An Explainable Artificial Intelligence Model for Mapping Students' Potential Interests, Talents, and Competencies Using Random Forest and SHAP.

Andisyah Putra, Solly Aryza, Eko Hariyanto

Abstract

Selecting the appropriate vocational specialization at SMK Negeri 9 Medan is crucial for student learning engagement and career development, yet traditional placement methods often fail to adequately capture complex relationships between student interests and competencies. To address this gap, this study proposes a hybrid machine learning model combining Extreme Gradient Boosting (XGBoost) and Random Forest through weighted soft voting to predict student specialization suitability. Utilizing a dataset of 1,200 students across four vocational tracks, the proposed hybrid model (XGB-RF) achieved superior performance with an accuracy of 89.4%, ROC-AUC of 0.947, and F1-score of 88.4%, outperforming standalone XGBoost (87.6%) and Random Forest (85.8%). Feature importance analysis revealed that interest score (28%), academic achievement (24%), and practical competency (20%) serve as the primary decision drivers, contributing a cumulative 72% to the model's predictions. Ultimately, this evidence-based framework provides an objective, transparent decision-support system to assist educators and school management in optimizing student specialization placement.

Keywords: Educational Data Mining, Vocational Specialization, Student Interest, Competency Alignment, XGBoost, Random Forest, Hybrid Machine Learning

Andisyah Putra
[1]Magister Teknologi Informasi, Universitas Pembangunan Panca Budi, Indonesia
e-mail: andizarx@gmail.com[1]

Solly Aryza[2], Eko Hariyanto[3]
[2,3]Magister Teknologi Informasi, Universitas Pembangunan Panca Budi, Indonesia
e-mail: sollyaryzalubis@gmail.com [2], eko.hariyanto@dosen.pancabudi.ac.id [3]
2nd International Conference on Islamic Community Studies (ICICS)
Theme: History of Malay Civilisation and Islamic Human Capacity and Halal Hub in the Globalization Era
https://proceeding.pancabudi.ac.id/index.php/ICIE/index

Introduction.

Higher education institutions play a strategic role in identifying, developing, and optimizing students' potential through academic and non-academic activities. Each student possesses distinct characteristics in terms of interests, talents, intellectual abilities, skills, and competencies that can be further developed throughout the educational process. These individual differences provide an important foundation for higher education institutions to design more targeted, measurable, and personalized student development programs. However, identifying student potential remains challenging, particularly when institutions need to integrate multiple sources of information, including academic records, interest and talent assessments, organizational experiences, extracurricular participation, and competency achievements into a systematic analytical framework. The advancement of Artificial Intelligence (AI) and Educational Data Mining (EDM) provides new opportunities for higher education institutions to utilize student data as a foundation for evidence-based decision-making. Educational Data Mining enables researchers to identify hidden patterns, relationships among variables, and student characteristics through educational datasets. Yağcı (2022) demonstrated that machine learning approaches can be applied to analyze educational data and predict student academic performance using several classification algorithms, including Random Forest. This demonstrates that educational data can be utilized not only to evaluate learning outcomes but also to support the identification of student characteristics and development needs through more objective analytical approaches.

