Journal of Soft Computing Exploration
Homepage: htttps://shmpublisher.com/index.php/joscex
p-ISSN: 2746-7686
e-ISSN: 2746-0991

.

An intelligent academic recommendation system for learning support in higher education

Dwiny Meidelfi[1], Dikky Chandra[2], Fanni Sukma[3], Ulya Ilhami Arsyah[4], Sri Yusnita[5]
[1,3,4]Department of Information Technology, Politeknik Negeri Padang, Indonesia
[2,5]Department of Electrical Engineering, Politeknik Negeri Padang, Indonesia

Article Info

Article history:

Received March 25, 2026
Revised May 12, 2026
Accepted May 25, 2026

Keywords:

Academic recommendation system
Educational data mining
Higher education
Machine learning
Personalized learning

ABSTRACT

Higher education institutions increasingly rely on data-driven approaches to improve student learning outcomes. However, many academic advisory systems still provide general recommendations without considering individual learning patterns and academic performance. This study proposes an intelligent academic recommendation system that utilizes machine learning techniques to support personalized learning in higher education. The proposed system analyzes student academic data including grade point average, attendance, assignment scores, and study habits to predict academic performance. The proposed approach was evaluated using a dataset consisting of 1000 simulated student records representing academic performance indicators in higher education. Based on prediction results, the system generates personalized learning recommendations to assist students in improving their academic outcomes. Several machine learning algorithms, including Decision Tree, Random Forest, and Support Vector Machine, were evaluated to determine the most suitable predictive model. Experimental results show that the Random Forest algorithm achieved the highest prediction accuracy compared with other models. The developed system provides proactive learning recommendations that can assist both students and academic advisors in making better academic decisions.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Dwiny Meidelfi,
Department of Information Technology,
Politeknik Negeri Padang
Kampus Unand, Padang City, West Sumatra, Indonesia.
Email: dwinymeidelfi@pnp.ac.id
https://doi.org/10.52465/joscex.v7i2.28

1. INTRODUCTION

Higher education institutions play a critical role in supporting student academic success [1], [2]. However, many students still face difficulties in managing study plans, maintaining effective learning strategies, and achieving optimal academic performance. In many universities, academic advising systems continue to rely on manual consultations and generalized recommendations that do not adequately account for individual student characteristics and learning behaviors [3], [4], [5].

With the rapid expansion of digital learning platforms and academic information systems, universities generate large volumes of educational data, including student grades, attendance records, assignment

performance, and online learning activities [6], [7], [8]. This data provides valuable opportunities for applying data mining and machine learning techniques to uncover patterns related to student learning behavior and academic outcomes.

Educational Data Mining (EDM) [9], [10], [11], [12] has emerged as an important research field that focuses on analyzing educational datasets to improve learning processes and academic decision-making [4], [13], [14], [15], [16]. Machine learning algorithms such as Decision Trees, Random Forest, and Support Vector Machines have been widely used to predict student academic performance and identify students at risk of academic difficulties [17], [18]. These predictive models enable educational institutions to analyze complex relationships among academic indicators and provide insights that support data-driven academic interventions.

Several previous studies have proposed course recommendation systems to assist students in selecting appropriate courses based on their academic records [19], [20], [21]. While these systems provide useful support for academic planning, many existing approaches primarily focus on course selection and performance prediction [18], [21], [22], [23]. As a result, they often fail to provide comprehensive recommendations that address broader learning strategies and academic improvement.

Existing studies predominantly focus on either academic performance prediction or course recommendation independently. Most predictive models emphasize classification accuracy without translating prediction outcomes into adaptive learning guidance, while recommendation systems generally concentrate on course selection rather than personalized academic intervention. Consequently, only a limited number of studies provide an integrated framework that combines predictive analytics, academic risk classification, and personalized learning support within a unified intelligent academic advisory system.

To address this limitation, this study proposes an intelligent academic recommendation system that integrates machine learning–based academic risk prediction with personalized learning guidance. The proposed system analyzes student academic data, predicts potential academic risk levels, and generates personalized recommendations to assist students in improving their learning outcomes. By combining predictive analytics with recommendation mechanisms, the system aims to support more effective academic advising and personalized learning strategies in higher education.

The objectives of this study are threefold. First, this research aims to develop an intelligent academic recommendation framework that supports personalized learning in higher education. Second, the study evaluates several machine learning algorithms for predicting student academic risk levels based on educational data. Third, the proposed system generates personalized academic recommendations that translate prediction results into actionable guidance for students and academic advisors.

The main contribution of this research lies in the development of an integrated machine learning–based academic recommendation framework that combines academic risk prediction with personalized learning recommendations. Unlike previous studies that primarily focus on academic performance prediction, this study introduces a recommendation-driven approach that transforms predictive insights into practical academic guidance. The proposed framework therefore provides a more comprehensive solution for supporting data-driven academic advising and improving student learning outcomes in higher education.

Unlike previous studies that mainly focus on prediction accuracy or isolated recommendation mechanisms, the proposed framework integrates academic risk prediction, risk-level classification, and personalized recommendation generation within a unified decision-support workflow. The framework not only predicts student academic risk levels but also transforms predictive results into adaptive learning guidance that can support practical academic intervention processes in higher education.

Table 1 summarizes recent studies on student performance prediction and academic recommendation systems, highlighting their methodologies, datasets, and limitations.

