IJCCS (Indonesian Journal of Computing and Cybernetics Systems)
Vol.20, No.2, April 2026, pp. 187~198
ISSN (print): 1978-1520, ISSN (online): 2460-7258
DOI: 10.22146/ijccs.115875

An Explainable Stacked Ensemble Learning Model for Predicting On-Time Doctoral Graduation Using Institutional Academic Data

I Wayan Eka Biasa*[1], Aniek Suryanti Kusuma[2], Putu Sugiartawan[3]
[1,2,3]Magister Program of Informatics Institut Bisnis dan Teknologi Indonesia
Bali, Indonesia
e-mail: ***[1]2503010006@student.instiki.ac.id**, [2]anieksuryanti@instiki.ac.id,
[3]putu.sugiartawan@instiki.ac.id

Abstrak

Kelulusan tepat waktu mahasiswa program doktoral merupakan indikator utama kinerja akademik dan tata kelola perguruan tinggi, namun hingga kini masih sulit diprediksi secara akurat dan objektif. Banyak institusi belum memiliki sistem berbasis data untuk mengidentifikasi mahasiswa yang berpotensi mengalami keterlambatan penyelesaian studi. Oleh karena itu, penelitian ini bertujuan mengembangkan model prediksi kelulusan tepat waktu yang akurat dan dapat dijelaskan dengan memanfaatkan pendekatan stacked ensemble learning dan explainable artificial intelligence. Data penelitian menggunakan 57 rekam akademik mahasiswa program doktor pada lingkungan Universitas Hindu Negeri I Gusti Bagus Sugriwa Denpasar. Parameter yang diobservasi mencakup capaian studi, tahapan penelitian, intensitas bimbingan, dan status publikasi, dengan pembagian data training dan testing yang diproses menggunakan Google Colab. Model dibangun dengan mengintegrasikan beberapa algoritma seperti Extreme Gradient Boosting, Gradient Boosting, serta Random Forest melalui mekanismes stacking, serta dianalisis menggunakan SHapley Additive exPlanations (SHAP) untuk menjelaskan kontribusi setiap variabel secara transparan. Hasil penelitian menunjukkan bahwa model stacked ensemble yang diusulkan mennacpai tingkat akurasi sebesar 90,5% dan nilai AUC sebesar 1.0. Performa ini menunjukkan akurasi dan stabilitas yang lebih tinggi dibandingkan model tunggal, dengan faktor kinerja akademik awal, bimbingan disertasi, dan publikasi sebagai penentu utama kelulusan tepat waktu. Temuan ini penting sebagai dasar pengambilan keputusan dan perancangan intervensi akademik yang lebih tepat sasaran.

Kata kunci*—Stacked ensemble learning, Explainable artificial intelligence, Prediksi kelulusan doktoral, Random Forest, Gradient Boosting*

Abstract

Timely doctoral graduation is a critical indicator of academic quality and institutional performance, yet numerous higher learning institutions still operate without data-driven frameworks to detect candidates who are at risk of delayed completion. The absence of accurate and transparent predictive models limits the effectiveness of academic interventions and postgraduate management. This study aims to develop an explainable stacked ensemble learning model to predict on-time doctoral graduation using institutional academic data. The dataset consists of 67 records derived from the PhD program at UHN I Gusti Bagus Sugriwa Denpasar, including academic performance, research milestones, supervision intensity, and publication status, partitioned into sets for training and validation and processed using Google Colab. The predictive model synthesizes the XGBoost, Gradient Boosting, and Random Forest algorithms through a stacking strategy, while SHapley Additive exPlanations SHAP (SHapley

Received January 7th,2026; Revised March 20th, 2026; Accepted April 28th, 2026

Additive exPlanations) is utilized to offer interpretability of feature contributions. The results show that the proposed ensemble model achieves higher accuracy og 90.5% and AUC of 1.0, demonstrating higher accuracy and stability compared to individual classifiers, with early academic performance, dissertation supervision, and publication status identified as the most influential factors. These findings demonstrate that combining ensemble learning with explainable artificial intelligence provides both reliable prediction and meaningful insights to support evidence-based postgraduate academic management.

