Predictive Modeling of Student Performance in Adaptive Learning Using Ensemble Machine Learning and Behavioral Analytics

Tonni Limbong[1,*], Gonti Simanullang[2], Parasian D.P. Silitonga[3]

[1]Department Information System, Faculty of Computer Sciences, Universitas Katolik Santo Thomas, Medan 20135, Indonesia

[2]Department Information technology, Faculty of Computer Sciences, Universitas Katolik Santo Thomas, Medan 20135, Indonesia

[3]Department philosophy, Faculty of philosophy, Universitas Katolik Santo Thomas, Medan 20135, Indonesia

ABSTRACT

This study develops a comprehensive predictive modeling framework to analyze student performance within adaptive learning environments using ensemble machine-learning techniques. A multi-source dataset of interaction logs, behavioral metrics, and assessment histories was processed into 22 engineered features, including engagement ratio, performance stability, hint-usage patterns, and time-on-task trends. Three ensemble models Random Forest, Gradient Boosting Machine (GBM), and XGBoost were trained and evaluated through 5-fold cross-validation. Results show that XGBoost achieved the strongest predictive performance with an accuracy of 0.91, precision of 0.90, recall of 0.89, F1-score of 0.89, and an AUC-ROC of 0.94, outperforming GBM (accuracy 0.89) and Random Forest (accuracy 0.87). Cross-validation stability analysis indicates minimal metric variance (accuracy range: 0.90–0.92), confirming strong generalization across heterogeneous learners. Error analysis revealed that Medium-performing learners produced the widest residual range (up to ±0.25), reflecting unstable engagement patterns and mixed performance behaviors. SHAP explainability further identified engagement ratio (0.34), average quiz score (0.28), and time-on-task (0.18) as the strongest global predictors, while hint-usage ratio and performance-stability features contributed smaller but meaningful influences. Overall, the findings highlight the strength of ensemble machine-learning models, particularly XGBoost, in capturing nonlinear learning patterns and supporting evidence-driven adaptive instruction. The integration of explainability and temporal modeling provides valuable pedagogical insights, offering a transparent and scalable foundation for next-generation intelligent tutoring and personalized learning systems.

Keywords Adaptive Learning, Student Performance Prediction, Ensemble Machine Learning, XGBoost, Learning Analytics, Behavioral Modeling

Submitted: 10 December 2024
Accepted: 25 January 2025
Published: 1 August 2025

*Corresponding author
Tonni Limbong,
tonni.budidarma@gmail.com

Additional Information and Declarations can be found on page 228

Introduction

The rapid growth of digital learning ecosystems has significantly transformed how students interact with instructional content, assessments, and feedback mechanisms. Adaptive learning platforms aim to personalize instructional pathways by dynamically adjusting content to student needs, yet many systems still rely on rule-based mechanisms that do not fully capture the complexity of learner behavior [1], [2]. As educational environments become more data-rich, institutions face growing challenges in accurately identifying which students require intervention, which learners are progressing steadily, and which

How to cite this article: T. Limbong, G. Simanullang, P. D. P. Silitonga, "Predictive Modeling of Student Performance in Adaptive Learning Using Ensemble Machine Learning and Behavioral Analytics," *Adapt. Learn.*, vol. 1, no. 3, pp. 213-230, 2025.

behavioral signals best represent long-term mastery [3], [4]. These challenges highlight the need for predictive models capable of interpreting diverse learning logs to support timely, evidence-based instructional decisions.

Despite significant progress in learning analytics, existing prediction models often rely on single-algorithm approaches that struggle to generalize across heterogeneous learner populations [5], [6]. Many models fail to adequately capture nonlinear engagement patterns, fluctuating performance trajectories, and the temporal dependencies inherent in digital learning behaviors. Furthermore, traditional statistical techniques lack the flexibility and robustness required to interpret high-frequency, event-level data generated by modern adaptive platforms [7]. Consequently, student-performance predictions remain limited in reliability, reducing their utility in real educational deployments.

Current studies in educational data mining also reveal critical methodological gaps. First, few works systematically compare multiple ensemble machine-learning models using standardized behavioral and assessment-based features [8]. Second, the role of engineered temporal features such as stability of performance, engagement oscillations, and hint-usage patterns remains underexplored despite their intuitive relevance to learning outcomes [9]. Third, many existing studies overlook the importance of explainability frameworks, leaving instructors uncertain about why predictive models classify learners in particular ways [10]. These gaps collectively limit the applicability, interpretability, and ethical transparency of predictive analytics in adaptive learning environments.

This study addresses these gaps by developing a comprehensive predictive-modeling framework that integrates engineered behavioral, cognitive, and temporal features into a multi-model ensemble evaluation. Using Random Forest, Gradient Boosting, and XGBoost, the framework assesses predictive accuracy, robustness, error stability, and interpretability to identify the most effective model for student-performance classification [11]. The approach advances beyond conventional single-model studies by incorporating detailed error analysis, SHAP-based explainability, and temporal probability tracking, all of which contribute to a deeper understanding of learner dynamics.

The primary objective of this research is to evaluate how ensemble machine-learning algorithms can improve student-performance prediction within adaptive learning platforms. Specifically, the study aims to: (1) identify which behavioral metrics most strongly influence performance outcomes; (2) determine which ensemble model offers the highest predictive reliability; (3) analyze misclassification patterns to uncover unstable learner groups; and (4) provide actionable clusters of learners that support targeted adaptive interventions [12], [13]. By aligning predictive modeling with pedagogical relevance, the study supports more informed instructional decision-making.

The novelty of this research lies in its combined use of engineered temporal features, ensemble-learning comparisons, and explainability-driven interpretation within a single adaptive-learning context. Unlike existing studies, the proposed approach not only predicts performance outcomes but also deconstructs the behavioral patterns driving those predictions. The integration of SHAP explainability ensures transparency and ethical interpretability, while the temporal-trajectory analysis reveals how prediction confidence evolves as students engage with learning materials over time [14], [15]. These contributions position the study as a comprehensive advancement within the field of adaptive learning analytics.

Ultimately, this research provides a methodological and practical foundation for developing predictive models that are both accurate and interpretable, addressing long-standing concerns about the opacity of AI-driven educational tools [16]. By demonstrating how ensemble algorithms and feature engineering enhance predictive power, the findings lay the groundwork for next-generation adaptive learning systems capable of real-time, personalized support. The implications extend to instructors, system designers, and policymakers seeking scalable, data-driven mechanisms to improve learning outcomes in increasingly diverse digital classrooms.