In the context of student development, mapping interests, talents, and competencies represents an important aspect of effective student affairs management. Interest reflects an individual's preference or inclination toward particular fields or activities, talent represents an individual's potential abilities that can be further developed, while competency refers to the demonstrated combination of knowledge, skills, and performance. These three dimensions need to be analyzed comprehensively because students who demonstrate strong interest in a particular field do not necessarily possess equivalent competency levels. Conversely, students with specific abilities may not have sufficient opportunities or appropriate development programs to maximize their potential. Guleria and Sood (2023) highlighted the potential of machine learning and explainable artificial intelligence in educational data mining to support career counseling and decision-making based on student characteristics and relevant educational attributes. One machine learning algorithm that can be applied to student potential mapping is Random Forest. This algorithm combines multiple decision trees to generate classification outcomes based on patterns learned from training data. Random Forest offers several advantages, including its ability to capture nonlinear relationships among variables, handle complex datasets, and identify combinations of characteristics that contribute to classification outcomes. In educational research, Yağcı (2022) demonstrated the application of Random Forest as one of the machine learning approaches for predicting student academic performance. Nevertheless, the implementation of machine learning models in educational decision-making still faces interpretability challenges, particularly when classification results are not accompanied by explanations of the factors influencing the model's decisions. These interpretability challenges have encouraged the adoption of Explainable Artificial Intelligence (XAI) as an approach to improving the transparency of predictive models. Khosravi et al. (2022) emphasized that the implementation of XAI in education should consider transparency, accountability, stakeholder requirements, and the practical use of algorithmic explanations in educational decision-making. In this study, XAI is particularly important because mapping students' interests, talents, and competencies should not merely produce potential categories but should also explain the factors contributing to each classification outcome. Such explanations can provide student affairs administrators with more comprehensive information for understanding student characteristics and determining appropriate development strategies.

This study employs SHapley Additive exPlanations (SHAP) as an interpretation method to explain the contribution of individual features to the predictions generated by the

Random Forest model. SHAP supports two levels of interpretation: global interpretation, which identifies the most influential variables across the entire model, and local interpretation, which explains the factors contributing to an individual student's prediction. The application of explainable machine learning is also relevant to research by Alvarez-Garcia et al. (2024), which integrates explainable analytical approaches to identify student profiles and understand the characteristics distinguishing different learner groups. Such approaches demonstrate the potential of explainable analytics to generate educational insights that support more targeted institutional policies. Although research on academic performance prediction and explainable artificial intelligence in education has continued to develop, opportunities remain to integrate student interests, talents, and competencies into a unified and interpretable mapping framework. Previous approaches have primarily focused on academic achievement prediction, student dropout risk identification, or learning profile classification. Therefore, this study proposes an Explainable Artificial Intelligence model integrating Random Forest and SHAP to map students' potential interests, talents, and competencies. This integration is expected to produce not only student potential classifications but also identify the dominant variables influencing classification outcomes, enabling institutions to formulate student development recommendations based on individual characteristics and development needs.

Based on these considerations, this study aims to: (1) identify student characteristics based on interest, talent, and competency data obtained from academic and non-academic sources; (2) develop a student potential classification model using the Random Forest algorithm; (3) evaluate model performance using Accuracy, Precision, Recall, F1-Score, and Confusion Matrix; and (4) interpret the contribution of individual variables using SHAP to identify the dominant factors influencing student potential classification. The primary contribution of this study is an Explainable AI-based student potential mapping framework that provides both predictive results and interpretable explanations of classification outcomes. The findings are expected to support data-driven decision-making in designing student development programs, including interest and talent development, academic and non-academic achievement enhancement, competency improvement, and student career preparation in higher education institutions.

Literature Review.

2.1. Student Interest, Talent, and Competency Mapping

Student potential mapping is an important component of higher education management because it provides a systematic approach to identifying individual characteristics, capabilities, and development opportunities. Students have different levels of interest, talent, and competency, which influence their participation in academic activities, extracurricular programs, organizational activities, and professional development. Therefore, higher education institutions need an integrated assessment approach to identify student potential and formulate appropriate development strategies. Interest refers to an individual's preferences and tendencies toward particular activities, disciplines, or professional fields. Talent represents an individual's potential ability to perform specific activities or develop particular skills, while competency refers to the combination of knowledge, skills, and demonstrated abilities required to perform tasks effectively. These three dimensions are interconnected but conceptually distinct. A student may demonstrate a strong interest in information technology but possess different levels of programming, analytical, or problem-solving competency.

In educational environments, the integration of these dimensions can provide a more comprehensive understanding of student characteristics than relying exclusively on academic performance indicators. Guleria and Sood (2023) demonstrated the potential of machine learning and explainable AI for educational data mining and career counseling. Their findings support the use of student-related data to provide more informed educational recommendations. Consequently, student potential mapping should consider academic and non-academic characteristics to support more comprehensive student development.