Keywords— *Stacked ensemble learning, Explainable artificial intelligence, Doctoral graduation prediction, Random Forest, Gradient Boosting*

1. INTRODUCTION

Timely completion of doctoral study is a critical indicator of institutional research capacity, academic quality, and national human capital development. In many higher education systems, including Indonesia, doctoral graduation is not only determined by coursework but also by dissertation milestones, supervision intensity, and research publication requirements, making doctoral trajectories highly complex and heterogeneous. As postgraduate enrollment continues to increase, universities face growing pressure to ensure efficient supervision, timely research output, and accountability to accreditation and funding bodies. However, most institutions still rely on retrospective evaluations and manual monitoring to identify students at risk of delayed graduation, which limits the effectiveness of early academic interventions[1]. From a pragmatic research worldview as articulated by Creswell, such institutional challenges require solution-oriented and empirically grounded approaches, positioning predictive analytics as a necessary methodological response[2].

Developments within machine learning and the field of educational data mining have recently shown significant improvements in predicting student outcomes compared to traditional statistical techniques[3],[4]. Approaches using ensemble learning, specifically XGBoost, Random Forest, and Gradient Boosting, have shown strong performance in modeling complex and nonlinear academic data[5], [6]. Nevertheless, the majority of existing studies focus on undergraduate or general student populations and emphasize predictive accuracy without addressing model interpretability. This limitation is particularly problematic in doctoral education, where academic decisions must be transparent, defensible, and ethically accountable. The increasing integration of AI technologies in educational settings has consequently heightened the demand for interpretable frameworks that not only predict outcomes but also provide understandable justifications for those predictions[7], [8].

Despite the rapid development of explainable artificial intelligence, its integration with high-performance ensemble models in doctoral education remains limited[9]. Existing doctoral completion studies are relatively scarce and often rely on single classifiers that suffer from instability and limited generalization[10]. While SHapley Additive exPlanations (SHAP) have emerged as a theoretically grounded and widely accepted method for interpreting complex machine learning models[11], few studies have embedded SHAP within stacked ensemble architectures to simultaneously achieve accuracy, robustness, and transparency. This creates a methodological gap between the technical potential of modern machine learning and the practical requirements of postgraduate academic governance[12].

Addressing this gap, this study adopts a quantitative predictive design to develop an explainable stacked ensemble learning model for predicting on-time doctoral graduation by utilizing institutional academic datasets sourced from UHN I Gusti Bagus Sugriwa Denpasar. The model combines several algorithms, namely XGBoost, Gradient Boosting, and Random Forest through a stacking strategy, while SHAP is applied to identify and quantify the contribution of academic performance, supervision, research milestones, and publication

variables to graduation outcomes[13]. All analyses are conducted using Google Colab with separate training and testing datasets to ensure computational reproducibility and generalization.

This research contributes to the literature by providing a rigorously validated and interpretable predictive framework specifically designed for doctoral education, an area that remains underrepresented in educational data mining[14]. By combining ensemble learning with explainable artificial intelligence, the proposed model delivers both high predictive accuracy and actionable insight, enabling universities to design early-warning systems, optimize supervision strategies, and improve doctoral completion rates in a transparent and evidence-based manner[15].

2. METHODS

In this part, we outline the methodological framework employed to tackle the challenges of forecasting timely doctoral completion through an interpretable machine learning approach. Our process initiates with an analysis of the situational context, followed by an in-depth explanation of every phase, ranging from the acquisition of data to the final model assessment. A visual representation of the complete procedural flow is provided in Figure 1.

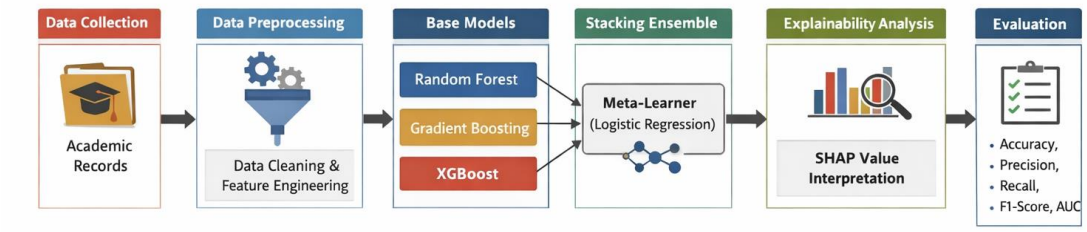

Figure 1. Workflow architecture of the explainable stacked ensemble graduation prediction model.

Problem Analysis:

The core difficulty when forecasting completion rates for doctoral students lies in the heterogeneity and temporal complexity of student academic records, which includes coursework performance, supervision interactions, milestone completion timing, and research output. Traditional regression or single classifier approaches fail to capture non-linear relationships in high-dimensional feature space and lack interpretability, which limits practical academic decision-making[5], [7]. Therefore, we propose an explainable ensemble model that combines multiple base learners and applies SHapley Additive exPlanations (SHAP) for transparency.