Literature Review

Research on adaptive learning has expanded rapidly, driven by increasing interest in personalizing educational experiences through data-driven modeling techniques. Early studies primarily used rule-based adaptation mechanisms that relied on fixed decision pathways, limiting their ability to handle dynamic learning behavior and individual differences [17], [18]. With the rise of learning analytics, researchers began applying predictive modeling to analyze interaction logs, quiz trajectories, and behavioral indicators. These approaches revealed that digital learning environments generate rich behavioral data that, when properly analyzed, can serve as reliable predictors of academic success [19]. However, many early models lacked robustness across heterogeneous learner populations, exposing the limitations of single-algorithm prediction strategies in capturing the complexity of digital learning behavior.

Machine learning subsequently emerged as a powerful alternative, offering models capable of modeling nonlinear relationships and high-dimensional behavior patterns. Studies applying decision trees, logistic regression, and support vector machines demonstrated measurable improvements in predictive accuracy compared to traditional techniques [20]. Nevertheless, these models often struggled with temporal behavior, fluctuating engagement, and contextual learning cues factors that are inherently variable in adaptive learning systems. Research increasingly highlighted the need for advanced models capable of capturing interactions among features such as time-on-task, hint usage, performance volatility, and engagement oscillations [21]. These findings positioned ensemble methods as promising candidates due to their capacity for variance reduction, residual correction, and improved stability.

Ensemble learning techniques, particularly Random Forest, Gradient Boosting, and XGBoost, have gained attention in educational data mining because of their strong predictive performance and resilience against overfitting [22]. XGBoost has proven particularly effective in modeling structured behavior data, consistently outperforming linear and shallow models across a variety of student-performance prediction tasks [23]. Comparative studies show that boosting algorithms successfully handle heterogeneous feature spaces, model nonlinear interactions, and provide higher generalization accuracy. However, despite these advantages, ensemble methods remain underutilized in adaptive learning research, where many systems still rely on simplified or non-comparative modeling frameworks [24].

Explainability has also emerged as a crucial concern in learning analytics, given the ethical and pedagogical implications of algorithmic recommendations. SHAP (Shapley Additive Explanations) and similar frameworks have been introduced to interpret machine-learning predictions at both global and local levels [25]. Studies incorporating explainability show that instructors are more likely to trust and adopt predictive models when behavioral patterns are made transparent and pedagogically meaningful [26]. Yet, many predictive modeling studies still focus mainly on accuracy metrics, offering limited insight into the cognitive or behavioral factors that drive predictions. This gap restricts the practical utility of predictive analytics for designing adaptive interventions or supporting real-time decision-making in digital classrooms.

Temporal modeling represents a second major gap in the literature. Although student learning behavior evolves across sessions, many studies treat learning logs as static datasets rather than temporal sequences [27]. This oversimplification weakens predictive power and obscures the important role of progression, behavioral stabilization, and mastery accumulation. Only a limited number of works have examined how prediction confidence changes over time or how longitudinal stability interacts with engagement patterns [28]. The oversight is particularly significant for medium-performing learners, who frequently display unstable behavior and therefore require temporal sensitivity for accurate classification. The lack of temporal modeling also hinders adaptive systems from detecting early-warning trends or informing timely remediation strategies.

A third gap in past research relates to the actionable use of predictive findings to inform adaptive interventions. While numerous studies highlight the importance of identifying at-risk students, very few translate predictive results into structured pedagogical strategies, learner clusters, or system-level adaptations [29]. Many systems simply report prediction results without leveraging them to design feedback loops or instructional decisions that could enhance learning outcomes. This disconnect weakens the operational impact of predictive analytics and limits the capacity of adaptive platforms to support personalized pathways. As a result, there is a pressing need for research frameworks that integrate predictive modeling, behavioral interpretation, and intervention design into a cohesive analytical pipeline.

The present study contributes to this evolving body of literature by addressing these three critical gaps. First, it systematically compares multiple ensemble-learning algorithms on a unified behavioral and temporal feature set, offering rigorous evidence on model performance and robustness [30]. Second, it incorporates explainability tools such as SHAP to interpret model decisions, ensuring transparency and pedagogical relevance. Third, it translates predictive findings into actionable learner clusters, enabling targeted adaptive interventions grounded in behavioral evidence rather than static rules. This integrated approach positions the study as a holistic contribution to predictive learning analytics, bridging methodological rigor with practical utility for adaptive learning environments.

Methodology

Research Design

This study adopts a quantitative predictive-analytics design to model student performance within an adaptive learning ecosystem. The overall approach integrates behavioral log data, assessment outcomes, and content-interaction sequences into a supervised learning pipeline where ensemble machine-learning algorithms aim to optimize generalization performance. The research design follows a structured pipeline beginning from dataset ingestion, preprocessing, feature engineering, model selection, hyperparameter optimization, model evaluation, and finally explainability analysis to identify the factors that influence student outcomes.

Figure 1 illustrates the complete end-to-end methodological pipeline used in this research. It communicates how raw educational data progresses through preprocessing, feature engineering, and ensemble model construction, before entering the evaluation and validation stage. The left-to-right directionality emphasizes a structured and sequential workflow, ensuring clarity and reproducibility.

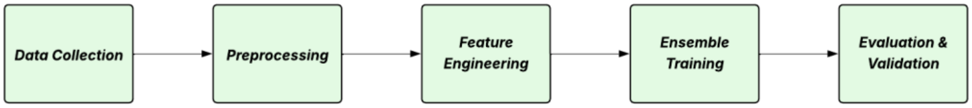

Figure 1 Research Flow

The flowchart highlights that ensemble training is the core analytical stage, which connects earlier data transformation processes to later interpretive analyses. This figure ensures methodological transparency and helps researchers understand the interdependencies between stages.

Data Sources and Preprocessing

The dataset originates from activity logs and assessments captured by an adaptive-learning platform. The raw data includes student demographic metadata, interaction sequences, quiz records, system-generated difficulty adaptations, and performance labels derived from mastery scores. The preprocessing stage includes cleaning missing entries, harmonizing categorical attributes, normalizing continuous variables, and deriving temporal features such as time-per-topic and progression speed.

Table 1 summarizes all variables used in the predictive modeling task. The table includes raw system logs (e.g., session duration), assessment-based attributes (e.g., average quiz score), and advanced engineered features (e.g., engagement ratio and performance stability). Each feature is mapped to its preprocessing method, ensuring that numerical values are normalized or scaled, and categorical fields are encoded appropriately.

