The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM)
ISSN: 2602-3199

The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 2024

Volume 28, Pages 277-285

ICBASET 2024: International Conference on Basic Sciences, Engineering and Technology

Utilizing the eXtreme Gradient Boosting Algorithm for Artificial Intelligence-supported Learning Analytics Application

Mustafa Cosar
Hitit University

Abstract: Recent technological advancements, including internet-based distance education and artificial intelligence-supported learning analytics, have significantly impacted the field of education. These advancements not only enhance the efficiency of education but also broaden access to learning while mitigating barriers to implementation. AI-supported learning analytics emerges as a pivotal tool for interpreting data gleaned from educational processes and stakeholders, thereby enhancing educational processes and outcomes. This tool streamlines the measurement, analysis, and evaluation of learning processes, encompassing a wide array of factors and parameters. Moreover, it contributes to the development of personalized and adaptive learning environments. In this study, a predictive model utilizing the XGBoost algorithm has been developed to analyze students' academic achievements. The model forecasts final exam grades based on various student characteristics, including age, participation rate, and exam scores. Evaluating the performance of the AI model involves metrics such as Mean Squared Error, Mean Absolute Error, and R^2 score. In findings indicate a strong prediction performance, with an R^2 score of 0.819. As a result of underscore the potential of AI-supported learning analytics as an effective tool for predicting and enhancing student academic performance.

Keywords: Artificial intelligence, Learning analytics, Machine learning, XGBoost algorithm

Introduction

Learning activity is a dynamic phenomenon that arises from the interaction of elements within education and is influenced by numerous parameters. During the emergence of this phenomenon, a plethora of data is generated from each stage and component. These data undergo various data processing processes such as organization, classification, and computation. As a result of these processes, they become manageable through data analysis, evaluation, and reporting activities.

The recommendation and utilization of Artificial Intelligence (AI) as a tool in the development and evaluation stages have demonstrated significant benefits in enhancing success, particularly in recent years. The concept of success varies across different domains, encompassing productivity, profitability, academic achievement, and accuracy. AI-supported Learning Analytics (LA) involves the utilization of AI technologies to understand, optimize, and enhance learning processes and student performance. Such an analytical process typically provides valuable contributions to educational institutions, teachers, and students.

Considering the existing resources, physical facilities, and student numbers in learning environments, the provision of personalized education may not seem feasible. However, it is essential to promptly address the negative impacts of these factors on student engagement, motivation, and success. Information technologies play a crucial role in improving these factors. AI technology can analyze data concerning all actors and factors involved in the process, thereby optimizing the process and outcomes. Additionally, as noted by Copgeven et al. (2023), AI techniques can be utilized to provide instant feedback to learners, enable learners to track their own learning processes, and offer messages to enhance learner motivation and engagement.

In today's world, where educational environments are increasingly becoming digitalized, the usage habits of learners in these environments, their navigation patterns, frequency of using learning resources, and their performance in exercises and exams have become valuable datasets. These datasets are processed, analyzed, and evaluated using machine learning methods. The findings obtained from these analyses are utilized as improvement factors for all components of education.

In this study, machine learning process using the XGBoost (eXtreme Gradient Boosting) algorithm was employed to predict final grades based on a dataset containing students' course information, demographic characteristics, exam scores, and final grades. The proximity rates between the predicted and actual grades were analyzed to draw certain conclusions. The article follows a structured format, with the second section dedicated to providing fundamental information supported by the literature on AI, XGBoost, and LA. The third section discusses the methodology of the application, including the dataset, the AI algorithm used, the developed model, and pseudo-codes in the Python language. The fourth section presents the findings obtained from running the application with tables and graphs. Finally, a general evaluation is made in the last section based on the findings, providing insights into the effectiveness of AI-supported XGBoost regression analysis in predicting student performance.

Literature Review

AI and data analytics are playing an increasingly important role in understanding and improving the learning processes in education. AI-supported LA has the potential to identify strengths and weaknesses in education by utilizing information derived from large datasets, and optimize educational programs accordingly. According to Chatti et al. (2012), Learning Analytics (LA) is a multidisciplinary research field that aims to develop methods for detecting and analyzing data models in educational environments. These methods are then utilized to enhance the overall learning experience.

