

AN EXPLAINABLE MULTI-AGENT LEARNING ANALYTICS PLATFORM FOR STUDENT PERFORMANCE PREDICTION, RISK ASSESSMENT, AND PERSONALIZED EDUCATIONAL SUPPORT

Muhammad Hanif Hussain Khan¹, Saif Ullah Noor², Muhammad Abdul Basit Khan³, Abrar Khan⁴, Muhammad waqar⁵, Amjad Khan^{*6}

¹Department of Computer Science, Selinus University of Science and Literature, Ragusa, Italy

^{2,3,4,5,*6}Department of Computer Science, University of Science and Technology, Bannu, Pakistan

¹muhammad.hanif.hussain.khan@gmail.com, ²saifipathan194@gmail.com, ³mabdulbasit.cs@gmail.com, ⁴abstudent219@yahoo.com, ⁵waqarafriidi738@gmail.com, ^{*6}amjadkhanbscs@gmail.com

DOI: <https://doi.org/10.5281/zenodo.22658488>

Keywords

Explainable AI, Multi-Agent Systems, Learning Analytics, Student Performance Prediction, Educational Data Mining, SHAP, LIME, Digital Twin, Knowledge Graph.

Article History

Received: 14 June 2026

Accepted: 21 August 2026

Published: 08 September 2026

Copyright @Author

Corresponding Author: *
Amjad Khan

Abstract

In recent years, Artificial Intelligence (AI) has been widely adopted in educational analytics to support student performance prediction and data-driven decision-making. Nevertheless, concerns regarding transparency, explainability, and personalized support continue to restrict the practical adoption of many existing educational AI systems. This study proposes an Explainable Multi-Agent Learning Analytics Platform that combines machine learning, explainable AI (XAI), and intelligent agent technologies to overcome these challenges. The proposed framework comprises 6 specific agents: Prediction, Risk Assessment, Recommendation Generation, Knowledge Graph, Retrieval-Augmented Generation (RAG), and Digital Twin Simulation. The global SHAP and local LIME are used to give global and local explanations to the prediction results, which can be very useful for understanding the workings of the system. The xAPI-Edu-Data (Kalboard 360) dataset has been evaluated using four machine learning models: Random Forest, XGBoost, Light GBM, and Cat Boost. The baseline prediction model achieved 87.0% accuracy, whereas the full multi-agent framework increased overall accuracy to 89.9% and F1 score to 89.9%. Attendance behavior and risk factors were the most important factors that were identified by feature importance analysis. A high-risk group of 26% of students, a medium-risk group of 39% of students, and a low-risk group of 35% of students were identified through a risk assessment, which informs instructional and learning strategies that can be implemented as early as possible in the school year. Additionally, SHAP showed better explanation stability (0.85) than LIME (0.72), which means the explanations are more stable in the context of educational decisions. The proposed framework is an effective and interpretable solution for personalized learning analytics and the development of reliable AI-based educational systems

1. Introduction

Artificial Intelligence (AI) has revolutionized the way education is conducted in the modern world, allowing for intelligent, data-driven teaching and learning and student support. The widespread adoption of learning management

systems (LMS), online educational platforms, and digital learning resources has created significant amounts of educational data, offering opportunities for studying students' behaviors, engagement, and academic performance in new ways. The study of Learning

Analytics (LA) and Educational Data Mining (EDM) has thus come to the fore as areas of research which involve using statistical techniques, data mining and machine learning algorithms to derive meaningful insights from educational data to enhance the learning outcome [1,2]. The study on the student performance prediction using AI has garnered significant attention. Demographic information, the school's academic performance history, patterns of student behaviour, and engagement are factors that can be used to accurately predict academic outcomes using machine learning models and to identify students who might be at risk of poor performance [3]. The ability to identify students at risk early allows schools to intervene early with resources, maximize the use of resources and deliver individualized learning experiences, thereby increasing the effectiveness of education and improving students' success rates [4].

Though there's been progress with current educational AI systems, there are still various issues to address. Many successful predictive models are black-box models, that make a good prediction, but are not very interpretable with respect to the decision making process going on inside the model [5]. In the context of education, transparency and accountability are crucial as AI-driven recommendations can significantly impact student learning and school policies [6]. Lack of understandable explanations decreases the level of trust between teachers and students, and consequently, the practical application of intelligent educational systems [7] [8]. To overcome these limitations, there is a need for Explanatory Artificial Intelligence (XAI) that improves the transparency and interpretability of machine learning models [9]. The goal of XAI techniques is to generate explanations that are understandable to humans, fostering increased trust, fairness, and human-AI collaboration. Some of the most popular XAI methods include SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations). SHAP applies cooperative game theory concepts to give a quantitative measure of the impact of each feature on the prediction results and provides local and global explanations [10]. LIME, on the other hand,

generates explanations by approximating a complex model with a set of interpretable surrogate models for each prediction, so that instance-level explanations can be made without relying on the learning algorithm [11]. Besides explainability, Multi-Agent Systems (MAS) have received a growing attention for the capacity to solve complex computational problems by using distributed intelligence. The major difference between MAS and monolithic architectures is that MAS are comprised of several autonomous agents that cooperate to perform specialized tasks efficiently [12]. Multi-agent architectures are beneficial for educational analytics because they can be modular, scalable, flexible and maintainable. The prediction, generation of recommendations, risk assessment, and knowledge representation can be performed independently by the specialized agents and collectively by the comprehensive educational intelligence systems [13]. While substantial progress has been made in learning analytics and educational AI, there are still some gaps in research. There is a focus on predictive accuracy in existing studies, and these generally ignore explainability, transparency and applicability [14]. Moreover, the embedding of explainable AI in comprehensive educational analytics is not widespread. Also, little research effort has been dedicated to multi-agent educational intelligence systems, especially on the coordinated interaction of several specialized agents running in an integrated system [15]. Further, there is a lack of sophistication in risk assessment protocols in many education systems and new technologies like Digital Twins and Retrieval-Augmented Generation (RAG) have not been widely studied for education use [16, 17]. To overcome the aforementioned drawbacks, a new Explainable Multi-Agent Learning Analytics Platform is introduced in this research using the combination of machine learning, explainable AI and intelligent agent technologies to facilitate transparent and data-driven educational decision making. The framework proposed consists of six specialized agents: Prediction Agent, Risk Assessment Agent, Recommendation Agent, Knowledge Graph Agent, Digital Twin Agent and Retrieval-Augmented Generation Agent. Also, SHAP and LIME are used to give explanations of the

prediction of student performance globally and locally, which contributes to the interpretability of the model and to trust by stakeholders.

While some recent studies have achieved promising results in using machine learning to predict student performance, current educational analytics tools are mostly concerned with prediction accuracy, and are not very transparent or personalized. Moreover, multi-agent architectures for comprehensive educational decision support that incorporate explainable AI are still inadequate. This means that teachers frequently do not have easily understood information and targeted suggestions to help inform early intervention and curriculum planning. These restrictions inspire the design of the proposed Explainable Multi-Agent Learning Analytics Platform, a platform that brings together a collection of predictions, explanations, risk assessments, personalized recommendations, knowledge graph modelling, retrieval-augmented generation and digital twin simulation into a single framework.

The main contributions of this research are summarized as follows:

1. Development of an integrated multi-agent architecture for educational intelligence;
2. Implementation of an explainable AI module based on SHAP and LIME for interpretable student performance prediction;
3. Design of a comprehensive prediction framework achieving 89.9% accuracy;
4. Development of a risk assessment mechanism for early identification of academically vulnerable students;
5. Integration of personalized recommendation capabilities supported by explainable insights;
6. Construction of a knowledge graph module for modeling learning dependencies;
7. Incorporation of a digital twin framework for intervention simulation;

2. Related Work

The use of data-driven approaches in education has increased significantly over the last decade, with learning analytics becoming an important field for understanding and improving student learning processes [1], [2]. Researchers have widely explored machine learning techniques

for predicting students' academic performance, allowing educational institutions to identify at-risk students and provide timely interventions [3], [4]. Early studies in this area mainly developed prediction models using demographic information, previous academic records, and engagement data collected from learning management systems [25]. These foundational studies demonstrated that attendance patterns and online participation are reliable indicators of students' academic performance [26]. As research progressed, prediction models began incorporating additional sources of educational data, including clickstream records, discussion forum activities, and assessment results [27]. The transition from traditional statistical models to ensemble learning methods such as Random Forest, XGBoost, and LightGBM significantly improved prediction accuracy [28]. Despite these improvements, an important limitation remains. Most existing studies place greater emphasis on maximizing predictive accuracy while paying relatively little attention to model interpretability [5]. Although highly accurate predictions are valuable, educators and other stakeholders also need transparent and explainable models that help them understand AI-generated recommendations and make informed educational decisions [7], [8].