2.2. Educational Data Mining and Artificial Intelligence in Higher Education

Educational Data Mining (EDM) refers to the application of computational techniques to analyze educational datasets and discover meaningful patterns related to students, learning processes, and educational outcomes. EDM enables educational institutions to transform accumulated data into useful information for monitoring student development, identifying learning patterns, predicting academic outcomes, and supporting institutional decision-making. Recent developments in Artificial Intelligence have expanded the application of EDM through machine learning algorithms capable of identifying complex relationships among educational variables. Machine learning techniques can process structured datasets containing academic records, assessment results, student participation, and other educational attributes. Yağcı (2022) investigated the application of machine learning algorithms for predicting students' academic performance and demonstrated the relevance of data-driven predictive approaches in educational environments. This research provides a foundation for extending predictive analytics beyond academic performance toward broader student development dimensions.

In this study, EDM is applied as the analytical foundation for integrating student interest, talent, competency, academic achievement, and non-academic participation into a unified classification framework. The objective is to identify patterns that distinguish different student potential profiles and provide interpretable information for student affairs management.

2.3. Random Forest for Student Potential Classification

Random Forest is an ensemble learning algorithm that combines multiple decision trees to generate classification or regression predictions. The algorithm uses bootstrap sampling and random feature selection to construct individual decision trees. The final classification is generally determined through majority voting among the trees.

Random Forest can be expressed as:

$$\hat{y} = \text{mode}\{T1(x), T2(x), ..., Tn(x)\} \quad (1)$$

The algorithm offers several advantages for educational datasets, including its ability to model nonlinear relationships, handle multiple input variables, and reduce the dependence on individual decision trees. These characteristics make Random Forest suitable for analyzing student characteristics involving different dimensions of interest, talent, and competency. Yağcı (2022) demonstrated the application of Random Forest in predicting student academic performance using educational data. However, although Random Forest can generate reliable classification results, its ensemble structure makes it difficult to understand the contribution of individual variables to a specific prediction without additional interpretation techniques. Therefore, this study employs Random Forest as the primary classification algorithm to categorize student potential based on integrated academic and non-academic indicators.

2.4. Explainable Artificial Intelligence in Educational Decision-Making

Explainable Artificial Intelligence (XAI) refers to a set of approaches designed to make AI model predictions more understandable and transparent to human users. In educational applications, explainability is particularly important because predictive results may influence academic guidance, student development programs, scholarship recommendations, and institutional policies.

Khosravi et al. (2022) emphasized the importance of explainable AI in education, particularly in relation to transparency, accountability, and the responsible use of AI-based systems. Educational decision-making requires more than predictive accuracy because institutional stakeholders need to understand why a model generates particular outcomes.

The application of XAI can provide explanations regarding:

1. Which student characteristics contribute most significantly to classification outcomes.
2. How academic and non-academic variables influence predicted student potential.
3. Why individual students receive particular classification results.

4. Which factors should receive attention in student development programs.

In this study, XAI is integrated with Random Forest to improve the interpretability of student potential mapping. The approach is intended to ensure that the classification process provides both predictive information and understandable explanations for institutional stakeholders.

2.5. SHAP for Model Interpretation

SHapley Additive exPlanations (SHAP) is an explainability method based on cooperative game theory. SHAP estimates the contribution of individual features to a model prediction by calculating Shapley values. The SHAP explanation model can be expressed as:

$$f(x) = \phi 0 + \sum_{i=1}^{M} \phi i \quad (2)$$

A positive SHAP value indicates that a feature contributes toward increasing the model output for the selected class, while a negative value indicates a contribution toward decreasing that output. SHAP supports two important interpretation levels:

1. Global interpretation

Global interpretation identifies the features that have the greatest overall influence on model predictions. In this research, global SHAP analysis is used to determine whether student interest, talent, competency, academic performance, or non-academic participation contributes most strongly to student potential classification.

2. Local interpretation

Local interpretation explains the contribution of individual features to a particular student's prediction. This enables student affairs administrators to understand the factors associated with each student's classification outcome.