AI is defined as the ability of machines to adapt to variable conditions, make intelligent predictions, solve problems, and perform behaviors that require a certain level of intelligence comparable to that of humans (Coppin, 2004). By incorporating AI into learning environments, intelligent tutoring systems are emerging. This system is described as a supportive and guiding system that provides feedback to and directs the student, the teacher, and the system itself before, during, and after the learning process (Fardinpour et al., 2014). In the design of this system, AI techniques and algorithms such as Decision Trees (DT), Artificial Neural Networks (ANNs), and Decision Networks are utilized (Talan, 2021).

According to Gligorea et al. (2023), AI technologies play a crucial role in optimizing learning paths, enhancing student engagement, and improving academic performance. Through a comprehensive literature review, the authors found evidence suggesting that AI has the potential to enhance students' test scores in various studies. The incorporation of AI and Machine Learning (ML) into e-learning platforms is emphasized as a key driver for personalized and effective educational experiences.

According to Moreno et al. (2020), AI-supported learning systems bring numerous benefits, including the creation of an enhanced learning environment, program flexibility, immediate feedback provision, control over students' experiences, and facilitation of accelerated development. These systems leverage their capacity to analysis large volumes of data, learn from tracks and experiences, make predictions, and provide personalized recommendations. By utilizing AI, it becomes possible to implement diverse teaching methods that cater to individual students' levels and overcome academic barriers (Tapalova & Zhiyenbayeva, 2022).

Liu and Yu (2023) summarized the overall architecture of a big data-based learning system in their study, as depicted in Figure 1. Upon careful examination of this architecture, it can be observed that the learning components layer, the information technology layer, and the data processing layer are in communication with each other. Additionally, it is understood that these layers interact continuously with the data management layer in the architecture.

In e-learning platforms, the utilization of AI and machine learning (ML) algorithms or methods is common for various purposes such as content personalization, academic performance prediction, knowledge gap identification, and dynamic assessments. Gligorea et al. (2023) recommend the use of AI algorithms in learning analytics processes to model students' behavior, predict their success, recognize complex patterns and optimize learning processes.

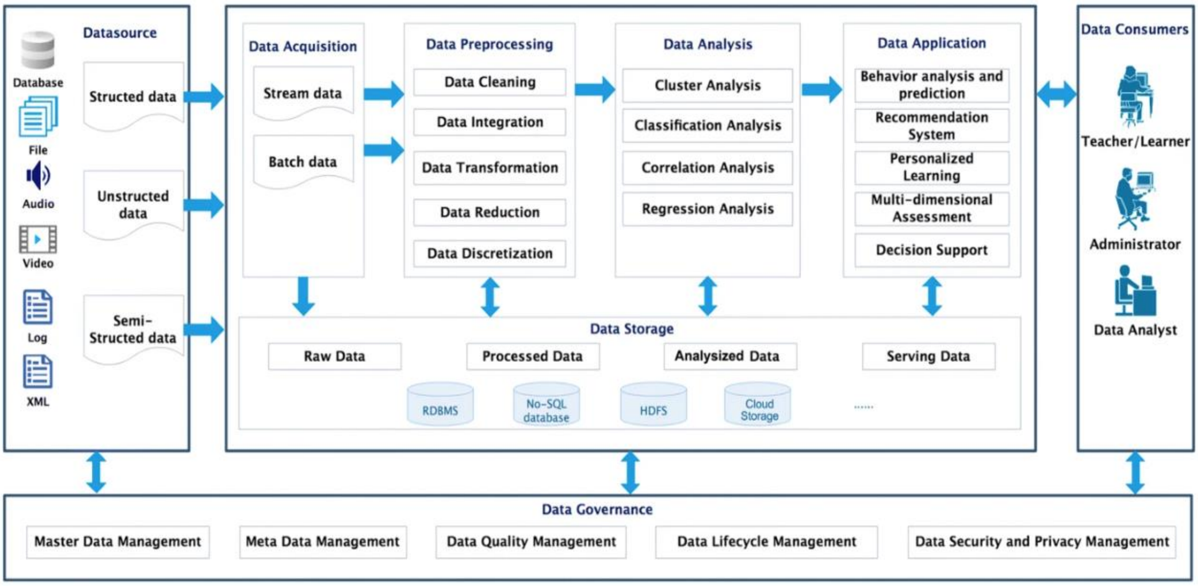

Figure 1. The architecture of big data based E-learning systems (Liu & Yu, 2023)

