JTAM (Jurnal Teori dan Aplikasi Matematika)
http://journal.ummat.ac.id/index.php/jtam
p-ISSN 2597-7512 | e-ISSN 2614-1175
Vol. 9, No. 3, July 2025, pp. 986-999

Naïve Bayes Algorithm: Analysis of Student Group Assignment Project Patterns in Mathematics Learning

Wardhani Utami Dewi[1*], Ira Vahlia[1], Nego Linuhung[1]
[1]Universitas Muhammadiyah Metro, Lampung, Indonesia
wardhaniutamidewi@ummetro.ac.id

Article History:
Received : 11-04-2025
Revised : 17-05-2025
Accepted : 22-05-2025
Online : 03-07-2025

Keywords:
Assignment;
Classification;
Student collaboration;
Naïve bayes;
Project.

ABSTRACT

Effective collaboration in mathematics learning is essential for developing students' critical thinking and problem-solving skills; however, identifying patterns that lead to successful group collaboration remains challenging. This study aims explicitly to identify and classify the patterns of student group assignment completion in the Logic and Sets course using the Naïve Bayes algorithm. Survey data from 65 mathematics education students were analyzed using a quantitative approach and machine learning techniques. Attributes such as group size, task completion time, participation, contribution strategies, and communication effectiveness were collected via structured questionnaires. Data analysis involved preprocessing, model training using Naïve Bayes, and validation through accuracy and posterior probability analysis. Results indicated that the Naïve Bayes model accurately distinguished groups with very good (A) and fairly good (B) performance, achieving 84.62% accuracy. Groups achieving an A grade typically featured balanced participation and open communication strategies, whereas groups graded B exhibited uneven participation and passive members. This research significantly contributes by demonstrating how data-driven predictive analytics can support instructors in monitoring and enhancing collaborative learning processes in mathematics courses. Future research could further refine predictive accuracy by incorporating additional factors such as leadership style and collaborative technologies, potentially integrating the model into learning management systems for real-time evaluation and intervention.

Crossref
https://doi.org/10.31764/jtam.v9i3.30459

——————————— ◆ ———————————

A. INTRODUCTION

Mathematics learning in higher education increasingly emphasizes collaborative and teamwork-based approaches, essential for fostering students' critical thinking, problem-solving skills, and effective interpersonal communication. Group assignments have become fundamental strategies in achieving these educational outcomes (Irma et al., 2023), significantly enhancing learning experiences and academic performance (Li & Tu, 2024; Liu et al., 2023).

The success and effectiveness of collaborative group tasks are influenced considerably by internal factors, including individual student contributions, internal group communication, task distribution strategies, and overall group management (Lorente et al., 2024; Subasman, 2024). Understanding and optimizing these collaborative patterns through systematic analytical approaches are essential to maximize educational benefits and improve learning quality (Song et al., 2024; Zamiri & Esmaeili, 2024).

Previous research related to the analysis of the effectiveness of group work has been carried out with various approaches, including research by Qureshi et al. (2023) which analyzes internal and external factors in the success of student collaboration using the SEM-PLS method. Meanwhile, a study conducted by Melguizo-Garín et al. (2022) Using multiple regression techniques to predict the relationship between group work competencies and project-based learning satisfaction in students. Recent research Ahmed et al. (2024) has also applied the data mining method with the Decision Tree algorithm to classify the pattern of student participation in group work. However, the use of probabilistic algorithms such as Naïve Bayes in the context of student group assignments in mathematics learning is still rare in the current literature.

Observations from Universitas Muhammadiyah Metro Mathematics Education Study Program indicate persistent challenges in completing group tasks, such as uneven division of responsibilities, ineffective communication, and poor time management, potentially compromising the quality of outcomes in the Logic and Sets course. This highlights the importance of employing predictive analytics to explore and address these issues effectively. Naïve Bayes, a probabilistic classification algorithm grounded in Bayes' theorem, is advantageous due to its computational efficiency (Nhu et al., 2020), ease of interpretation, and ability to handle categorical data effectively (El Barakaz et al., 2021; Tan et al., 2022). Given its relevance in recognizing patterns and predicting group performance based on observable attributes, Naïve Bayes serves as a suitable methodological choice.

Naïve Bayes is one of the probabilistic-based classification algorithms that works with the principle of Bayes' Theorem (Ekong et al., 2024; Nakhipova et al., 2024; Vu et al., 2022), and has been widely used in various fields such as text classification, behavioral prediction, and risk detection. According to An et al. (2023) and Bahtiar et al. (2023) the main advantage of this algorithm lies in its ability to handle categorical data efficiently, as well as provide fairly accurate classification results even with a limited amount of data. In addition, Naïve Bayes is very easy to implement and the results are relatively easy to interpret by non-technical users, such as educators or Education practitioners (Albreiki et al., 2022; Nurjulaiha et al., 2025; Xu & Babaian, 2021).

In the context of mathematics learning, particularly group assignments, the Naïve Bayes algorithm enables probabilistic analysis of various factors influencing successful student collaboration (Nakhipova, Kerimbekov, Umarova, ibrahim Bulbul, et al., 2024). By assuming independence between features, this algorithm can recognize dominant patterns emerging in group work and predict group outcomes based on observed attributes, making it highly relevant and suitable for this study (Uddin et al., 2022; Ujwal & Malik, 2023). Moreover, integrating machine learning algorithms like Naïve Bayes into educational evaluation frameworks provides educators with actionable insights, enabling targeted interventions to enhance collaboration dynamics and overall academic outcomes, as well as facilitating early identification of groups needing additional support. This research explicitly aims to identify and classify patterns of student collaboration in completing group assignments within the Logic and Sets course using the Naïve Bayes algorithm, thereby contributing insights to enhance collaborative learning practices in higher education.