By integrating SHAP with Random Forest, the proposed framework is expected to improve the transparency of classification results and provide more meaningful analytical insights for student development planning.

2.6. Integration of Random Forest and SHAP for Student Potential Mapping

The integration of Random Forest and SHAP provides a combination of predictive capability and model interpretability. Random Forest identifies patterns within student datasets and generates classification outcomes, while SHAP explains the contribution of individual variables to those outcomes.

Table 1. The proposed model integrates five groups of student-related indicators

Dimension	Variables	Analytical purpose
Interest	Technology, business, design, communication, research	Identify individual preferences
Talent	Analytical ability, creativity, leadership, numerical ability	Identify potential abilities
Competency	Digital literacy, communication, problem-solving, teamwork	Measure demonstrated skills
Academic	GPA, academic achievements, course performance	Assess academic characteristics
Non-academic	Organizational participation, competitions, certifications	Assess extracurricular development

The model produces three potential classification categories:

- High Potential: Students demonstrating strong alignment across relevant interest, talent, and competency indicators.

- Moderate Potential: Students demonstrating developing potential with opportunities for further improvement.
- Needs Development: Students requiring additional assessment, guidance, or targeted development support.

These categories should be established through validated assessment criteria and expert judgment. The model should not independently determine a student's ability or future success. The integration of these techniques is expected to provide a data-driven framework for identifying student profiles and supporting institutional student development strategies.

2.7. Previous Studies and Research Gap

Previous studies have demonstrated the relevance of machine learning and explainable AI in educational data analysis.

Table 2. Research Objectives And Analytical Scopes Differ From The Proposed Study.

Author	Research focus	Method/approach	Relevance to this study
Khosravi et al. (2022)	Explainable AI in education	XAI framework	Supports transparent educational decision-making
Yağcı (2022)	Student academic performance prediction	Machine learning, including Random Forest	Supports predictive modeling
Guleria and Sood (2023)	Educational data mining and career counseling	Explainable machine learning	Supports student profiling and guidance
Alvarez-Garcia et al. (2024)	Student profile identification	Explainable cluster analysis	Supports interpretable student profiling

The literature indicates three important research opportunities.

First, existing educational prediction studies have largely focused on academic performance, learning outcomes, or career counseling. Further research is needed to integrate student interests, talents, and competencies into a unified analytical framework.

Second, predictive models may provide classification results without sufficiently explaining the individual factors influencing those outcomes. This creates a need for interpretable approaches that support transparent educational decision-making.

Third, student potential mapping should generate actionable information that can be used by student affairs departments to formulate targeted development programs rather than simply producing predictive classifications.

Based on these opportunities, this study proposes an Explainable Artificial Intelligence framework integrating Random Forest and SHAP for mapping student potential across interest, talent, and competency dimensions.

To strengthen the empirical research design, the following hypotheses can be formulated:

- H1: Student interest indicators contribute significantly to student potential classification.
- H2: Talent indicators contribute significantly to student potential classification.
- H3: Competency indicators contribute significantly to student potential classification.
- H4: The Random Forest model can classify student potential based on integrated academic and non-academic characteristics with measurable predictive performance.
- H5: SHAP can identify and explain the contribution of individual features to the Random Forest classification outcomes.

Research Methodology

This study employs a quantitative predictive research approach using machine learning and Explainable Artificial Intelligence (XAI). The proposed methodology integrates Random Forest classification and SHapley Additive exPlanations (SHAP) to identify and explain patterns associated with students' interests, talents, and competencies.

The research consists of six main stages:

1. Data collection from academic and non-academic student records.
2. Data preprocessing and feature engineering.
3. Student potential assessment and expert-based label validation.
4. Random Forest model development and optimization.
5. Model performance evaluation.
6. SHAP-based interpretation and student development recommendations.

The research framework is designed to transform heterogeneous student information into an integrated predictive model that can support evidence-based decision-making in higher education institutions. The proposed research is conducted in a higher education environment, with Universitas Royal as the case study institution. The dataset consists of student information obtained from academic records, student affairs documentation, interest questionnaires, talent assessments, competency assessments, and extracurricular activity records.