Prediction algorithms can range from simple to highly complex and have their own limitations. This process requires collaboration among individuals with different skills, such as data analysts, education experts, and content specialists. This team needs to organize input data, information on the accuracy of prediction algorithms, and which predictions should be reported to stakeholders. As an example of AI algorithms used in predicting academic achievement in LA processes, Veluri et al. (2021) suggest the following algorithms: Support Vector Machine, NaiveBayes Classifier, DT, k-Nearest Neighbors (kNN), Logistics Regression, and ANNs.

Ifenthaler and Yau (2020) emphasize that Learning Analytics (LA) has emerged as a methodology that integrates education and data science to evaluate the effectiveness of online learner interactions. Its main objectives are to inform instructional design and provide support for student success. Similarly, Castellanos-According to Reyes et al. (2023), there is a general consensus that in level of higher education, LA primarily focuses on assisting student success.

Drachsler and Greller (2012) categorize LA studies into two groups based on their objectives: reflection and prediction. Reflection studies enable students or teachers to monitor their performance and progress by comparing old and new data within their own dataset. Prediction studies, on the other hand, focus on identifying existing problems or situations in advance. Broos et al. (2020) highlight that the primary goal of reflection studies is to enable individuals, such as students or other stakeholders, to monitor their performance and evaluate themselves based on the reflection of their performance using both old and new data (Önder, Öztaş & Akçapınar, 2023).

Er (2023) states that there is sufficient evidence in their study to suggest that ML algorithms can accurately predict students who are possibility of school dropping out based on their online interaction data. The predictive power of these models relies on advanced mathematical calculations required by the underlying algorithms. However, this complexity makes it challenging for end-users to interpret the results. For example, while dropout prediction can provide instructors with a highly reliable report on students at risk of dropping out, it may not provide sufficient information about the reasons behind this prediction. This can lead to a delay in timely intervention. The author saw that among 1087 students in a sample data set, 351 had dropped out of school. He tried to predict this situation with AI. He calculated the prediction accuracy of the study as 0.80.

In their study, Hasib and colleagues (2022) conducted a comparative analysis using Logistic Regression, K-NN, SVM, XGBoost, and NBC algorithms. For the XGBoost algorithm, they found an accuracy value of 0.98. XGBoost algorithm has shown promising results in predicting student performance with the support of AI. However, it should be noted that the accuracy metric used for calculating prediction accuracy is typically employed for multi-class predictions. Since the target variable in the dataset represents a continuous value, it is suggested to calculate metrics such as R-squared (R^2), Mean Absolute Error (MAE), Variance, Mean Absolute Percentage Error (MAPE) and Mean Squared Error (MSE) instead of accuracy.

Felicia and Ferren (2022) performed an analysis with the Student Performance dataset using Generalized Linear Model, Random Forest and NBC algorithms. According to the results of the research, the algorithm with the highest accuracy was Naive Bayes with 48.59%. Zhang et al. (2021) in their study with the same data set using Gradient Boosting DT (GBDT), XGBoost, Adaptive Boosting (AdaBoost), Random Forest and DT algorithms, found that random had the highest score in accuracy values. They found it to be 0.90308 in the forest algorithm. The low accuracy values of these models may be due to the characteristics of the data set not being suitable for these models.

Contrary to these studies I have come across in the literature, in this research, an analysis was made with MSE, MAE and R2 values rather than accuracy value. Performance of predictive models; It should be noted that it depends on model parameters, course type, participant characteristics, data set structure and many other features. Therefore, transferring models between courses appears to be quite difficult, even for different runs of the same course. For this reason, model and application parameters are likely to produce different results in different courses (Er, 2023; Gasevic et al., 2016).