B. METHODS

This study employs a quantitative approach with exploratory and predictive characteristics, aligned with the objective to identify student collaboration patterns and predict group assignment success using machine learning algorithms. This approach enables systematic data analysis and the application of probabilistic classification models to support data-driven decision-making. The population comprises all students in the Mathematics Education Study Program at Universitas Muhammadiyah Metro who have completed the Logic and Sets course. A sample of 65 students was selected from semesters 2, 4, and 6 to ensure a representative cross-section of academic levels, providing a comprehensive variation of collaboration patterns. This sample size is deemed adequate based on minimal data requirements for training and testing machine learning models within this study's context.

Data were collected through survey techniques using questionnaires that were compiled to identify various variables related to the group's task completion patterns. The questionnaire includes aspects such as the frequency of contributions, forms of communication, division of tasks, completion time, and the role of each group member. The questionnaire was distributed online using a digital platform, namely google form to facilitate the collection of data from all respondents. The survey results data will be analyzed using a machine learning-based Naïve Bayes algorithm, which is a probabilistic classification method based on Bayes' Theorem. This analysis aims to model and classify the pattern of student group assignment completion into several categories based on the attributes obtained from the questionnaire. The analysis process includes the data preprocessing stage, the division of training data and test data, the training of the Naïve Bayes model, and the evaluation of classification accuracy using metrics such as confusion matrix and prediction accuracy. The following attributes used in this study are presented in Table 1.

Table 1. Attributes of the Extraction Results from the Research Questionnaire

No.	Attribute Name	Data Type	Example Values/Categories
1	Number of Group Members	Numerical	2, 3, 4, etc.
2	Total Task Completion Time	Numerical (Hours)	3, 5, 8, etc.
3	Distribution of Processing Time	Categorical	- Divided evenly- Dominated by several members- Based on task section
4	Individual Contributions of Each Member	Numerical (%)	25%, 30%, 45%, etc.
5	Attendance of Inactive Members	Categorical (binary)	- Yes- No
6	Strategies to Address Differences in Contribution	Categorical	- Re-discussion- Encouraging/sanctions- No special effort
7	Group Task Outcome Assessment	Categorical	- Excellent (A/A-)- Fairly Good (B+/B)
8	The Effectiveness of Group Communication	Numerical (ordinal)	1 (Ineffective), 2 (Less effective), 3 (Moderately effective), 4 (Highly effective)
9	Obstacles During the Work Process	Categorical (binary)	- Yes- No
10	Suggestions for Improvement	Categorical	- A fairer division of tasks- Improved communication- Improved time management

All analysis using *R-Studio* software and the stages of analysis using machine learning with the Naïve Bayes algorithm in this study are:

1. Preprocessing Data

At this stage, the raw data of the survey results that have been collected is processed so that it is ready for analysis. This preprocessing process includes:

a. Data Cleaning
   Delete duplicate data or incomplete data and address inconsistent data (Mirzaei et al., 2022).
b. Encoding (Pengkodean Data)
   Convert categorical data into numerical formats, for example:
   1) Encoding labels (Yes=1, No=0, Very Good=4, Good=3, etc.).
   2) One-hot encoding for attributes with unordered categories (e.g. "contribution strategy").

c. Data Normalization
   Align numerical data ranges, such as time attributes or member contributions, using min-max normalization (Bisht & Kumar, 2023).

2. Data Splitting

The data that has been cleaned is divided into two parts, namely: training data is used to train the Naïve Bayes model (Afdhaluzzikri et al., 2022; Iedwan et al., 2024). In this study, the training data is 80% of the total data. Then test data, to test the performance of the Naïve Bayes model that has been trained. The testing data in this study is 20% of the total data.

3. Naïve Bayes Model Training

Naïve Bayes is a probability-based classification algorithm that uses Bayes' Theorem assuming independence between predictor variables (C. Dewi et al., 2023; Resti et al., 2023; Sabzevari & Eslamian, 2023). Mathematically, the Naïve Bayes classification is determined by the formula (1):

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \tag{1}$$

$P(Y|X)$: posterior probability (class-based attribute)YX
$P(X|Y)$: probability likelihood (probability of an attribute X given class Y)
$P(Y)$: prior probability (class probability in general Y)
$P(X)$: Probability of evidence (normalization of input data).

4. Model Testing

Test data is used to measure classification performance by inputting test data attributes into the Naïve Bayes model. Then the model will predict the class Y for each test data based on the highest probability. Compare the prediction results with the actual class.

5. Evaluation of the Naïve Bayes Model

An evaluation is carried out to find out how well the model can classify the data correctly. This evaluation uses the following metrics:

a. Confusion Matrix
   The confusion matrix presents the number of correct classifications (True Positive and True Negative) as well as the wrong classifications (False Positive and False Negative) (Chicco et al., 2021; W. U. Dewi et al., 2023; Tangirala, 2020).
b. Prediction Accuracy
   Measure the percentage of correct predictions out of the total predictions.

The research steps are described through the workflow presented in Figure 1.

