Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)
Vol. 8, No. 3, November 2025, pp. 667 – 676
p-ISSN: 2614-3372 | e-ISSN: 2614-6150

Predicting Student Learning Outcomes in Vocational Computer and Network Engineering Using Naïve Bayes

[1*]Lailam Baridah, [2]Raissa Amanda Putri
[1,2]Department of Information System, State Islamic University of North Sumatra, Medan, Indonesia
Email: [1]lailambaridah84@gmail.com, [2]raissa.ap@uinsu.ac.id

Article Info
Article history:
Received Sep 16th, 2025
Revised Nov 06th, 2025
Accepted Nov 14th, 2025

Keyword:
Educational Data Mining
Naïve Bayes
Prediction System
Student Learning Outcomes
Vocational Education

ABSTRACT

This study applied the Naïve Bayes algorithm to predict student learning outcomes in the Basic Computer and Network Engineering subject at SMKN 1 Sipispis. A quantitative approach was employed, using data from 311 students, which consisted of both academic variables (assignments, midterm exams, and final exams) and non-academic variables (attendance, attitude, and learning interest). The dataset was preprocessed by cleaning, encoding, and splitting into training and testing sets using several ratios (90/10, 80/20, 70/30, and 60/40). The Naïve Bayes model was trained and evaluated using accuracy, precision, recall, and F1-score metrics. The best performance was achieved with the 80/20 data split, yielding an accuracy of 74.6%, demonstrating the model's ability to capture probabilistic relationships between academic and non-academic factors. These findings indicate that the Naïve Bayes algorithm can effectively classify student performance levels such as Fair, Good, and Excellent, providing a reliable foundation for an automated decision support system. The developed web-based system can help teachers identify students at risk of declining performance early, enabling more adaptive and data-driven educational interventions.

Corresponding Author:
Lailam Baridah,
Department of Information System,
State Islamic University of North Sumatra
Jl. Lap. Golf No.120, Kp. Tengah, Pancur Batu District, Deli Serdang Regency,
North Sumatra – Indonesia, 20353.
Email: lailambaridah84@gmail.com

DOI: http://dx.doi.org/10.24014/ijaidm.v8i3.38333

1. INTRODUCTION

Technological advancements in the education sector have opened opportunities for the integration of Artificial Intelligence (AI) and Machine Learning (ML) to enhance the quality of learning and assessment processes [1]. Among the various algorithms available, the Naïve Bayes classifier is widely utilized due to its computational efficiency, ease of implementation, and competitive accuracy compared to other models. This algorithm operates on probabilistic principles under the assumption of independence among predictor variables, making it highly suitable for predicting student learning outcomes based on diverse indicators [2].

SMKN 1 Sipispis is the only vocational high school in Sipispis District, Serdang Bedagai Regency, and plays an important role in preparing competent human resources in the field of Computer and Network Engineering (TKJ). However, the school faces challenges in accurately predicting student learning outcomes. The inability to identify students' academic performance at an early stage often leads to delays in providing guidance and interventions for those experiencing academic decline [3]. Moreover, the current manual evaluation process makes teachers' assessments less efficient and potentially subjective. Based on these challenges, this study aims to develop a web-based system for predicting student learning outcomes using the Naïve Bayes algorithm [4]. The proposed system integrates academic variables (assignment scores, midterm exams, and final exams) and non-academic variables (attendance, attitude, and learning interest) to produce a more comprehensive and objective prediction model. Through the implementation of the Naïve Bayes algorithm, this research is expected to help teachers identify at-risk students early, enhance the effectiveness of learning evaluations, and provide a more accurate basis for decision-making to improve student learning outcomes in the Basic Computer and Network Engineering subject at SMKN 1 Sipispis [5].

The application of the Naïve Bayes algorithm in the field of education has been proven effective in various prior studies; however, most of these works still exhibit several limitations that form clear research gaps. For instance, Hudzaifah et al. (2024) achieved an accuracy of 91.70% in predicting MTCNA certification outcomes, but their study focused solely on technical certification contexts, not on predicting students' academic performance [6]. Oktavia and Anggreini (2024) reported an accuracy of 97.24% in identifying PIP scholarship recipients based on socio-economic variables, yet their model did not address academic performance prediction [7]. Similarly, Zega et al. (2024) applied Naïve Bayes to assess programming proficiency among Informatics students using questionnaire data, which was effective but limited in scope and excluded academic or behavioral learning factors [8]. Ranny A.C. Walangare (2022) combined academic and non-academic factors, but did not perform a comparative analysis with alternative algorithms to evaluate Naïve Bayes' relative performance [9]. In addition, Ricky Gunawan et al. (2019) classified academic data using Naïve Bayes, yet did not integrate the model into an automated system that teachers could utilize directly [10].