Method

In the study, a predictive model using the XGBoost AI algorithm was developed to forecast student academic scores. The model was trained using the Student Performance dataset. The final exam grades were predicted by taking into account various factors such as the students' ages, genders, family information, attendance rates, and academic performance data such as Grade 1, Grade 2, and Grade 3 (Final Score) within the dataset. For this purpose, an XGBoost decision tree model was built, taking into account the dataset features. Metrics such as MSE, MAE, and R2 Score employed to evaluate the accuracy and predictive power of the model. Additionally, a scatter plot was created to demonstrate the impact of actual scores on the predicted final scores.

Dataset

These data are taken from the dataset called Student Performance, which was collected by Cortez and Silva (2008) from secondary students of two Portuguese schools using school reports and surveys. The dataset used in the study comprises various types of data, including student exam grades, demographic information, social factors, and school-related data. The dataset encompasses a total of 395 students, providing a comprehensive set of information for analysis and modeling purposes. Students' grades in Mathematics were used to determine their academic performance. The data set is available on the website as the UCI Machine Learning Repository (UCI, 2024).

Model Architecture and Model Training

XGBoost is a widely used gradient boosting algorithm that has achieved impressive results in many areas of ML. The application of meta-learning to recommend hyperparameters for XGBoost can greatly reduce the time and computational resources needed to optimize the model. By leveraging meta-learning techniques, the process of finding the most effective hyperparameters for XGBoost can be automated, leading to more efficient and effective model optimization. Recently, XGBoost algorithm has been increasing in popularity due to its convenience, scalability, and ability to work well with both regression and ranking problems using the Gradient Boosting structure (Chen & Guestrin, 2016; Morinho et al., 2024).

In XGBoost, multiple decision trees are used to build the model. In the model, each tree is a collection of weak decision trees that sequentially refine the residues of previous trees. The hyperparameter *n_estimators* control the number and type of trees used in the ensemble. The layer depth of the tree is controlled by the *max_depth* hyperparameter. Formula 1 defines the basic formula of this algorithm.

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \in F, \tag{1}$$

Each tree is trained to correct the prediction errors of previous trees. This is done by optimizing a loss function based on the negative gradient of the error function (e.g., mean square error). According to Formula 1; $\hat{y}_i$ is the predicted value for the i-th observation (Predicted final exam grade); f_k is the k th regression tree; x_i is the input features for the i-th observation (student feature vector for this application); F is the set of all possible regression trees and K is the model is the total number of trees used. These trees are used to predict the final exam grade based on the student's characteristics. Finally, the predictions of all trees are combined to obtain the total prediction.

Hyperparameter tuning using the *GridSearchCV* method can optimize the model's performance by adjusting crucial parameters that have a significant impact on the model's behavior. These parameters include the number of trees, learning rate, and tree depth. Algorithm 1 in Python presents the general structure of the model as pseudo code, allowing for the systematic exploration of hyperparameter combinations.

Algorithm 1. XGBoost Model Application for Learning Anatilics

1. Import necessary libraries: pandas, xgboost numpy and matplotlib libraries
   from sklearn[.metrics, .model, .preprocessing] import [LabelEncode, GridSearchCV]
2. Loading of sample dataset
   student-mat.csv
3. Preprocessing (names and others features) and Split the dataset (train/test)
   X = data.drop("G3", axis=1)
   y = data["G3"]
   X_encoded = X.apply(label_encoder.fit_transform)
4. Create a XGBoost_AI model:
   xgb_model = xgb.XGBRegressor()
   …
   param_grid = {
   'max_depth': [3, ., .],
   'n_estimators': [100, ., .],
   'learning_rate': [0.05, ., .],
   }
5. Train the model:
   model.fit(X_train, y_train)
6. Make predictions:
   y_pred = xgb_model.predict(X_test)
7. Model's performance values
   mse =…
   mae =…
   r2 =…
8. Evaluate the model's performance
   *xgb_model = xgb.XGBRegressor(**best_params)*
9. Visualize predictions and true values:
   print("Mean Squared Error (MSE):", mse)
10. Graphs of all models and metrics with the plot( ):
    plt.scatter, plt.show().