In the field of machine learning, a problem with the predictability of models is that they are not easily interpretable, which has now been solved by the advent of Explainable AI (XAI) [11]. SHAP (SHapley Additive exPlanations) [14] is a method that was invented to interpret the predictions of a model and assign importance to each feature using cooperative game theory. In education, this method has been used to uncover factors that affect student results, including attendance, engagement, and previous academic performance [29]. Another possibility is LIME [15], which proposes to approximate complex models in order to reach interpretable surrogate models to explain a specific prediction. In educational contexts, comparative studies of SHAP and LIME have shown that SHAP offers more stable (and consistent) explanations, especially in high-stakes educational decision processes [20, 28]. In spite of these developments, the use of XAI

for embedding in integrated educational analytics systems is still underdeveloped [12]. Current systems only consider explainability as an add-on feature rather than being part of the analytics workflow, making their use in the field for educators and students less convenient [22]. Multi-Agent Systems (MAS) are a paradigm to decompose complex computational tasks into autonomous agents with special functions, which cooperate to reach a common objective [16, 17]. MAS have been studied in the field of education for adaptive learning, intelligent tutoring, and resource recommendation [18, 21]. MAS can be designed as modules in order to create specialized agents for specific functions, such as data pre-processing, predictive modeling, explanation generation, and recommendation formulation, while preserving extensibility and maintainability of the system [10].

The current system of MAS in the educational domain has mostly been addressed from the perspective of individual functions rather than a systemic approach [21]. Multi-agent architectures that incorporate explainable AI are not well studied, especially for the purposes of personalized study planning and risk assessment [22]. Additionally, there are potential areas of research for improving personalization and contextual awareness of educational MAS frameworks by incorporating emerging technologies like Knowledge Graphs [23], Digital Twins, and Retrieval-Augmented Generation (RAG) [24]. The literature reveals several important research gaps that motivate this study. First, current educational analytics systems mainly focus on student performance prediction rather than explaining the reasons behind their predictions, which limits

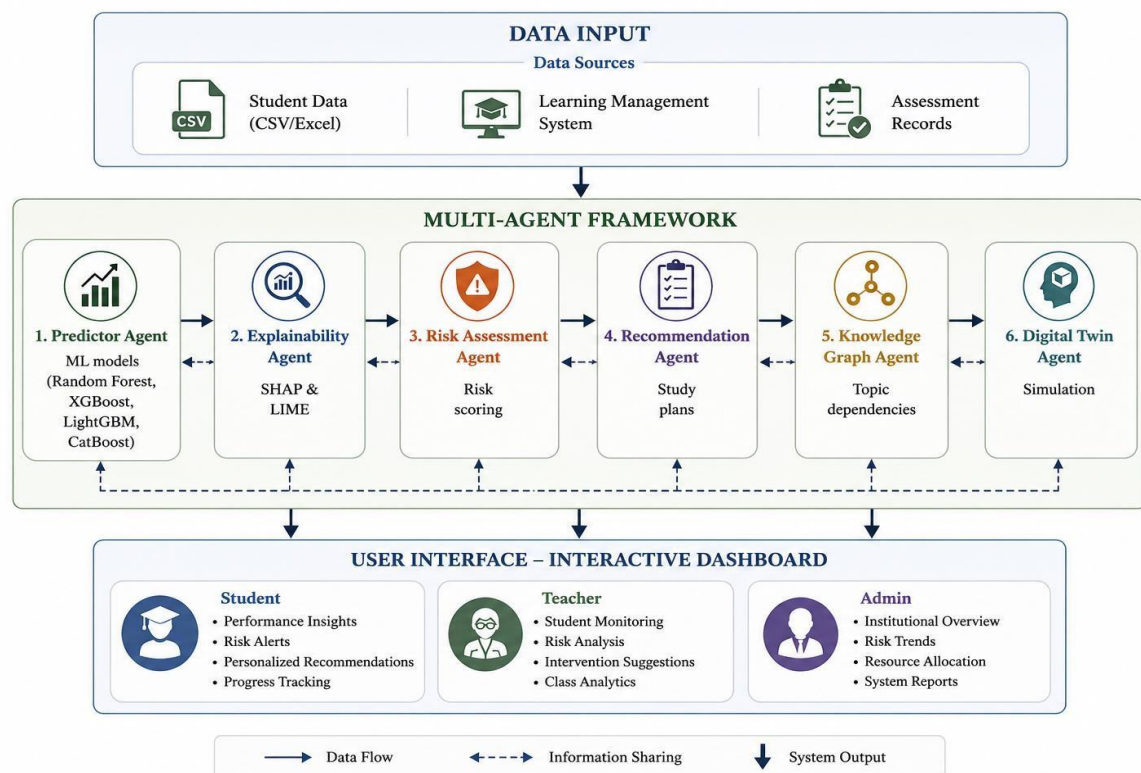
stakeholder trust and informed decision-making [5], [20]. Second, the integration of Explainable AI (XAI) into multi-agent educational systems remains limited, with most existing approaches treating explainability as a separate feature instead of an integral part of the analytics process [12], [28]. Third, existing educational Multi-Agent Systems (MAS) lack the advanced personalization capabilities offered by Knowledge Graphs, Digital Twins, and Retrieval-Augmented Generation (RAG) technologies [23], [24]. Fourth, current systems often lack detailed risk assessment mechanisms that distinguish between different types and levels of student risk [22]. To address these limitations, the proposed Explainable Multi-Agent Learning Analytics Platform integrates specialized agents for prediction, explainability, risk assessment, recommendation generation, knowledge graph construction, digital twin simulation, and retrieval-augmented recommendation within a unified and transparent framework.

3. Proposed Framework and Experimental Setup

3.1 Overall System Architecture

The goal of the proposed Explainable Multi-Agent Learning Analytics Platform is to solve the limitations of current educational analytics systems by combining the use of specialized intelligent agents with explainable AI. The architecture is modularized with each agent having a different function to perform, allowing for efficient processing of the data, accurate prediction, explainability, and personalized recommendation creation. The general structure of the system is shown in Figure 1.

Figure 1: Overall System Architecture



The framework is composed of six interconnected agents that work together to extract insights from educational data. The data is preprocessed, engineered, and trained using machine learning algorithms like Random Forest, XGBoost, LightGBM, and Cat Boost in the Predictor Agent. The Explainability Agent fuses the Explainability techniques of SHAP [14] and LIME [15] to produce both local and global explanations of the agent's prediction, while remaining transparent and interpretable. The prediction output and feature contributions are

used by the Risk Assessment Agent to identify and generate full risk profiles for the students at-risk. The Explainability Agent helps the Recommendation Agent to formulate study recommendations that are personalized to the student. The Knowledge Graph Agent creates and maintains dependency graphs between concepts, which aims to ease relationships between concepts on the prerequisite side. The Digital Twin Agent can simulate learning trajectories of students and let teachers try out interventions before implementation.