From these findings, three major research gaps can be identified. First, most previous studies relied solely on academic or administrative data, without incorporating non-academic variables such as attendance, attitude, and learning interest, despite these factors significantly influencing students' academic outcomes. Second, prior works have not implemented predictive models in a web-based system, limiting their practical usability for teachers in real-time decision-making. Third, there is a lack of comparative algorithmic analysis, leaving insufficient empirical evidence regarding the relative strengths of Naïve Bayes in predicting student performance, particularly in vocational school contexts [11].

To address these gaps, this study introduces three key novelties. First, it integrates both academic and non-academic variables to construct a more holistic and representative prediction model for vocational school students. Second, it develops a web-based automatic prediction system capable of performing real-time classification, allowing teachers to easily access and utilize the results without manual computation. Third, it includes a comparative analysis with alternative algorithms, providing empirical validation of the reliability and efficiency of Naïve Bayes compared to other predictive models [12].

Thus, this research not only extends the application of the Naïve Bayes algorithm in educational prediction but also highlights its novelty through multidimensional data integration, web-based system implementation, and comparative model evaluation [13]. These innovations are expected to assist teachers in early identification of at-risk students, enable timely academic interventions, and enhance the effectiveness of learning evaluation and decision-making in vocational education, particularly in the Basic Competence of Computer and Network Engineering (TKJ) subject at SMKN 1 Sipispis.

2. RESEARCH METHOD

This research employs a quantitative approach, utilizing an experimental method to analyze and predict student learning outcomes in the Computer and Network Engineering (TKJ) program by applying the Naïve Bayes algorithm. The research stages consist of data collection, preprocessing, data splitting, model training, testing, evaluation, and prediction result generation. The workflow of this process follows the flowchart shown earlier [14]. The research methodology is illustrated in Figure 1.

The research process begins with the data collection stage, which involves obtaining archived student grade records from the Computer and Network Engineering program. The dataset consists of academic variables such as assignment scores, midterm exam (UTS), and final exam (UAS), as well as non-academic variables, including attendance, attitude, and learning interest [15]. The target variable to be predicted is the student learning outcome, categorized into six classes: Low, Poor, Fair, Average, Good, and Excellent.

The next stage is data preprocessing, which aims to prepare the dataset for analysis using the Naïve Bayes algorithm [16]. Several preprocessing activities are conducted, including data cleaning to remove irrelevant attributes such as student names or IDs, and data standardization to ensure consistent formatting across all attributes. Non-numeric categorical variables are converted into numeric form using Label Encoding; for example, learning outcome categories are encoded as follows: Low = 0, Poor = 1, Fair = 2, Average = 3, Good = 4, and Excellent = 5. In addition, missing value handling is performed by either removing or replacing incomplete records to avoid bias and preserve model accuracy.

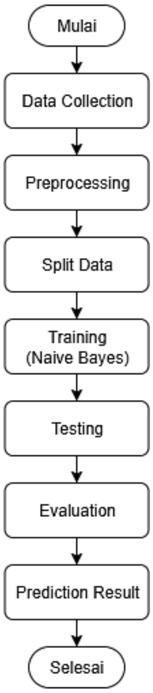

Figure 1. Stages of the Naïve Bayes Algorithm

After the preprocessing stage, the dataset was divided into several subsets with ratios of 90/10, 80/20, 70/30, and 60/40 using the hold-out method. This division was performed randomly while maintaining the balance of the six learning outcome categories to ensure that both subsets were representative and unbiased [17]. Mathematically, the data split can be expressed as equation 1.

$$D = D_{train} \cup D_{test}, D_{train} \cap D_{test} = \emptyset \tag{1}$$

which indicates that the training and testing data are mutually exclusive and non-overlapping.

The model training phase utilizes the Naïve Bayes algorithm, which is based on Bayes' Theorem [18]. The theorem can be expressed as equation 2.

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \tag{2}$$

In this equation, H represents the hypothesis or class label (the category of student learning outcomes: Low, Poor, Fair, Average, Good, Excellent), and X denotes the observed attributes such as assignments, midterms, final exams, attendance, attitude, and learning interest [19]. Here, $P(H|X)$ refers to the probability of a hypothesis H given data X, $P(X|H)$ is the likelihood of observing data X under class H, $P(H)$ is the prior probability of class H, and $P(X)$ represents the total probability of the data. The model predicts the class with the highest posterior probability using Equation 3.

$$H^* = \arg\max_{H} P(H) \prod_{i=1}^{n} P(X_i|H) \tag{3}$$

Once the model has been trained, the model testing phase is conducted to assess its ability to make accurate predictions on unseen data. The testing dataset is fed into the trained model, and the predicted results are compared against the actual class labels to evaluate predictive accuracy [20]. Following testing, the model evaluation stage is conducted to comprehensively assess the model's performance. The accuracy is calculated using Equation 4.