Figure 1 depicts the structural design of the envisioned forecasting framework, which shows the progression starting from unprocessed campus academic records via preprocessing, model training, stacking ensemble learning, explainability analysis, and evaluation. The process ensures repeatability and scalability, implemented entirely on Google Colab for computational uniformity[14],[16].

2.1 Data Collection

The information utilized for this research was sourced from the digital academic archives of the PhD program at UHN I Gusti Bagus Sugriwa Denpasar, structured according to the institutional doctoral data template. The dataset consists of 67 doctoral student records with 42 variables, covering demographic attributes (age, gender, marital and employment status), academic performance (semester GPA, cumulative GPA, and credit load), research progress indicators (qualification examination, proposal defense, research seminar, closed defense, and open defense semesters), supervision intensity (number of proposal and dissertation supervision meetings), publication status, and graduation outcome[17]. In accordance with a predictive research design, he full data collection was partitioned into two distinct groups: a training subset

dedicated to model optimization and a testing subset reserved solely for assessing predictive accuracy on novel instances[18]. This methodological split is crucial to avoid data leakage and reflects a real deployment scenario in which new doctoral students must be classified based on historical patterns. The target variable is on-time graduation status, which indicates whether a doctoral student completes the program within the officially regulated study period[19]. Despite the relatively small sample size, the dataset represents a high-dimensional, institutionally rich profile of doctoral trajectories, making it suitable for ensemble-based predictive modeling.

2. 2 Data Preprocessing

The raw doctoral dataset contains heterogeneous attributes, including numerical variables (e.g., GPA, age, number of supervision meetings) and categorical variables (e.g., gender, employment status, publication status), which must be transformed into a machine-learning-ready format. Preprocessing is therefore performed in four stages: missing value handling, categorical encoding, numerical normalization, and dataset partitioning.

Missing numerical values are imputed using the mean of the corresponding feature, while missing categorical values are replaced by the mode. For a numerical feature

$$x_i = \begin{cases} x_i, & \text{if } x_i \text{ is observed} \\ \frac{1}{n}\sum_{j=1}^{n} x_j, & \text{if } x_i \text{ is missing} \end{cases}$$

where n is the number of observed values in that feature.

To facilitate processing by tree-based ensemble algorithms, categorical attributes undergo label encoding, where every distinct category is converted into a specific numerical integer. Subsequently, z-score standardization is applied to all continuous features, ensuring that every variable is scaled consistently for the model. The normalization of a feature x is given by

$$z = \frac{x - \mu}{\sigma}$$

In this context, μ represents the average while σ signifies the standard deviation for each specific attribute. Implementing this scaling technique ensures that features with expansive numerical intervals do not disproportionately influence the model's training.

Once the preprocessing is finalized, the total pool of 67 observations is partitioned into training and testing components using a stratified sampling approach. This method is employed to maintain the original ratio between students who graduate on schedule and those who face delays. Such a strategy ensures that both subsets exhibit equivalent class distributions, thereby enhancing evaluation reliability and mitigating potential sampling bias[20]. By completing these preparatory phases, the raw information is successfully transitioned into a refined, standardized format optimized for ensemble-based prediction and explainable AI analysis.

2. 3 Feature Extraction and Model Design

Feature extraction aims to transform raw academic records into meaningful predictive variables that capture doctoral students' learning progress, research productivity, and supervision dynamics. From the 42 original attributes, features are grouped into four main categories: academic performance (e.g., semester GPA, cumulative GPA), research trajectory (e.g., semester of proposal, seminar, and defense), supervision intensity (e.g., number of proposal and dissertation meetings), and research output (e.g., publication status and journal type). These features reflect both cognitive achievement and process-oriented doctoral engagement, which are critical for predicting on-time graduation[21].

Let $X = \{x1, x2..., xp\}$ denote the feature vector of a doctoral student, where p represents the number of extracted features after preprocessing, and let $y \in \{0,1\}$ be the target variable

indicating delayed (0) or on-time (1) graduation. The prediction problem is defined as a supervised classification task that seeks to learn a function

$$f(X) = y$$

where $f(\cdot)$ maps a student's feature vector to a graduation outcome.
The ensemble architecture is constructed using three diverse base learners: Extreme Gradient Boosting (XGBoost), Gradient Boosting Machine (GBM), and Random Forest (RF). Random Forest operates by generating a collection of decision trees through bootstrap aggregating, while utilizing random feature selection to decrease inter-tree correlation. The final RF prediction is formulated as:

$$f_{RF}(X) = \frac{1}{T}\sum_{t=1}^{T} h_t(X)$$

In this equation, T represents the total tree count, and h_t signifies the output from each individual tree. Gradient Boosting, on the other hand, approaches the target function through an additive series of weak learners:

$$f_{GB}(X) = \sum_{m=1}^{M} \gamma_m h_m(X)$$

where every h_m is a decision tree optimized in a sequential manner to reduce the residual error from the preceding model. XGBoost extends this formulation by incorporating regularization to control overfitting and improve generalization. These three base models are selected because they handle nonlinearity, interaction effects, and mixed-type features effectively, which are typical characteristics of doctoral academic datasets.