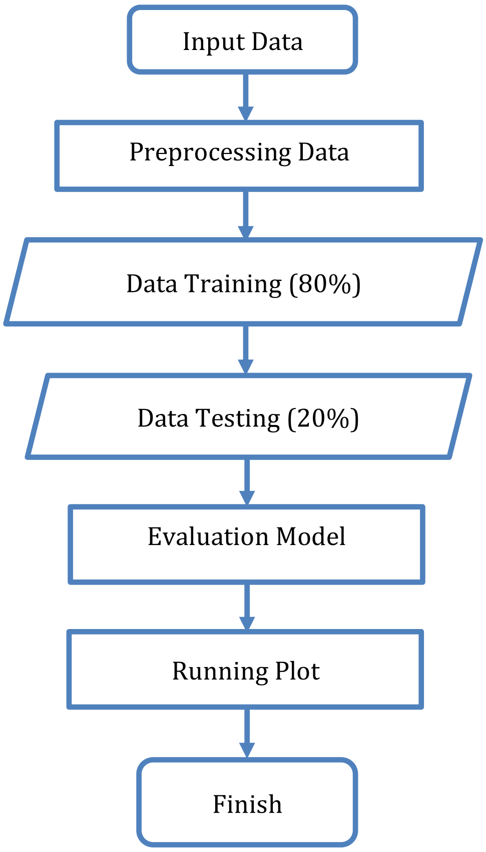

Figure 1. Workflow algoritma Naïve Bayes

C. RESULT AND DISCUSSION

The data used in this study involved a number of important attributes that reflect aspects of student collaboration in completing group assignments in the Logic and Sets course, such as the number of group members, time allocation, individual contributions, and communication effectiveness. The target class was divided into two categories: groups with excellent grades (class A) and groups with fairly good grades (class B). The data distribution showed that approximately 38.46% of the groups were classified as class A, while the remaining 61.54% were in class B.

In this study, the target variable (Y) is the assessment of student group assignment results, categorized into two classes: Class A (A+ and A-) representing excellent to very good performance, and Class B (B+ and B-) representing fairly good or needing improvement. This classification distinguishes between groups with optimal collaboration and those with less effective performance. A Naïve Bayes classification model was developed using survey data covering collaboration aspects such as group size, time spent, contribution distribution, group strategy, and communication effectiveness. The model calculates the probability of each observed attribute, individually or in combination, to predict the final grade category of the group assignment. The first step in the classification process is to look at the prior probabilities of each class, which shows the proportion of groups that obtained A and B scores in the training data. The results are presented in Figure 2.

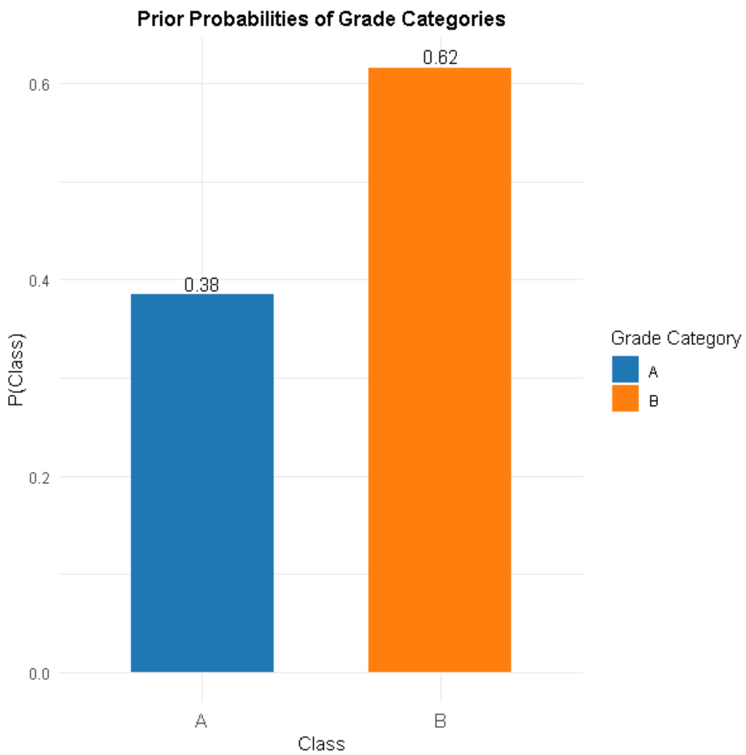

Figure 2. Prior Probability of Group Task Value Categories

Based on Figure 2 this probability shows that in the training data, around 38.46% of the student group is in the A grade category, while the other 61.54% is in the B category. Once the prior distribution of each class is known, the next step in applying the Naïve Bayes algorithm is to calculate the conditional probabilities of each predictor attribute to the target class. This conditional probability describes how likely an attribute is to appear in each value category (A or B). The following conditional probabilities of each attribute are presented in Figure 3.