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Test\ Data} \times 100\% \tag{4}$$

In addition to accuracy, precision, recall, and F1-score are also computed to provide deeper insight into model performance. These metrics are defined as equation 5 [22]:

$$Precision = \frac{TP}{TP+FP}, Recall = \frac{TP}{TP+FN}, F1 - score = 2.\frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

where TP(True Positive) represents correctly predicted positive cases, FP(False Positive) denotes incorrectly predicted positive cases, and FN(False Negative) refers to positive cases that were incorrectly classified as negative [21].

The final stage of this research is the generation of prediction results, in which the system produces predictive outputs based on both academic and non-academic variables. For example, if the input data are Assignment = 85, Midterm = 88, Final Exam = 90, Attendance = High, Attitude = Good, and Learning Interest = High, then the output prediction would be Learning Outcome = Good. These prediction results are displayed through a web-based system, enabling teachers to easily identify students at risk of academic decline and provide early interventions to improve learning outcomes.

The methodological framework of this study was developed by referring to and expanding upon previous research that employed the Naïve Bayes algorithm in the field of education. Several prior studies have demonstrated the algorithm's effectiveness in performing classification and prediction tasks using educational data. Hudzaifah et al. (2024) conducted research on the application of Naïve Bayes to predict technical certification outcomes. The findings revealed that Naïve Bayes achieved high accuracy with efficient computational performance, as the algorithm operates based on simple yet effective probabilistic principles. Meanwhile, Oktavia and Anggreini (2024) utilized the Naïve Bayes algorithm to classify education aid recipients based on socioeconomic factors. Their study emphasized that Naïve Bayes performs well in processing categorical data, such as parental income and education level, without requiring complex training processes. Furthermore, Zega et al. (2024) applied the Naïve Bayes algorithm in academic assessment to predict student achievement levels based on assignment and exam scores. Their results indicated that the model could assist teachers in identifying students who might experience a decline in academic performance. Similarly, Walangare (2022) employed the Naïve Bayes algorithm to classify student performance by considering variables such as attendance, exam scores, and classroom participation. Both studies reinforced the reliability of Naïve Bayes in processing academic data to produce accurate and interpretable predictions.

Based on these prior studies, it can be concluded that the Naïve Bayes algorithm has proven effective in various educational contexts involving student performance prediction and academic classification. Therefore, this research continues that line of study by applying the Naïve Bayes algorithm to predict student learning outcomes in the basic subjects of the Computer and Network Engineering (TKJ) program, integrating both academic variables (such as assignment, midterm, and final exam scores) and non-academic factors (such as attendance, attitude, and learning interest).

3. RESULTS AND ANALYSIS

3.1. Data Collection

The data used in this study consists of 311 student records from SMK Negeri 1 Sipispis. This dataset contains both academic and non-academic information of the students. The collected variables include assignment scores, midterm scores, final exam scores, attendance, attitude, and learning interest. The student learning outcome variable is used as the prediction target in this study. The details are presented in Table 1.

Table 1. Dataset

No	Student Name	Assignment	Midterm	Final Exam	Attendance	Attitude	Learning Interest	Learning Outcome
1	Aditya Pratama Damanik	83	85	85	Average	Fair	Average	Fair
2	Afgan zailani	84	85	86	Average	Fair	Average	Fair
3	Alvin damanik	86	86	89	Average	Fair	Average	Fair
4	Ananda silfia saragih	88	89	90	High	Good	High	Good
…	…	…	…	…	…	…	…	…
311	Nilam Lestari Aquino	98	88	82	High	Fair	Average	Fair

Based on the correlation analysis shown in Figure 2 (Correlation Matrix), the relationships among the variables indicate moderate positive correlations, ranging from 0.29 to 0.64. This suggests that all attributes contribute positively to student learning outcomes, although the strength of their influence varies. The attitude variable shows the strongest correlation with learning outcomes (r = 0.64), followed by attendance (r = 0.48) and learning interest (r = 0.48). In contrast, the assignment variable has the weakest correlation value (r = 0.30), indicating that student attitudes and engagement-related factors play a more significant role compared to purely academic aspects.