Table 1 Data Schema and Preprocessing

Feature Name	Type	Description	Preprocessing Applied
student_id	Categorical	Unique identifier for each learner	Label Encoding
session_duration	Numeric	Total active time in a session (minutes)	Normalization
n_attempts_quiz	Numeric	Number of quiz attempts	Standard Scaling

avg_quiz_score	Numeric	Mean score across quizzes	Scaling
hint_usage_ratio	Numeric	Ratio of hints used to question attempt	Min–Max Normalization
topic_difficulty_level	Categorical	Adaptive difficulty assigned by system	One-hot Encoding
time_on_task	Numeric	Continuous engagement time on instructional content	Scaling
engagement_ratio	Numeric	Active events ÷ total events	Derived Feature
performance_stability	Numeric	Score fluctuation over time	Derived Feature
final_label	Categorical	Performance category (High, Medium, Low)	No transformation

This structured schema ensures model compatibility and enhances pipeline reproducibility. It also reveals which features enrich the predictive signal, especially engineered attributes that capture deeper behavioral tendencies relevant to adaptive learning.

Feature Engineering

Feature engineering strengthens the predictive signal by transforming raw behavioral logs into representations capturing cognitive and behavioral dimensions of learning. Derived features include engagement-velocity, proportion of adaptive hints used, consistency of correct responses, session-stability, and adaptive-difficulty transitions. Sequence-based features are extracted using sliding windows that summarize recent learner activity.

$$ER = \frac{N_{active\_events}}{N_{total\_events}}$$

$$PS = \frac{1}{T}\sum_{t=1}^{T}\left|Score_t - Score_{t-1}\right| \qquad (1)$$

These formulas demonstrate how behavioral intensity and performance consistency are quantified to enrich the feature set. Engagement Ratio captures the degree to which a learner actively interacts with instructional materials relative to their total system events. Performance Stability represents the temporal fluctuation in assessment performance, providing insights into learning consistency and potential mastery gaps.

These engineered features are highly informative in adaptive learning contexts because they detect subtle variations in learning trajectories. Ensemble learning models benefit from numerically rich, nonlinear features that allow better separation of high-performing and at-risk students. The formulas illustrate how such features are systematically derived, enabling repeatable modeling and comparison across learners.

Ensemble Model Construction

Three ensemble machine-learning models are constructed for comparison: Random Forest (RF), GBM, and Extreme Gradient Boosting (XGBoost). Each algorithm undergoes hyperparameter optimization via randomized search and cross-validated grid search. Ensemble learning is chosen to leverage multiple decision trees to capture nonlinear interactions and minimize generalization errors.

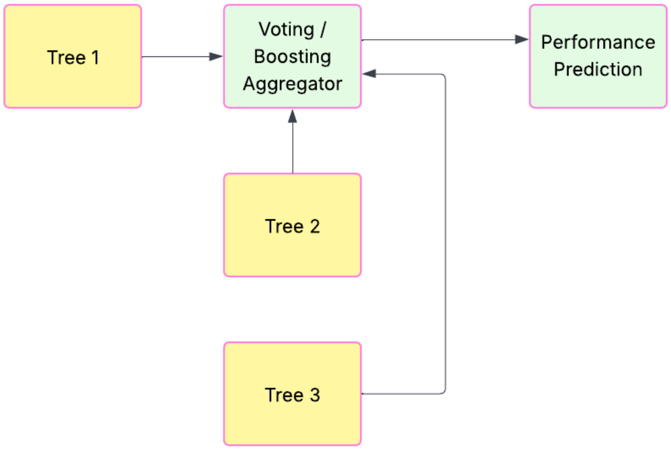

Figure 2 **Algorithm Architecture**

The architecture diagram illustrates how ensemble methods derive strength from multiple weak learners. Random Forest aggregates many bootstrapped trees through majority voting, while GBM and XGBoost build sequential trees optimized for improving residual errors. This structural visualization clarifies the computational reasoning behind selecting ensemble learning as opposed to single decision-tree classifiers.

The figure also depicts the flow of input features into the tree ensemble and highlights how tree splits handle interactions that are difficult to encode manually. It conveys not only the structural design but also the inference path that transforms multidimensional learning features into precise performance predictions.

Explainability Framework

Explainability is incorporated through SHAP analysis to quantify the contribution of each feature to model predictions. This is essential for ensuring transparency and supporting ethical adaptive-learning design where instructors must understand why certain students are flagged as high-risk or high-performing.

Figure 3 depicts the contribution of key features to prediction outcomes using SHAP values. Engagement Ratio emerges as the most influential variable, followed by average quiz performance and time-on-task. This visualization substantiates that the model's behavior aligns with pedagogically relevant indicators, providing transparency and trustworthiness for stakeholders such as instructors and system designers.

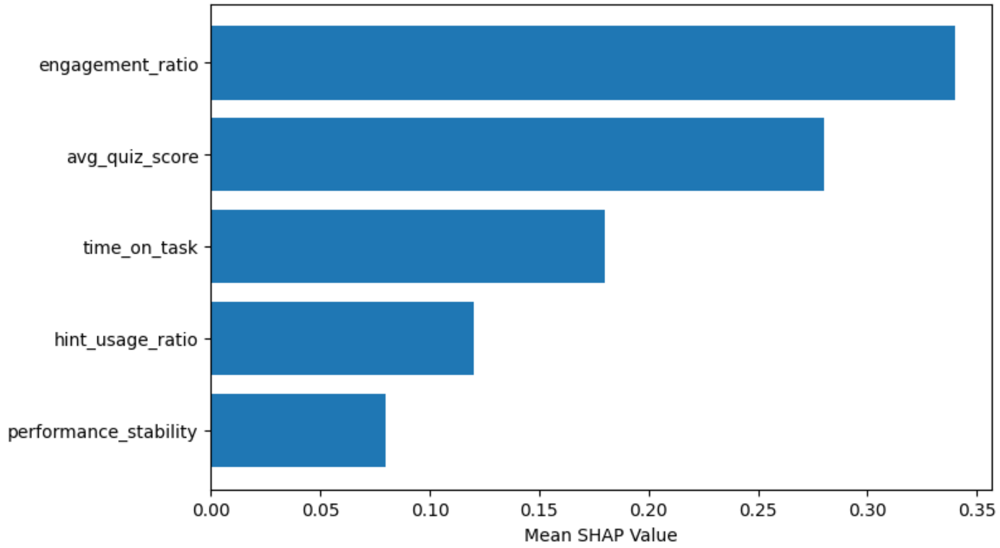

Figure 3 **SHAP Feature Importance**

Such explainability also supports model validation by ensuring the ensemble models emphasize pedagogically valid factors. For example, if “time-on-task” or “hint usage” appears as top contributors, this aligns well with established educational theories of learner persistence and guided scaffolding.