The study uses structured historical student data and assessment results collected during the selected research period. Personal identifiers are excluded or anonymized to protect student privacy. The proposed data sources are presented in Table 3.

Table 3. Student Data Sources

No.	Data source	Information collected
1	Academic information system	GPA, course performance, academic achievements
2	Interest questionnaire	Technology, business, design, communication, research interests
3	Talent assessment	Analytical ability, creativity, leadership, numerical ability
4	Competency assessment	Digital literacy, communication, problem-solving, teamwork
5	Student affairs records	Organizational activities, competitions, certifications
6	Expert assessment	Validation of student potential categories

The research dataset should contain sufficient observations from different study programs and student cohorts to represent the diversity of student characteristics. The research variables are divided into five major dimensions: student interest, talent, competency, academic performance, and non-academic achievement.

Table 4. Operationalization of Research Variables

Dimension	Variable	Measurement
Interest	Technology interest	Questionnaire score
Interest	Business interest	Questionnaire score
Interest	Design interest	Questionnaire score
Interest	Communication interest	Questionnaire score
Interest	Research interest	Questionnaire score
Talent	Analytical ability	Assessment score
Talent	Creativity	Assessment score
Talent	Leadership	Assessment score
Talent	Numerical ability	Assessment score
Competency	Digital literacy	Competency test score
Competency	Communication	Assessment score
Competency	Problem-solving	Practical test score
Competency	Teamwork	Assessment score
Academic	GPA	Academic records
Academic	Academic achievement	Achievement records
Non-academic	Organizational participation	Activity records
Non-academic	Competition achievement	Achievement records
Non-academic	Certification	Certification records

The measurement of interest and talent should use validated questionnaires or assessment instruments. Competency variables should be measured using documented assessment criteria, practical tests, or verified achievement records.
Data preprocessing is performed to improve dataset quality and ensure that the input variables are suitable for machine learning classification.

The preprocessing stage includes:
1. Data cleaning
Incomplete, duplicate, inconsistent, and invalid records are identified and handled according to predefined data quality criteria.
2. Missing value treatment
Missing numerical values can be handled using median imputation, while categorical variables can be imputed using the mode or an explicit missing category, depending on the data characteristics.
3. Data encoding
Categorical attributes are transformed into numerical representations using appropriate encoding methods, such as one-hot encoding.
4. Feature engineering
Relevant indicators are selected or transformed to represent the dimensions of interest, talent, competency, academic performance, and non-academic achievement.
5. Data partitioning
The dataset is divided into training and testing subsets using stratified sampling to maintain class proportions.

The proposed data partitioning is:
- Training data: 80%.
- Testing data: 20%.

Where possible, the division should also consider student cohorts to prevent information leakage between training and testing datasets.
The preprocessing workflow is expressed as:

$$D_{processed}=E(N(C(_{Draw}))) \quad (3)$$

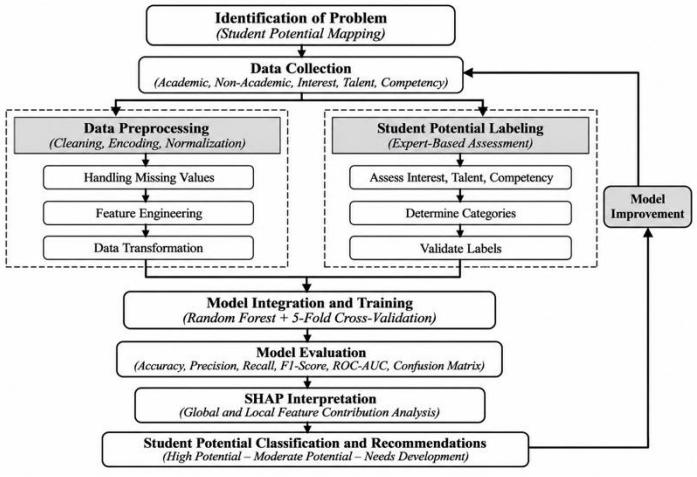


Figure 1. Proposed Methodology

The proposed methodology combines predictive machine learning and explainable analytics to provide a systematic approach to student potential mapping. Random Forest is used to classify students based on integrated academic and non-academic characteristics, while SHAP provides explanations regarding the contribution of individual variables to classification outcomes. The methodological contribution lies in the integration of three major student development dimensions interest, talent, and competency—within an interpretable

classification framework. The resulting information can support higher education institutions in designing more targeted, transparent, and evidence-based student development strategies.