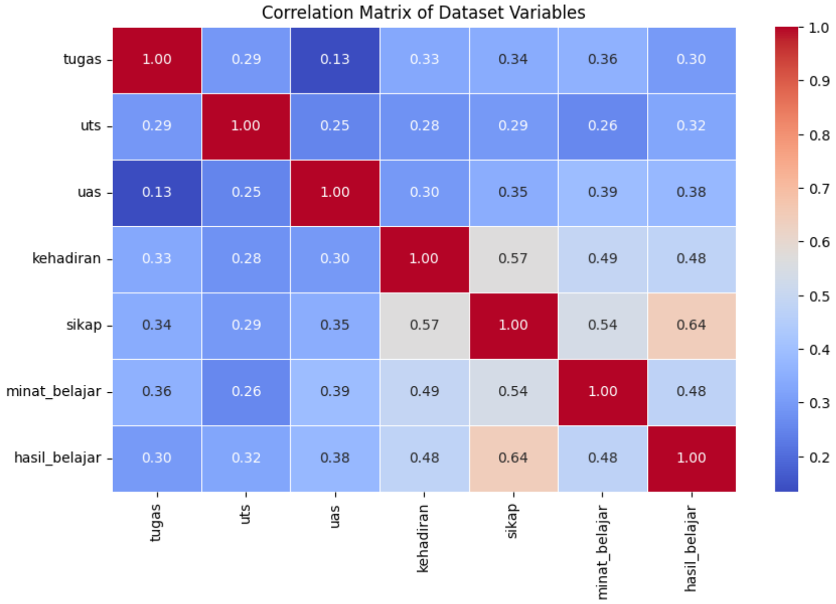

Figure 2. Correlation Matrix

Table 2. Correlation Values Between Variables and Learning Outcomes

Attribute	Correlation with Learning Outcome
Learning Outcomes	1.000000
Attitude	0.642268
Attendance	0.476948
Learning Interest	0.475833
Midterm	0.376925
Final Exam	0.324321
Assigment	0.298065

The summary of correlations with the target variable is presented in Table 2, confirming that non-academic factors, particularly attitude and attendance, have a greater impact on learning outcomes. Therefore, integrating both academic and non-academic variables in predictive modeling provides a more comprehensive understanding of student performance.

3.2. Preprocessing Data

The data preprocessing stage was carried out to clean and prepare the dataset before model training [23]. The first step was to standardize column names to lowercase and remove extra spaces, followed by deleting irrelevant columns such as No and Student Name. Numerical columns (assignments, midterm, final exam) were converted into numeric data types, while categorical columns (attendance, attitude, learning interest, learning outcome) were cleaned from inconsistent spellings and standardized using an ordinal mapping with the following order: low (0), poor (1), fair (2), average (3), good (4), and high (5). Rows containing missing values were removed to avoid interference during training. Since the Naïve Bayes algorithm is probability-based, normalization was not required as the data values were already within a consistent range. The details are presented in Table 3.

Table 3. Data Normalization Results

Assignment	Midterm	Final Exam	Attendance	Attitude	Learning Interest	Learning Outcome
83	85	85	3	2	3	2
84	85	86	3	2	3	2
86	86	89	3	2	3	2
88	89	90	5	4	5	4
…	…	…	…	…	…	…
98	88	82	5	2	3	2

3.3. Split Data

In this study, the dataset was divided into several proportions to evaluate the model's consistency and to minimize the potential for overfitting. Four different data split ratios were used, namely 90/10, 80/20, 70/30, and 60/40, applying the hold-out method. This technique was chosen to allow the model to learn effectively from the training data while maintaining sufficient unseen data for testing and validation. The splitting process was carried out randomly while maintaining the balance of the six learning outcome categories Low, Poor, Fair, Average, Good, and High, in both training and testing datasets. The total dataset used in this study consists of 311 student records from SMK Negeri 1 Sipispis. The details of the number of records used for training and testing in each split ratio are presented in Table 4.

Table 4. Split Data Training and Testing

Split Ratio	Training Data	Testing Data	Total Data
90/10	280	31	311
80/20	249	62	311
70/30	218	93	311
60/40	187	124	311

From Table 3, it can be seen that all split configurations maintain a consistent total of 311 records, differing only in the proportion allocated for training and testing. These variations were used to analyze how different data partitions affect the model's learning stability and predictive performance, which are further discussed in the evaluation section.

3.4. Model Training

The model was trained using Gaussian Naïve Bayes on 80% of the dataset (training set) [24]. All categorical features, such as attendance, attitude, and learning interest, were converted into ordinal scales according to their respective levels, while numerical features (assignments, midterm exam, and final exam) ranged between 0–100 and were therefore not normalized. The training process was conducted using the model.fit(X_train, y_train) command on the training data to build a probabilistic model that links the features to the learning outcome classes (Low, Poor, Fair, Average, Good, Excellent). The trained model was then saved for use in the evaluation stage.

