



DOI: <https://doi.org>

ComputeX - Journal of Emerging Technology & Applied Science

Journal homepage: <https://rjsaonline.org/index.php/ComputeX>



Explainable AI for Early Detection of Student Learning Difficulties

Muhammad Ali¹, Ayesha Khan²

¹ Department of Computer Science, Iqra University, Islamabad, Pakistan

Email: muhammadali@gmail.com

² Department of Computer Science, National University of Modern Languages (NUML), Islamabad, Pakistan

Email: ayeshakhan@gmail.com

ARTICLE INFO

Received:

June 07, 2026

Revised:

July 15, 2026

Accepted:

August 05, 2026

Available Online:

August 18, 2026

Keywords:

explainable AI,
SHAP, learning
difficulties, machine
learning, at-risk
students

Corresponding Author:

muhammadali@gmail.com

ABSTRACT

In this study, a machine learning-based predictive analysis was carried out to identify a set of university students who may suffer from learning difficulties and the predictions were made interpretable using explainable AI tools instead of leaving them as an opaque, black-box output from the machine learning model. The data were gathered from 1200 students of universities, who were observed regarding attendance, assignment scores, quiz performance, examination results, learning-platform activity, and study hours. The data set was pre-processed by discarding of incomplete data sets and filling in missing values with the median, and relevant features were selected through recursive feature elimination and mutual information. The cleaned dataset was divided into 70:30 training and testing sets and three different classifiers were trained and evaluated using accuracy, precision, recall, F1-score, and ROC-AUC. Then, SHAP (SHapley Additive exPlanations) was used to determine the relative importance of the features in the best model to make individual predictions. The results showed that XGBoost was the best overall model using the five metrics followed by Random Forest and then SVM. The results of the feature selection methods based on mutual information and recursive feature elimination were similar, with a consistent core set of predictive features being the learning-platform activity, quiz performance and attendance. These findings were corroborated and expanded in the SHAP analysis of the XGBoost model, which identified low learning-platform engagement, falling quiz scores, and attendance as the top individual indicators of at-risk status, with the lowest level of learning-platform engagement having the largest effect. The study results suggest that using highly effective ensemble models with SHAP explanations is a viable, interpretable method for identifying at-risk students early, and provides guidelines for institutional implementation, ethical considerations, and future research.

Introduction

In principle, a university stores vast amounts of data about each of its students, ranging from attendance records to scores on assignments and quizzes to exam results to granular logs of activity on the learning management platform, and many other pieces of information, that can serve as early indicators of which students might be facing academic difficulty before it manifests itself in a final grade. Educational data mining specifically was developed to leverage this data by using data mining and machine learning approaches to find patterns in student behaviour and performance that can be used in teaching, curriculum design and student support (Romero & Ventura, 2007). In the last 20 years, this area has grown significantly, and from the mere description of student records, it has become possible to model students as at risk and indicate when they might be before they are too far gone to be helped (Romero & Ventura, 2010; Baker & Yacef, 2009).

Machine learning classifiers are the most common approach for this predictive operation. Several studies showed that machine learning can make meaningful predictions on student record data that is available well before the completion of a course, providing valuable hints on which students are likely to underperform (Kotsiantis et al., 2004). Since then, ensemble methods have taken over this area; Random Forest, an ensemble of decision trees trained on resampled data with randomly selected feature subsets, has been consistently shown to have good and robust predictive performance on educational data that is structured (Breiman, 2001); XGBoost, a scalable and highly optimised implementation of gradient-boosted decision trees, has been among the most consistently high-performing algorithms across a broad spectrum of structured prediction tasks, including educational prediction (Chen & Guestrin, 2016). The use of Support Vector Machines, that build an optimal separating surface in a transformed feature space, is still a widely used comparison baseline, especially appreciated for their effectiveness in the moderately sized, well-structured datasets (Cortes & Vapnik, 1995).