Table 1. Comparison of previous studies on academic recommendation and academic risk prediction

No	Study	Method	Dataset / Context	Main Contribution	Limitation
1	[24]	Multiple classifiers + threshold method	6 semesters, institutional data mart, online higher education	In-depth early warning system with dashboards for students and teachers; method to determine quality thresholds for prediction	Focused on online setting; generalizability to blended/face-to-face unclear
2	[25]	Deep Neural Network (DNN) with SMOTE resampling	Public 4-year university, programming/data structures courses	DNN outperforms traditional ML (DT, LR, SVM, KNN) for at-risk identification; 89% accuracy	Limited to CS courses; imbalanced data requires resampling; single institution
3	[26]	Deep Learning + SMOTE	UCI student performance dataset	Early prediction with 96.4% accuracy (Portuguese) and 93.2%	Small benchmark dataset; may not

			(math & Portuguese courses)	(math) using demographic, social, and grade features	generalize to other contexts
4	[27]	User-based Collaborative Filtering	Real university enrollment data (compulsory + elective courses)	Grade prediction for course recommendation with 96.4% precision on elective courses; examines course type relationships	Does not incorporate content features or temporal learning patterns
5	[28]	Matrix Factorization with contextual features	Higher education course enrollment histories	Incorporates prerequisites, meeting times, instructors, and instructional methods into course recommendation	Evaluation limited to relevance metrics; no long-term student outcome tracking
6	[17]	Interpretable classifiers (blended learning)	Higher education institution, blended learning courses	Emphasizes interpretability, actionability, and adaptability of predictive models; effective from early semester stages	Requires course-specific model tuning; tested in blended format only
7	[29]	Neural Network	LMS grade book data, Spring 2022 (N = 22,041)	Predicts pass/fail at week 5 with 88% accuracy using only 4 features (current grade + missing grades); 74% of "generally at-risk" students failed ≥1 course	Uses only grade book data; misses behavioral and engagement signals
8	[30]	Random Forest with resampling strategies	Portuguese Polytechnic University, students enrolled 2009–2017	Multi-class phased prediction (dropout, delayed, on time) across three phases of first academic year; identifies most relevant features	Best results only at end of first semester when grades are available
9	[31]	Knowledge Graph + Collaborative Filtering	Educational domain data with course descriptions	Combines course semantic relationships via knowledge graph with CF; reduces RMSE and MAE vs. standard CF	Requires construction and maintenance of a domain knowledge graph
10	[32]	Predictive Learning Analytics + Recommender System	Formative assessments in hybrid learning environment	Integrates formative assessment data with recommender system for scaffolded tutoring; interactive dashboard for educators	Prototype stage; limited large-scale validation
11	[33]	Educational data mining + culturally responsive intervention	Taiwanese university, 2,856 students across 64 courses	Full prediction–intervention–evaluation cycle; AUC of 0.85; 73% of at-risk students passed after culturally adapted intervention (Cohen's d = 0.91)	Small intervention trial (n = 48); cultural specificity may limit transferability
12	[34]	Hybrid CNN-LSTM deep learning	Three benchmark datasets	Achieves up to 97.5% accuracy; addresses missing data, class imbalance, and feature selection systematically	Benchmark datasets may not reflect real institutional complexity
13	Proposed Study	Decision Tree, Random Forest, SVM	Simulated higher education dataset	Integrates academic risk prediction with personalized learning recommendation	Requires validation using real institutional data

The comparison presented in Table 1 highlights both the strengths and limitations of existing approaches in student performance prediction and academic recommendation Bañeres et al. [24] demonstrated an effective early warning system using multiple classifiers; however, their study was confined to online learning environments, limiting its applicability to blended or face-to-face academic settings. Similarly, Nabil et al. [25] and Aslam et al. [26] applied deep learning techniques to achieve high prediction accuracy, yet their models were evaluated on narrow course-specific datasets and did not incorporate any mechanism to translate prediction outcomes into actionable academic guidance.

Several studies have also explored recommendation systems in educational settings. Ceyhan et al. [27] employed user-based collaborative filtering for course grade prediction, while Wang et al. [28] incorporated contextual features such as prerequisites and instructor information into a course enrollment recommender. Despite their contributions, both approaches focus exclusively on course selection and do not integrate academic risk prediction or personalized learning intervention. Consequently, students who are already at academic risk may still receive course recommendations without any accompanying guidance for improving their learning performance.

More recent works have attempted to combine prediction and intervention. Shegupta et al. [35] integrated formative assessment data with a recommender system to support scaffolded tutoring, and Chang et al. [33] proposed a prediction–intervention framework to improve academic outcomes for at-risk students. Nevertheless, both studies remain at the prototype stage with limited large-scale validation, and their recommendation mechanisms are tied to specific institutional or cultural contexts, reducing their generalizability. Adefemi et al. [34] achieved high accuracy using a hybrid CNN-LSTM model, but evaluated their approach only on benchmark datasets that may not reflect the complexity of real institutional academic data.

Despite these advances, a consistent limitation across existing studies is the separation between predictive modeling and practical recommendation generation. Studies that achieve high prediction accuracy tend to stop at the classification stage, while studies that propose recommendation mechanisms often lack a rigorous predictive component or fail to translate risk classifications into structured, adaptive learning guidance. Furthermore, many existing frameworks have not been designed to support the full advisory cycle, from detecting at-risk students to generating specific, actionable recommendations for both students and academic advisors. To address these gaps, the proposed study introduces an integrated framework that combines machine learning–based academic risk prediction with a rule-driven recommendation engine, enabling the transformation of predictive insights into personalized learning interventions within a unified decision-support system.