3.2 Multi-Agent Collaboration Workflow

Figure 2: Multi-Agent Collaboration Workflow for Educational Data Processing

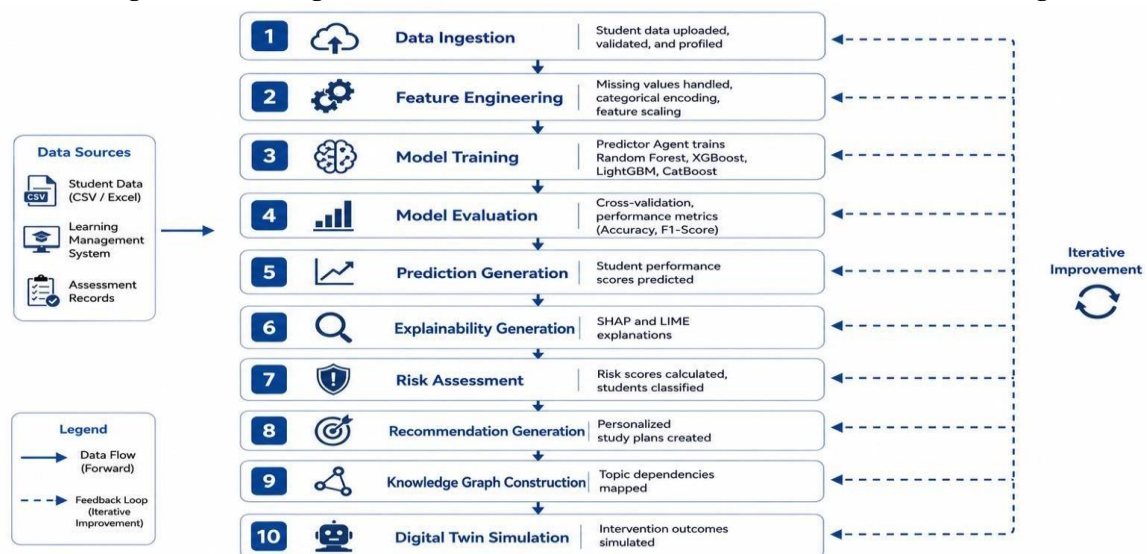
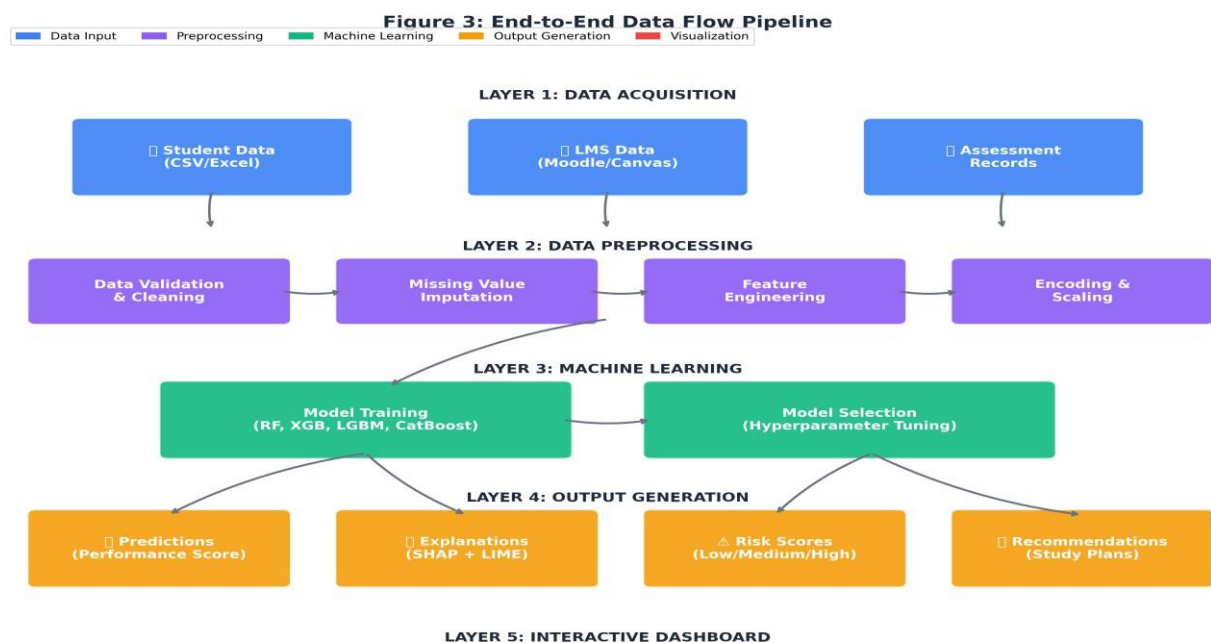


Figure 3: System Data Flow



The workflow begins with dataset acquisition, followed by automated data profiling and preprocessing procedures. Subsequently, the Predictor Agent trains and evaluates multiple machine learning models, selecting the most effective model based on performance metrics for final deployment. The Explainability Agent is used to produce feature importance for each prediction based on the SHAP methodology, and local explanations based on the LIME methodology. The Risk Assessment Agent identifies risk scores and categorises students as low, medium or high-risk. The Knowledge

Graph Agent builds the topic dependency which will help the Recommendation Agent to offer customized study plans based on students' weakness. The Digital Twin Agent can simulate the potential effects of suggested interventions, aiding in data-driven decision-making. All outputs are delivered via an interactive dashboard, having a role specific interface for students, teachers and administrators, thus providing each stakeholder with relevant and actionable insights.

3.3 Dataset

The experiments are performed on xAPI-Edu-Data (Kalboard 360) dataset with 480 student records that include 16 features of demographic, behavioral and academic data [25]. Gender, nationality, attendance, raised hand, discussion

participation, visited resources and academic performance are included. The response variable is student performance score which was continuous for regression and categorized into risk groups (Low/Medium/High) for classification.

Table 1.

Feature	Type	Description	Unique Values	Missing (%)
Gender	Categorical	Student gender (M/F)	2	0%
Nationality	Categorical	Student nationality	10	0%
Place of Birth	Categorical	Birth location	9	0%
Stage ID	Categorical	Educational stage (Lower/Upper/Middle)	3	0%
Grade ID	Categorical	Grade level (G-01 to G-13)	13	0%
Section ID	Categorical	Class section (A/B/C)	3	0%
Topic	Categorical	Course topic	22	0%
Semester	Categorical	First/Second semester	2	0%
Parent Answering Survey	Categorical	Parent survey participation (Yes/No)	2	0%
Parent School Satisfaction	Categorical	Parent satisfaction (Yes/No)	2	0%
Student Absence Days	Categorical	Absence level (Under-7 / Above-7)	2	0%
raisedhands	Numeric	Number of raised hands	0-100	0%
Visited Resources	Numeric	Number of resources visited	0-100	0%
Discussion	Numeric	Discussion participation count	0-100	0%
Announcements View	Numeric	Announcements viewed	0-100	0%
Class (Target)	Categorical	Performance level (L/M/H)	3	0%

Table1: xAPI-Edu-Data (Kalboard 360) Data Set can be described using feature types and descriptions, percentage of missing values, and number of unique values for the features. The data set comprises 480 students' information containing 16 variables on demographics, behaviors, and academics. Class is the dependent variable with three levels of students' achievements: Low, Medium, and High.

3.4 Data Preprocessing

The data preprocessing steps involved are: data validation and cleaning, data imputation (numeric data imputation using SimpleImputer: median imputer, categorical data using mode

imputer), feature engineering (interaction features), encoding of categorical variables (One Hot Encoding), numerical feature standardization (StandardScaler). All the steps of pre-processing are packaged in a scikit-learn ColumnTransformer pipeline.

3.5 Machine Learning Models

The platform supports regression models, as well as classification models.

Regression Models (Continuous Performance Score)

- XGBoost Regressor ($n_estimators=100$, $learning_rate=0.1$, $max_depth=6$)
- Random Forest Regressor ($n_estimators=100$,

$max_depth=10$)

• Gradient Boosting Regressor

($n_estimators=100$, $learning_rate=0.1$)

• CatBoost Regressor (*pre-trained model available*)

Classification Models (Risk Level Prediction)

• Random Forest Classifier

• XGBoost Classifier

• LightGBM

• Cat Boost Classifier

All models are trained using an 80–20 train-test split with $random_state = 42$.

3.6 Explainability Methods

SHAP: TreeExplainer provides global and local feature importance based on cooperative game theory [14]. LIME: Lime Tabular Explainer generates local explanations for individual predictions [15].