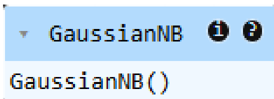

Figure 3. Model Naive Bayes

3.5. Evaluation Model

The model evaluation stage aims to assess how well the Naïve Bayes algorithm predicts student learning outcomes based on the prepared test data. In this stage, evaluation metrics such as the confusion matrix, accuracy, precision, recall, and F1-score are used to measure the model's prediction performance against the actual data. The following figure presents the average confusion matrix across the tested data split scenarios: [25].

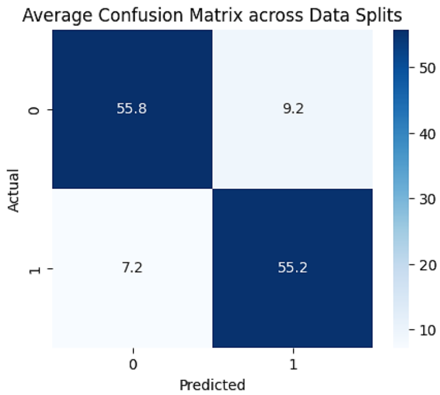

Figure 4. Confusion Matrix

Based on the confusion matrix above, the Naïve Bayes model demonstrates strong classification performance in predicting student learning outcomes. The high proportion of correct predictions (True Positives and True Negatives) and the relatively small number of incorrect predictions (False Positives and False Negatives) indicate that the model consistently recognizes patterns across different data split proportions. The comparison of evaluation metrics for each data split scenario is summarized in the following table 5.

Table 5. Confusion Matrix Result

Data Split	Train Accuracy	Accuracy	Precision	Recall	F1-Score
90/10	0.7168	0.6875	0.7155	0.6875	0.6875
80/20	0.7258	0.7460	0.7729	0.7460	0.7460
70/30	0.7327	0.7447	0.7518	0.7446	0.7446
60/40	0.7688	0.6960	0.7372	0.6969	0.6960

The following figure illustrates the comparison of training and testing accuracy across all data split scenarios, as shown in Figure 5.

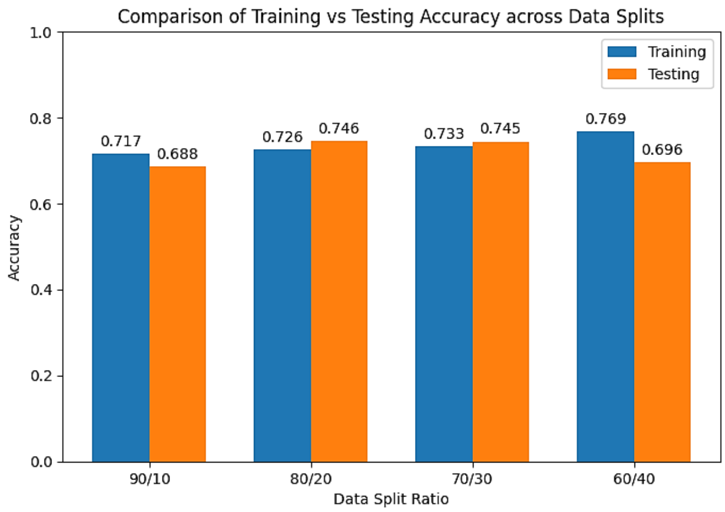

Figure 5. Comparison of Training and Test Data Accuracy

As shown in the chart, the difference between training and testing accuracy values remains relatively small across all split configurations. The highest accuracy was achieved with the 80/20 split (0.746), while the lowest was observed with the 90/10 split (0.687). This result suggests that the proportion of training and testing data influences the model's performance stability.

Overall, the evaluation results demonstrate that the Naïve Bayes model performs effectively and consistently in classifying student learning outcomes. This indicates that the model can serve as a reliable foundation for developing a decision support system to automatically monitor and evaluate students' academic performance.

3.6. Prediction Result

The final stage of this study presents the prediction outcomes of the Naïve Bayes model based on the variable assignments, midterm exams (UTS), final exams (UAS), attendance, attitude, and learning interest. As shown in the table above, the model produced prediction results consistent with the actual data. For instance, data with academic scores (assignments, UTS, UAS) in the range of 83–86 and relatively low to moderate non-academic scores (attendance, attitude, learning interest) were predicted as "Cukup" (Fair). Meanwhile, data with higher academic scores (assignments 88–90, UTS 89–92, UAS 90–92) and good non-academic performance were classified as "Baik" (Good). The Prediction Result can be seen in Table 6.