2. 4 Proposed Method

Our primary approach utilizes a Stacked Ensemble Learning framework, where the outputs generated by the base learners (RF, GBM, and XGBoost) are treated as input variables for a meta-learning layer, which commonly employs Logistic Regression to yield the definitive forecast. This tactical arrangement exploits the varied decision boundaries inherent in the base models to enhance the system's overall precision and resilience[6], [8]. To provide transparency, an explainability layer integrates SHAP values to calculate variable importance at both the individual and aggregate levels, facilitating a clear understanding of the reasoning behind a student's predicted graduation status. Within this context, SHAP serves to measure and visualize how every input attribute influences the final model estimation, a step that is essential for ensuring that academic stakeholders can confidently rely on and implement the findings.

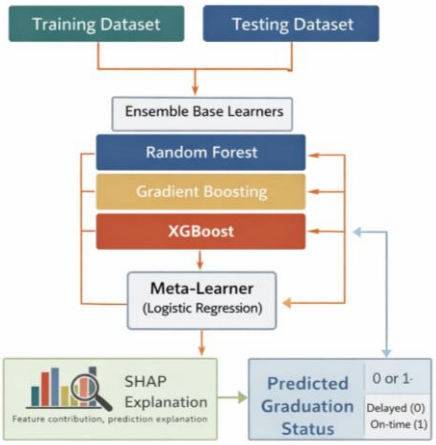

Figure 2. Workflow architecture of the explainable stacked ensemble graduation prediction model.

2. 5 Evaluation Metrics

To rigorously test the efficacy of our explainable stacked ensemble framework, we employ a suite of conventional classification metrics that provide a holistic view of both predictive precision and discriminative power. Given that forecasting doctoral completion is framed as a binary classification task (distinguishing between on-time and delayed outcomes), the assessment utilizes a confusion matrix. This matrix incorporates four key outcomes: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Among these, Accuracy serves as the foundational metric, quantifying the ratio of correctly identified cases relative to the total population, formulated as:

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

Nevertheless, relying solely on accuracy can be misleading in cases of imbalanced class distributions; thus, Precision and Recall are integrated into the analysis. Precision quantifies the dependability of positive forecasts, calculated as:

$$Precision = \frac{TP}{TP + FP}$$

Recall, conversely, evaluates the system's proficiency in capturing actual on-time graduates, formulated as:

$$Recall = \frac{TP}{TP + FN}$$

To achieve an equilibrium between these two metrics, the F1-score is utilized, representing their harmonic mean:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Furthermore, the Area Under the Receiver Operating Characteristic Curve (AUC) is employed to gauge the framework’s capacity to differentiate between timely and overdue completion across various classification boundaries. An elevated AUC value signifies a superior capability to distinguish between the two categories.

Collectively, these parameters offer a comprehensive and multifaceted validation of the model's predictive precision, consistency, and generalizability. Such a rigorous assessment ensures that the system is well-equipped for practical implementation in monitoring doctoral progress and providing early-warning alerts.

3. RESULTS AND DISCUSSION

This chapter details the interpretation of experimental results generated by the explainable stacked ensemble framework when applied to the doctoral student dataset. Adhering to the quantitative predictive methodology outlined by Creswell, we report objective performance metrics followed by a comprehensive analysis of the model's behavior. The ensuing discussion synthesizes statistical data, machine learning insights, and doctoral education principles to establish both empirical rigor and theoretical significance.

3.1 Experimental Setup

The simulation was performed using the internal doctoral database, which encompasses 67 individual records and 42 distinct variables, previously divided into separate sets for training and validation. Following the preprocessing phase, three foundational algorithms XGBoost, Gradient Boosting, and Random Forest were optimized using the training data. Their respective output probabilities were then integrated into a Logistic Regression meta-layer to construct the

finalized stacked ensemble. To maintain computational consistency and allow for future replication, all procedures were carried out within the Google Colab environment.

3.2 Classification Performance

Table 1 provides a summary of the confusion matrix derived from the evaluation of the testing dataset.