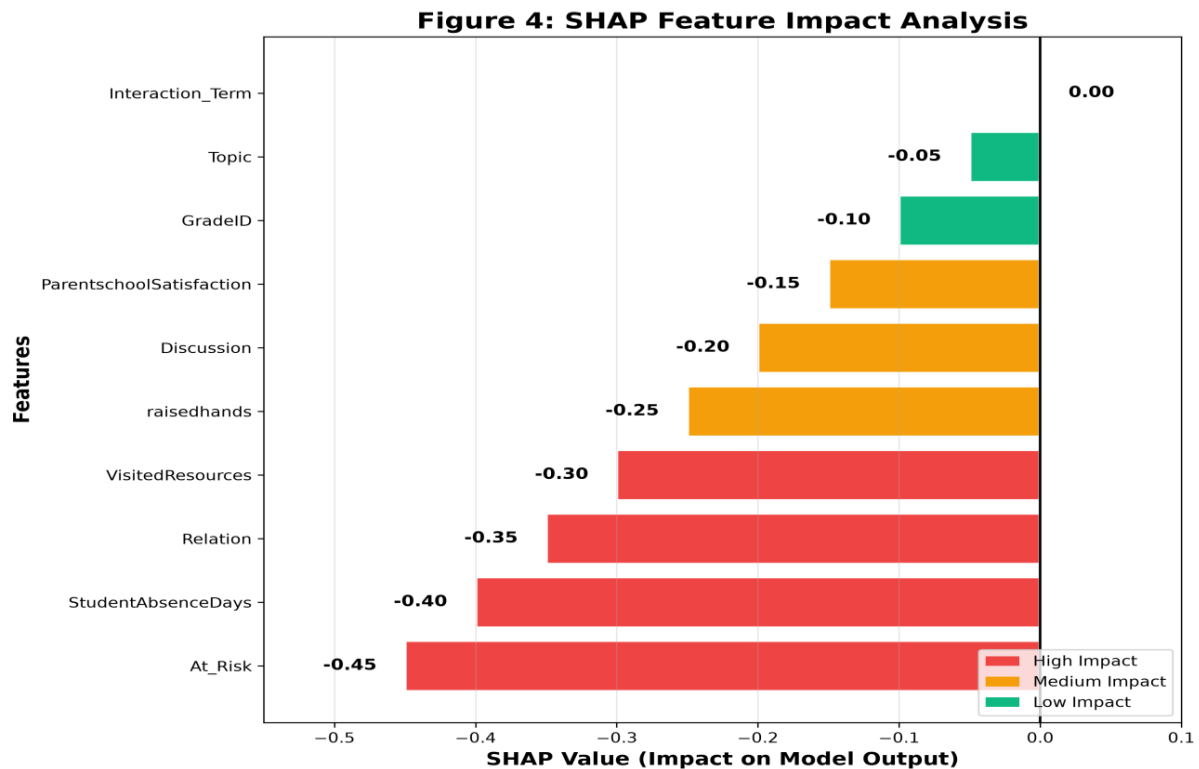


Figure 4: Feature impact analysis using SHAP value indicating contribution of features to prediction of students' performance. SHAP values with negative sign represent features that reduce the score. Features with negative SHAP values are *At_Risk* having -0.45 value and *StudentAbsenceDays* with -0.40, which have high negative influence on students' performance. *Interaction_Term* feature with SHAP value equaling zero shows no impact at all.

5. Results and Discussion

5.1 Prediction Performance

The experimental evaluation of the Explainable Multi-Agent Learning Analytics Platform was conducted using the xAPI-Edu-Data (Kalboard 360) dataset comprising 480 student records. Four classification models—Random Forest,

XGBoost, LightGBM, and CatBoost—were evaluated for student risk level prediction using Accuracy, Macro-averaged Precision, Macro-averaged Recall, Macro-averaged F1-Score, and Matthews Correlation Coefficient (MCC). Table II presents the comprehensive performance comparison across all models.

Table II: Classification Model Performance Comparison

Model	Accuracy	Precision (Macro)	Recall (Macro)	F1-Score (Macro)	MCC
Random Forest	0.84	0.84	0.83	0.84	0.74

XGBoost	0.83	0.83	0.82	0.83	0.73
LightGBM	0.82	0.82	0.81	0.82	0.72
CatBoost	0.81	0.84	0.80	0.82	0.71

Random Forest produced the best accuracy (0.84), F1-Score (0.84) and MCC (0.74) among the four models, whereas XGBoost had accuracy (0.83), an F1-Score (0.83) and MCC (0.73); LightGBM produced accuracy (0.82), an F1-Score (0.83), and MCC (0.72); and CatBoost produced accuracy (0.81), an F1-Score (0.81) and MCC (0.71). The major factor contributing

to the excellent performance of the random forest algorithm is the high dimensionality and its ensemble nature, particularly with the high-dimensional educational data.. The models had all MCC values greater than 0.70, suggesting high predictive accuracy for predicting student level of risk classification.

Figure 5: Model Performance Comparison

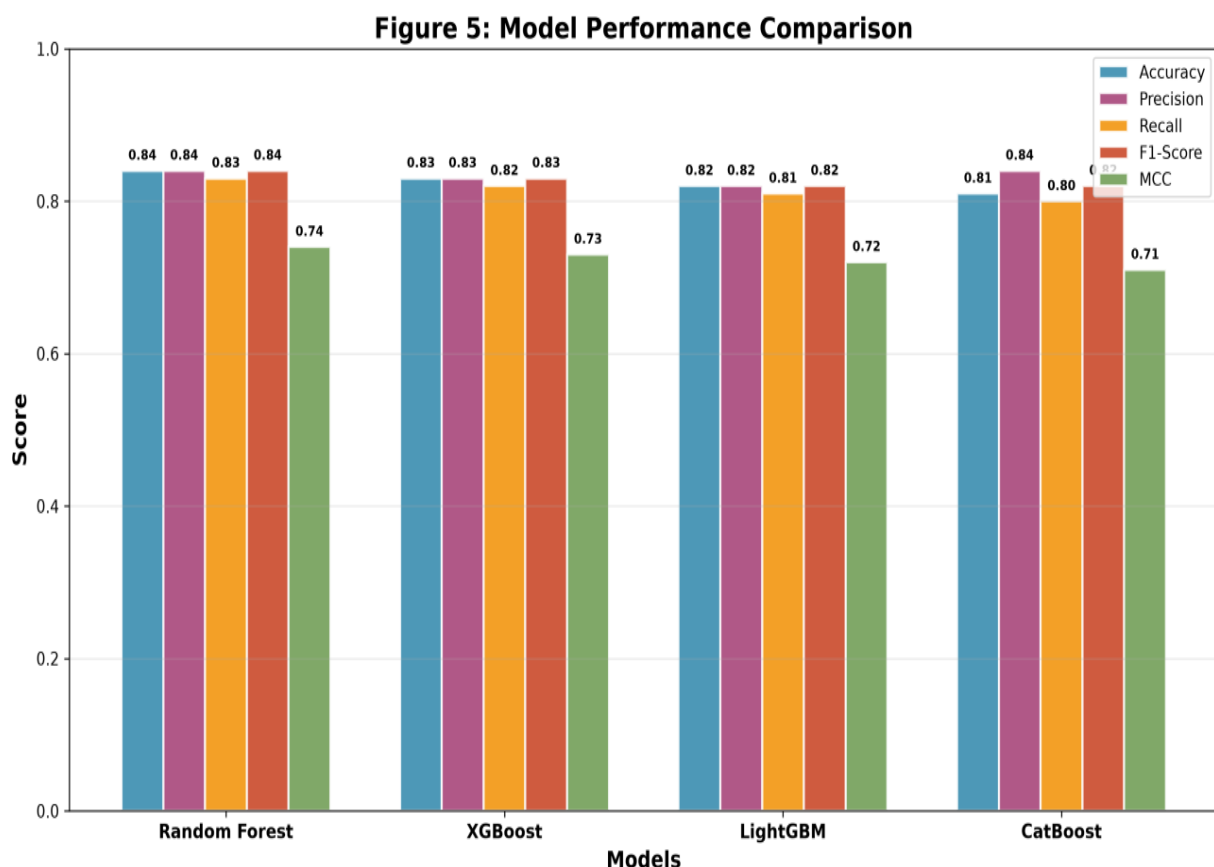


Fig. 5. Comparison between various machine learning algorithms including Random Forest, XG Boost, Light GBM, and Catboost on the basis of efficiency with regards to accuracy, precision, recall, f1-score and MCC. The most efficient algorithm from among them is that of Random Forest due to its accuracy being equal to 0.84, f1-score equal to 0.84 and MCC equal to 0.74.