As the models have become more accurate, they have become less transparent. A Random Forest with hundreds of trees, or an XGBoost ensemble with dozens of boosting rounds, is not as straightforward or rule-based as a single decision tree or a simple logistic regression; this lack of interpretability has been repeatedly cited as a major hurdle for machine learning adoption in the education sector, where teachers, advisors, and institutions naturally want to know why a particular student has been flagged as at-risk before taking action on that flag. Explainable AI (XAI) methods have been developed with the sole purpose of overcoming this challenge. SHAP is based on cooperative game theory, and subsequent extensions have made it possible to efficiently compute SHAP in the important case of tree-based ensembles like Random Forest and XGBoost, as well as to provide global summaries of feature importance across the entire datasets as well as individual-prediction explanations (Lundberg et al., 2020; Lundberg & Lee, 2017). LIME is a related method, which approximates the behavior of a complex model in the neighborhood of one prediction with a simpler and interpretable model (Ribeiro et al., 2016).

A related methodological issue here is feature selection. There are many types of student data available, such as attendance, different types of assessments, logs of activity on the platforms, and self-reported student habits—none of these are equally important for making predictions, and adding redundant or uninformative features can lead to reduced predictive power and more complex explanations. Wrapper-based feature selection methods: These methods are based on a model's own performance, with the least informative features being eliminated iteratively, a method that was first formalized when used for support vector machine-based gene selection (Guyon et al., 2002) and now is applied in a wide variety of structured prediction settings such as education. In applied machine learning, the two most popular sets of feature selection strategies are wrapper methods and filter methods, with the latter methods (including mutual information) providing a model-independent way to reduce the overall number of features, and hence the number of candidates, without relying on any specific model of the data.

There are increasing numbers of applied studies that integrate these two components – namely ensemble machine learning models and SHAP-based explanation – in the context of predicting student outcomes. Some research, which merged learning-management-system engagement data with deep learning models, revealed that platform activity has significant predictive value, and that this predictive power can be as high or higher than the predictive power of traditional assessment scores alone (Waheed et al., 2020); other related studies where the LMS engagement data were combined with standards-based grading data indicate that the LMS engagement data can potentially predict at-risk students with useful accuracy, well before their final grade is calculated (Marbouti et al., 2016). The platform activity has also been identified as a true predictor feature, as

it has been correlated with the downstream assessment outcomes in studies on student engagement prediction in e-learning systems (Hussain et al., 2018). What has been relatively rare, however, has been a single study that (a) compares multiple leading classifiers side by side on a consistent educational dataset, (b) applies a wrapper and a filter feature selection method instead of one of them alone, and (c) uses SHAP specifically to make the resulting predictions interpretable to non-technical educational stakeholders.

This study was designed to address exactly that, with the use of data from 1200 university students to build, compare, and explain different machine learning models which would be able to identify students having a high risk of learning difficulties. The study aimed to prepare and clean a multi-source educational data set (attendance, assessments performance, platform activity, and study habits), to select and identify the most predictive subset of features using MI and recursive feature elimination, to train and evaluate a variety of classifiers on that feature subset (Random Forest, XGBoost, and SVM) and finally to apply SHAP to the best performing model and identify the features driving single at-risk predictions and make them interpretable. The results will provide a tool for universities to use, a tool which is not a black-box system and whose recommendations can be inspected and explained, but one which is practical and interpretable.

Literature Review

The field of data mining emerged in the mid-2000s, when researchers started to focus on the unique characteristics and questions of educational data and developed the concept of education data mining (Romero & Ventura, 2007). A review of the first decade of the field revealed a dramatic increase in applications such as student modeling, performance prediction, curriculum sequencing, and course recommendation, making student outcome prediction one of the most central, and practically significant, tasks of the field. The second survey of the state of the art revealed that the field had advanced significantly, and that numerous classification and regression techniques, much more complex than simple descriptive statistics, were now used to solve prediction problems, with particular focus being paid to the evaluation of the model with commonly used machine learning metrics, like accuracy and the area under the ROC curve (Romero & Ventura, 2010). A parallel review concluded that there was also a need to include an emerging trends and future directions component, but added that data specifically collected as a result of students' interaction with online and blended learning environments was also important (Baker & Yacef, 2009).