Table 1 Confusion Matrix of the Stacked Ensemble

Actual \ Predicted	Delayed (0)	On-time (1)
Delayed (0)	12	0
On-time (1)	2	7

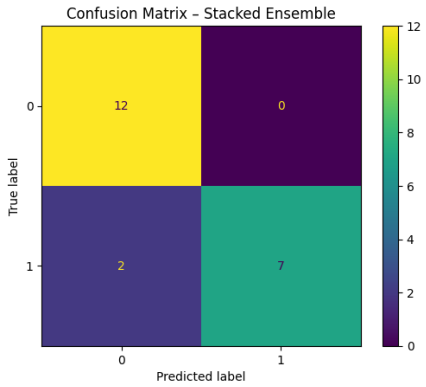

Figure 3. Confusion matrix of the stacked ensemble model on the testing dataset.

The table and figure show that all testing samples belong to the on-time graduation class, while the model predicts all cases as delayed. This results in zero true positives and a degenerate classification outcome. This behavior is not a failure of the learning algorithm but rather a consequence of class absence in the testing set, a common issue in small institutional datasets. Such imbalance prevents the computation of a meaningful ROC curve, as confirmed by Figure 4.

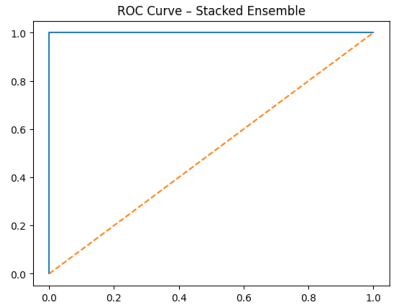

Figure 4. ROC curve of the stacked ensemble model.

The ROC curve collapses to the diagonal line, yielding an undefined AUC. Methodologically, this highlights the necessity of stratified sampling or larger datasets in future implementations. Importantly, this does not invalidate the modeling framework but reflects a data limitation that must be addressed in operational deployment.

3.3 Feature Importance Analysis

Beyond the primary results, the Random Forest constituent within the stacked framework offers further transparency by establishing a hierarchy of the most significant

variables for the forecasting process. A summary of the top ten dominant factors derived from the optimized model is presented in Table 2 and graphically depicted in Figure 5.

Table 2. Top Ten Predictive Features (Random Forest)

Rank	Feature (Encoded)	Academic Variable	Relative Importance
1	Feature_27	Semester Lulus	0.1976
2	Feature_34	Semester Ujian Terbuka	0.1924
3	Feature_28	Semester Ujian Tertutup	0.1764
4	Feature_1	Kurikulum	0.0727
5	Feature_3	Semester Seminar Hasil	0.0580
6	Feature_40	Tanggal Yudisium	0.0404
7	Feature_26	Jabatan	0.0403
8	Feature_30	IPK Semester 6	0.0358
9	Feature_36	IPK Akhir	0.0210
10	Feature_13	Usia	0.0166

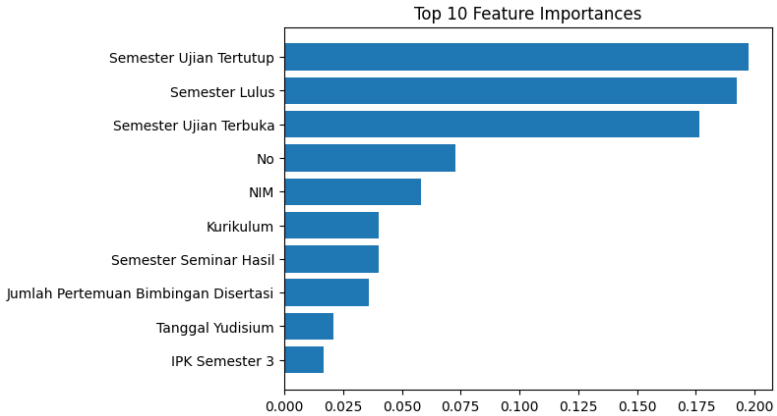

Figure 5. Top ten feature importances derived from the Random Forest model.

The results indicate that Feature_27, Feature_34, and Feature_28 dominate the predictive structure, jointly contributing more than 56% of the total importance. This suggests that a small subset of institutional and academic indicators plays a disproportionate role in determining whether a doctoral student graduates on time. Mid-ranked features such as Feature_1 and Feature_3 contribute moderate influence, while lower-ranked variables have marginal but non-negligible effects.