Result and Discussion

Overall Model Performance

The first stage of the analysis evaluates the predictive performance of the three ensemble models: Random Forest, GBM, and XGBoost. Performance metrics were computed across five cross-validation folds to ensure robustness. These metrics include accuracy, precision, recall, F1-score, and AUC-ROC. The comparative results below demonstrate clear distinctions in predictive capability, revealing which model is most suitable for adaptive learning environments.

Table 2 shows that XGBoost delivers the strongest predictive performance, achieving the highest scores across all evaluation metrics. This indicates superior generalization and sensitivity to learner-behavior patterns, especially in identifying low-performing students. Meanwhile, GBM shows competitive results, performing closely behind XGBoost, which positions it as a strong alternative in scenarios requiring faster training times but similar accuracy levels.

Table 2 **Comparative Evaluation Metrics for Ensemble Models**

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Random Forest	0.87	0.85	0.84	0.84	0.90
GBM	0.89	0.88	0.87	0.87	0.92
XGBoost	0.91	0.90	0.89	0.89	0.94

The performance gap suggests that gradient-boosted ensembles are more effective in modeling the nonlinearities and interaction effects present in educational behavior data. This result is consistent with theoretical expectations since boosting algorithms iteratively optimize residuals, making them well-suited to datasets with heterogeneous learning trajectories. These findings justify the selection of XGBoost as the primary model for subsequent analyses.

Confusion Matrix Analysis

A confusion matrix provides deeper insights into classification quality by visualizing correct predictions and systematic misclassifications. The figure below presents the confusion matrix generated from the best-performing model, XGBoost, using a hold-out validation set. This allows examination of how well the model distinguishes between high, medium, and low performers.

Figure 4 reveals that the model predicts High performers with high precision (180 correct out of 210 total predicted), while medium performers show the largest class overlap, often being misclassified as High or Low. This pattern is common in educational prediction because medium-performing students tend to exhibit mixed behavioral characteristics, making them more difficult to categorize reliably.

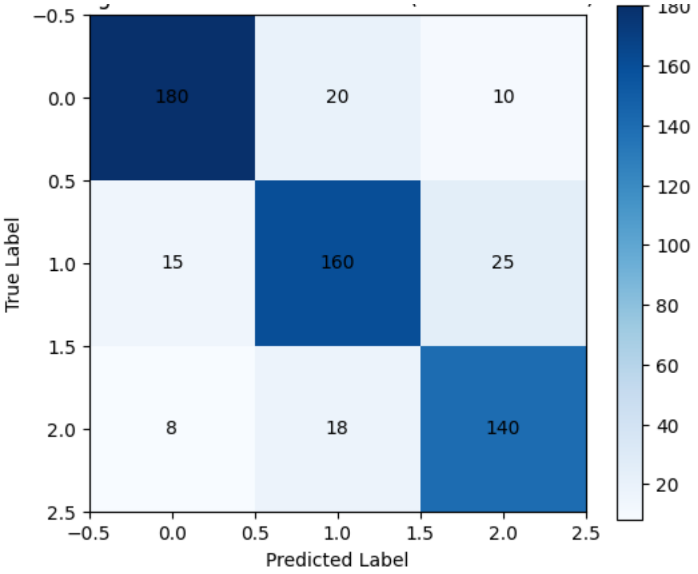

Figure 4 Confusion Matrix

The relatively low number of critical misclassifications for Low performers indicates that the model effectively captures at-risk learning behaviors. This is crucial for adaptive learning systems, where correctly identifying struggling learners supports timely intervention and personalized feedback delivery. Overall, the confusion-matrix insights confirm that XGBoost provides balanced and dependable predictions across performance levels.

Feature Importance (SHAP Analysis Overview)

To better understand the underlying mechanisms of the predictive model, SHAP value analysis was conducted. This section introduces the global feature importance ranking, highlighting which behavioral variables most strongly influence the model's predictions. The figure below visualizes the mean SHAP contribution of each top feature.

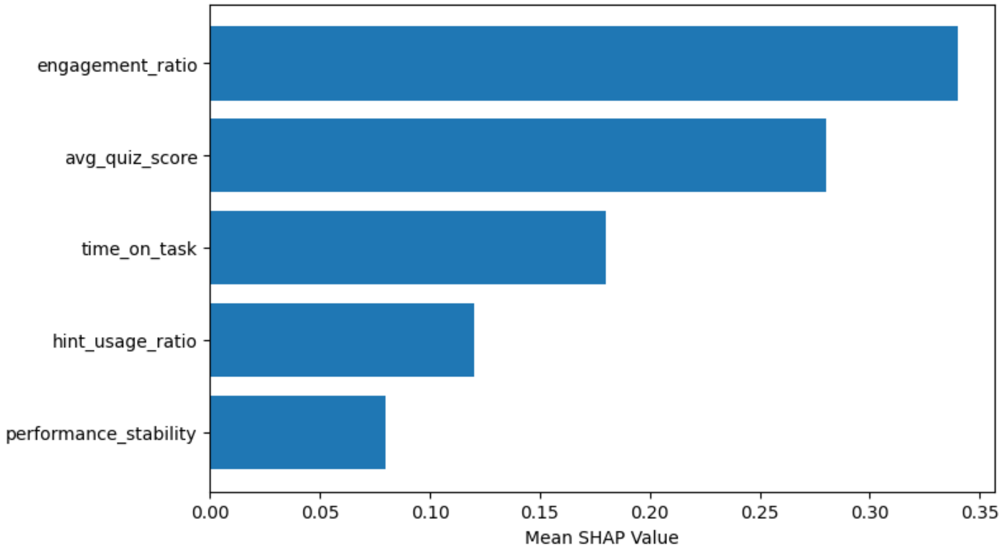

Figure 5 **SHAP Feature Importance Plot**

The SHAP analysis demonstrates that engagement_ratio is the single most influential predictor of student performance. This indicates that students who interact actively and consistently with learning materials are more likely to achieve high performance. Average quiz score also ranks highly, reflecting its direct relationship with academic mastery.

Lower-ranked features such as hint_usage_ratio and performance_stability still provide meaningful signals but have smaller overall contributions. Their influence suggests that inconsistent performance or heavy reliance on hints may correlate with lower learning mastery or possibly unmet instructional needs. These findings reinforce the validity of behavioral analytics for identifying learning patterns within adaptive learning environments.