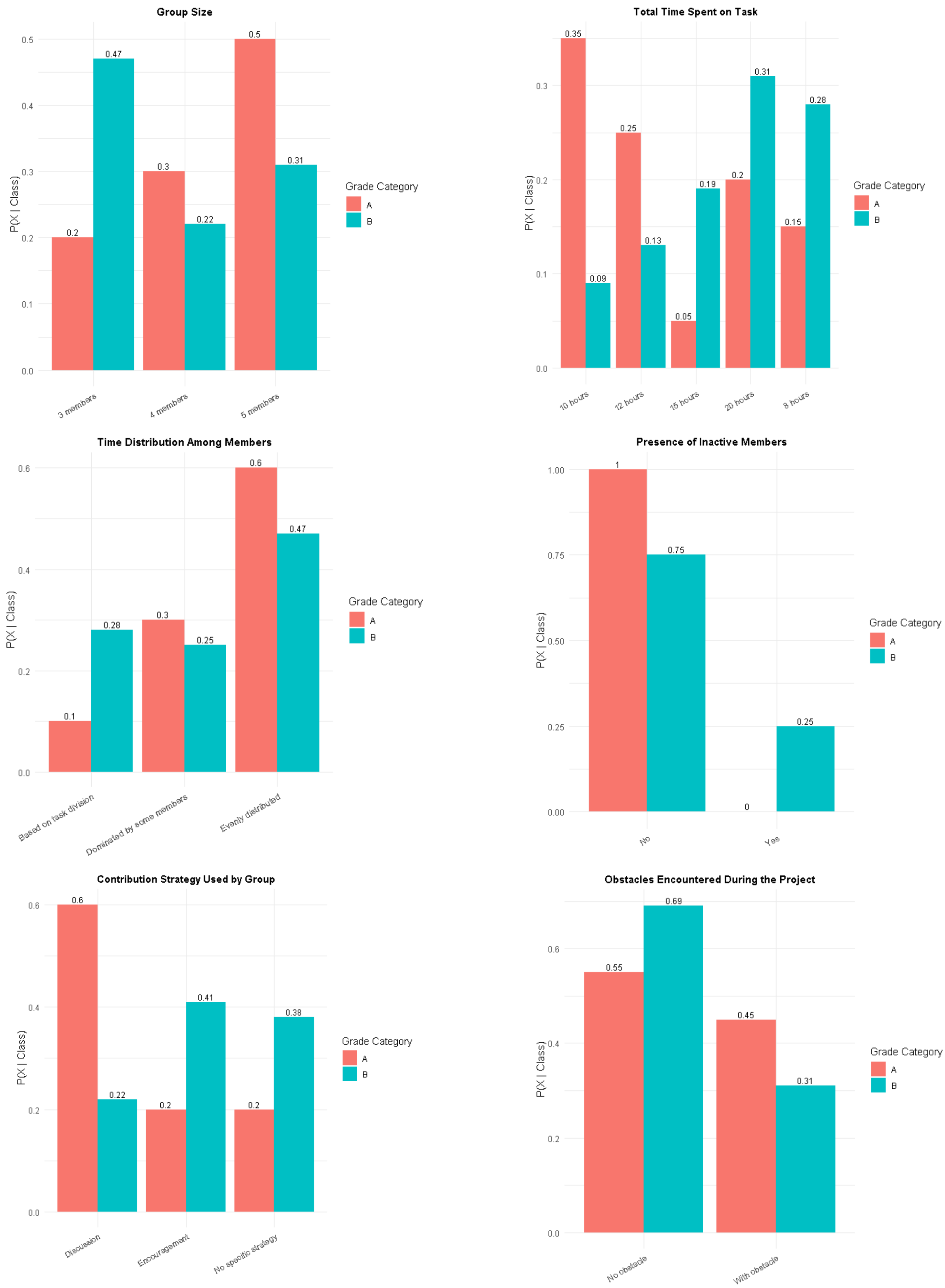

Figure 3. The conditional probability of each attribute

Based on Figure 3 the results of the conditional probability calculation of various attributes to the category of group assignment scores found a number of patterns that distinguish student groups with the final results of categories A (very good) and B (quite good). First, in terms of the number of group members, the group that obtained an A score tended to consist of five people (50%), while the group with a B score consisted of three people (47%). This shows that the larger the number of members, the higher the potential for effective collaboration, with the group management doing well. In terms of total time to complete tasks, the group with a grade of A tends to complete tasks in a relatively optimal time, which is 10 to 12 hours. Meanwhile, the group with a B value showed a less stable time pattern, either too short (8 hours) or too long (15–20 hours), indicating a possible lack of efficiency or planning.

In terms of time distribution between members, the group with the dominant A value divides the time equally (60%), while group B shows a tendency to divide time-based on task parts or even dominated by certain members. This emphasizes the importance of a balanced division of responsibilities in collaborative tasks. The participation attributes of group members also play an important role; all groups with a grade of A reported no inactive members (100%), while the group with a grade of B had 25% of cases of passive members. This confirms that the full activity of all members is the key to the success of the collaboration. Furthermore, from the strategy of overcoming inequality of contributions, the A value group tends to rely on open discussions (60%) to divide tasks fairly. In contrast, group B was more likely to only give encouragement or not even make any special efforts. This shows that communication and negotiation skills in the group affect the achievement of results.

Interestingly, in terms of communication effectiveness, the B group rated the communication as very effective (score 4), compared to the A group which was the majority at a score of 3. This shows that the perception of communication alone is not enough to determine success if it is not accompanied by a balanced distribution of work and contribution. Finally, in the attributes of obstacles in the work process, the group with a score of B reported no problems (69%) compared to group A (55%). This is likely due to differences in the level of reflectivity between groups; High-scoring groups may be more critical in assessing their internal dynamics and constraints.

Overall, these patterns suggest that the success of group tasks is not only influenced by a single factor, but is the result of a combination of factors such as member activeness, work distribution strategies, time-sharing, and problem-solving approaches. The Naïve Bayes model can help identify these patterns probabilistic and can be used as a basis for recommendations in the formation and coaching of student study groups. This classification process is then tested on the test data (test set) to find out the extent to which the model can recognize the patterns contained in the data. The evaluation of the model performance was carried out using the confusion matrix in Figure 4 and Figure 5. which shows the results of the model's classification in distinguishing the actual class and the prediction class.