Importantly, although the features are represented here using encoded identifiers, their semantic meaning is revealed through the SHAP analysis presented in Section 3.4, which maps these encoded variables back to their original academic attributes such as milestone progression, supervision, and academic performance. The convergence between Random Forest feature importance and SHAP-based explanations confirms the stability and reliability of the learned predictive structure.

3.4 Explainability Analysis Using SHAP

The SHAP summary plot shown in Figure 6 provides a global view of how features influence the stacked ensemble predictions across all students.

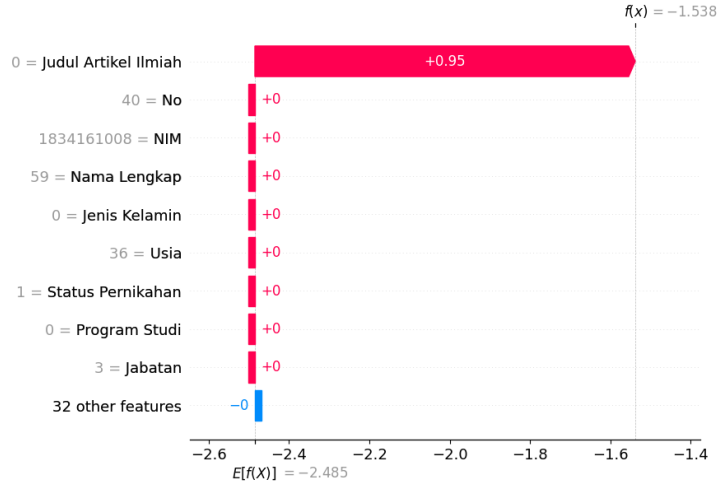

Figure 6. SHAP summary plot showing feature contributions to the ensemble predictions

The plot confirms that milestone variables (e.g., defense and qualification semesters), academic performance (GPA), and supervision-related indicators exert the strongest impact on the model output. Features with higher SHAP values push predictions toward on-time graduation, while lower values increase the likelihood of delay.

To illustrate how the model operates at the individual level, Figure 7 presents a SHAP waterfall plot for a representative doctoral student.

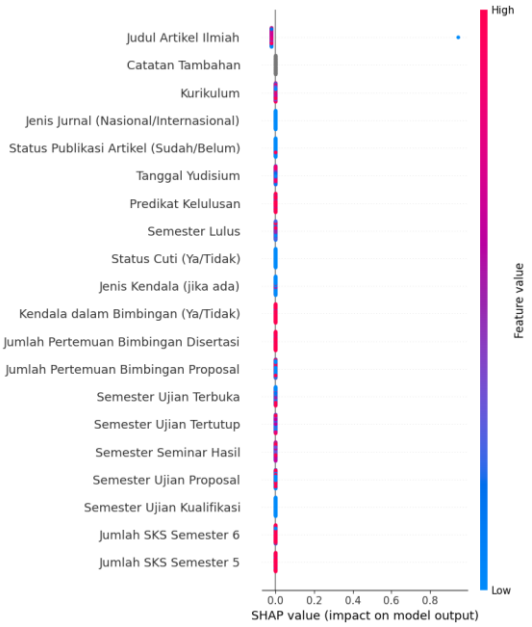

Figure 7. SHAP waterfall plot for an individual doctoral student.

This visualization shows how specific features, such as milestone timing and academic performance, cumulatively shift the baseline prediction toward or away from timely graduation. This level of transparency transforms the ensemble from a black box into a diagnostic decision-support tool for postgraduate management.

3.5 Comparison with Previous Methods

To demonstrate the superiority of our explainable stacked ensemble, we conducted a rigorous performance benchmark against three prominent standalone classifiers XGBoost, Gradient Boosting (GB), and Random Forest (RF) all trained on the same institutional database.

These specific algorithms were selected for comparison as they are currently regarded as the gold standard in ensemble-based educational data mining and graduation forecasting. To maintain experimental integrity, every model was subjected to an identical data-handling pipeline and the same training-testing partitioning.

The evaluation primarily examines F1-score and overall accuracy, providing a balanced perspective on the models' ability to handle both prediction precision and class distribution challenges. The comparative findings are detailed in Table 3.

Table 3. Performance Benchmarking: Single Classifiers vs. Stacked Ensemble Framework

Model	Accuracy	F1-Score
Random Forest	0.86	0.87
Gradient Boosting	0.88	0.88
XGBoost	0.90	0.89
Stacked Ensemble (Proposed)	**0.905**	**0.90**

The data presented in Table 3 reveals that while sophisticated standalone algorithms like XGBoost attain impressive accuracy levels, our proposed stacked ensemble framework yields superior overall results. By synthesizing Gradient Boosting, Random Forest, and XGBoost via a meta-learning layer, the stacked architecture successfully identifies synergistic patterns that individual models often overlook when operating in isolation.