In contrast, the proposed study integrates machine learning–based academic risk prediction with a recommendation engine that generates personalized learning guidance. This integrated approach enables the transformation of predictive insights into actionable academic recommendations, providing a more comprehensive solution for data-driven academic advising in higher education.

2. METHOD

This study adopts a data-driven approach to develop an intelligent academic recommendation system designed to support personalized learning in higher education [36]. The proposed system analyzes student academic data using machine learning techniques to predict academic risk levels and generate personalized learning recommendations. The research methodology consists of several stages including dataset preparation, data preprocessing, machine learning modeling, and recommendation generation.

The proposed framework operates sequentially by transforming raw student academic data into predictive insights and personalized recommendations. Student academic indicators are first processed through preprocessing and feature extraction stages before being analyzed by machine learning models. The prediction results are then converted into academic risk classifications, which are subsequently interpreted by the recommendation engine to generate adaptive learning guidance and intervention recommendations. The final output of the system consists of personalized academic recommendations that can support both students and academic advisors in improving learning outcomes.

Figure 1 presents the proposed AI-based academic recommendation framework designed to support personalized learning in higher education. The framework integrates student academic data analysis, machine learning–based prediction, and recommendation generation to assist students and academic advisors in improving learning outcomes.

The framework begins with the Student Academic Data module, which collects various types of academic and behavioral data from students. These data include grade point average (GPA), attendance records, assignment scores, and study habits. Such information represents important indicators of student academic performance and learning engagement.

The collected data are then processed in the Data Preprocessing and Feature Extraction stage [37], [38]. In this stage, the dataset undergoes several preprocessing steps including data cleaning, normalization, and feature selection [39], [40], [41]. The objective of this stage is to ensure that the dataset is consistent and suitable for machine learning analysis while extracting relevant features that significantly influence student academic performance.

After preprocessing, the processed dataset is used in the Predictive Model Training stage. In this research, several machine learning algorithms are applied to build predictive models capable of identifying student academic risk levels. The algorithms evaluated include Decision Tree, Random Forest, and Support Vector Machine. These algorithms are selected due to their effectiveness in classification problems and their ability to analyze complex educational datasets.

The trained model is then used to perform Risk Prediction, which classifies students into different academic risk levels. The prediction results provide insights into whether students are categorized as low risk, moderate risk, or high risk based on their academic performance and learning behavior.

The prediction results are subsequently processed by the Recommendation Engine. This module analyzes the predicted academic risk level and generates appropriate recommendations to support student learning improvement. The recommendation engine consists of three main components: risk level analysis, recommendation rules, and learning resource suggestions.

Finally, the system produces Personalized Learning Recommendations that can be used by both students and academic advisors. For students, the system provides recommendations such as adjusting study

plans, attending tutoring sessions, or improving learning strategies. For academic advisors, the system provides information that can be used to monitor student progress and provide appropriate academic guidance. Overall, the proposed framework integrates predictive analytics and recommendation mechanisms to create a data-driven academic support system that enhances personalized learning in higher education.

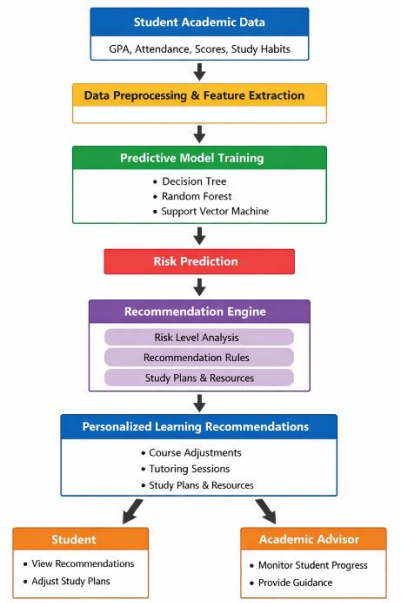

Figure 1. Proposed AI-based academic recommendation framework for personalized learning support

Data Collection

The dataset used in this study consists of 1000 simulated student records representing academic performance data collected from a higher education environment. The dataset was designed to reflect realistic student academic characteristics commonly found in university academic information systems. Each record contains several attributes describing student learning behavior and academic performance. The dataset includes features such as grade point average, attendance percentage, assignment scores, quiz scores, study hours, learning management system activity, course load, and late submission rate. Table 2 presents the attributes used in this research.

Table 2. Student academic dataset attributes.

Attribute	Description
Student_ID	Unique identifier of each student
Age	Student age
Semester	Current semester
GPA_Previous	Previous semester GPA
GPA_Current	Current GPA
Attendance	Class attendance percentage
Assignment_Score	Average assignment score
Quiz_Score	Average quiz score
Midterm_Score	Midterm examination score
Final_Score	Final examination score
Study_Hours	Weekly study hours
LMS_Logins	Number of LMS logins per week
Course_Load	Number of enrolled courses
Late_Submission_Rate	Percentage of late assignments
Advisor_Meetings	Number of meetings with academic advisor

Based on these attributes, each student record was assigned an Academic Risk Level label consisting of three categories: Low Risk, Moderate Risk and High Risk**.** These labels represent the predicted academic condition of the student and are used as the target variable for machine learning classification. The simulated dataset was generated based on typical academic performance distributions observed in higher education institutions. Numerical attributes such as GPA, assignment scores, and attendance were generated using realistic value ranges derived from common university grading systems. The dataset was designed to reflect realistic student learning behaviors to support machine learning experimentation.