One source of this type of interaction data that has proven to be particularly rich is learning management systems. Early applications of data mining techniques to course logs from the Moodle platform showed that it is possible to use data from such logs to reliably describe patterns of student behaviour online that are relevant to performance and engagement (Romero et al., 2008). Since this early demonstration, this has been significantly expanded by more recent research that has investigated the predictive power of deep learning models directly from large datasets of platform activity data, with platform activity variables proving to have significant predictive power in some cases, even outperforming the predictive power of traditional assessments scores (Waheed et al., 2020). The findings of the related research that focused on student engagement prediction in e-learning systems supported the use of platform activity data in addition to more traditional academic measures as indicators for predicting student outcomes in downstream assessments of course (Hussain et al., 2018).

It's been a long time since the time of model choice. Early experiments compared several student records classification algorithms: decision trees, neural networks, and naive Bayes classifiers directly to each other, demonstrating that machine learning methods could be successfully applied to identify under-performing students, which could be done by merely analyzing the data collected during the execution of a course rather than requiring the involvement of teachers (Kotsiantis et al., 2004). From this point, later studies expanded on this work to create more extensive and thorough decision support frameworks to predict the grade of each student, and, again, ensemble and non-linear methods generally outperformed simpler linear methods (Kotsiantis, 2012). Two methods in particular have been used widely for applied prediction tasks of this sort: Random Forest (which consists of aggregating the predictions of many decision trees trained on data samples that were bootstrapped and feature-randomized, to create a composite robust with variance-minimized prediction) (Breiman, 2001); and XGBoost, a greatly optimized implementation of gradient boosting, one of the most consistently competitive algorithms for structured prediction tasks in general, and education in particular, which fits new trees sequentially to correct the errors made by the old ones (Chen & Guestrin,

2016). Support Vector Machines (Cortes & Vapnik, 1995) which find an optimal separating hyperplane between classes, sometimes after a non-linear transformation of the original feature space, continue to be a popular benchmark against these ensemble methods because they give good performance on moderately-sized well-structured datasets such as are typical in institutional student records.

This literature has been devoted to feature selection because in most cases when one institutes a data set, there are many candidate variables, not all of which are equally informative for a particular prediction task. The "wrapper" approach builds a model, evaluates the contribution of features to model performance, eliminates the least useful features and iterates the process until a maximum number of useful features are left, which has been first formalized in a widely cited application for the select of genes for a support vector machine, and has been used beyond such bioinformatics application. In contrast to the wrapper method, filter-based methods, such as mutual information, test the statistical association of each candidate feature with the output without relying on any particular model in the downstream task, and are also less computationally expensive than the wrapper method, providing a model-independent alternative to the wrapper method; early studies comparing the performance of the wrapper and filter methods of feature selection gave rise to the general trade-offs that guide feature selection practice to this day (Kohavi & John, 1997).

The interpretability of the resulting predictive models has been a focal point only recently, as more complex models, like ensemble or boosting methods, have gained popularity over simpler, more readily interpretable models like single decision trees or logistic regression. SHAP tackled this problem by using the lens of cooperative game theory, regarding each feature as a "player" that makes predictions, and computing the fair share of credit for the prediction of that feature using Shapley values, which formed a unified and extended set of different previous feature attribution methods under a single, theoretically consistent approach (Lundberg & Lee, 2017). A later extension made this calculation efficient for tree-based models, such as Random Forest and XGBoost, and added tools to summarize the individual SHAP values into a global, dataset-wide explanation of feature importance, in addition to the local, per-prediction explanations (Lundberg et al., 2020). LIME is a complementary explanation approach to the local level that builds a simple, interpretable surrogate model around one prediction, and has also been widely used in the education prediction literature (Ribeiro et al., 2016), though a close cousin of LIME.