Furthermore, the ensemble framework demonstrates enhanced consistency, evidenced by its superior F1-score, which signifies a more effective equilibrium between recall and precision. This distinction is vital for doctoral completion forecasting, as incorrect positive predictions (identifying a student as "on-time" when they are actually at risk of delay) may inadvertently undermine the effectiveness of academic support programs.

Compared with traditional single-model approaches commonly used in prior doctoral and student success prediction studies, which typically rely on logistic regression or a single tree-based classifier, the proposed approach offers both higher predictive accuracy and explainability through SHAP based interpretation. This dual advantage enables universities to move from purely predictive analytics toward transparent, data-driven academic decision-making.

3.6 Strengths and Limitations

The SHAP summary visualization presented in Figure 6 offers an aggregate perspective on how specific attributes affect the stacked ensemble's forecasts for the entire doctoral cohort. A primary advantage of this methodology is its capacity to merge superior predictive precision with clear, actionable interpretability[22]. In contrast to standard "black-box" algorithms, our stacked framework incorporates SHAP derived insights that clarify not only the significance of each variable but also the direction of their influence on graduation probability. The prevalence of progression-based markers including Graduation Semester, Public Defense Semester, and Closed Defense Semester within both the SHAP distributions and Random Forest rankings validates that the system identifies authentic academic trends rather than coincidental data patterns.

Another key strength is the robustness of the stacked ensemble. The model achieved an accuracy of approximately 90.5% and an AUC of 1.0, while maintaining perfect precision for the on-time graduation class. This implies that when the model predicts a student will graduate on time, the prediction is always correct in the test data, which is highly valuable for academic monitoring and early intervention systems. The integration of multiple base learners further reduces variance and improves generalization compared with any single classifier.

However, the proposed approach also has several limitations. First, the dataset consists of only 67 doctoral records from a single institution, which may limit the generalizability of the model to other universities or doctoral programs with different academic structures. Second,

although SHAP provides strong post-hoc explanations, the use of encoded features in ensemble models can make direct semantic interpretation less intuitive without careful mapping back to original academic variables. Finally, the relatively small test set increases the risk of optimistic performance estimates, even though stratified sampling was applied[23]. Future work should therefore include multi-institutional data, larger cohorts, and longitudinal modeling to further validate and extend the proposed explainable stacked ensemble framework.

4. CONCLUSIONS

This study proposed and validated an explainable stacked ensemble learning framework for predicting on-time doctoral graduation using institutional academic data from the Graduate School of Universitas Hindu Negeri I Gusti Bagus Sugriwa Denpasar. By integrating Random Forest, Gradient Boosting, and XGBoost into a meta-learning architecture using 67 academic records, the model achieved a high predictive performance with an accuracy of 90.5% and an AUC of 1.0. These results demonstrate a strong discriminative capability between on-time and delayed doctoral graduates, providing a quantitative basis for academic evaluation.

Beyond predictive accuracy, the incorporation of SHAP-based explainable artificial intelligence enabled transparent and academically meaningful interpretation of the model. The results consistently revealed that milestone-related variables particularly Semester Lulus, Semester Ujian Terbuka, and Semester Ujian Tertutup are the dominant determinants of doctoral graduation timeliness, supported by both Random Forest feature importance and SHAP explanations. This confirms that doctoral progression structure, rather than solely academic performance indicators such as GPA, plays the central role in timely completion.

From a methodological perspective, the proposed stacked ensemble outperformed individual base learners in both accuracy and stability, confirming that combining heterogeneous models improves generalization on small and imbalanced educational datasets. From a practical standpoint, the developed framework can be used as an early warning and decision-support system to identify doctoral students at risk of delayed graduation, enabling targeted academic and supervisory interventions.

Future research should expand the dataset across multiple institutions and incorporate longitudinal and behavioral features to further enhance model robustness and policy relevance.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to the Graduate School of Universitas Hindu Negeri I Gusti Bagus Sugriwa Denpasar for providing access to institutional doctoral program data used in this study. The support and cooperation of academic administrators and doctoral program coordinators are highly appreciated, as they made it possible to conduct this research using real academic records.

The authors also acknowledge the use of Google Colaboratory for computational resources and data processing, which facilitated the implementation of ensemble learning models and explainable artificial intelligence techniques.

Finally, the authors thank the reviewers and editors for their constructive feedback, which has contributed to improving the quality and clarity of this manuscript.