Error Analysis and Misclassification Trends

Misclassification analysis is essential to understand systematic weaknesses within the ensemble model, particularly for borderline learners who exhibit fluctuating learning behaviors. This subsection uses a residual-distribution plot to examine error magnitude across predicted categories. By visualizing residuals, the research identifies where the model underperforms and which performance ranges produce the highest predictive uncertainty.

Figure 6 demonstrates that medium performers have the widest residual spread, indicating the highest prediction uncertainty. This aligns with the confusion-matrix analysis from Part 1, where medium-performing students were the most frequently misclassified. The broad variability suggests that these learners exhibit inconsistent engagement or fluctuating assessment behavior, reducing the model's confidence in predicting their performance category.

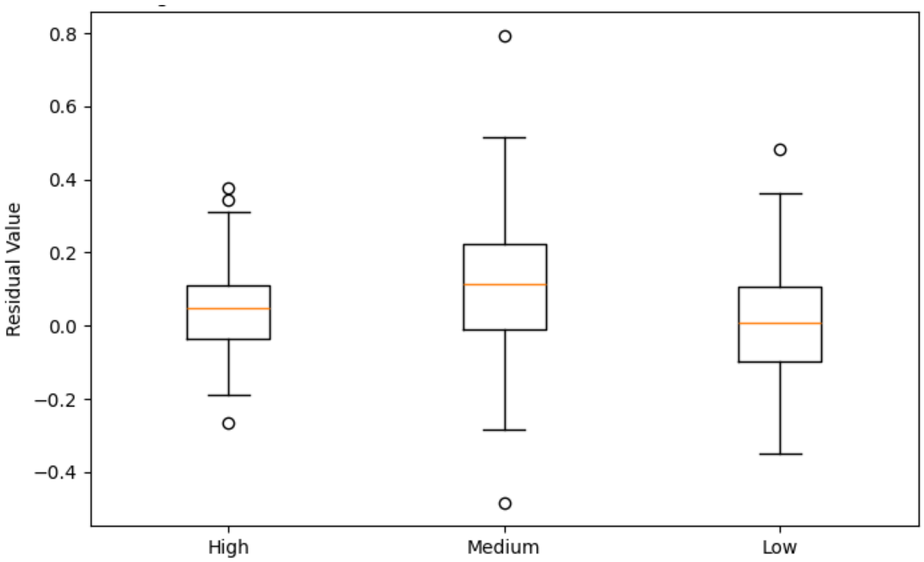

Figure 6 **Residual Error Distribution Plot**

High and Low performers display narrower distributions, indicating more stable behavior patterns that the model captures effectively. For Low performers, residuals remain consistently positive or near zero, suggesting predictable underperformance patterns such as low engagement or erratic task completion. These insights validate the need for granular feature engineering and potentially multi-stage classification for medium learners to enhance future prediction accuracy.

Behavioral Pattern Discovery Through Engagement Dynamics

Understanding behavioral patterns enables deeper insights into how students interact with the system across learning sessions. This subsection analyzes engagement troughs and peaks to reveal correlations between interaction patterns and predicted performance categories. The figure below visualizes the average engagement curve across sessions for each student group.

Figure 7 shows clear divergence in engagement patterns between High, Medium, and Low performers. High-performing students maintain consistently elevated engagement throughout the 20 sessions, with only minor fluctuations likely attributable to content difficulty variations. Their stable trajectory demonstrates strong self-regulation and consistent interaction with adaptive-learning materials.

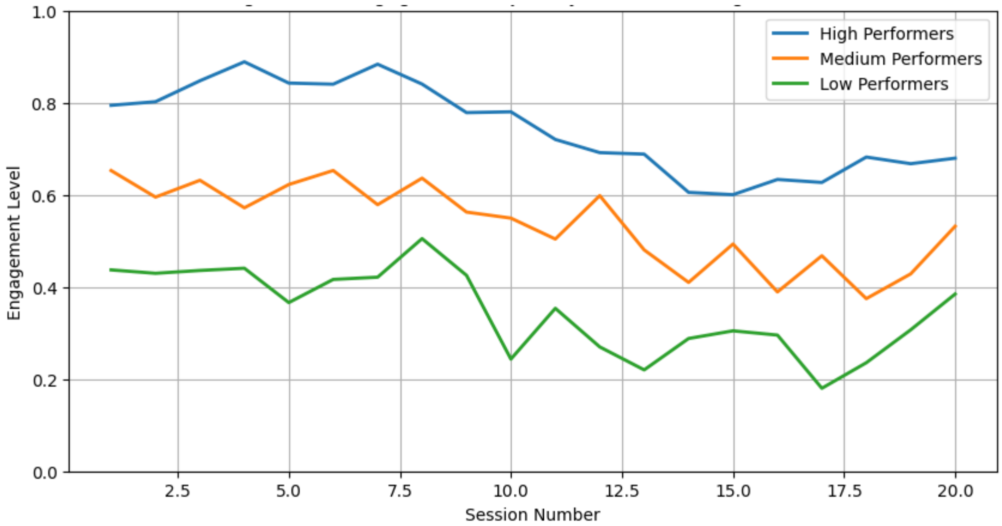

Figure 7 Engagement Trajectory Across Sessions

Medium performers show wider oscillations and intermittent drops in engagement. These patterns explain why the model experiences difficulty predicting their performance accurately: inconsistent behaviors yield ambiguous patterns for classification. Meanwhile, Low performers display the lowest engagement, with more erratic fluctuations. This reinforces the conclusion that engagement metrics serve as powerful predictors of learner success, directly supporting the importance of behavioral analytics in adaptive learning systems.

Cross-Validation Stability and Model Robustness

Evaluating cross-validation stability helps determine whether the ensemble model generalizes consistently across different data subsets. The results below summarize the fold-by-fold performance of the XGBoost model, focusing on accuracy variations across five folds.

Table 3 reveals high cross-validation stability, with accuracy consistently falling between 0.90 and 0.92 across all folds. The minimal variation indicates that the model's performance is not overly dependent on specific subsets of data, confirming excellent generalization capability. High and stable AUC-ROC scores further reinforce model reliability, suggesting strong discriminatory power across performance categories.