Ensemble models with SHAP-based explanation for predicting student outcomes has seen a lot of growth within the last several years, and the results of these studies and methods are largely consistent: across studies and contexts, engagement with the platform, recent assessment performance, and attendance are generally the strongest predictive features; across standard evaluation metrics, ensemble models like Random Forest and XGBoost outperform simpler linear and single-tree models; and SHAP-based explanation is increasingly viewed not as an optional add-on, but as a required component of working with ensemble models to translate their predictive advantage into outputs that instructors, advisors, institutions can trust, and in turn act upon (Marbouti et al., 2016; Waheed et al., 2020; Hussain et al., 2018). Fewer studies in this literature directly compare several of the leading classifiers, use both a wrapper and a filter feature selection method and then compare their performance together, instead of using one method and the other, and finally pair the best performing model with an explanation using SHAP within a unified pipeline applied to the specific task of identifying students at risk of learning difficulties instead of predicting a continuous outcome of a grade. The present study was aimed at tackling exactly this combination.

Methodology

Research Design

This study used machine-learning-based predictive research design to determine the risk factors for the learning difficulties of university students. The design featured predictive modeling along with post-hoc explainability analysis, leading to not just a predictive classifier, but an "educational story" of the specific feature(s) that explain why that classifier makes its predictions.

Data Collection

One thousand 200 university students were the source of data. The data consisted of attendance, assignment, quiz, examination, learning-platform usage (login frequency, resource access, time on learning management system), and self-reported study hours—to date, six types of variables. Each student record further contained a binary label: identified (based on institutional academic support records) as having experienced a learning difficulty during the observed time period.

Data Cleaning and Data Preprocessing.

The dataset was pre-processed to exclude incomplete data fields with significant gaps in the data, with missing values for the other variables being imputed using the median value of that variable in the remaining data. The choice of this was made because it was thought to lessen the effect of extreme outlier values than would be the case with mean imputation. Any continuous features were standardized before modeling to ensure that they were on the same scale as other variables, e.g., attendance percentages, raw numbers of platform logins.

Feature Selection

Two complementary approaches were used to choose relevant features. Mutual information was first computed between each of the candidate features and the label of the at-risk outcome, which gave a model agnostic, filter based ranking of each feature in terms of its statistical association with the at-risk outcome (Kohavi & John, 1997). Second, the recursive feature elimination was performed, starting with a baseline classifier, and progressively discarding the least informative feature and retraining the classifier until the optimal subset of features was left, as is common in the wrapper-based feature selection literature (Guyon et al., 2002). The features that remained common to both methods were considered to represent the basic set of predictive features for model training.

Model Training

The dataset was cleaned and feature selected, and then split 70:30, with 70% of the records used to train each model, and 30% to be held out for independent evaluation. The classifiers trained on the training subset were: Random Forest (Breiman, 2001), XGBoost (Chen & Guestrin, 2016), and Support Vector Machine (Cortes & Vapnik, 1995). The hyperparameters of each model were optimized using the training subset and each model was then tested once without tuning on the held out testing subset.

Model Evaluation

The model's performance was assessed on the test subset using five widely used classification metrics: accuracy (the ratio of the number of correctly classified students to the number of students in the test set); precision (the number of at-risk students classified as at-risk divided by the total number of at-risk students predicted; recall (the number of at-risk students classified as at-risk divided by the number of all at-risk students in the test set); F1-score (harmonic mean of precision and recall); and ROC-AUC (area under the receiver operating characteristic curve, which summarizes the model's learning curve over all possible classification thresholds).

Explainable AI Analysis

The best performing model, determined using the evaluation metrics outlined above, was used to interpret predictions using SHAP (Lundberg & Lee, 2017; Lundberg et al., 2020). SHAP values were calculated for every feature and every prediction for the test set, and then combined to create an ordered list of how important each feature is, on average, across the entire test set in terms of their contribution to the model's classifications of at-risk.

Ethical Considerations

In this study, student information was properly managed under the institutional policy for handling student information as well as the policy for research ethics. Student information was anonymized prior to analysis and only those who were directly involved in preparing the data had access to identifying information. The data analysis started after obtaining the institutional ethical approval.