In the code example given above, there are 395 students in the student-math.csv file. A portion of these students is assigned to the training set, while the remaining students are allocated to the test set. This distinction was made using the *train_test_split* function so that the data set was 80% training and 20% testing. These additional features can be utilized to enhance the overall power of the model or to emphasize a specific combination of features.

Results and Discussion

In this application, calculations were made through a prediction model using the attributes of the student data set. Metrics were employed to evaluate the performance of the model. Mathematical metric measurements and a graphical scatter plot were created to show the impact of all attributes on the final scores. When the Python code of the application is run, the parameter values are calculated as in Table 1 below. In Table 1, for each student in the first column, the actual final grade versus the estimated final grade obtained when the model is run is given. It can be seen that these notes are close to each other. Table 2 gives the values of these grade predictions calculated on the basis of metrics.

Table 1. Student-based grade predictions of the prediction model

Student	Actual Score	Predicted Score
78	10	8.693336
371	12	11.427962
248	5	6.250742
55	10	10.154843
390	9	9.746037
…	…	…
364	12	11.213716
82	6	6.671789
114	9	10.471691
3	15	13.825548
18	5	6.275213

Table 2. Calculated metric values of the prediction model

Metrics/Model	XGBoost
MSE	3.714846907185788
MAE	1.165801862749872
R^2 Score	0.818832560655874
S^2 (Variance)	20.989616397866733

The calculated MSE value of 3.71 suggests good model accuracy. MAE is a metric that measures the average absolute difference between model predictions and actual values. The calculated MAE value of 1.17 indicates that, on average, the model predicts approximately 1.17 units away from the actual values. The Model's R^2 value of 0.82 suggests that approximately 82% of the variance in the target variable is explained by the independent variables. Overall, the model shows a strong predictive power as reflected in the MSE, MAE, and R^2 scores, indicating its ability to capture the relationship between the data.

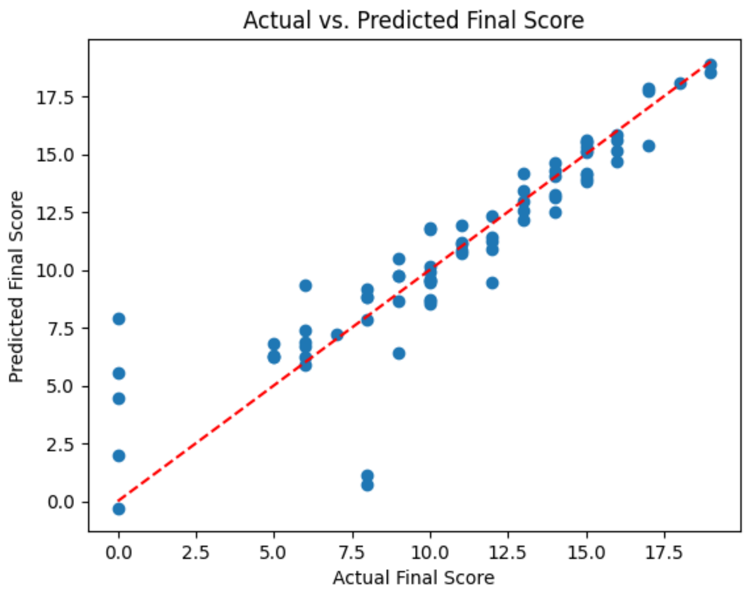

Figure 2. Actual vs. predicted final score plot

In Figure 2, the graph displays the distribution of the actual and predicted final results of the sample. The x horizontal axis is the axis of actual final scores, the Y vertical axis is the axis of predicted final scores. The presence of a red dashed line aids in visualizing a perfect prediction scenario, where the actual and predicted values align. The concentration of data points around this red line indicates the level of compatibility between the actual and predicted values.

In the graph in Figure 3, the predicted final scores are shown on the x-axis, and the residuals (differences between actual and predicted scores) are shown on the y-axis. The red dashed line indicates the point at which the residuals are zero (i.e., a perfect prediction is made). This chart shows how accurate your model's predictions are and where they are inaccurate.