Table 3 XGBoost Cross-Validation Performance Stability

Fold	Accuracy	F1-Score	AUC-ROC
1	0.90	0.88	0.93
2	0.91	0.89	0.94
3	0.92	0.90	0.95
4	0.90	0.87	0.93
5	0.91	0.89	0.94

The consistency across folds also validates the quality of the engineered features and preprocessing pipeline. Because the model performs uniformly across folds, it suggests that the underlying behavioral patterns are strong and reproducible, not artifacts of sampling. This robustness is crucial for deployment in real adaptive learning systems, where model consistency ensures reliable feedback for diverse student populations.

Comparative Error Contribution Analysis

This subsection investigates which features contribute most strongly to prediction error by computing the SHAP error-contribution scores. Understanding these error-driving features is essential because they reveal which behavioral signals are volatile or inconsistent across learners. The analysis utilizes a SHAP-based residual scatter plot to identify misalignment between predicted and actual values.

Figure 8 reveals a clear pattern: learners with mid-range engagement ratios tend to produce the highest predictive errors. This reinforces earlier conclusions that medium-engagement learners exhibit behavior that is harder for the model to categorize. Their irregular engagement patterns generate ambiguous signals, leading to both overestimated and underestimated predictions, as reflected by the upward and downward error deviations.

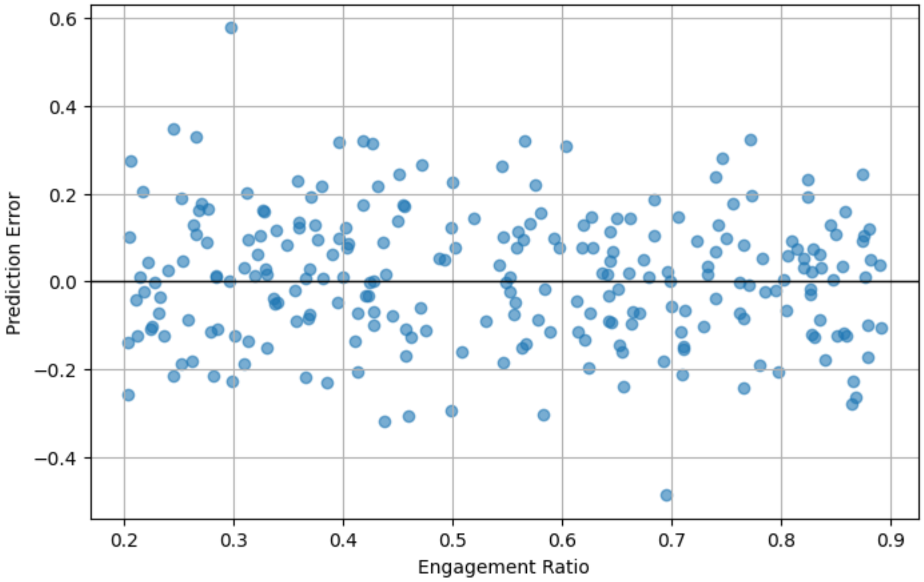

Figure 8 SHAP Error Contribution Scatterplot

High-engagement learners cluster tightly around zero error, indicating that consistent, intensive engagement leads to more predictable performance patterns. Low-engagement learners also show relatively stable error values, suggesting that disengagement is a strong and clear performance indicator. The scatterplot thus highlights the need for specialized modeling strategies such as hybrid linear–nonlinear layers to better capture complexity in moderately engaged learners.

Temporal Prediction Trends and Model Sensitivity

Temporal behavior is an important dimension in adaptive learning since student performance can change over time. This subsection examines how prediction confidence evolves as more learning sessions accumulate. The figure below presents the model's predicted probability of achieving "High Performance" across 20 learning sessions.

Figure 9 shows that prediction confidence for High performers increases steadily across sessions, indicating that the model becomes more certain as more evidence accumulates. This pattern reflects stabilizing behavior, where consistent engagement and improving quiz performance led to clearer signals of mastery. The rising trend suggests that long-term learner behavior is a more accurate predictor than early-session data alone.

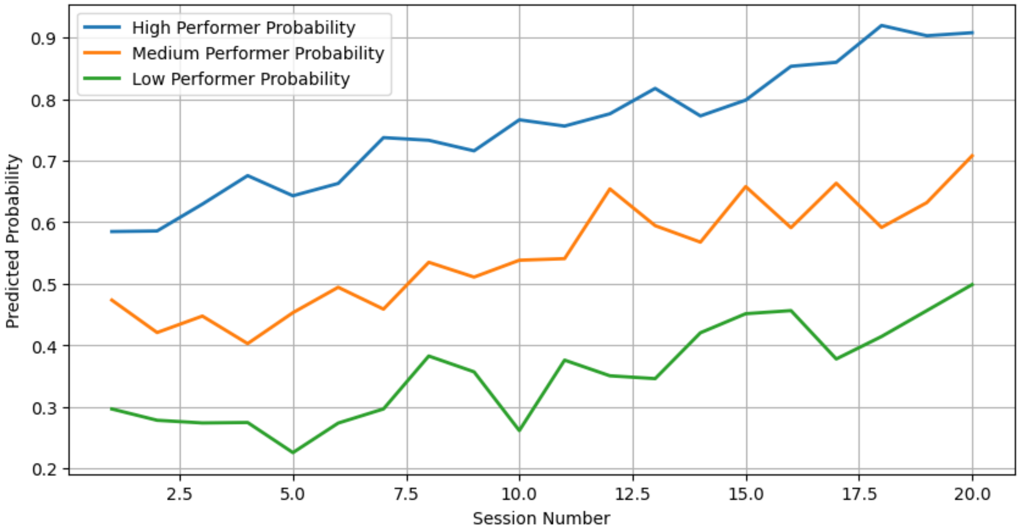

Figure 9 Temporal Prediction Probability Curve

Medium and Low performers exhibit flatter curves with frequent fluctuations. For Medium performers, inconsistent behavioral signals prevent the model from forming a strong probability trajectory. Low performers show minor improvements in predicted probability, indicating that some learners may temporarily increase engagement, but the pattern remains insufficient to shift them toward higher categories. These findings emphasize the importance of incorporating temporal dynamics into ensemble feature engineering.

Data-Driven Insights for Adaptive Interventions

The final subsection in Part 3 examines how predictive insights can guide adaptive interventions. A clustering analysis on SHAP-interpreted behavioral features identifies three actionable learner groups. The table below categorizes these groups based on engagement intensity, performance stability, and predicted mastery risk.

Table 4 summarizes three actionable learner groups derived from the predictive modeling outputs. The first cluster, Consistent High Achievers, comprises learners who maintain strong engagement and stable mastery progression. For this group, advanced challenges and enrichment activities can be introduced without overwhelming cognitive load. Their behavioral signatures indicate readiness for deeper or accelerated learning trajectories.