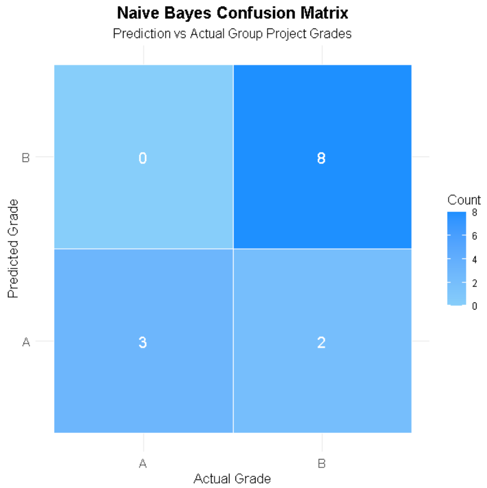

Figure 4. Confusion Matrix

Figure 5. Visualization of actual versus predicted data

Figure 4 and Figure 5 together show that out of a total of 13 test data, as many as 3 groups with an actual value of A were correctly predicted (True Positive), while 8 groups with a value of B were also correctly classified (True Negative). However, there were 2 groups with an A value that were incorrectly classified as B (False Negative), and no prediction error was found for category B (False Positive = 0). The visual pattern in Figure 3 confirms that the model is more dominant in making predictions for class B, which is reflected in the high number of B predictions and is spread for both actual data A and B. In contrast, predictions for class A only appear in actual data A and never appear in actual data B. After the results of the model classification through the confusion matrix are known, the next step is to evaluate the model's performance more quantitatively using a number of evaluation metrics. These metrics include accuracy, recall, specificity, precision, and F1-score values that provide a more comprehensive picture of the model's ability to correctly classify data, for both high (A) and medium (B) value categories, as shown in Figure 6.

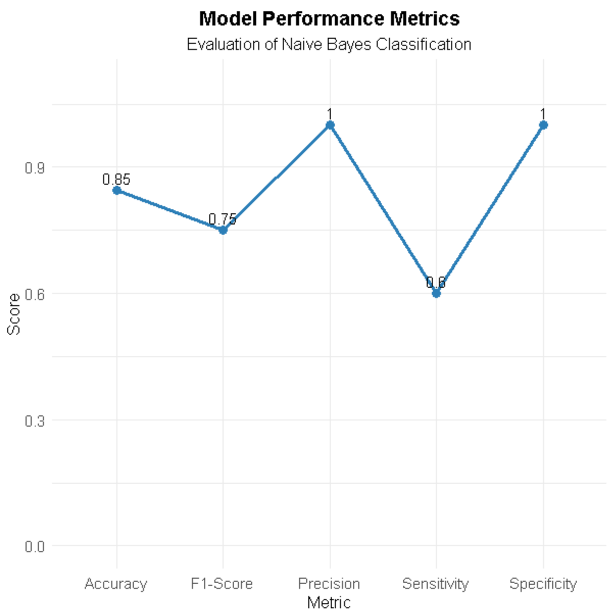

Figure 6. Model evaluation

Graph in Figure 6 shows the results of the Naïve Bayes model performance evaluation based on five main metrics, namely accuracy, F1-score, precision, sensitivity, and specificity. The model has an accuracy of 0.85, which signifies that about 85% of the test data is correctly classified. Precision and specificity both reached a value of 1.00, indicating that the model was very accurate in identifying class B and did not make predictive errors against that class. The F1 score of 0.75 reflects a good balance between precision and recall (sensitivity) which is still quite good overall. However, the sensitivity of the model is only 0.60, which means that the model is less than optimal in recognizing all data from class A. This shows that although the model is very accurate in predicting when declaring data as class A, there are still many data A that are actually classified as B. To gain a deeper understanding of the model's confidence level in each of the predictions produced, an analysis of the posterior probabilities of each observation was performed. This posterior probability visualization provides information on how confident the model is in classifying each data into a specific value category, as shown in Table 7.