Table 6. Prediction Result

No	Student Name	Assignment	Midterm	Final Exam	Attendance	Attitude	Learning Interest	Learning Outcome	Prediction Encoded	Prediction Label
1	Aditya Pratama Damanik	83	85	85	3	2	3	2	2	Fair
2	Afgan Zailani	84	85	86	3	2	3	2	2	Fair
3	Alvin Damanik	86	86	89	3	2	3	2	2	Fair
4	Ananda Silfia Saragih	88	89	90	5	4	5	4	4	Good
5	Andika Pratama Saragih	90	92	92	5	4	5	4	4	Good
…	…	…	…	…	…	…	…	…	…	…

These results demonstrate that the Naïve Bayes model successfully captures the relationship patterns between academic and non-academic factors in predicting student learning outcomes. Therefore, this model can be effectively applied as a decision-support tool in schools to help teachers identify students at risk of performance decline and provide early learning interventions.

3.7. Implementation System

This page functions to display and manage student training data while executing the entire prediction model process. Student data, including academic and non-academic information, is first entered and stored in the database, then processed through several preprocessing stages such as standardization, conversion of categorical attributes into numerical form, and data splitting into training and testing sets with specific ratios. The Naïve Bayes model is then trained using the training data to form probabilistic patterns, while the testing data is used to evaluate model performance using metrics such as accuracy, precision, recall, F1-score, and the confusion matrix. The Student Training Data Page and the system implementation are shown in Figure 6.

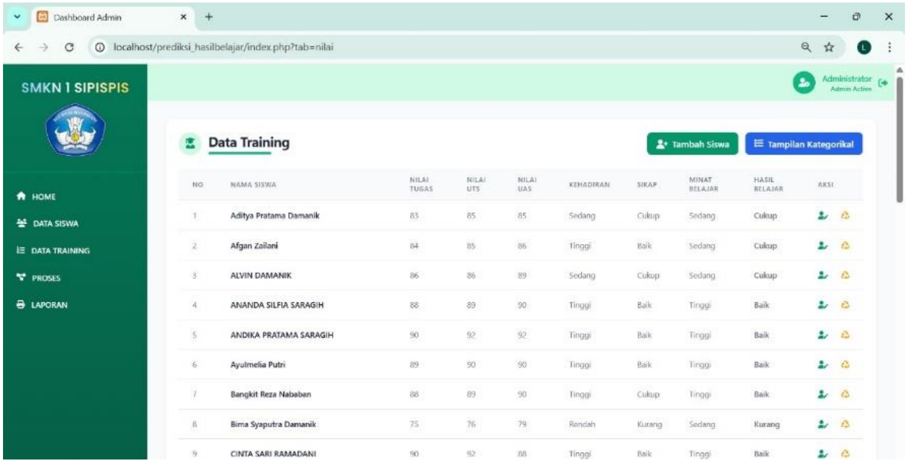

Figure 6. Student Training Data Page

The evaluation results are presented in tables and charts to facilitate model performance analysis. Once validated, the system can perform real-time predictions of new student learning outcomes, store the results in a log, and update the training data for future periods. Thus, this page not only manages student data but also integrates the entire machine learning workflow from data input to evaluation and prediction as part of a Naïve Bayes-based academic decision support system.

3.8. Discussion

Based on the testing outcomes using various data split ratios (90/10, 80/20, 70/30, and 60/40), it was observed that the 80:20 split achieved the best performance, with an accuracy of 0.746, followed by the 70:30 ratio with 0.744 accuracy. This indicates that using 80% of the data for training and 20% for testing provides the optimal balance between the model's learning ability and its generalization to new data. These findings align with machine learning theory, which states that using too little training data can lead to underfitting, where the model fails to capture key patterns, while using too much training data may cause overfitting, reducing generalization. Therefore, the 80:20 ratio proved to be the most effective configuration for the student learning outcome dataset used in this study.

When compared to previous studies that also applied the Naïve Bayes algorithm in the education domain, the results show a similar trend, where the best performance was typically achieved with training data ratios between 70% and 80%. This reinforces that Naïve Bayes is suitable for predicting learning outcome categories, particularly when the dataset includes both academic attributes (assignments, midterm, and final exams) and non-academic factors (attendance, attitude, and learning interest). In terms of implications, this study highlights the significance of non-academic factors such as student attitude and learning interest, which show a strong correlation with learning outcomes. The developed system can serve as a decision support tool for teachers or academic advisors to monitor and evaluate student performance more comprehensively.

However, this study has several limitations, including a relatively small dataset size and the use of only one machine learning algorithm (Naïve Bayes). Future research should consider expanding the dataset, implementing cross-validation techniques, and comparing performance with other algorithms, such as Decision Tree or Random Forest, to achieve a more comprehensive analysis.