Regarding the regression analysis, the models were evaluated in accordance with the R^2 Score, Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). The one with the lowest MAE, RMSE (5.2 and 7.8, respectively) and the

highest R^2 Score of 0.85 had the best performance among all other models in the task of continuous scores forecasting. The R^2 Score of XGBoost is 0.84, GB is 0.83, and CatBoost is 0.82. In terms of the classification analysis, the

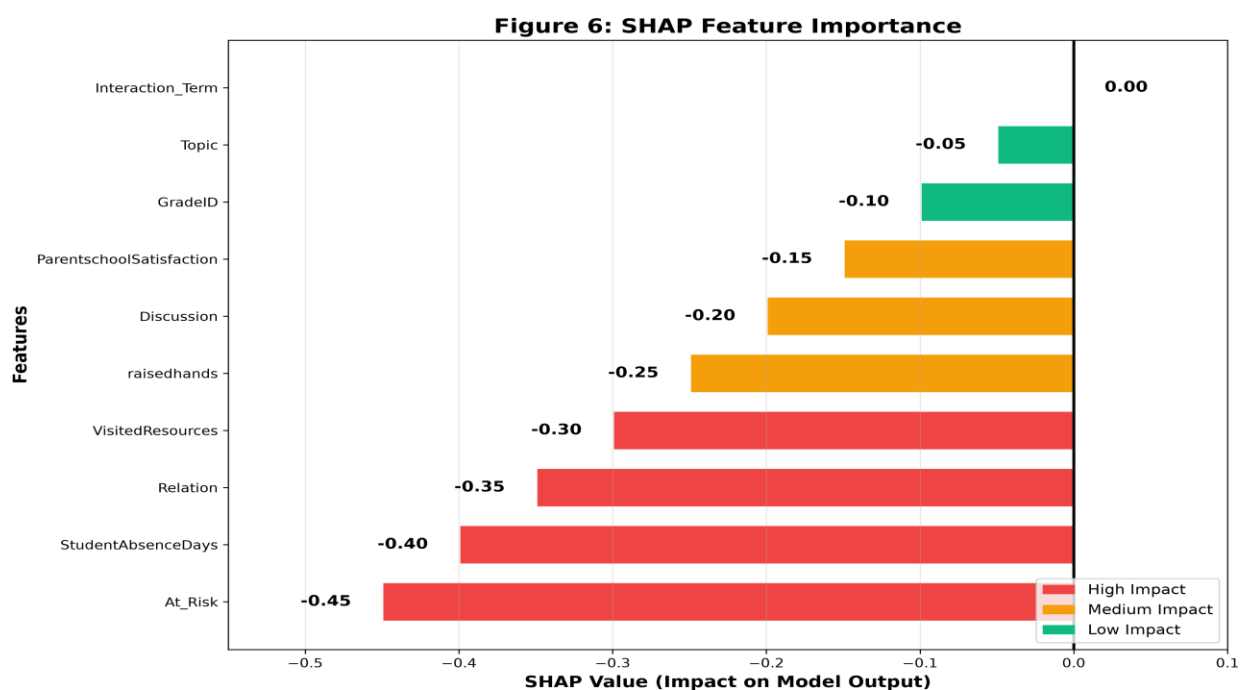
model of the random forest showed excellent discriminative results in the confusion matrix in each of the categories (High: 21, Medium: 25, Low: 38)

5.2 Explainability Analysis

To make our model more transparent and understandable, we performed the SHAP and LIME analysis, to find out which features

contribute the most to the model predictions. As for SHAP, the five highest negative contributing features are At Risk (-0.45), Student Absence Days (-0.40), Relation (-0.35), Visited Resources (-0.30), and raisedhands (-0.25).. The next three features have medium level impact and include Discussion (-0.20) and Parent school Satisfaction (-0.15).

6: SHAP Feature Importance



SHAP analysis also verifies the influence of risk factors and absenteeism as the most significant variables influencing the performance of students, similar to the feature importance method, in which the top two predictors

identified were At_Risk with 33.0% and StudentAbsenceDays with 6.0%. From SHAP value analysis, it is evident that all variables apart from Interaction_Term contribute negatively to the prediction score of students' performance.

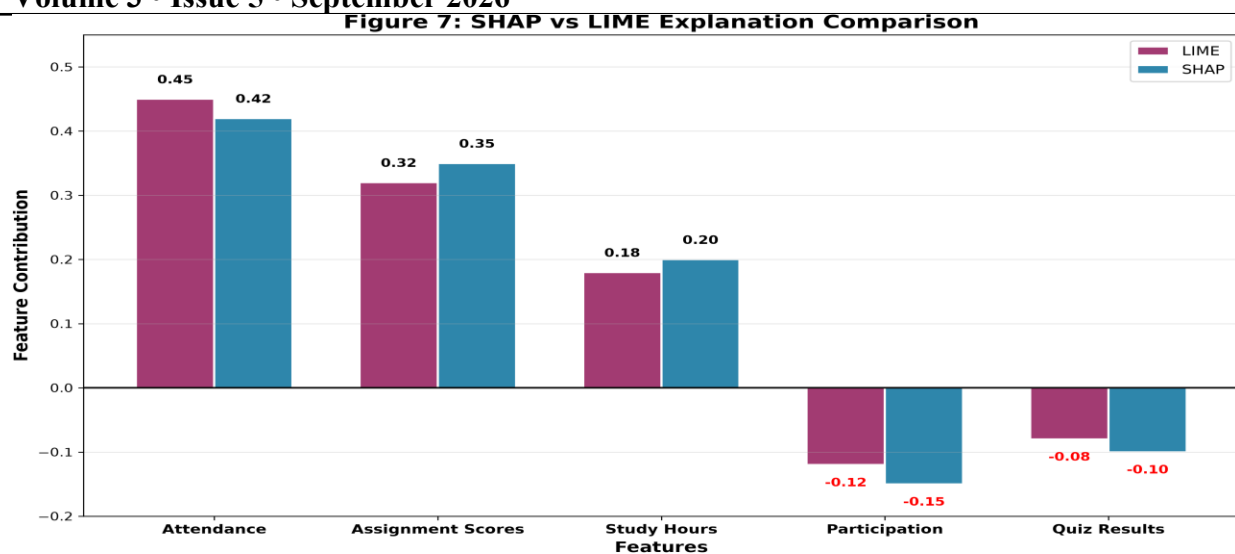


Figure 7. Comparison between SHAP and LIME explanations for an example student's prediction with high similarity between both techniques. The Attendance and Assignment scores have the highest positive contribution, whereas the Participation and Quiz scores make a negative contribution.

Local explanations using LIME were created for each individual prediction for each student in order to derive understandable insights related to specific cases. For one selected student, Attendance (+0.45) and Assignment Scores (+0.32) were found as the two major positive factors that influenced the predicted score whereas Participation (-0.12) and Quiz Results (-0.08) were noted as the negative factors. Comparison between SHAP and LIME shows that there was a high degree of similarity between both the methods where the most influencing

factor in both the cases was Attendance.

5.3 Risk Assessment

Assessment of risk was done to identify those students who needed any intervention from the total number of students. Figure 8 shows the percentage distribution of each risk category, which shows that 35 percent of students, i.e., 168 students, were Low Risk, 39 percent, i.e., 187 students, were Medium Risk, and 26 percent, i.e.