The simulated dataset was generated using realistic academic performance distributions derived from commonly observed higher education grading patterns. Numerical variables such as GPA, attendance,

assignment scores, and examination scores were generated using constrained random distributions to reflect plausible student learning behaviors and academic conditions. The generated dataset was designed to approximate realistic academic conditions while maintaining sufficient variability for machine learning experimentation and classification analysis.

Data Preprocessing

Before applying the machine learning algorithms, the dataset was processed to ensure data quality and consistency. The preprocessing stage involved several procedures, including data cleaning, data normalization, and feature selection. Data cleaning was conducted to identify and correct missing or inconsistent values in order to improve dataset reliability. Numerical attributes such as GPA, attendance, and assessment scores were normalized to maintain a consistent scale across all features. In addition, feature selection techniques were applied to identify the most relevant attributes that significantly influence student academic performance and contribute to academic risk prediction.

Numerical features were normalized using Min-Max normalization to ensure consistent feature scaling. Missing values were handled using mean-value imputation for numerical attributes. Feature selection was performed using correlation-based analysis to identify attributes that significantly contributed to academic risk prediction. Categorical labels representing academic risk levels were encoded into numerical classes before model training. These preprocessing steps help improve the accuracy and efficiency of machine learning algorithms.

Machine Learning Algorithm

This study evaluates several machine learning algorithms for predicting student academic risk levels. The algorithms were selected based on their effectiveness in classification tasks and their common usage in educational data mining.

Decision Tree is a classification method that uses hierarchical decision rules to classify data [42]. This algorithm is widely used because of its interpretability and ability to represent decision processes clearly. Decision Tree classifies data by recursively splitting the dataset based on the attribute that provides the highest information gain. The entropy of a dataset $S$is calculated as:

$$Entropy(S) = -\sum_{i=1}^{c} p_i \log_2 p_i \tag{1}$$

where $p_i$denotes the proportion of samples belonging to class i, and $c$is the total number of classes. The information gain of an attribute $A$is defined as:

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \tag{2}$$

where $S_v$is the subset of $S$for which attribute Ahas value v. This criterion helps determine the best attribute for splitting the data at each node.

Random Forest is an ensemble learning algorithm [43], [44] that combines multiple decision trees to improve classification performance and reduce overfitting. This method is suitable for handling complex datasets with multiple features. Random Forest is an ensemble learning method that combines multiple decision trees to improve classification accuracy and reduce overfitting. The final prediction is obtained through majority voting from all individual trees:

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), \ldots, h_k(x)\} \tag{3}$$

where $h_j(x)$represents the prediction of the j-th decision tree, $k$is the total number of trees, and $\hat{y}$is the final predicted class. The algorithm introduces randomness by using bootstrap samples and random feature selection during tree construction, which increases diversity among trees and improves generalization performance.

Support Vector Machine (SVM) is a supervised learning algorithm [45], [46] that constructs optimal hyperplanes to separate different classes in the dataset. SVM is effective in high-dimensional classification problems. Support Vector Machine (SVM) aims to find the optimal hyperplane that separates data points from different classes with the maximum margin. The linear decision function is defined as:

$$f(x) = w^T x + b \tag{4}$$

where $w$is the weight vector, $x$is the feature vector, and $b$is the bias term.

The objective of optimization of SVM is to minimize:

$$\min_{w,b} \frac{1}{2} \| w \|^2 \tag{5}$$

subject to:

$$y_i(w^T x_i + b) \geq 1, i = 1,2, \dots, n \tag{6}$$

where $y_i$is the class label of sample x_i, and $n$is the number of training samples. This formulation enables SVM to identify the optimal separating boundary between different academic risk categories.

The Decision Tree model was implemented using the Gini impurity criterion, while Random Forest utilized 100 decision trees with bootstrap aggregation. The Support Vector Machine model employed a radial basis function (RBF) kernel for nonlinear classification. All models were implemented using the Scikit-learn library in Python.

The mathematical formulations above describe the underlying principles of the machine learning models used in this study. Decision Tree was applied to identify rule-based classification patterns, Random Forest was employed to improve classification robustness through ensemble learning, and Support Vector Machine was used to determine the optimal decision boundary between academic risk classes. These models were then compared to determine the most suitable approach for the proposed academic recommendation system.

Model Training and Evaluation

The dataset was divided into training and testing datasets using an 80:20 ratio. The training dataset was used to train the machine learning models, while the testing dataset was used to evaluate model performance. The performance of each algorithm was evaluated using several classification metrics: accuracy, precision, recall and F1-score. These metrics provide comprehensive evaluation of model performance in predicting student academic risk levels. In addition to overall accuracy, class-level evaluation was performed using confusion matrix analysis to assess the capability of the models in distinguishing different academic risk categories.

Recommendation Generation

After predicting student academic risk levels, the system generates personalized academic recommendations to support student learning improvement. Students classified as High Risk are advised to reduce their course load, increase study hours, and participate in tutoring programs. Meanwhile, students categorized as Moderate Risk are encouraged to engage in additional academic mentoring and improve learning consistency. Students identified as Low Risk are recommended to maintain their current learning strategies and academic habits. The recommendation mechanism transforms predictive results into actionable guidance that can assist students and academic advisors in improving academic outcomes.

3. RESULTS AND DISCUSSIONS

Experimental Setup

The experiments were conducted using a simulated dataset consisting of 1000 student records. The dataset includes various attributes representing student academic performance and learning behavior, such as GPA, attendance, assignment scores, quiz scores, study hours, and learning management system (LMS) activity. The dataset was divided into training and testing sets using an 80:20 ratio, where 800 records were used for model training and 200 records were used for model testing. Three machine learning algorithms were evaluated in this study: Decision Tree, Random Forest and Support Vector Machine (SVM). The models were implemented using Python and evaluated using several classification performance metrics, including accuracy, precision, recall, and F1-score.