4. CONCLUSION

This study concludes that the Naïve Bayes algorithm can effectively predict student learning outcomes in the Basic Computer and Network Engineering subject at SMKN 1 Sipispis by integrating both academic and non-academic variables. The model achieved the highest accuracy of 74.6% with an 80/20 data split, demonstrating stable and consistent classification performance across different data proportions. The results indicate that non-academic factors such as attendance, attitude, and learning interest significantly influence student achievement alongside academic variables. The novelty of this research lies in three aspects: (1) the integration of academic and non-academic variables to provide a more comprehensive prediction model, (2) the implementation of a web-based automated prediction system that enables real-time classification and teacher accessibility, and (3) the empirical evaluation of the Naïve Bayes algorithm's effectiveness across multiple data split scenarios. For future work, it is recommended to expand the dataset size, incorporate additional influencing factors such as socio-economic conditions and learning environment, and compare performance with other algorithms such as Decision Tree or Random Forest. Moreover, integrating the Naïve Bayes-based system into school information management platforms would enable continuous and real-time monitoring of student progress, supporting data-driven interventions and improving learning outcomes in vocational education.

REFERENCES

[1] T. S. Nurazizah, "Artificial Intelligence dan Machine Learning dalam Kehidupan Manusia," 2025. [Online]. Available: https://www.researchgate.net/publication/389464205

[2] A. Ashari Muin, "Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)," *Jurnal Ilmiah Ilmu Komputer*, vol. 2, no. 1, 2016, [Online]. Available: http://ejournal.fikom-unasman.ac.id

[3] D. El, R. Purba, and Z. A. Matondang, "Web Quality Evaluation Penerapan Model View Controller Design Pattern pada Pengembangan Sistem Informasi SMK Negeri 1 Sipispis," 2021.

[4] I. Magdalena, H. N. Fauzi, and R. Putri, "Pentingnya Evaluasi Dalam Pembelajaran Dan Akibat Memanipulasinya," 2020. [Online]. Available: https://ejournal.stitpn.ac.id/index.php/bintang

[5] D. Khairy, N. Alharbi, M. A. Amasha, M. F. Areed, S. Alkhalaf, and R. A. Abougalala, "Prediction of student exam performance using data mining classification algorithms," *Educ Inf Technol (Dordr)*, vol. 29, no. 16, pp. 21621–21645, Nov. 2024, doi: 10.1007/s10639-024-12619-w.

[6] S. M. Hudzaifah *et al.*, "Implementasi Algoritma Naïve Bayes dalam Memprediksi Tingkat Kelulusan Siswa pada Sertifikasi Mikrotik Certified Network Associate (MTCNA)," 2024. [Online]. Available: https://journal.stmiki.ac.id

[7] I. P. Oktavia and N. L. Anggreini, "Implementasi Algoritma Naive Bayes Dengan Metode Klasifikasi Dalam Menentukan Siswa Penerima Bantuan Program Indonesia Pintar (Studi Kasus : SMPN 3 Cihampelas)," 2024.

[8] I. K. Zega, N. Medina, D. Aprillia S, and Y. Yennimar, "Implementasi Algoritma Naïve Bayes untuk Memprediksi Kemampuan Pemrograman Mahasiswa Teknik Informatika Menggunakan Dataset Kuesioner," *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 4, no. 11, pp. 377–389, Dec. 2024, doi: 10.52436/1.jpti.483.

[9] R. Angella, C. Walangare, and B. Sujatmiko, "Penerapan Algoritma Naive Bayes Dalam Sistem Pendukung Keputusan Pemilihan Peminatan Konsentrasi Berdasarkan Nilai Akademik Berbasis Web Pada Program Studi S1 Pendidikan Teknologi Informasi," 2022.

[10] R. Gunawan, S. Hadi Wijoyo, and S. A. Wicaksono, "Klasifikasi Hasil Belajar Peserta Didik Pada Jurusan Teknik Komputer dan Jaringan (TKJ) di SMK Negeri 3 Malang Menggunakan Algoritme Naïve Bayes," 2019. [Online]. Available: http://j-ptiik.ub.ac.id

[11] A. A. Permana, R. Taufiq, R. Destriana, and A. Nur'aini, "Implementasi Algortima Naïve Bayes Untuk Prediksi Kelulusan Mahasiswa," vol. 13, pp. 65–70, 2024, [Online]. Available: http://jurnal.umt.ac.id/index.php/jt/index

[12] J. Teknologi Informasi Mura Randi Estian Pambudi, H. Purnomo, and R. Irawan, "Pemanfaatan Data Mining Untuk Prediksi Prestasi Akademik Siswa Berdasarkan Pola Kehadiran, Aktivitas Belajar Mengguakan Naive Bayes Logistic Regression," 2024.

[13] I. I. Sepulau, K. Kusumanto, and N. L. Husni, "Student Behavior Monitoring System in Classroom Environment Using YOLOv8," *Indonesian Journal of Artificial Intelligence and Data Mining*, vol. 8, no. 2, p. 437, Aug. 2025, doi: 10.24014/ijaidm.v8i2.37086.