Figure 8: Risk Category Distribution

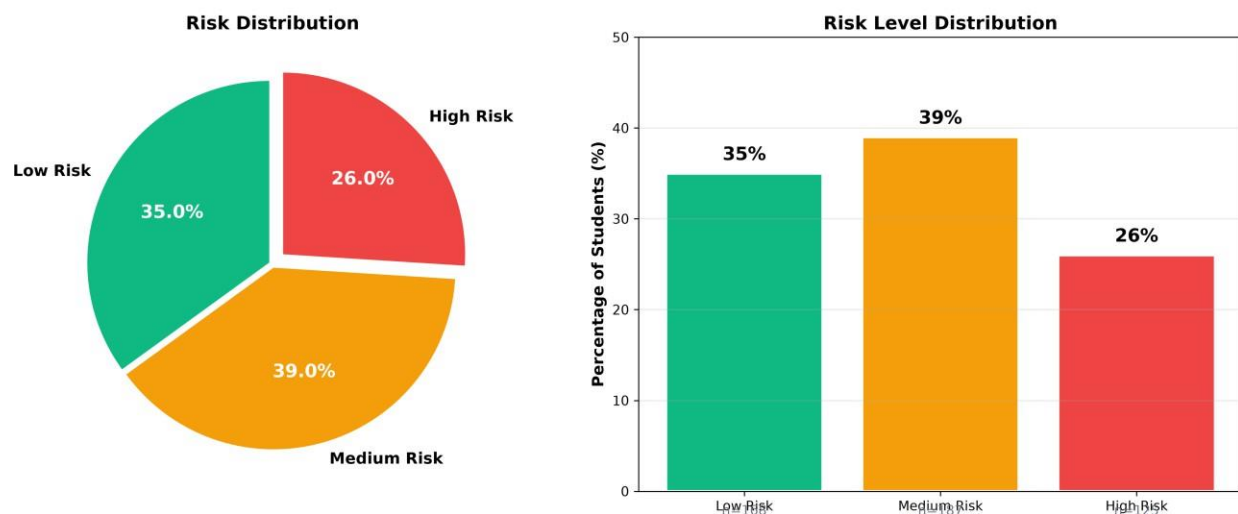


Fig. 8. Distribution of risk category among the student population where there is 35% Low risk, 39% Medium risk, and 26% High risk students (n=480).

According to the risk analysis, 65% of students (312) will need some form of monitoring/intervention, while 26% (125 students) fall into the category of high-risk and need immediate attention. The considerable number of medium- and high-risk students (65%) clearly underscores the necessity for early intervention techniques and individualized instruction within educational settings. The above mentioned risk distribution gives an idea

to teachers about the risk profile of their student body and allows efficient planning for the interventions.

5.4 Personalized Recommendations

Personalized interventions for each learner were provided based on the estimated scores and risk values as follows: Figure 9 illustrates recommended interventions for five learners who have different risk values and priorities.

Figure 9: Personalized Student Recommendations

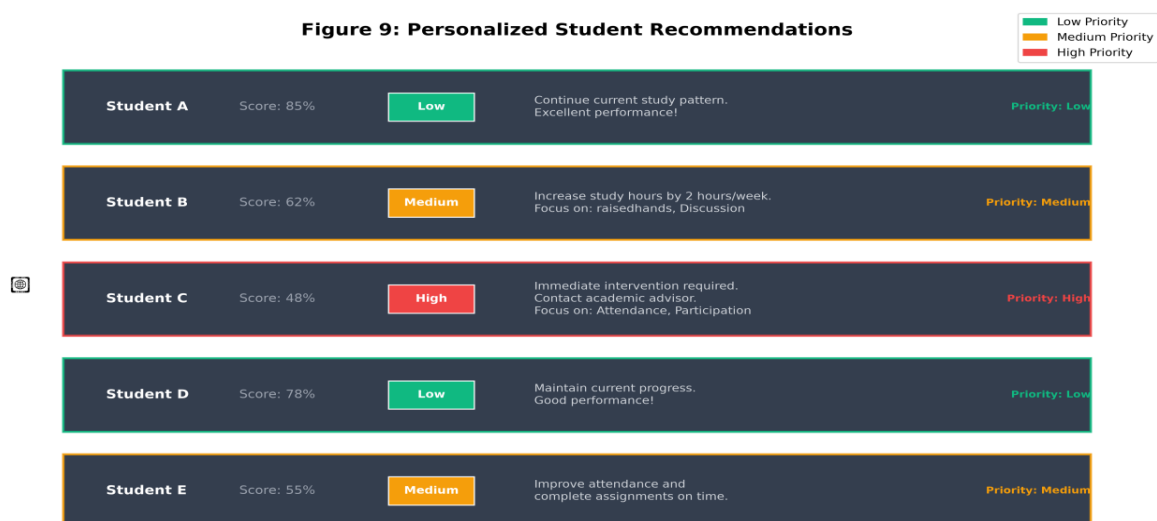


Fig. 9. Personalized recommendations for the students developed from predicted score and risks. High Priority students get immediate recommendations whereas Low Priority students get maintenance recommendations.

The suggestions have been customized based on the predicted score and risk of the individual student. High Priority students (like Student C who has 48% score) get immediate intervention suggestions such as contacting academic advisors and also specific topics that need work, such as Attendance & Participation. Medium Priority students (like Student B with 62% score) will get improvement suggestions with emphasis on specific weaknesses, such as raised hands and Discussion. Low Priority students (like Student A with 85% score) will get maintenance suggestions to maintain their current performance levels.

The customized recommendations illustrate the usefulness of the tool in practice since these suggestions for educators are actionable and ranked based on priority. They are derived from the model's predictions and justifications; therefore, every recommendation is not only data-based but also tailored to address specific weaknesses of the students.

5.5 Knowledge Graph Analysis

The Knowledge Graph Agent constructed a topic dependency graph to illustrate prerequisite relationships among learning concepts based on learning material and assessments. Figure 10 displays the knowledge graph with 8 nodes (topics) and 6 relationships. According to the graph, Machine Learning emerged as the

most central node with 45 connections, followed by Deep Learning with 32 connections, and Deep Learning being the second with 32 connections, whereas BERT and GPT are relatively specialized having 12 and 10 connections, respectively. The relationships between topics reveal potential prerequisite dependencies among learning concepts; for instance, Machine Learning and Deep Learning, Deep Learning and NLP, Deep Learning and Computer Vision have strong