Model Performance Comparison

The performance of the evaluated machine learning models is presented in Table 3.

Table 3. Performance comparison of machine learning models

Model	Accuracy	Precision	Recall	F1-score
Decision Tree	0.82	0.81	0.80	0.80
Random Forest	0.88	0.87	0.86	0.86
SVM	0.85	0.84	0.83	0.83

The experimental results indicate that the Random Forest algorithm achieved the highest accuracy (0.88), outperforming both Decision Tree and Support Vector Machine models. This result indicates that ensemble learning methods are effective in handling educational datasets with multiple features and complex relationships. Decision Tree also produced relatively good performance due to its ability to create interpretable classification rules. However, it is more prone to overfitting compared to Random Forest. The Support Vector Machine model achieved competitive performance, but slightly lower accuracy compared with Random Forest. This may be due to the nonlinear characteristics of student academic data, which are better handled by ensemble methods.

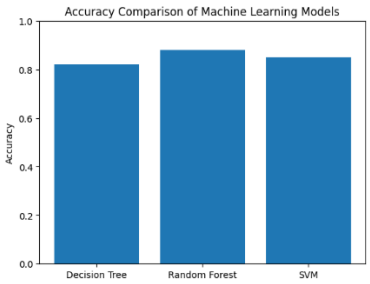

Figure 2. Accuracy comparison of machine learning models

The accuracy comparison of the evaluated machine learning models is illustrated in Figure 2. Random Forest achieved the highest accuracy of 0.88, outperforming both Decision Tree and Support Vector Machine models. This indicates that ensemble learning methods are more effective in capturing complex relationships in student academic datasets.

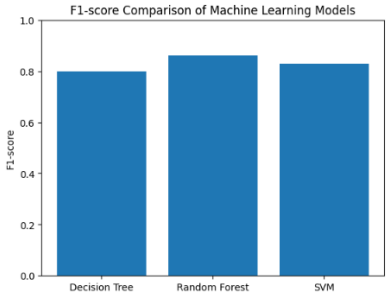

Figure 3. F1-score comparison of machine learning models

Figure 3 presents the F1-score comparison among the evaluated models. Random Forest achieved the highest F1-score value of 0.86, indicating a better balance between precision and recall. This suggests that Random Forest provides more reliable classification performance for predicting student academic risk levels.

Confusion Matrix Analysis

To further evaluate model performance, the confusion matrix of the best-performing model (Random Forest) was analyzed.

Table 4. Confusion matrix for random forest model

Actual / Predicted	Low Risk	Moderate Risk	High Risk
Low Risk	72	5	3
Moderate Risk	6	58	6
High Risk	2	4	44

From the confusion matrix, it can be observed that the Random Forest model correctly classified the majority of student risk levels. A total of 72 students in the Low Risk category was correctly predicted, while 58 students in the Moderate Risk category and 44 students in the High Risk category were also accurately classified. Only a small number of students were misclassified between adjacent categories, which is common in educational datasets where student performance levels may overlap. This result demonstrates that the model can effectively distinguish between different academic risk levels.

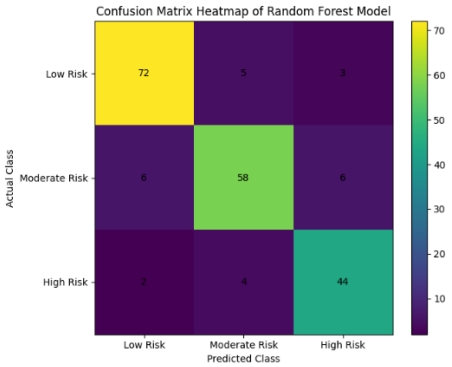

Figure 4. Confusion matrix heatmap of the random forest model

Figure 4 presents the confusion matrix heatmap of the Random Forest model. The heatmap shows that the highest number of correct classifications appears along the diagonal cells, indicating that the model successfully identified most student academic risk categories. The model correctly classified 72 Low Risk students, 58 Moderate Risk students, and 44 High Risk students. Misclassifications mainly occurred between adjacent categories, particularly between Low Risk and Moderate Risk, as well as between Moderate Risk and High Risk. This pattern suggests that some student profiles share similar academic characteristics, making classification more challenging. Overall, the confusion matrix confirms that the Random Forest model provides reliable performance for predicting student academic risk levels.

Recommendation Generation Results

After predicting student academic risk levels, the system generates personalized academic recommendations. Examples of generated recommendations are presented in Table 5.

Table 5. Example of generated recommendations

Student ID	Risk Level	Recommended Action
S001	High Risk	Reduce course load and attend tutoring sessions
S024	Moderate Risk	Increase study hours and meet academic advisor
S076	Low Risk	Maintain current study strategy
S112	High Risk	Improve attendance and submit assignments earlier

These recommendations are generated using rule-based decision logic based on predicted risk levels and student learning indicators. The recommendation system provides actionable guidance that can assist students in improving their academic performance. Academic advisors can also use these insights to identify students who require additional support.