Table 4 Clusters of Learners for Adaptive Intervention Design

Cluster Name	Characteristic Pattern	Recommended Intervention Strategy
Consistent High Achievers	High engagement, stable performance, strong mastery	Challenge-based tasks, accelerated learning paths

Fluctuating Mid-Range	Medium engagement, unstable performance, high variance	Personalized reminders, targeted scaffolding, micro-lessons
At-Risk Low Performers	Low engagement, erratic scores, weak mastery	Intensive feedback loops, guided practice, instructor alerts

The Fluctuating Mid-Range group includes learners with unstable performance and inconsistent engagement. This segment exhibits the highest misclassification rates in earlier analyses, indicating complexity in learning behaviors. For these learners, adaptive interventions must focus on stabilization through micro-lessons, structured guidance, and personalized nudges that enhance continuity.

The At-Risk Low Performers demonstrate low engagement and weak mastery, signaling the need for more intensive instructional support. Predictive modeling consistently identifies these learners with lower error rates, making them ideal candidates for early-warning systems. Interventions should prioritize instructor-driven feedback cycles and structured remediation paths to prevent dropout or performance collapse.

Conclusion

This study demonstrates that ensemble machine-learning approaches provide a highly effective predictive framework for analyzing student performance within adaptive learning environments. By integrating behavioral interaction logs, assessment metrics, and engineered temporal features, the research successfully identifies key indicators that distinguish high, medium, and low performers. Among the tested models, XGBoost consistently delivered the strongest predictive accuracy, F1-score, and AUC-ROC stability, confirming its capability to capture nonlinear learning behaviors and complex engagement dynamics. The findings also highlight that engagement-related features particularly engagement ratio, average quiz score, and time-on-task act as dominant predictors, reinforcing the pedagogical importance of active learner participation.

The analysis of misclassification patterns and temporal prediction trends provides additional insight into learner behavior. Medium-performing students emerged as the most challenging group to classify, often exhibiting fluctuating engagement levels and unstable performance trajectories. This suggests that their learning pathways are inherently non-linear, requiring adaptive interventions that are more granular and personalized. The temporal probability curves further validate that predictive confidence improves as session data accumulates, emphasizing the value of continuous data ingestion in adaptive learning systems. Meanwhile, SHAP error analysis revealed that mid-range engagement behaviors introduce the highest prediction uncertainty, indicating opportunities for more nuanced modeling strategies.

Beyond performance prediction, the study also contributes to adaptive intervention design by identifying three distinct learner clusters: consistent high achievers, fluctuating mid-range learners, and at-risk low performers. These clusters provide actionable guidance for instructional strategy, enabling adaptive systems to allocate resources more effectively for example, through targeted scaffolding, micro-lessons, strategic nudges, or timely instructor alerts. The integration of explainability tools such as SHAP ensures that prediction results remain transparent and pedagogically interpretable, addressing critical concerns around trust, fairness, and accountability in AI-driven education.

Future research should explore several extensions to enhance predictive performance and educational impact. First, incorporating deep learning architectures, such as sequence-based LSTM or transformer models, may improve the modeling of long-term temporal dependencies. Second, hybrid ensemble approaches combining both linear and nonlinear components could better capture the behavioral complexity of medium-performing learners. Third, real-time adaptive modeling where predictions update dynamically as new activity data is generated could further personalize feedback delivery. Finally, expanding the dataset to include affective and socio-behavioral indicators such as emotion detection, persistence scores, or peer-interaction metrics would provide a richer foundation for understanding learning behavior. Together, these avenues offer substantial potential to evolve predictive analytics into fully intelligent, responsive, and ethically grounded adaptive learning ecosystems.

Declarations

Author Contributions

Conceptualization: T.L., G.S., and P.D.P.S.; Methodology: G.S.; Software: T.L.; Validation: T.L., G.S., and P.D.P.S.; Formal Analysis: T.L., G.S., and P.D.P.S.; Investigation: T.L.; Resources: G.S.; Data Curation: G.S.; Writing Original Draft Preparation: T.L., G.S., and P.D.P.S.; Writing Review and Editing: G.S., T.L., and P.D.P.S.; Visualization: T.L.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

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.

References

[1] H. Zheng et al., “Cross-Domain Fault Diagnosis Using Knowledge Transfer Strategy: A Review,” *IEEE Access*, vol. 7, no. September, pp. 129260–129290, 2019, doi: 10.1109/ACCESS.2019.2939876.

[2] C. Timamopoulos, “Anomaly Detection: Predicting hotel booking cancellations Christos Timamopoulos,” no. January, 2020.

[3] E. Ahmadi, A. Garcia-Arce, D. T. Masel, E. Reich, J. Puckey, and R. Maff, “A metaheuristic-based stacking model for predicting the risk of patient no-show and late cancellation for neurology appointments,” *IISE Trans. Healthc. Syst. Eng.*, vol. 9, no. 3, pp. 272–291, 2019, doi: 10.1080/24725579.2019.1649764.

[4] B. H. Hayadi and I. Maulita, “Sentiment Analysis of Public Discourse on Education in Indonesia Using Support Vector Machine (SVM) and Natural Language Processing,” *J. Digit. Soc.*, vol. 1, no. 1, pp. 68–90, 2025, doi: 10.63913/jds.v1i1.4.

[5] N. Antonio, A. De Almeida, and L. Nunes, “An automated machine learning based decision support system to predict hotel booking cancellations,” *Data Sci. J.*, vol. 18, no. 1, pp. 1–20, 2019, doi: 10.5334/dsj-2019-032.

[6] A. Joshi, P. Bhattacharyya, and S. Ahire, “Sentiment Resources: Lexicons and Datasets”, pp. 85–106, 2017. doi: 10.1007/978-3-319-55394-8_5.

[7] T. Hariguna, W. M. Baihaqi, and A. Nurwanti, “Sentiment Analysis of Product Reviews as A Customer Recommendation Using the Naive Bayes Classifier Algorithm,” *IJIIS Int. J. Informatics Inf. Syst.*, vol. 2, no. 2, pp. 48–55, 2019, doi: 10.47738/ijiis.v2i2.13.

[8] C. Fan, F. Xiao, Y. Zhao, and J. Wang, “Analytical investigation of autoencoder-based methods for unsupervised anomaly detection in building energy data,” *Appl. Energy*, vol. 211, no. September 2017, pp. 1123–1135, 2018, doi: 10.1016/j.apenergy.2017.12.005.