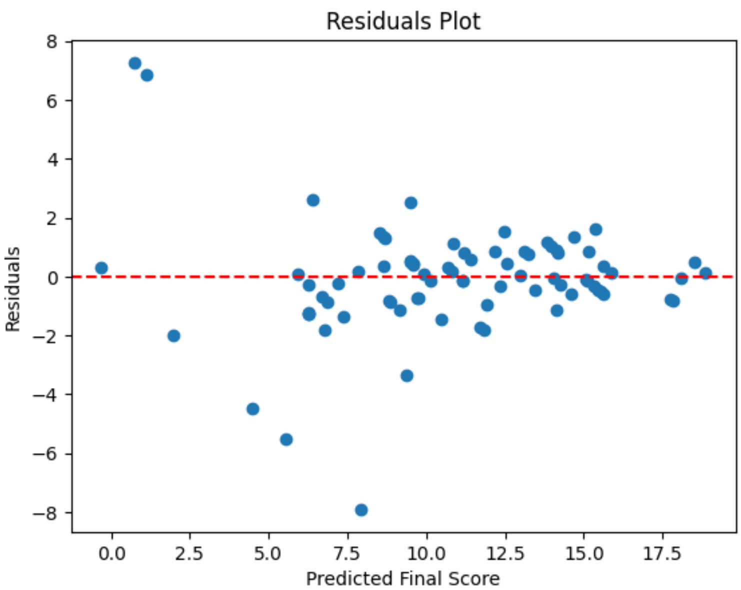

Figure 3. Residuals plot graph

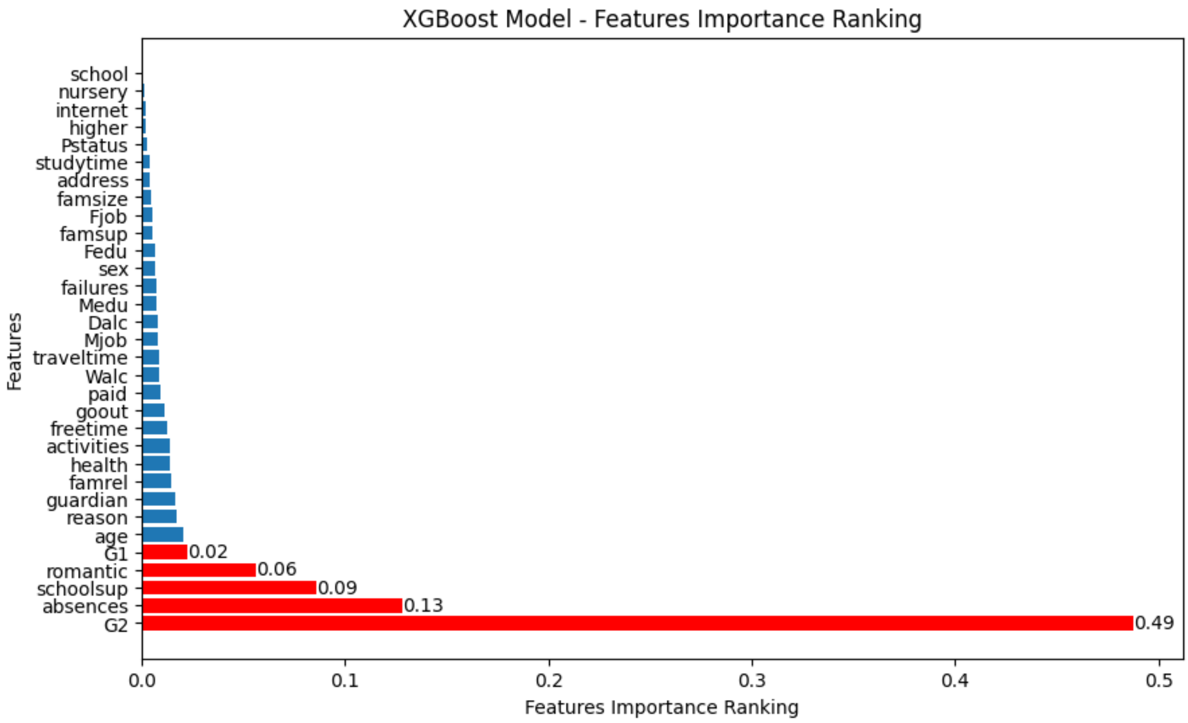

Figure 4. Weighted Effect of different features by model

Model coefficients are presented graphically in Figure 4 to help determine the effect of each feature or variable on the target variable. This analysis is important to determine which features have more weight and how important they are in the predictions. Looking at Figure 4, you can see how much impact all of these features have on the XGBoost model. According to Figure 4, the 5 features that have the most weight in predicting the final score are; It is understood that there are G2, absences, schoolsub, romantic and G1. This information can help better understand the factors that influence student performance and optimize their effects.

Conclusion

This LA application, the XGBoost model proved highly successful in predicting student performance. Among the attributes used, G2, Absences, and Schoolsub were found to have a significant impact on final exam grades.

While the G1 score and certain demographic characteristics also influenced performance, their effects were comparatively less pronounced. The utilization of AI and ML techniques has revolutionized student data analysis, enabling adaptive learning systems to create personalized profiles and gain insights into individual strengths and needs. By leveraging AI algorithms, these systems can tailor learning content, adjust task difficulty, and provide targeted interventions, resulting in improved learning outcomes and enhanced engagement, motivation, and information retention for learners.

The impact of information technologies and AI on educational environments goes beyond just LA and extends to the utilization of immersive technologies like internet of things, cloud computing, virtual reality and augmented reality for enhancing the learning experience. Such technologies especially help medical, engineering, design and architecture students to connect with reality and visualize objects and concepts from abstract to concrete. Radu (2012) emphasizes that thanks to these technologies, the content is enriched, learning improves and motivation increases.