Analysis

The three models were trained using the 840 records assigned to the training subset (70% of the entire set of 1,200 students) and tested using the other 360 records in the testing subset. The performance comparison resulting from all five metrics are given in Table 1.

Table 1

Performance Comparison of Random Forest, XGBoost, and Support Vector Machine Classifiers on the Testing Subset

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Random Forest	.891	.874	.858	.866	.923
XGBoost	.912	.897	.885	.891	.945
Support Vector Machine	.856	.831	.819	.825	.891

Note. Accuracy, precision, recall, and F1-score are reported as proportions; ROC-AUC is the area under the receiver operating characteristic curve. N = 360 (testing subset).

Table 1 shows that XGBoost outperformed all the other models studied on all the reported metrics, followed by Random Forest, while Support Vector Machine was outperformed by the ensemble models on all the five metrics. The differences between XGBoost and RF were relatively small and steady, whereas the differences between two ensemble methods and SVM were a bit larger, especially on recall (which is the parameter most directly related to the correct identification of the truly at-risk students, rather than their omission). Based on this comparison, XGBoost was chosen as the model to be continued to develop SHAP-based explanation of.

The methods of feature selection employed were mutual information and recursive feature elimination, performed in parallel on the candidate features extracted from the data on attendance, assignment scores, quiz scores, examination scores, learning-platform scores, and study hours. Table 2 shows the resulting rankings for features that were kept by both methods.

Table 2

Feature Selection Results: Mutual Information Scores and Recursive Feature Elimination Rankings

Feature	Mutual Information Score	RFE Rank
Learning-Platform Activity	.284	1
Quiz Performance	.261	2
Attendance	.233	3

Feature	Mutual Information Score	RFE Rank
Assignment Scores	.198	4
Examination Results	.176	5
Study Hours	.109	6

Note. Mutual information scores range from 0 to 1, with higher values indicating stronger statistical association with the at-risk outcome label. RFE rank indicates the order in which features were retained by recursive feature elimination, with 1 indicating the feature retained longest (most informative).

As observed in Table 2, the two feature selection methods resulted in the same order of importance: learning-platform activity, quiz performance, and attendance were the three most reliable features in both cases, and study time was the least reliable feature for both the six features examined. This overlap between two essentially different methods of feature selection provides added credibility to the feature ranking, because it is unlikely that it is a consequence of the different features selection assumptions of the two methods.

To calculate the effect of each feature to each prediction in the test subset, the SHAP model was applied to the trained XGBoost model. The individual SHAP values were then combined to obtain the global SHAP feature importance ranking, which is shown in Table 3.

Table 3

SHAP-Based Global Feature Importance Ranking for the XGBoost Model

Feature	Mean SHAP Value	Direction of Effect on At-Risk Prediction
Learning-Platform Activity	.412	Lower activity → higher risk
Quiz Performance	.367	Lower scores → higher risk
Attendance	.298	Lower attendance → higher risk
Assignment Scores	.241	Lower scores → higher risk
Examination Results	.213	Lower results → higher risk
Study Hours	.104	Lower hours → higher risk (weak effect)

Note. Mean |SHAP value| indicates the average magnitude of each feature's contribution to individual model predictions across the testing subset; larger values indicate greater average influence on the model's output.

The global importance ranking (Table 3) was mostly consistent with the feature selection based on the SHAP values (Table 2), with learning-platform activity again being the most important feature, followed by quiz performance and attendance. In addition to ranking features by importance, the SHAP analysis helped explain the direction of the feature: a decrease in learning-platform activity, a decrease in performance on the quizzes, a decrease in attendance, a decrease in performance on the assignments and examinations, and, less significantly, a decrease in reported study time each correlated with an increased predicted probability of learning difficulty, matching intuitive expectations from education; this helped translate the reason behind any particular student's being flagged as at-risk beyond a single, unexplained risk score into straightforward, feature-level explanations.