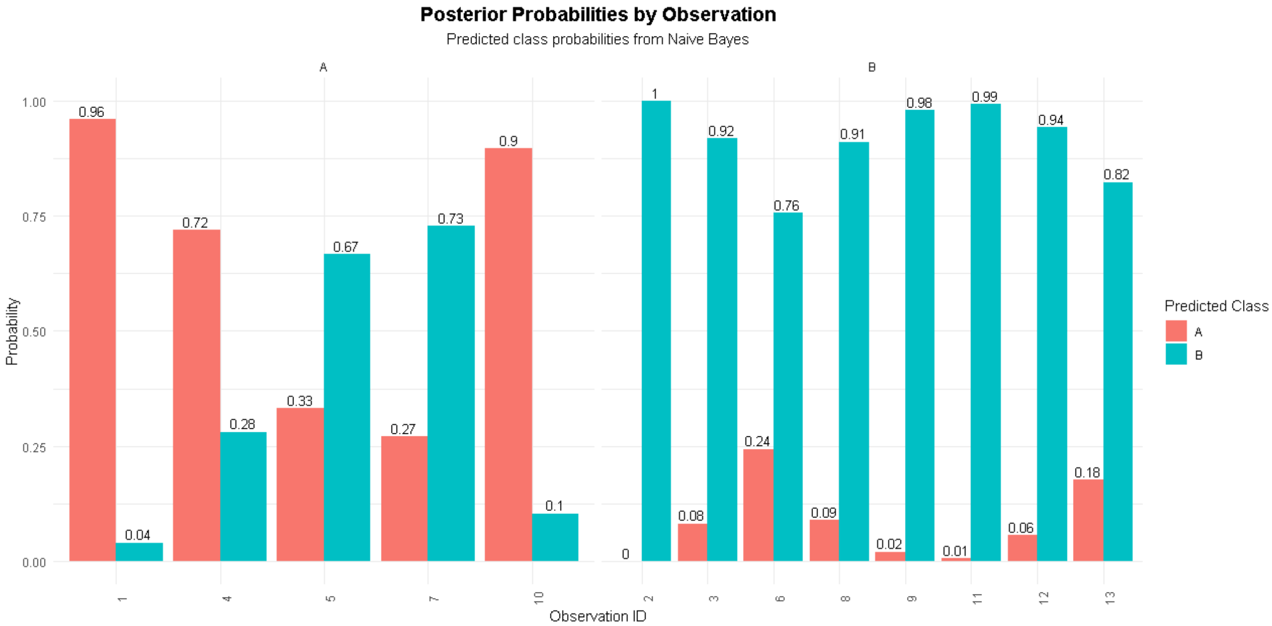

Figure 7. Posterior Opportunities

The posterior probability visualization of the Naïve Bayes model shows the model's confidence level in classifying each observation in the test data into the category of A or B values. Most of the observations have a high and clear probability of one of the classes, such as observations 1 and 10 which show a probability of more than 0.90 for class A, as well as observations 2, 3, 6, and 8 are very confident of getting into class B with a probability above 0.90. This indicates that the model is quite confident in making predictions on most of the data.

However, some observations show a more balanced probability distribution between classes A and B, such as observations 4, 5, and 7. These cases reflect model uncertainty, which means that the characteristics of such observations are not very strong in favor of one of the classes dominantly. This information is important in the context of decision-making, as observations with a probability close to the threshold could be candidates for further evaluation or manual intervention. Thus, posterior analysis not only helps to understand the classification results but also provides insight into the level of confidence and potential ambiguity in the model's decisions.

The results of the study showed that the Naïve Bayes algorithm was able to classify the pattern of completion of student group assignments in the Logic and Set courses with 84.62% accuracy. Key findings showed that the A-rated group generally consisted of five members (50%), divided their time evenly (60%), fully active (100%), and completed tasks in optimal time (10–12 hours). Meanwhile, the B-rated group mostly consists of three members (47%), with too short or long working hours (8–20 hours), unbalanced work distribution, and some passive members (25%). Contribution strategies also differentiate the two: group A uses more open discussions (60%), while group B is less likely to have any specific efforts (38%). Although group B rated their communication to be very effective (69% on a score of 4), this was not always in line with good outcomes, demonstrating the importance of systematic work in completing logic-based tasks.

The connection with the Logic and Associations course is important because this project requires structural, argumentative, and collaborative thinking of the core competencies in the course. The Naïve Bayes model succeeded in uncovering the probabilistic relationship between factors that affect the success of the group while showing the weakness of the model in detecting superior groups (sensitivity class A = 0.60). This research is in line with the findings (Winantu & Khatimah, 2023), (Choi et al., 2020)and (Lin, 2021) which both highlight the importance of interaction and contribution strategies. The implication of this study is the availability of predictive models that lecturers can use to monitor group dynamics and provide timely interventions. Its contribution lies in the use of data-driven predictive analytics to improve the quality of collaborative learning in formal logic-based courses

D. CONCLUSION AND SUGGESTIONS

This study demonstrates that the successful completion of group assignments in the Logic and Sets course is strongly influenced by balanced task distribution, active participation of all members, and effective communication strategies within the group. The quality of collaboration is a key factor in achieving high group performance outcomes. By utilizing the Naïve Bayes algorithm, dominant patterns of group collaboration were identified, providing valuable insights for educators to effectively monitor and support group dynamics. The model achieved an overall accuracy of 84.62% in classifying group success, though further refinement is needed to improve sensitivity in detecting high-performing groups. Future research is recommended to incorporate additional variables such as leadership styles and collaborative technologies to enrich the analysis and develop more robust predictive models that can be integrated into learning management systems to support collaboration and academic achievement in real time.

Future research is recommended to incorporate additional variables that may influence the success of group collaboration, such as leadership styles and collaborative technologies. Leadership styles can be measured using quantitative instruments like the Leadership Style Inventory, which assesses transformational, transactional, and laissez-faire leadership, given their significant roles in group motivation and coordination. Meanwhile, the use of collaborative technologies such as online project management platforms can be evaluated through surveys measuring frequency of use and communication effectiveness, which are believed to enhance coordination and transparency within group work. Including these variables is expected to produce a more comprehensive predictive model and provide more targeted recommendations for collaborative learning strategies in higher education.