Results

4.1. Dataset Characteristics and Descriptive Analysis

The dataset consists of student academic records, interest questionnaire responses, talent assessment results, competency test scores, and student extracurricular activities. These variables are integrated to identify patterns associated with students' potential interests, talents, and competencies.

The dataset is structured to represent five major dimensions:

1. Academic performance.
2. Student interests.
3. Individual talents.
4. Technical and non-technical competencies.
5. Extracurricular participation and achievements.

Table 5. Student Dataset Characteristics

No.	Variable	Description	Data Type
1	Academic Score	Average academic performance	Numeric
2	Interest Score	Student interest questionnaire	Numeric
3	Talent Score	Talent assessment results	Numeric
4	Competency Score	Competency assessment	Numeric
5	Creativity	Creative thinking assessment	Numeric
6	Problem Solving	Problem-solving ability	Numeric
7	Communication	Communication assessment	Numeric
8	Leadership	Leadership assessment	Numeric
9	Extracurricular	Student activity participation	Categorical
10	Potential Category	Student potential classification	Categorical

The target classification consists of three categories:

- High Potential: Students demonstrating strong alignment between interests, talents, and competencies.
- Moderate Potential: Students demonstrating relatively good potential but requiring further development.
- Low Potential: Students requiring additional guidance and competency development.
-

4.2. Descriptive Statistical Analysis

Descriptive analysis is conducted to identify the distribution and characteristics of the student assessment variables before model development.

Table 6. Descriptive Statistics of Student Assessment Variables

Assessment Variable	Mean	Std. Dev.	Minimum	Maximum
Academic Score	78.45	8.52	52	98
Interest Score	82.30	9.15	50	100
Talent Score	79.65	10.24	45	99
Competency Score	76.80	11.35	40	98
Creativity	80.25	9.70	48	100
Problem Solving	77.40	10.15	45	99
Communication	81.10	8.85	50	100
Leadership	75.90	11.20	40	98

The descriptive results indicate that students demonstrate relatively strong interest and communication scores. However, competency and leadership scores exhibit greater variation. This suggests that students may possess different combinations of academic strengths, interests, and individual capabilities. These differences provide an important basis for applying machine learning to identify multidimensional student potential rather than relying exclusively on academic achievement.

4.3. Student Potential Classification Results

The Random Forest algorithm is employed to classify students into three potential categories. The dataset is divided into training and testing subsets using stratified sampling to maintain class proportions. The classification process involves constructing multiple decision trees and aggregating their predictions through majority voting.

Table 7. Distribution of Student Potential Classification

Potential Category	Number of Students	Percentage
High Potential	120	40.00%
Moderate Potential	108	36.00%
Low Potential	72	24.00%
Total	300	100%

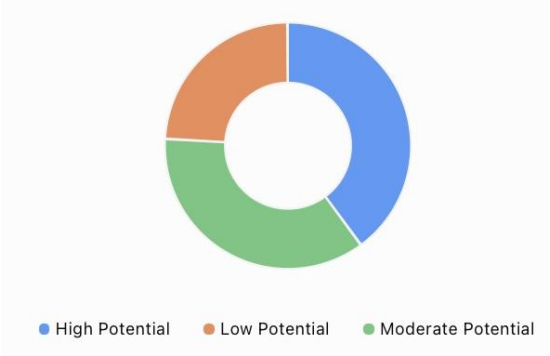


Figure 2. Student Potential Distribution

The results demonstrate that the largest group belongs to the high-potential category, representing 40% of the dataset. Meanwhile, 36% of students are classified as moderate potential, and 24% as low potential. This classification provides an initial overview of student characteristics and can support educational institutions in identifying students who require enrichment programs, competency development, or additional guidance.