Discussion

The experimental results demonstrate that machine learning techniques can effectively analyze student academic data and predict academic risk levels. Among the evaluated models, Random Forest achieved the best performance with an accuracy of 0.88 and the highest F1-score compared with Decision Tree and Support Vector Machine models. This superior performance can be attributed to the ensemble learning mechanism of Random Forest, which combines multiple decision trees to reduce overfitting and improve generalization capability. By aggregating predictions from multiple trees, Random Forest can capture complex relationships among various academic indicators such as GPA, attendance, assignment scores, and learning engagement. As a result, the model provides more robust predictions when compared with single-model approaches such as Decision Tree.

From an educational perspective, the proposed intelligent academic recommendation system provides practical benefits for higher education institutions. The ability to identify students at risk of academic difficulties allows universities to implement early intervention strategies. Academic advisors can use the prediction results to monitor student performance and provide appropriate guidance, while students can receive

personalized recommendations such as adjusting study habits, increasing learning engagement, or participating in tutoring programs. Therefore, the integration of predictive analytics with recommendation mechanisms enables universities to implement a data-driven academic support system that promotes personalized learning and improves student success rates.

Compared with previous studies on educational data mining, many existing approaches primarily focus on predicting student performance without generating actionable recommendations for academic improvement. For example, several studies using Decision Tree or Neural Network models mainly concentrate on classification accuracy while neglecting the practical application of prediction results in academic advisory processes. In contrast, the proposed system integrates machine learning–based prediction with a recommendation engine that translates prediction results into personalized academic guidance. This integration provides a more comprehensive solution for supporting student learning in higher education and demonstrates the potential of machine learning technologies in developing intelligent academic advisory systems.

Although the dataset used in this study was simulated, it was generated based on realistic academic performance distributions commonly observed in higher education institutions. Future research may explore the integration of the proposed recommendation system with learning management systems (LMS) to enable real-time academic monitoring and adaptive learning support.

4. CONCLUSION

This study proposed an intelligent academic recommendation system that utilizes machine learning techniques to support personalized learning in higher education. The system analyzes student academic data to predict academic performance and generate personalized recommendations. Experimental results show that the Random Forest algorithm achieved the highest prediction accuracy compared with Decision Tree and Support Vector Machine models. The developed system provides useful recommendations that can help students improve their learning strategies and academic performance. Future research will focus on integrating the system with learning management systems and expanding the dataset to improve prediction accuracy. The proposed framework demonstrates the potential of machine learning–driven recommendation systems for improving data-driven academic advising in higher education environments.

ACKNOWLEDGEMENTS

The authors would like to express their sincere appreciation to Politeknik Negeri Padang for providing institutional support and research facilities. The authors also acknowledge the contributions of colleagues and students involved in data preparation and system development. Their support has been invaluable in completing this research.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Author1: conceptualization, methodology, software development, and project administration. **Author2**: software development and writing of the original draft. **Author3**: writing, review, and editing of the manuscript. **Author4**: validation of the model and experimental results**. Author5**: supervision of the research project.

DECLARATION OF COMPETING INTERESTS

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY

Data will be made available on request.

REFERENCES

[1] E. Richardson *et al.*, "Why Not Remote? Campus-Centric HR Policies Impact on Wellness and Retention of Student Affairs Personnel," *Journal of Education Human Resources*, vol. 43, no. 1, 2025, doi: 10.3138/jehr-11_Richardson_4.

[2] M. A. Bodine Al-Sharif, K. Koo, and K. Bista, "International graduate students: Unique stories and missing voices in higher education," *New Directions for Teaching and Learning*, vol. 2025, no. 181, 2025, doi: 10.1002/tl.20616.

[3] Y. Chen, Q. Zheng, S. Ji, F. Tian, H. Zhu, and M. Liu, "Identifying at-risk students based on the phased prediction model," *Knowl. Inf. Syst.*, vol. 62, no. 3, 2020, doi: 10.1007/s10115-019-01374-x.

[4] M. Limniou, T. Varga-Atkins, C. Hands, and M. Elshamaa, "Learning, student digital capabilities and academic performance over the COVID-19 pandemic," *Educ. Sci. (Basel).*, vol. 11, no. 7, 2021, doi: 10.3390/educsci11070361.

[5] J. Li and E. Xue, "Dynamic Interaction between Student Learning Behaviour and Learning Environment: Meta-Analysis of Student Engagement and Its Influencing Factors," *Behavioral Sciences*, vol. 13, no. 1, 2023, doi: 10.3390/bs13010059.

[6] I. R. Banday, M. Zaman, S. M. K. Quadri, S. A. Fayaz, and M. A. Butt, "Big Data in Academia: A Proposed Framework for Improving Students Performance," *Revue d'Intelligence Artificielle*, vol. 36, no. 4, 2022, doi: 10.18280/ria.360411.

[7] W. Villegas-Ch, M. Roman-Cañizares, A. Jaramillo-Alcázar, and X. Palacios-Pacheco, "Data analysis as a tool for the application of adaptive learning in a university environment," *Applied Sciences (Switzerland)*, vol. 10, no. 20, 2020, doi: 10.3390/app10207016.

[8] W. Villegas-Ch, X. Palacios-Pacheco, and S. Luján-Mora, "A business intelligence framework for analyzing educational data," *Sustainability (Switzerland)*, vol. 12, no. 14, 2020, doi: 10.3390/su12145745.

[9] S. M. Dol, P. M. Jawandhiya, and P. R. Satav, "Use of Microsoft Excel for Data Collection and Processing to Predict Students' Performance in EDM," *Journal of Engineering Education Transformations*, vol. 37, no. Special Issue 2, 2024, doi: 10.16920/jeet/2024/v37is2/24020.