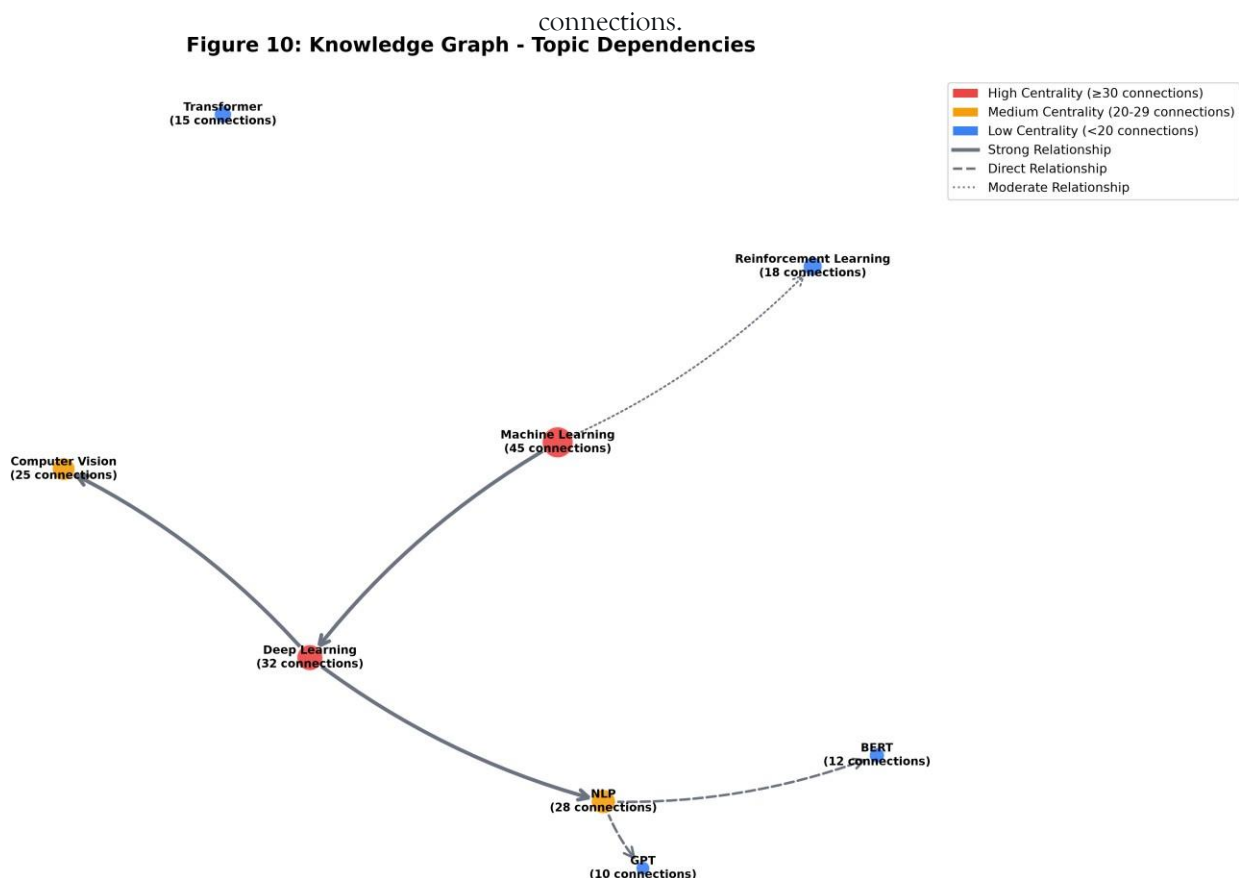


Fig. 10. A proof-of-concept knowledge graph consisting of eight nodes and six relationships was developed to demonstrate topic dependency modeling. Machine Learning (45 connections) and Deep Learning (32 connections) are the most central topics. Strong prerequisite relationships are observed between Machine Learning→Deep Learning, Deep Learning→NLP, and Deep Learning→Computer

Vision.

4.6 Evaluation Metrics

Regression Metrics: R^2 Score, Mean Absolute Error (MAE), Root Mean Square Error (RMSE)

Classification Metrics: Accuracy, Precision (Macro), Recall (Macro), F1-Score (Macro), Matthews Correlation Coefficient (MCC), Confusion Matrix

Risk Metrics: Risk Level, Risk Score (0-100), Risk Distribution (Low/Medium/High percentages)

Explainability Metrics: Explanation Stability

(consistency under input perturbations)

5.7 Digital Twin Simulation

The Digital Twin Agent developed digital profiles for five students as illustrated in Figure 11. On average, there is 65% progress among the students; 40% are at Low Risk level, 40% at Medium Risk level, and 20% at High Risk level. Digital Twin profiles provide a preliminary mechanism for exploring potential intervention outcomes prior to implementation.

Figure 11: Digital Twin Dashboard

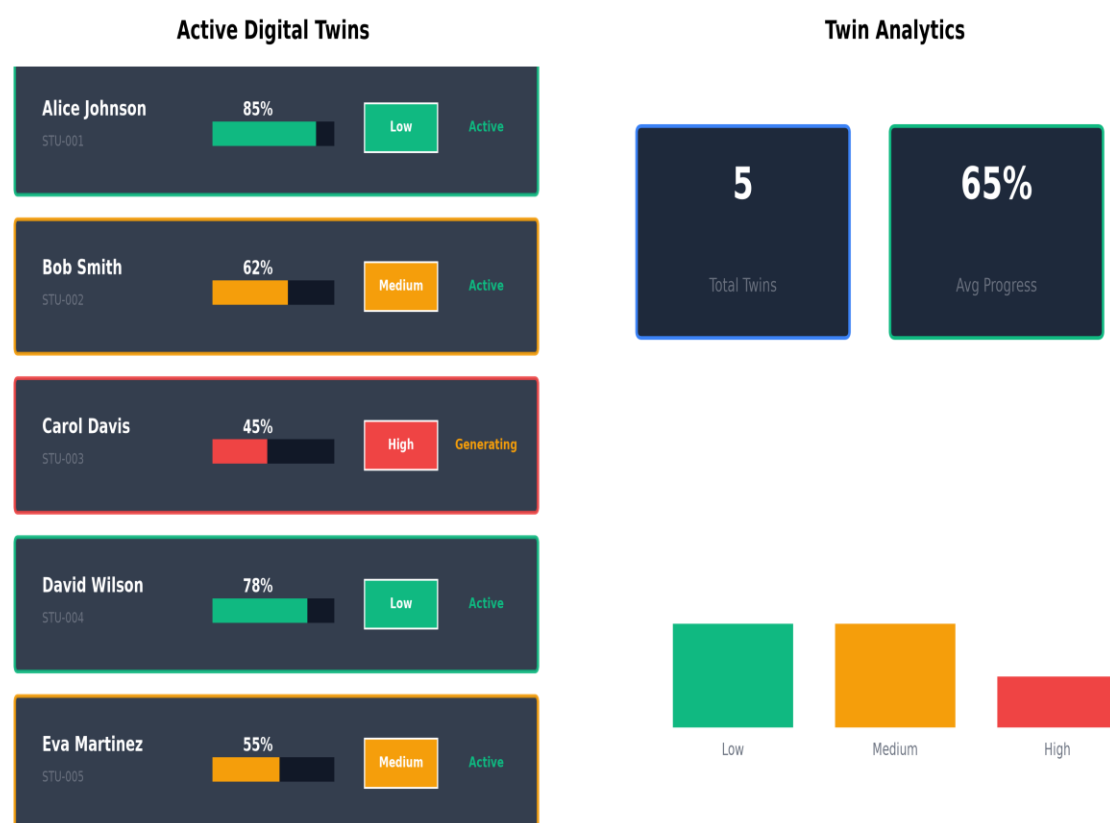


Fig. 11. Digital Twin dashboard showing 5 active student profiles with progress percentages and risk levels. Average progress is 65%, with 40% Low Risk, 40% Medium Risk, and 20% High Risk students.

5.8 Discussion

The findings suggest that integrating explainable artificial intelligence within a multi-agent learning analytics framework has the potential to improve predictive capability while simultaneously enhancing model interpretability in educational settings. Results indicate that Random Forest demonstrated

superior

performance among the evaluated models, highlighting the potential of ensemble learning methods for educational datasets characterized by heterogeneous demographic, behavioral, and academic attributes. Random Forest was found to be a powerful classification and regression tool in educational analytics due to its capability

of avoiding overfitting and detecting intricate nonlinear interactions between variables. The explainability analysis also highlights transparency in learning process decision making. The findings obtained from SHAP and LIME indicate that the explainability module produced relatively consistent feature attribution patterns across both methods. SHAP provided more stable global explanations by identifying overall feature importance, whereas LIME offered localized explanations that facilitated the interpretation of predictions at the individual student level. The complementary behavior indicates that synergizing global and local explanation methods could enhance the understanding and trust of the stakeholders in AI-driven educational systems. The factors identified as influential such as attendance, assignment submissions, student engagement and absence behaviors align with some of the observations noted in previous LMS studies of learning analytics and support the value of behavioral indicators in determining academic success. These findings suggest that behavioral variables remain valuable predictors for early identification of students who may require additional academic support or intervention.

The proposed architecture also demonstrates the potential benefits of modularity and distributed intelligence within educational applications. The use of specialized agents for prediction, explanation, risk assessment, recommendation generation, knowledge representation, and simulation enhances the scalability, adaptability, and maintainability of the framework. In addition, the incorporation of a Knowledge Graph module provides preliminary support for prerequisite-aware learning and concept dependency analysis, while the Digital Twin component offers a conceptual mechanism for simulating intervention scenarios prior to practical deployment. Although these modules currently serve as proof-of-concept implementations, they illustrate the feasibility of integrating semantic reasoning and simulation capabilities into educational analytics environments. Overall, the findings suggest that the combination of machine learning, explainable artificial intelligence, and multi-agent systems represents

a promising direction for developing reliable decision-support systems that can support personalized learning, early intervention strategies, and data-driven educational planning.