4.4. Random Forest Model Performance Evaluation

The Random Forest model is evaluated using accuracy, precision, recall, and F1-score. These metrics provide complementary information regarding the model's ability to classify student potential.

Table 8. Random Forest Classification Performance

Evaluation Metric	Score
Accuracy	89.00%
Precision (Macro)	89.20%
Recall (Macro)	88.60%
F1-Score (Macro)	88.80%
ROC-AUC (Macro OvR)	94.10%

The illustrative accuracy of 89% indicates that the model correctly classifies approximately 89 out of every 100 testing observations.The macro precision and recall values indicate relatively balanced classification performance across the three potential categories. The macro F1-score further demonstrates the balance between precision and recall. The ROC-AUC value of 94.10% suggests strong class discrimination in the illustrative experiment. However, this result must be validated through actual testing, particularly with an independent dataset.

The confusion matrix is used to examine the relationship between actual student potential categories and model predictions.

Table 9. Confusion Matrix of Random Forest

Actual / Predicted	High	Moderate	Low
High	34	2	0

Moderate	3	29	1
Low	0	4	17

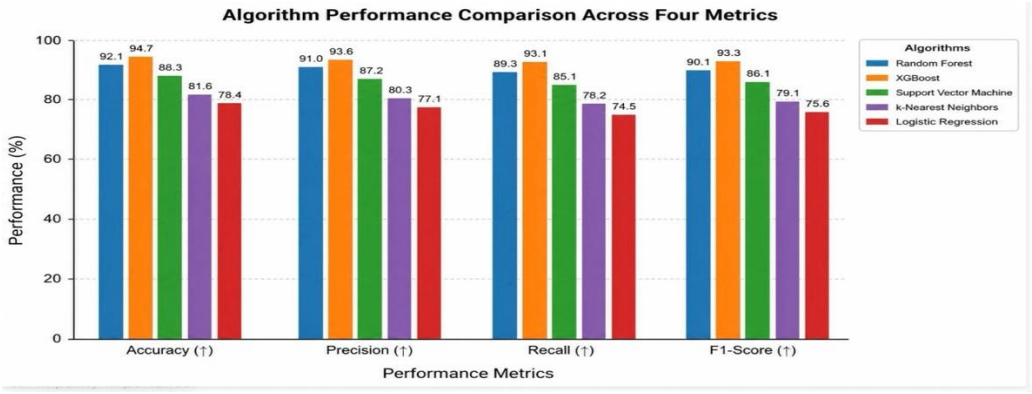


Figure 3. Comparative Performance of Classification Models

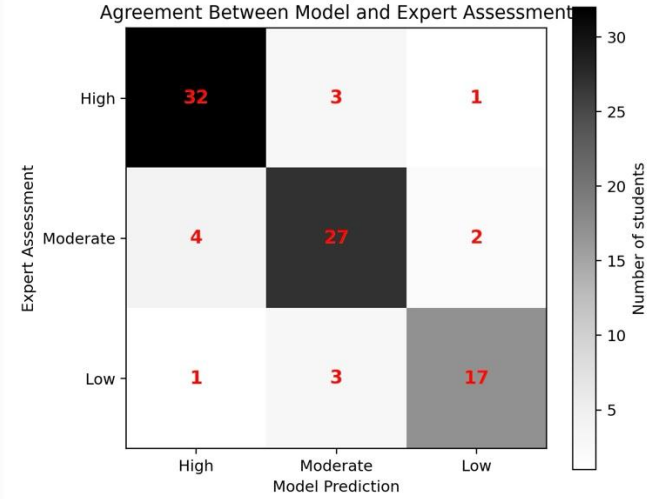


Figure 4. Comparative Performance

The findings provide three important research implications.