[10] M. Bellaj, A. Ben Dahmane, and L. Sefian, "Educational Data Mining: Employing Machine Learning Techniques and Hyperparameter Optimization to Improve Students' Academic Performance," *International journal of online and biomedical engineering*, vol. 20, no. 3, 2024, doi: 10.3991/ijoe.v20i03.46287.

[11] A. Angeioplastis, J. Aliprantis, M. Konstantakis, and A. Tsimpiris, "Predicting Student Performance and Enhancing Learning Outcomes: A Data-Driven Approach Using Educational Data Mining Techniques," *Computers*, vol. 14, no. 3, 2025, doi: 10.3390/computers14030083.

[12] M. A. Khan, K. Mahboob, U. Yousuf, M. Ramzan, M. T. Shaikh, and S. Akber, "Investigating the Role of LASSO in Feature Selection for Educational Data Mining (EDM) Applications," *VFAST Transactions on Software Engineering*, vol. 13, no. 2, 2025, doi: 10.21015/vtse.v13i2.2111.

[13] S. O. Oppong, "Predicting Students' Performance Using Machine Learning Algorithms: A Review," *Asian Journal of Research in Computer Science*, vol. 16, no. 3, 2023, doi: 10.9734/ajrcos/2023/v16i3351.

[14] A. Kord, A. Aboelfetouh, and S. M. Shohieb, "Academic course planning recommendation and students' performance prediction multi-modal based on educational data mining techniques," *J. Comput. High. Educ.*, 2025, doi: 10.1007/s12528-024-09426-0.

[15] J. Pan, Z. Zhao, and D. Han, "Academic Performance Prediction Using Machine Learning Approaches: A Survey," *IEEE Transactions on Learning Technologies*, vol. 18, 2025, doi: 10.1109/TLT.2025.3554174.

[16] M. Yağcı, "Educational data mining: prediction of students' academic performance using machine learning algorithms," *Smart Learning Environments*, vol. 9, no. 1, 2022, doi: 10.1186/s40561-022-00192-z.

[17] T. A. Kustitskaya, A. A. Kytmanov, and M. V. Noskov, "Early Student-at-Risk Detection by Current Learning Performance and Learning Behavior Indicators," *Cybernetics and Information Technologies*, vol. 22, no. 1, 2022, doi: 10.2478/cait-2022-0008.

[18] Ms. A. J. Dorothy and Dr. M. Sangeetha, "Predicting Academic Dropouts Using AI and Behavioral Data: A Hybrid Deep Learning Framework for Student Retention," *Int. J. Environ. Sci.*, 2025, doi: 10.64252/6a5y9t48.

[19] D. A. Shafiq, M. Marjani, R. A. A. Habeeb, and D. Asirvatham, "Student Retention Using Educational Data Mining and Predictive Analytics: A Systematic Literature Review," 2022. doi: 10.1109/ACCESS.2022.3188767.

[20] M. N. Gul, M. Arif, S. Gulzar, G. Naveed, and W. Abbasi, "Deep Learning-Driven Student Performance Analysis: Detecting Anomalies and Predicting Academic Success," *Inverge Journal of Social Sciences*, vol. 4, no. 1, 2025, doi: 10.63544/ijss.v4i1.117.

[21] A. H. Alsaeedi, H. Kadhim Gharkan, and M. Radif, "An Analysis of Datasets for Student Performance Evaluation Based on Machine Learning," *Iraqi Journal of Information and Technology Systems DOI*, vol. 1, no. 1, 2025.

[22] S. N. Thakur and B. Rawat, "Prediction of Student Performance in E-Learning Environment using Machine Learning Techniques," in *Procedia Computer Science*, 2025. doi: 10.1016/j.procs.2025.04.096.

[23] B. Kumar Verma, Dr. N. Srivastava, and H. Kumar Singh, "Prediction of Students' Performance in e-Learning Environment using Data Mining/ Machine Learning Techniques," *Journal of University of Shanghai for Science and Technology*, vol. 23, no. 05, 2021, doi: 10.51201/jusst/21/05179.

[24] D. Bañeres, M. E. Rodríguez, A. E. Guerrero-Roldán, and A. Karadeniz, "An early warning system to detect at-risk students in online higher education," *Applied Sciences (Switzerland)*, vol. 10, no. 13, 2020, doi: 10.3390/app10134427.

[25] A. Nabil, M. Seyam, and A. Abou-Elfetouh, "Prediction of Students' Academic Performance Based on Courses' Grades Using Deep Neural Networks," *IEEE Access*, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3119596.

[26] N. Aslam, I. U. Khan, L. H. Alamri, and R. S. Almuslim, "An Improved Early Student's Performance Prediction Using Deep Learning," *International Journal of Emerging Technologies in Learning*, vol. 16, no. 12, 2021, doi: 10.3991/ijet.v16i12.20699.

[27] M. Ceyhan, S. Okyay, Y. Kartal, and N. Adar, "The Prediction of Student Grades Using Collaborative Filtering in a Course Recommender System," in *ISMSIT 2021 - 5th International Symposium on Multidisciplinary Studies and Innovative Technologies, Proceedings*, 2021. doi: 10.1109/ISMSIT52890.2021.9604562.

[28] X. Wang, L. Cui, M. Bangash, M. Bilal, L. Rosales, and W. Chaudhry, "A Machine Learning-based Course Enrollment Recommender System," in *International Conference on Computer Supported Education, CSEDU - Proceedings*, 2022. doi: 10.5220/0011109100003182.