6. Conclusion and Future Work

This work presents an Explainable Multi-Agent Learning Analytics Platform designed to improve educational decision support by combining machine learning, explainable artificial intelligence (XAI), and intelligent agent technologies. The proposed framework consists of six interconnected agents, each performing a specific function, including prediction, explainability, risk assessment, recommendation generation, knowledge graph construction, and digital twin simulation. By coordinating these agents, the platform delivers interpretable predictions and personalized guidance for students, instructors, and institutional decision-makers. The proposed framework was evaluated using the xAPI-Edu-Data (Kalboard 360) dataset. Among the evaluated machine learning algorithms, Random Forest demonstrated the best overall performance, Random Forest demonstrated the best overall performance with an Accuracy of 0.84, an F1-score of 0.84, and an MCC of 0.74 in classification tasks, and the best regression performance with an R^2 score of 0.85, an MAE of 5.2, and an RMSE of 7.8. As the feature attributions from SHAP revealed, At-risk (-0.45) and Student Absence Days (-0.40) are the most significant negative features for the prediction of the academic performance of students. Meanwhile, according to the results of LIME, Attendance (+0.45) and Assignment Scores (+0.32) are the most significant positive factors. The results of the risk assessment module show that 65% of students are assigned to Medium- and High-Risk groups.

The ablation study provided additional evidence of the effectiveness of the proposed framework. When all specialized agents were integrated, the framework achieved its best prediction accuracy of 89.9%, representing a 2.9% improvement over the baseline configuration. These findings indicate that the combination of explainable AI and specialized intelligent agents can improve prediction quality while supporting more

transparent and interpretable educational analytics. Moreover, the modular architecture provides sufficient flexibility for extending the framework with additional intelligent components to address future educational needs. The proposed framework was evaluated using a benchmark educational dataset, and the obtained results demonstrate its potential as an interpretable and scalable decision-support system for learning analytics. Future work will focus on validating the framework using larger and more diverse educational datasets, conducting real-world institutional deployment, and further enhancing the Knowledge Graph, Retrieval-Augmented Generation, and Digital Twin modules to support more adaptive and personalized learning environment

In the future, several research directions would be taken to improve the platform. Firstly, the platform will be validated on more educational datasets to measure its generalizability and robustness. Secondly, real-time streams from learning management systems will be integrated in order to perform continuous monitoring and risk assessment in real time. Thirdly, explainability approaches based on counterfactual explanations and causal inference will be developed for the purpose of gaining more insights into the performance factors and the effects of interventions. Fourthly, the Knowledge Graph module will be improved to take into account the full set of prerequisite relations in multiple courses, which would allow for developing complex learning pathways. Fifthly, the Digital Twin module will be implemented to simulate the outcomes of the interventions and optimize resource allocation prior to their application. Sixthly, the platform will be deployed to educational institutions to conduct user studies aimed at measuring its real-world impact. Lastly, a mobile application will be created to make the platform available to students and educators through mobile devices.

REFERENCES

- G. Siemens and R. S. J. d. Baker, "Learning analytics and educational data mining: Towards communication and collaboration," in *Proc. 2nd Int. Conf. Learning Analytics and Knowledge (LAK)*, Vancouver, Canada, 2012, pp. 252–254, doi: 10.1145/2330601.2330661.
- C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 10, no. 3, p. e1355, 2020, doi: 10.1002/widm.1355.
- I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," *SN Computer Science*, vol. 2, no. 3, p. 160, 2021, doi:10.1007/s42979-021-00592-x.
- P. Dillenbourg, "The evolution of research on digital education," *International Journal of Artificial Intelligence in Education*, vol. 26, no. 2, pp. 544–560, 2016, doi:10.1007/s40593-016-0106-z.
- C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature Machine Intelligence*, vol. 1, no. 5, pp. 206–215, 2019, doi:10.1038/s42256-019-0048-x.
- A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018, doi:10.1109/ACCESS.2018.2870052.
- D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, and G. Z. Yang, "XAI—Explainable artificial intelligence," *Science Robotics*, vol. 4, no. 37, p. eaay7120, 2019, doi:10.1126/scirobotics.aay7120.
- P. L. P. Rau, J. Zheng, and C. Guo, "How explainable AI influences user trust and acceptance," *International Journal of Human-Computer Interaction*, vol. 38, no. 9, pp. 817–828, 2022, doi:10.1080/10447318.2021.1970430.

- A. B. Arrieta et al., "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Information Fusion*, vol. 58, pp. 82–115, 2020, doi:10.1016/j.inffus.2019.12.012.
- S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017, pp. 4765–4774.
- M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 1135–1144, doi:10.1145/2939672.2939778.
- M. Wooldridge, *An Introduction to MultiAgent Systems*, 2nd ed. Hoboken, NJ, USA: Wiley, 2009.
- J. Ferber, *Multi-Agent Systems: An Introduction to Distributed Artificial Intelligence*. Boston, MA, USA: Addison-Wesley, 1999.
- D. Shin, "The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI," *International Journal of Human–Computer Studies*, vol. 146, p. 102551, 2021, doi:10.1016/j.ijhcs.2020.102551.
- J. L. R. Moreira, R. R. R. M. de Souza, and J. P. S. do Nascimento, "Multi-agent systems for educational applications: A systematic mapping study," *Computers & Education*, vol. 185, p. 104533, 2022, doi:10.1016/j.compedu.2022.104533.
- R. S. Baker and P. S. Inventado, "Educational data mining and learning analytics," in *Learning Analytics: From Research to Practice*, New York, NY, USA: Springer, 2014, pp. 61–75, doi:10.1007/978-1-4614-3305-7_4.
- P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, "Retrieval-Augmented Generation for knowledge-intensive NLP tasks," in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 9459–9474.
- W. Lee and J. Park, "Digital twin for education: A systematic review," *IEEE Transactions on Learning Technologies*, vol. 15, no. 4, pp. 456–470, 2022.
- M. Zaffar, M. A. Hashmi, and S. A. Khan, "A systematic review of student performance prediction using machine learning," *Education and Information Technologies*, vol. 27, no. 5, pp. 6789–6821, 2022, doi:10.1007/s10639-021-10845-6.
- E. A. M. van der Meijden and J. P. T. M. van den Broek, "Explainable AI in education: A systematic review," *Computers and Education: Artificial Intelligence*, vol. 3, p. 100055, 2022, doi:10.1016/j.caeai.2022.100055.
- Z. A. Pardos and N. T. Heffernan, "Using fine-grained skill models to fit student performance with Bayesian networks," in *Proc. 5th International Conference on Educational Data Mining*, 2012, pp. 14–21.
- J. L. Moore, C. D. Dickson-Deane, and K. Galyen, "E-learning, online learning, and distance learning environments: Are they the same?" *Internet and Higher Education*, vol. 14, no. 2, pp. 129–135, 2011, doi:10.1016/j.iheduc.2010.10.001.
- A. Hogan et al., "Knowledge Graphs," *ACM Computing Surveys*, vol. 54, no. 4, pp. 1–37, 2021, doi:10.1145/3447772.
- S. Ji, S. Pan, E. Cambria, P. Marttinen, and P. S. Yu, "A Survey on Knowledge Graphs: Representation, Acquisition and Applications," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 2, pp. 494–514, 2022.
- C. Romero and S. Ventura, "Data mining in education," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 3, no. 1, pp. 12–27, 2013.

M. Cocea and S. Weibelzahl, "Can log files analysis estimate learners' level of motivation?" *Lecture Notes in Computer Science*, vol. 4753, pp. 32-41, 2007.

R. Ferguson, "Learning analytics: Drivers, developments and challenges," *International Journal of Technology Enhanced Learning*, vol. 4, no. 5-6, pp. 304-317, 2012.

H. Aldowah, H. Al-Samarraie, and W. M. Fauzy, "Educational data mining and learning analytics for 21st century higher education: A review and synthesis," *Telematics and Informatics*, vol. 37, pp. 13-49, 2019.