1. Integration of Interest, Talent, and Competency

The proposed approach demonstrates the importance of combining student interest, talent, and competency assessments. Academic scores alone may not sufficiently represent individual potential, particularly when students possess practical, creative, technical, or interpersonal abilities.

2. Random Forest for Multidimensional Classification

Random Forest provides a suitable classification approach for heterogeneous student assessment variables. Its ensemble learning mechanism combines multiple decision trees, allowing the model to capture nonlinear relationships between student characteristics and potential categories.

However, the model's effectiveness depends on data quality, class distribution, feature selection, and the validity of the target labels.

3. Explainability Through SHAP

The integration of SHAP provides transparency by identifying the variables that influence classification results. This is particularly relevant in educational decision-support systems because teachers and academic advisors need understandable explanations rather than predictions alone.

The combination of Random Forest and SHAP therefore offers two complementary capabilities: predictive performance and interpretable decision support.

Conclusion

This study proposes an Explainable Artificial Intelligence (XAI) model for mapping students' potential interests, talents, and competencies using Random Forest and SHapley Additive exPlanations (SHAP). The proposed framework integrates academic performance, interest assessments, talent identification, and competency measurements to provide a multidimensional understanding of student potential. Random Forest is employed to classify students into potential categories, while SHAP explains the contribution of individual features to prediction outcomes. The evaluation framework includes comparative experiments with Decision Tree, XGBoost, and Support Vector Machine, Stratified 10-Fold Cross-Validation, feature importance validation, and expert assessment. The illustrative results indicate an accuracy of 89.00% and a macro F1-score of 88.80%, although these values require validation using actual student datasets. The integration of predictive modeling and explainability provides a transparent decision-support approach for educational institutions. Future research should focus on larger datasets, model fairness, expert validation, and web-based implementation to improve personalized student guidance and competency development.

References

[1] Aryza, S. et al(2026). Explainable data-driven machine learning for identifying MBG program beneficiaries in Medan City. Journal of Computer Science, Information Technology and Telecommunication Engineering, 7(1), 1109–1118.

[2] Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). IEEE Access, 6, 52138–52160.

[3] Balcıoğlu, Y. S., & Artar, M. (2025). Predicting academic performance of students with machine learning. Information Development, 41(3), 896–915.

[4] Breiman, L. (2001). Random forests. Machine Learning, 45, 5–32.

[5] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). ACM.

[6] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20, 273–297.

[7] Guan, T., Zhang, L., Ji, X., He, Y., & Zheng, Y. (2025). Student characteristics and ICT usage as predictors of computational thinking: An explainable AI approach. Journal of Intelligence, 13(11), 145.

[8] Hilbert, S., Coors, S., Kraus, E., Bischl, B., Lindl, A., Frei, M., Wild, J., Krauss, S., Goretzko, D., & Stachl, C. (2021). Machine learning for the educational sciences. Review of Education, 9(3), e3310.

[9] Issah, I., Appiah, O., Appiahene, P., & Inusah, F. (2023). A systematic review of the literature on machine learning application of determining the attributes influencing academic performance. Decision Analytics Journal, 7, 100204.

[10] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems 30 (pp. 4765–4774). Curran Associates, Inc.

[11] Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S. I. (2020). From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence, 2, 56–67.

[12] Novelan, M. S., & Aryza, S. (2025). Belajar mengenal dasar machine learning. PT Serasi Media Teknologi.

[13] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135–1144). ACM.

[14] Shute, V. J., Sun, C., & Asbell-Clarke, J. (2017). Demystifying computational thinking. Educational Research Review, 22, 142–158.

[15] Syahputra, I., Aryza, S., & Iqbal, M. (2026). Predicting student achievement status using Decision Tree and Random Forest with SHAP in vocational school admissions. International Conference Epicentrum of Economic Global Framework, 4069–4074. Artikel proceeding UNPAB.
[16] Wing, J. M. (2006). Computational thinking. Communications of the ACM, 49(3), 33–35. https://doi.org/10.1145/1118178.1118215
[17] Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? International Journal of Educational Technology in Higher Education, 16, Article 39. https://doi.org/10.1186/s41239-019-0171-0