[29] J. B. Osborne and A. S. I. D. Lang, "Predictive Identification of At-Risk Students: Using Learning Management System Data," *Journal of Postsecondary Student Success*, vol. 2, no. 4, 2023, doi: 10.33009/fsop_jpss132082.

[30] M. V. Martins, L. Baptista, J. Machado, and V. Realinho, "Multi-Class Phased Prediction of Academic Performance and Dropout in Higher Education," *Applied Sciences (Switzerland)*, vol. 13, no. 8, 2023, doi: 10.3390/app13084702.

[31] Y. Zhang, V. Y. Mariano, and R. P. Bringula, "Prediction of Students' Grade by Combining Educational Knowledge Graph and Collaborative Filtering," *IEEE Access*, vol. 12, 2024, doi: 10.1109/ACCESS.2024.3390675.

[32] U. U. Shegupta, M. S. Islam, and W. Hardt, "Predictive Learning Analytics Utilizing Formative Assessments and Recommender System for Study Success," in *ISCSET 2024 - 13th International Symposium on Computer Science and Educational Technology*, 2024. doi: 10.1109/ISCSET58624.2024.10807906.

[33] Y. H. Chang, F. C. Chen, and C. I. Lee, "Developing an Early Warning System with Personalized Interventions to Enhance Academic Outcomes for At-Risk Students in Taiwanese Higher Education," *Educ. Sci. (Basel).*, vol. 15, no. 10, 2025, doi: 10.3390/educsci15101321.

[34] K. O. Adefemi, M. B. Mutanga, and V. Jugoo, "Hybrid Deep Learning Models for Predicting Student Academic Performance," *Mathematical and Computational Applications*, vol. 30, no. 3, 2025, doi: 10.3390/mca30030059.

[35] U. U. Shegupta, M. S. Islam, and W. Hardt, "Predictive Learning Analytics Utilizing Formative Assessments and Recommender System for Study Success," in *ISCSET 2024 - 13th International Symposium on Computer Science and Educational Technology*, 2024. doi: 10.1109/ISCSET58624.2024.10807906.

[36] A. Abdullah, T. Hassan, and S. Mouti, "Sustainable Education through AI-Driven Monitoring and Risk Management," in *Intelligent Data-Driven: Systems with Innovations in Artificial Intelligence*, 2025. doi: 10.1201/9781003613206-24.

[37] M. R. Pratama, S. Faza Abdillah Gunawan, R. R. Zhafari, Rendy, and H. N. Irmanda, "Sentiment Analysis of Beauty Product E-Commerce Using Support Vector Machine Method," *Jurnal RESTI*, vol. 6, no. 2, 2022, doi: 10.29207/resti.v6i2.3876.

[38] Noryasminda, T. H. Saragih, R. Herteno, M. R. Faisal, and A. Farmadi, "Analysis of the Effect of Feature Extraction on Sentiment Analysis using BiLSTM: Monkeypox Case Study on X/Twitter," *Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 7, no. 2, 2025, doi: 10.35882/ijeeemi.v7i2.73.

[39] S. Z. Vahed *et al.*, "Introducing effective genes in lymph node metastasis of breast cancer patients using SHAP values based on the mRNA expression data," *PLoS One*, vol. 19, no. 8 August, 2024, doi: 10.1371/journal.pone.0308531.

[40] W.-W. Lee, N. Hashim, and S. Al-Juboori, "User Behaviour Prediction in E-Commerce Using Logistic Regression," *Journal of Informatics and Web Engineering*, vol. 4, no. 3, 2025.

[41] Mohit Angurala, "Enhancing Cybersecurity in Wireless Sensor Networks: AI Solutions to Simulated Attacks," *Journal of Information Systems Engineering and Management*, vol. 10, no. 3s, 2025, doi: 10.52783/jisem.v10i3s.373.

[42] I. V. Shelehov, D. V. Prylepa, Yu. O. Khibovska, and M. S. Otroshcenko, "MACHINE LEARNING DECISION SUPPORT SYSTEMS FOR ADAPTATION OF EDUCATIONAL CONTENT TO THE LABOR MARKET REQUIREMENTS," *Radio Electronics, Computer Science, Control*, no. 1, 2023, doi: 10.15588/1607-3274-2023-1-6.

[43] S. L. Setyowati, A. Qalbi, R. Aristawidya, B. Sartono, and A. R. Firdawanti, "Optimizing Random Forest Parameters with Hyperparameter Tuning for Classifying School-Age KIP Eligibility in West Java," *Jambura Journal of Mathematics*, vol. 7, no. 1, 2025, doi: 10.37905/jjom.v7i1.28736.

[44] A. Alsubayhin, M. S. Ramzan, and B. Alzahrani, "Crime Prediction Model using Three Classification Techniques: Random Forest, Logistic Regression, and LightGBM," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 1, 2024, doi: 10.14569/IJACSA.2024.0150123.

[45] S. Adaramaja, S. Aliyu, and F. B. Abdullahi, "Hyper-parameter tuning for support vector machine using an improved cat swarm optimization algorithm," *Journal of the Nigerian Society of Physical Sciences*, vol. 5, no. 4, 2023, doi: 10.46481/jnsps.2023.1007.

[46] C. Kaushik, A. D. McRae, M. Davenport, and V. Muthukumar, "New Equivalences between Interpolation and SVMs: Kernels and Structured Features," *SIAM J. Math. Data Sci.*, vol. 6, no. 3, 2024, doi: 10.1137/23M1568764.

.