[14] R. Rahayu, "Algoritma Naive Bayes," 2023. [Online]. Available: https://www.researchgate.net/publication/376713713

[15] M. Salsabila, L. Lindawati, and M. Fadhli, "Development of a CNN-Based Mental Health Consultation Application Integrating Facial Expressions and DASS-42 Questionnaire," *Indonesian Journal of Artificial Intelligence and Data Mining*, vol. 8, no. 2, p. 417, Aug. 2025, doi: 10.24014/ijaidm.v8i2.37525.

[16] S. K. A. Firdausi, "4 Tahap Preprocessing Data, Beserta Penjelasan & Studi Kasus," https://dibimbing.id/blog/detail/mengenal-apa-itu-tahap-preprocessing-data.

[17] N. S. Fauziah and R. D. Dana, “Implementasi Algoritma Naive bayes dalam Klasifikasi Status Kesejahteraan Masyarakat Desa Gunungsari,” *Blend Sains Jurnal Teknik*, vol. 1, no. 4, pp. 295–305, Mar. 2023, doi: 10.56211/blendsains.v1i4.234.

[18] M. Saiful, S. Samsuddin, and Moh. F. Wajdi, “Implementasi Algoritma Naive Bayes Untuk Memprediksi Predikat Ketuntasan Belajar Siswa Pasca Pandemi Covid 19,” *Infotek: Jurnal Informatika dan Teknologi*, vol. 4, no. 1, pp. 29–38, Jan. 2021, doi: 10.29408/jit.v4i1.2982.

[19] D. Dahri, F. Agus, and D. M. Khairina, “Metode Naive Bayes Untuk Penentuan Penerima Beasiswa Bidikmisi Universitas Mulawarman,” *Jurnal Informatika Mulawarman*, vol. 11, no. 2, p. 29, 2016.

[20] N. S. Ramadan, D. Darwis, L. Ratu, K. Kedaton, and B. Lampung, “Perbandingan Metode Naïve Bayes Dan Svm Untuk Sentimen Analisis Masyarakat Terhadap Serangan Ransomware Pada Data KIP-K,” *Jurnal Sistem Informasi dan Informatika (Simika)*, vol. 8, no. 1, 2025.

[21] R. Nurhidayat and K. E. Dewi, “KOMPUTA : Jurnal Ilmiah Komputer dan Informatika Penerapan Algoritma K-Nearest Neighbor Dan Fitur Ekstraksi N-Gram Dalam Analisis Sentimen Berbasis Aspek,” vol. 12, no. 1, 2023, [Online]. Available: https://www.kaggle.com/datasets/hafidahmusthaanah/skincare-review?select=00.+Review.csv.

[22] U. I. Wahyuni, R. Kurniawan, A. Info, K. Kunci, and K. Kemampuan, “Algoritma K-Nearest Neighbor Classification Sebagai Sistem Pengelompokan Kemampuan Akademik Siswa Berbasis Web,” vol. 5, no. 2, pp. 589–600, 2025, doi: 10.51454/decode.v5i2.1246.

[23] Y. Y. Zandroto, A. V. Vitianingsih, A. L. Maukar, N. K. Hikmawati, and R. Hamidan, “Sentiment Analysis of BCA Mobile App Reviews Using K-Nearest Neighbour and Support Vector Machine Algorithm,” *Indonesian Journal of Artificial Intelligence and Data Mining*, vol. 8, no. 2, p. 448, Aug. 2025, doi: 10.24014/ijaidm.v8i2.37773.

[24] N. D. Prasojo and M. Z. Abdillah, “Perancangan Sistem Pakar Untuk Mendeteksi Kerusakan Laptop Dengan Metode Naive Bayes Berbasis Web,” 2024.

[25] W. I. Rahayu, C. Prianto, and E. A. Novia, “Perbandingan Algoritma K-Means dan Naïve Bayes Untuk Memprediksi Prioritas Pembayaran Tagihan Rumah Sakit Berdasarkan Tingkat Kepentingan pada PT. Pertamina (PERSERO),” 2021.

BIBLIOGRAPHY OF AUTHORS

Lailam Baridah is currently an information systems undergraduate student at the State Islamic University of North Sumatra, Indonesia. She is interested in deep learning about data mining, namely data processing, methods such as Naive Bayes and visualizing data with the Python programming language. She can be contacted at the email address: lailambaridah84@gmail.com

Raisa Amanda Putri received her S.Kom degree from STMIK Mikroskil University, Medan, North Sumatra in 2011. Then she continued her education by earning an M.T.I degree from Bina Nusantara University in 2015. Currently, she serves as a lecturer in information systems at the State Islamic University of Sumatra. north. One of his research interests is data mining. She can be contacted at the email address: raissa.ap@uinsu.ac.id