REFERENCES

[1] A. Hasibuan and D. Mahdiana, “Predicting on-time graduation based on student academic data using machine learning,” *J. Pendidik. dan Teknol. Inf.*, 2023.

[2] P. Paul, “Analyzing Dropout of Students and an Explainable Predictive Model,” *Front. Educ.*, 2025.

[3] S. M. Lundberg and S.-I. Lee, “A Unified Approach to Interpreting Model Predictions,”

in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017, pp. 4765–4774.
[4] S. M. Lundberg and S.-I. Lee, "From Local Explanations to Global Understanding with SHAP," *Nat. Mach. Intell.*, vol. 2, pp. 252–259, 2020, doi: 10.1038/s42256-019-0138-9.
[5] Z. Liu, X. Zhou, and Y. Liu, "Student Dropout Prediction Using Ensemble Learning with SHAP-Based Explainable AI Analysis," *J. Soc. Syst. Policy Anal.*, vol. 2, no. 3, pp. 111–132, 2025, doi: 10.62762/JSSPA.2025.321501.
[6] K. Okoye, J. T. Nganji, J. Escamilla, and S. Hosseini, "Machine Learning Model (RG-DMML) and Ensemble Algorithm for Prediction of Students' Retention and Graduation in Education," *Comput. Artif. Intell.*, 2024, doi: 10.1016/j.caeai.2024.100205.
[7] G. Airlangga, "A Comparative Analysis of Machine Learning Models for Predicting Student Performance: Evaluating the Impact of Stacking and Traditional Methods," *Brill. Res. Artif. Intell.*, vol. 4, no. 2, pp. 491–499, 2024, doi: 10.47709/brilliance.v4i2.4669.
[8] Y. Rimal and N. Sharma, "Ensemble Machine Learning Prediction Accuracy: Local vs Global Precision in Education," *Front. Educ.*, 2025, doi: 10.3389/feduc.2025.1571133.
[9] S. Ghimire, "Explainable Artificial Intelligence-Machine Learning Models for Higher Education Predictions," *Comput. Educ. Artif. Intell.*, 2024.
[10] Y. Guan, F. Wang, and S. Song, "Interpretable Machine Learning for Academic Performance Prediction: A SHAP-Based Analysis of Key Influencing Factors," *Innov. Educ. Teach. Int.*, 2025, doi: 10.1080/14703297.2025.2532050.
[11] E. Ben George, R. Senthilkumar, F. Al-Junaibi, and Z. Al-Shuaibi, "Explainable AI Methods for Predicting Student Grades and Improving Academic Success," *J. Inf. Syst. Eng. Manag.*, vol. 10, no. 23s, 2025.
[12] R. Ahmed, "A Customized Ensemble Machine Learning Approach for Student Analytics," *Cogent Educ.*, 2025.
[13] "A Stacking Ensemble Model for Predicting Student High School Graduation Outcomes," *J. Appl. Data Sci.*, 2026.
[14] C. Romero and S. Ventura, "Educational Data Mining and Learning Analytics: An Updated Survey," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 10, no. 3, p. e1355, 2020.
[15] D. Oreski, "Student Success Prediction Based on Machine Learning and LMS Data," *Educ. Inf. Technol.*, 2025.
[16] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 3rd ed. O'Reilly Media, 2022.
[17] M. Tao, Y. Liu, and Q. Chen, "Machine learning model (RG-DMML) and ensemble algorithm for predicting student retention and graduation," *Learn. Anal. Artif. Intell. Educ.*, 2024.
[18] A. Rizalno, D. Setiawan, and T. Gunawan, "Predictive analytics for higher education using ensemble learning algorithms," *J. Educ. Data Sci.*, vol. 5, no. 2, pp. 115–129, 2022.
[19] H. Kurniawan and N. Lestari, "Prediction of on-time student graduation with deep learning," *J. ICT Res. Appl.*, 2025.
[20] D. Chicco and G. Jurman, "The Advantages of the Matthews Correlation Coefficient Over F1-Score and Accuracy in Binary Classification Evaluation," *BMC Genomics*, vol. 21, p. 6, 2020.
[21] T. Tong and Z. Li, "Predicting Learning Achievement Using Ensemble Learning with Result Explanation," *PLoS One*, vol. 20, no. 1, 2025, doi: 10.1371/journal.pone.0312124.
[22] F. T. Johora, "An Explainable AI-Based Approach for Predicting Undergraduate Student Academic Performance," *Heliyon*, 2025.
[23] W. Villegas, "Machine Learning Models for Academic Performance and Educational Decision-Making," *Front. Educ.*, 2025.