On the other hand, some basic problems are encountered in LA with such AI technologies, such as not being able to establish the appropriate model, not being able to calculate the correct metrics, not being able to ensure data confidentiality, the emergence of advanced hardware and software needs, and the system's inability to make accurate predictions with very little data at the beginning. Additionally, integration and compatibility problems also arise during the implementation of such innovations. In order to cope with such negativities, good coordination should be ensured between educators, designers and practitioners.

This study places significant emphasis on the importance of AI-supported LA in education and demonstrates its effectiveness in predicting student performance with AI algorithms. In particular, it shows that information obtained through the use of large data sets is a valuable resource for improving decision-making processes in education. It may be a good idea to initially start with a simple regression model, and then improve the model with AI algorithms and improve prediction rates. Future research may focus on further developing AI-assisted LA by using larger data sets and applying different ML techniques.

Scientific Ethics Declaration

The author declares that the scientific ethical and legal responsibility of this article published in EPSTEM journal belongs to the author.

Acknowledgements or Notes

* This article is an updated and extended version of the abstract presented orally at the International Conference on Basic Sciences, Engineering and Technology (www.icbaset.net) held in Alanya/Turkey on May 02-05, 2024.

References

Broos, T., Pinxten, M., Delporte, M., Verbert, K. & De Laet, T. (2020). Learning dashboards at scale: Early warning and overall first year experience. *Assessment & Evaluation in Higher Education*,*45*(6), 855-874.

Castellanos-Reyes, D., Koehler, A.A., & Richardson, J.C. (2023). The i-SUN process to use social learning analytics: A conceptual framework to research online learning interaction supported by social presence. *Frontiers in Communication*,8,

Chatti, M. A., Dyckhoff, A. L., Schroeder, U. & Thus, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, *4*(5-6), 318-331.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining* (pp. 785-794).

Coppin, B. (2004). *Artificial intelligence illuminated* (p.739). Massachusetts, USA: Jones & Bartlett Publishers.

Cortez, P., & Silva, A.M.G. (2008). Using data mining to predict secondary school student performance. In A. Brito, J. Teixeira (Eds.), *Proceedings of 5th Future Business Technology Conference (FUBUTEC 2008)* (pp. 5-12).