REFERENCES

Afdhaluzzikri, A., Mawengkang, H., & Sitompul, O. S. (2022). Perfomance analysis of Naive Bayes method with data weighting. *Sinkron: Jurnal Dan Penelitian Teknik Informatika*, *6*(3), 817–821. https://doi.org/10.33395/sinkron.v7i3.11516

Ahmed, A., Nipa, F. A., Bhuyian, W. U., Mushfique, K. M., Shahin, K. I., Nguyen, H.-H., & Farid, D. M. (2024). Students' performance prediction employing Decision Tree. *CTU Journal of Innovation and Sustainable Development*, *16*(Special issue: ISDS), 42–51. https://doi.org/10.22144/ctujoisd.2024.321

Albreiki, B., Habuza, T., & Zaki, N. (2022). Framework for automatically suggesting remedial actions to help students at risk based on explainable ML and rule-based models. *International Journal of Educational Technology in Higher Education*, *19*(1), 49. https://doi.org/10.1186/s41239-022-00354-6

An, Q., Rahman, S., Zhou, J., & Kang, J. J. (2023). A comprehensive review on machine learning in healthcare industry: classification, restrictions, opportunities and challenges. *Sensors*, *23*(9), 4178. https://doi.org/10.3390/s23094178

Bahtiar, S. A. H., Dewa, C. K., & Luthfi, A. (2023). Comparison of Naïve Bayes and Logistic Regression in Sentiment Analysis on Marketplace Reviews Using Rating-Based Labeling. *Journal of Information Systems and Informatics*, *5*(3), 915–927. https://doi.org/10.51519/journalisi.v5i3.539

Bisht, K., & Kumar, A. (2023). A method for fuzzy time series forecasting based on interval index number and membership value using fuzzy c-means clustering. *Evolutionary Intelligence*, *16*(1), 285–297. https://doi.org/10.1007/s12065-021-00656-0

Chicco, D., Tötsch, N., & Jurman, G. (2021). The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Mining*, *14*(13), 1–22. https://doi.org/10.1186/s13040-021-00244-z

Choi, Y., Farnadi, G., Babaki, B., & Van den Broeck, G. (2020). Learning fair naive bayes classifiers by discovering and eliminating discrimination patterns. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*(06), 10077–10084. https://doi.org/10.1609/aaai.v34i06.6565

Dewi, C., Chen, R.-C., Christanto, H. J., & Cauteruccio, F. (2023). Multinomial Naïve Bayes classifier for sentiment analysis of internet movie database. *Vietnam Journal of Computer Science*, *10*(04), 485–498. https://doi.org/10.1142/S2196888823500100

Dewi, W. U., Nisa, K., & Usman, M. (2023). Implementation Of Decision Tree And Support Vector Machine On Raisin Seed Classification. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, *12*(1), 1309–1321. https://doi.org/10.24127/ajpm.v12i1.6873

Ekong, B., Edet, A., Udonna, U., Uwah, A., & Udoetor, N. (2024). Machine Learning Model for Adverse Drug Reaction Detection Based on Naive Bayes and XGBoost Algorithm. *British Journal of Computer, Networking and Information Technology*, *7*(2), 97–114. https://doi.org/10.52589/BJCNIT-35MFFBC6

El Barakaz, F., Boutkhoum, O., & El Moutaouakkil, A. (2021). A hybrid naïve Bayes based on similarity measure to optimize the mixed-data classification. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, *19*(1), 155–162. https://doi.org/10.12928/telkomnika.v19i1.18024

Iedwan, A. S., Mauliza, N., Pristyanto, Y., Hartanto, A. D., & Rohman, A. N. (2024). Comparative Performance of SVM and Multinomial Naïve Bayes in Sentiment Analysis of the Film'Dirty Vote'. *Scientific Journal of Informatics*, *11*(3), 839–848. https://doi.org/10.15294/sji.v11i3.10290

Irma, I., Suarlin, S., & Rahman, N. (2023). An Examination of Social Science Learning Outcomes Using Group Work and Individual Assignments: Educational Strategies in Focus. *Journal of Social Sustainability Impact*, *1*(2), 1–12. https://ojs.unpatompo.ac.id/index.php/jssi/article/view/259

Li, M.-M., & Tu, C.-C. (2024). Developing a Project-Based Learning Course Model Combined with the

Think–Pair–Share Strategy to Enhance Creative Thinking Skills in Education Students. *Education Sciences*, *14*(3), 233. https://doi.org/10.3390/educsci14030233

Lin, L. (2021). Smart teaching evaluation model using weighted naive bayes algorithm. *Journal of Intelligent & Fuzzy Systems*, *40*(2), 2791–2801. https://doi.org/10.3233/JIFS-189320

Liu, P., Cao, F., Yao, Q., Xu, L., & Dai, H. (2023). Effectiveness of educational and training intervention for improving hand hygiene compliance amongst nurses: a systematic review and meta-analysis of randomised controlled trials. *Scottish Medical Journal*, *69*(3), 110–121. https://doi.org/10.1177/00369330231178909

Lorente, S., Arnal-Palacián, M., & Paredes-Velasco, M. (2024). Effectiveness of cooperative, collaborative, and interdisciplinary learning guided by software development in Spanish universities. *European Journal of Psychology of Education*, *39*(4), 4467–4491. https://doi.org/10.1007/s10212-024-00881-y

Melguizo-Garín, A., Ruiz-Rodríguez, I., Peláez-Fernández, M. A., Salas-Rodríguez, J., & Serrano-Ibáñez, E. R. (2022). Relationship between group work competencies and satisfaction with project-based learning among university students. *Frontiers in Psychology*, *13*(811864), 1–5. https://doi.org/10.3389/fpsyg.2022.811864