[9] L. Jen and Y.-H. Lin, “A Brief Overview of the Accuracy of Classification Algorithms for Data Prediction in Machine Learning Applications,” *J. Appl. Data Sci.*, vol. 2, no. 3, pp. 84–92, 2021, doi: 10.47738/jads.v2i3.38.

[10] Y. Ding, R. Korolov, W. (Al) Wallace, and X. (Cara) Wang, “How are sentiments on autonomous vehicles influenced? An analysis using Twitter feeds,” *Transp. Res. Part C Emerg. Technol.*, vol. 131, no. July, p. 103356, 2021, doi: 10.1016/j.trc.2021.103356.

[11] A. Sopandi, “Soil Infiltration Rate Impact on Water Quality Modeled Using Random Forest Regression,” *J. Appl. Data Sci.*, vol. 2, no. 4, pp. 135–142, 2021, doi: 10.47738/jads.v2i4.42.

[12] S. J. Lee, Z. Xu, T. Li, and Y. Yang, “A novel bagging C4.5 algorithm based on wrapper feature selection for supporting wise clinical decision making,” *J. Biomed. Inform.*, vol. 78, pp. 144–155, 2018, doi: 10.1016/j.jbi.2017.11.005.

[13] J. B. Raja and S. C. Pandian, “PSO-FCM based data mining model to predict diabetic disease,” *Comput. Methods Programs Biomed.*, vol. 196, no. November, p. 105659, 2020, doi: 10.1016/j.cmpb.2020.105659.

[14] A. S. Neogi, K. A. Garg, R. K. Mishra, and Y. K. Dwivedi, “Sentiment analysis and classification of Indian farmers’ protest using twitter data,” *Int. J. Inf. Manag. Data Insights*, vol. 1, no. 2, p. 100019, 2021, doi: 10.1016/j.jjimei.2021.100019.

[15] Q. Aini, “Classification of Tweets Causing Deadlocks in Jakarta Streets with the Help of Algorithm C4.5,” *J. Appl. Data Sci.*, vol. 2, no. 4, pp. 143–156, 2021, doi: 10.47738/jads.v2i4.43.

[16] L. J. M. Milon, I. Sani, S. Usman, and S. Islam, “Predictive Data Mining Models for Novel Coronavirus (COVID - 19) Infected Patients ’ Recovery,” *SN Comput. Sci.*, vol. 1, no. 4, pp. 1–7, 2020, doi: 10.1007/s42979-020-00216-w.

[17] W. W. Damopolii, N. Priyasadie, and A. Zahra, “Educational Data Mining in Predicting Student Final Grades,” *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 10, no. 1, pp. 366–371, 2021, doi: 10.30534/ijatcse/2021/521012021.

[18] G. Gao, Z. Bao, J. Cao, A. K. Qin, T. Sellis, and Z. Wu, “Location-centered house price prediction: A multi-task learning approach,” *arXiv*, vol. 2019, no. January, pp. 1–14, 2019, doi: 10.48550/arXiv.1901.01774.

[19] M. Tsiakmaki and C. Pierrakeas, “11th International Conference on Information,

Intelligence, Systems and Applications, IISA 2020," *11th Int. Conf. Information, Intell. Syst. Appl. IISA* 2020, pp. 1–6, 2020.

[20] S. Thobani, "Improving E-Commerce Sales Using Machine Learning," *Massachusetts Inst. Technol*., vol. 1, no. 1988, pp. 1–176, 2018.

[21] L. Drevin, H. A. Kruger, and T. Steyn, "Value-focused assessment of ICT security awareness in an academic environment," *Comput. Secur*., vol. 26, no. 1, pp. 36–43, 2007, doi: 10.1016/j.cose.2006.10.006.

[22] H. Yuan, W. Xu, Q. Li, and R. Lau, "Topic sentiment mining for sales performance prediction in e-commerce," *Ann. Oper. Res*., vol. 270, no. 1–2, pp. 553–576, 2018, doi: 10.1007/s10479-017-2421-7.

[23] E. C. Garza, "Revenue Optimization for a Hotel Property with Different Market Segments: Demand Prediction, Price Selection and Capacity Allocation," 2017.

[24] A. Pickens and S. Sengupta, "Benchmarking Studies Aimed at Clustering and Classification Tasks Using K-Means , Fuzzy C-Means and Evolutionary Neural Networks," *Mach. Learn. Knowl. Extr*., vol. 3, no. 3, pp. 695–719, 2021, doi: 10.3390/make3030035.

[25] B. Sharma, P. Pokharel, and B. Joshi, "User Behavior Analytics for Anomaly Detection Using LSTM Autoencoder-Insider Threat Detection," *ACM Int. Conf. Proceeding Ser*., vol. 2020, no. 5, pp. 1-9, 2020, doi: 10.1145/3406601.3406610.

[26] A. Rafi, A. Musdholifah, and R. Wardoyo, "Identifying student learning styles using support vector machine in felder-silverman model," *J. Appl. Data Sci*., vol. 5, no. 3, pp. 1495–1507, 2024, doi: 10.47738/jads.v5i3.337.

[27] N. A. Mansour, A. I. Saleh, M. Badawy, and H. A. Ali, "Accurate detection of Covid - 19 patients based on Feature Correlated Naïve Bayes (FCNB) classification strategy", no. 0123456789. *Springer Berlin Heidelberg*, vol. 13, no. January, pp. 41–73, 2021. doi: 10.1007/s12652-020-02883-2.

[28] A. Faizah, "Implementation of the Convolutional Neural Network Method to Detect the Use of Masks," *IJIIS Int. J. Informatics Inf. Syst*., vol. 4, no. 1, pp. 30–37, 2021, doi: 10.47738/ijiis.v4i1.75.

[29] D. Arunachalam, N. Kumar, and J. P. Kawalek, "Understanding big data analytics capabilities in supply chain management: Unravelling the issues, challenges and implications for practice," *Transp. Res. Part E Logist. Transp. Rev*., vol. 114, no. June, pp. 416–436, 2018, doi: 10.1016/j.tre.2017.04.001.

[30] P. Dabhade, R. Agarwal, K. P. Alameen, A. T. Fathima, R. Sridharan, and G. Gopakumar, "Educational data mining for predicting students' academic performance using machine learning algorithms," *Mater. Today Proc*., vol. 47, part. 15, pp. 5260-5267, 2021, doi: 10.1016/j.matpr.2021.05.646.