Copgeven, N. S., Ozkaya, H., & Aydın, S. (2023). Acık ve uzaktan ogrenmede yapay zeka destekli oyunlastırma. *Acıkogretim Uygulamaları ve Arastırmaları Dergisi*, *9*(1), 386-407

Drachsler, H., & Greller, W. (2012). The pulse of learning analytics understandings and expectations from the Stakeholders. *LAK '12: Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 120-129).

Er, E. (2023). KACD'lerde okuldan ayrılanları tahmin etmeye ve anlamaya yonelik acıklanabilir bir makine ogrenimi yaklasımı. *Kastamonu Egitim Dergisi*, *31*(1), 153-164.

Fardinpour, A., Pedram, M. M., & Burkle, M. (2014). Intelligent learning management systems: Definition, features and measurement of intelligence. *International Journal of Distance Education Technologies (IJDET)*, *12(*4), 19-31,

Felicia, & Ferren. (2022). Exploring secondary school performance by using machine learning algorithms. *Journal of Educational Analytics*, *1*(1), 41-60.

Gasevic, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: the effects of instructional conditions in predicting academic success. *Internet and Higher Education*, 28–84.

Gligorea, I., Cioca, M., Oancea, R., Gorski, A. T., Gorski, H., & Tudorache, P. (2023). Adaptive learning using artificial intelligence in e-Learning: A literature review. *Education Sciences*, *13*(12),1216.

Hasib, K. M., Rahman, F., Hasnat R., & Alam, M. G. R. (2022). A machine learning and explainable ai approach for predicting secondary school student performance. *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)* (pp.399-405). Las Vegas, NV, USA.

Ifenthaler, D., & Yau, J. Y. K. (2020). Utilising learning analytics to support study success in higher education: A systematic review. *Educational Technology Research and Development, 68,* 1961-1990.

Liu, M., & Yu, D. (2023). Towards intelligent e-learning systems. *Education and Information Technologies*, *28,* 7845–7876.

Marinho, L.T., Nascimento, D. C., & Pimentel, B. A. (2024). Optimization on Selecting XGBoost hyperparameters using meta-learning. *Expert Systems*, e13611.

Moreno-Guerrero, A. J., Lopez-Belmonte, J., Marin-Marín, J. A., & Soler-Costa, R. (2020). Scientific development of educational artificial intelligence in Web of Science. *Future Internet*, *12*(8), 124.

Onder, A., Oztas, G. S., & Akcapınar, G. (2023). Ogrenme analitigi surecine yonelik modellere genel bir bakıs: Kavramsal bir cerceve onerisi. *Acıkogretim Uygulamaları ve Arastırmaları Dergisi,9*(1), 92-117,

Radu, I. (2012). Why should my students use AR? A comparative review of the educational impacts of augmented-reality. *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp.313-314). IEEE.

Shuai Z., Jie C., Wenyu Z., Qiwei X., & Jiaxuan S. (2021). Education data mining application for predicting students' achievements of Portuguese using ensemble model. *Science Journal of Education*, *9*(2), 58-62.

Talan, T. (2021). *Egitimde dijitallesme ve yeni yaklasımla* (1th ed., p.363). Istanbul, Turkey: Efe Akademi Publishing

Tapalova, O., & Zhiyenbayeva, N. (2022). Artificial intelligence in education: AIED for personalised learning pathways. *The Electronic Journal of e-Learning (EJEL)*, *20*, 639-653,

UCI. (2024). *Student performance. UCI machine learning repository*. Retrieved from https://archive.ics.uci.edu/dataset/320/student+performance

Veluri, R. K., Patra, I., Naved, M., Prasad, V. V., Arcinas, M. M., Beram, S. M., & Raghuvanshi, A. (2022). Learning analytics using deep learning techniques for efficiently managing educational institutes. *Materials Today: Proceedings*, *51*(8), 2317-2320,

Author Information

Mustafa Cosar
Department of Computer Engineering, Hitit University
19030, Corum/Turkiye
Contact e-mail: *mustafacosar@gmail.com*

To cite this article:

Cosar, M. (2024). Utilizing the eXtreme gradient boosting algorithm for artificial intelligence-supported learning analytics application. *The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 28,* 277-285.