Mirzaei, A., Carter, S. R., Patanwala, A. E., & Schneider, C. R. (2022). Missing data in surveys: Key concepts, approaches, and applications. *Research in Social and Administrative Pharmacy*, *18*(2), 2308–2316. . https://doi.org/10.1016/j.sapharm.2021.03.009

Nakhipova, V., Kerimbekov, Y., Umarova, Z., ibrahim Bulbul, H., Suleimenova, L., & Adylbekova, E. (2024). Integration of Collaborative Filtering Into Naive Bayes Method to Enhance Student Performance Prediction. *International Journal of Information and Communication Technology Education (IJICTE)*, *20*(1), 1–18. https://doi.org/10.4018/IJICTE.352512

Nakhipova, V., Kerimbekov, Y., Umarova, Z., Suleimenova, L., Botayeva, S., Ibashova, A., & Zhumatayev, N. (2024). Use of the naive Bayes classifier algorithm in machine learning for student performance prediction. *International Journal of Information and Education Technology*, *14*(1), 92–98. https://doi.org/10.1016/j.sapharm.2021.03.009

Nhu, V.-H., Shirzadi, A., Shahabi, H., Singh, S. K., Al-Ansari, N., Clague, J. J., Jaafari, A., Chen, W., Miraki, S., & Dou, J. (2020). Shallow landslide susceptibility mapping: A comparison between logistic model tree, logistic regression, naïve bayes tree, artificial neural network, and support vector machine algorithms. *International Journal of Environmental Research and Public Health*, *17*(8), 2749. https://doi.org/10.3390/ijerph17082749

Nurjulaiha, S., Kurniawan, R., Dikananda, A. R., & Suprapti, T. (2025). Optimizing Naïve Bayes Algorithm Through Principal Component Analysis To Improve Dengue Fever Patient Classification Model. *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, *4*(2), 1005–1015. https://doi.org/10.59934/jaiea.v4i2.798

Qureshi, M. A., Khaskheli, A., Qureshi, J. A., Raza, S. A., & Yousufi, S. Q. (2023). Factors affecting students' learning performance through collaborative learning and engagement. *Interactive Learning Environments*, *31*(4), 2371–2391. https://doi.org/10.1080/10494820.2021.1884886

Resti, Y., Irsan, C., Neardiaty, A., Annabila, C., & Yani, I. (2023). Fuzzy discretization on the multinomial naïve Bayes method for modeling multiclass classification of corn plant diseases and pests. *Mathematics*, *11*(8), 1761. https://doi.org/10.3390/math11081761

Sabzevari, Y., & Eslamian, S. (2023). Bayesian theory: Methods and applications. In *Handbook of Hydroinformatics* (pp. 57–68). Elsevier. https://doi.org/10.1016/B978-0-12-821285-1.00026-9

Song, C., Shin, S.-Y., & Shin, K.-S. (2024). Implementing the dynamic feedback-driven learning optimization framework: a machine learning approach to personalize educational pathways. *Applied Sciences*, *14*(2), 916. https://doi.org/10.3390/app14020916

Subasman, I. (2024). Evaluation of The Effectiveness of Group Cooperation In The Context of Problem Base Learning (PBL): A Mixed Method Approach. *Jurnal Konseling Pendidikan Islam*, *5*(1), 11–23. https://doi.org/10.32806/jkpi.v5i1.99

Tan, Q., Mu, X., Fu, M., Yuan, H., Sun, J., Liang, G., & Sun, L. (2022). A new sensor fault diagnosis method for gas leakage monitoring based on the naive Bayes and probabilistic neural network classifier. *Measurement*, *194*, 111037. https://doi.org/10.1016/j.measurement.2022.111037

Tangirala, S. (2020). Evaluating the impact of GINI index and information gain on classification using decision tree classifier algorithm. *International Journal of Advanced Computer Science and Applications*, *11*(2), 612–619. https://doi.org/10.14569/IJACSA.2020.0110277

Uddin, S., Ong, S., & Lu, H. (2022). Machine learning in project analytics: a data-driven framework and case study. *Scientific Reports*, *12*(1), 15252. https://doi.org/10.1038/s41598-022-19728-x

Ujwal, U. J., & Malik, S. (2023). A hybrid weight based feature selection algorithm for predicting students' academic advancement by employing data science approaches. *International Journal of Education and Management Engineering*, *13*(5), 1. https://doi.org/10.5815/ijeme.2023.05.01

Vu, D.-H., Vu, T.-S., & Luong, T.-D. (2022). An efficient and practical approach for privacy-preserving Naive Bayes classification. *Journal of Information Security and Applications*, *68*, 1–8. https://doi.org/10.1016/j.jisa.2022.103215

Winantu, A., & Khatimah, C. (2023). Perbandingan Metode Klasifikasi Naive Bayes Dan K-Nearest Neighbor Dalam Memprediksi Prestasi Siswa. *INTEK : Jurnal Informatika Dan Teknologi Informasi*, *6*(1 SE-Articles), 58–64. https://doi.org/10.37729/intek.v6i1.3006

Xu, J. J., & Babaian, T. (2021). Artificial intelligence in business curriculum: The pedagogy and learning outcomes. *The International Journal of Management Education*, *19*(3), 100550. https://doi.org/10.1016/j.ijme.2021.100550

Zamiri, M., & Esmaeili, A. (2024). Methods and technologies for supporting knowledge sharing within learning communities: A systematic literature review. *Administrative Sciences*, *14*(1), 17. https://doi.org/10.1016/j.ijme.2021.100550