pendidikan, **7**(1),

1-9 (2021).
[4] Y. S. Nuraeni and D. Irawati, Procuratio : Jurnal Ilmiah Manajemen, **10**(3), 303–314, (2022).
[5] A. M. Al-Smadi, A. Abugabah, and A. A. Smadi, Procedia Computer Science, **201**(2022), 383–389 (2022).
[6] L. Vicent and M. Segarra , "Learning Management System," in *Multimedia in Education: Adaptive Learning and Testing*, edited by I. Cheng; L.V. Safont; A. Basu; R. Goebel (World Scientific, Singapore, 2010) pp. 21–48.
[7] S. T. Sianturi, and U. L. Yuhana, IPTEK The Journal of Technology and Science, **33**(2), 94-104 (2022).
[8] Widiyawati, and Y. Anistyasari, Jurnal IT-EDU, **05**(01), 512–519 (2020).
[9] E. O. Istiqomah, A. Atiqoh, and Suhari, Jurnal Teknologi Pendidikan: Jurnal Penelitian dan Pengembangan Pembelajaran, **8**(1), 100-110 (2023).
[10] H. Dhika, F. Destiawati, M. Sonny, and Surajiyo, In 3rd International Conference on Learning Innovation and Quality Education (ICLIQE 2019), **397**, 944–952 (2020).
[11] E. Ikhsan, Jurnal Sistem Informasi, **10**(2), 414-422 (2021).
[12] K. Lavidas, S. Papadakis, A. Filippidi, C. Karachristos, A. Misirli, A. Tzavara, V. Komis, and N. Karacapilidis, Sustainability, **15**(7), 1–12 (2023).
[13] G. Akçapinar and A. Bayazit, Elementary Education Online, **18**(1), 406–415 (2019).
[14] J. Calderon-Valenzuela, K. Payihuanca-Mamani, and N. Bedregal-Alpaca, (IJACSA) International Journal of Advanced Computer Science and Applications, **13**(1), 321–328 (2022).
[15] M. M. Fakhri, D. Fadhilatunisa, Rosidah, M. F. B, M. A. Satnur, and F. Fajrin, Chemistry Education Review (CER), **5**(2), 157-169 (2022).
[16] D. H. Simbolon, R. H. Perangin, and K. Sebayang, Jurnal Curere, **5**(2), 92-98 (2021).
[17] H. Sokout, and T. Usagawa, SHS Web Conferences, **102**, 1-8 (2021).
[18] L. Muflikhah, D. E. Ratnawati, and R. R. M. Putri, *Data Mining* (Malang, UB Press, 2018).
[19] E. Mariati, A. Lestari, and W. Widiatry, Jurnal Teknologi Informasi: Jurnal Keilmuan Dan Aplikasi Bidang Teknik Informatika, **14**(2), 112-118 (2020).
[20] D. A. Putra, and M. Kamayani, Prosiding Seminar Nasional Teknoka, **5**, 34–40, (2020).
[21] D. Triyansyah, and D. Fitrianah, InComTech: Jurnal Telekomunikasi dan Komputer, **8**(3), 163-182 (2018).
[22] E. Etriyanti, D. Syamsuar, and Y. N. Kunang, Telematika, **13**(1), 56–67 (2020).
[23] W. Nugraha, R. Sabaruddin, and S. Murni, Techno.Com, **23**(2), 319–327 (2024).
[24] B. Nugroho, and A. Denih, KOMPUTASI: Jurnal Ilmiah Ilmu Komputer dan Matematika, **17**(2), 381-387 (2020).
[25] F. Kasyidi, R. Ilyas, and N. M. Annisa, MIND (Multimedia Artificial Intelligent Networking Database) Journal, **6**(2), 194–204 (2021).
[26] S. Widodo, H. Brawijaya, and S. Samudi, Sinkron: jurnal dan penelitian teknik informatika, **6**(4), 2407-2414 (2022).
[27] M. H. Ariansyah, E. N. Fitri, and S. Winarno, Jurnal Teknik Informatika (Jutif), **4**(3), 493-501 (2023).
[28] D. M. W. Powers, Journal of Machine Learning Technologies, **2**(1), 37-63, (2011).
[29] S. Shrestha, and M. Pokharel, International Journal of Informatics and Communication Technology (IJ-ICT), **10**(1), 9-18 (2021).