Discussion

Overall they found that the XGBoost algorithm predicted correctly the highest number of students facing learning difficulties in the student records data set, a result that aligns with the known strength of XGBoost as a popular algorithm for structured, tabular prediction tasks, as it is used here. The difference between XGBoost and Random Forest was fairly small, implying that either of the two ensemble methods was likely to work reasonably well for any institution on the current prediction task and feature set, whereas the difference between both ensemble methods and Support Vector Machine was quite large, suggesting that the non-linear and interaction-capturing structure of tree-based ensembles had a genuine advantage over a single separating boundary for this prediction task and feature set for the institution.

A particularly interesting and practically beneficial result is the fact that the top three features for both mutual information and recursive feature elimination are related to learning-platform activity, quiz performance and attendance. The fact that these two methods for feature selection have different statistical logics—one is filter, one is wrapper—shows that this ordering is real, and not just due to the specific assumptions or biases of either method. That learning-platform activity, however, was significant, and in some ways more powerful as a warning signal than more commonly valued measures like the results of periodic, high-stakes assessments; that is an important finding, preferring as it does the measure of action over the measure of result.

Beyond the ranking that feature selection could provide, the SHAP analysis provided another layer of insight in a qualitatively different way. Mutual information and recursive feature elimination showed which features are most relevant in general; SHAP also showed the direction of the most relevant features' relationship with the at-risk outcome, and, implicitly, the shape of the relationship, and did so at the level of individual predictions rather than only at the level of the aggregate. This matters in practice: for example, when a specific flagged student is viewed by an academic advisor, in theory, they will not only know that the student has been flagged, but which specific factors – a recent drop in platform activity, missed quizzes, declining attendance – have been most significant to that particular flag; thereby, in theory, providing a more actionable and more feature-level explanation of the risk score which can inform a targeted conversation or intervention.

When taken together, the model comparison, feature selection, and SHAP results present a consistent, reinforcing narrative: a well-tuned ensemble model can flag at-risk students with good accuracy based on routinely collected institutional data, and the features driving good accuracy are heavily weighted to ongoing assessment performance and day-to-day student engagement with the educational system, rather than any singular dramatic event, addressing the black-box concern that has traditionally hindered the adoption of machine learning tools of this type in the education context.

Conclusion

The purpose of this study was to create and articulate a machine-learning system that can identify early in the student's trajectory at the university whether there is a risk of developing a learning difficulty based on information obtained from the presence/absence of students, the scores of the assignments, the scores of the quizzes, the exam scores, the level of activity on the learning platform, and the study hours. The dataset was cleaned and features with the highest amount of information and most significant using both mutual information and recursive feature elimination were selected, followed by training and comparing three classifiers, where XGBoost showed the best performance in terms of accuracy, precision, recall, F1-score and ROC-AUC. When using SHAP analysis on the resulting XGBoost model, it became clear that learning-platform

activity, quiz performance and attendance were the most important factors in predicting at risk, and the direction of each of these factors' influence – both on the individual prediction and at the dataset level. The results in this domain are more representative of the capability to achieve good predictive accuracy and good interpretability, not the need to sacrifice one for the other, and show the special value of learning-platform usage data as an early, day-to-day indicator of academic risk, to complement the predictive power of periodic assessment scores, which in this study's illustrative results outperformed it.

Recommendations

Based on these results, we recommend universities who are considering implementing predictive analytics for early identification of at-risk students to look for models with better performance on this type of structured and multi-source educational data, like XGBoost or Random Forest, instead of simpler linear models; and that any such model deployed be accompanied by an explainability technique like SHAP, or similar, rather than simply deployed as an unexplained risk score. This study also suggests that institutions should invest in a consistent collection and integration of learning-platform activity data, especially because it appears to be the strongest predictive indicator present; otherwise, institutions that are primarily dependent on periodic assessment data may be missing out on an earlier warning signal – and one that is more informative, in some aspects. It is also suggested that any use of such a system should be accompanied by explicit institutional rules about communicating through SHAP-based explanations to advisors, instructors, and students themselves, so flagged students can receive specific, actionable feedback instead of an unexplained label, and so the system helps rather than supplants the professional judgment of academic staff. Finally, the findings of this study are not generalizable to other institutions, other disciplines, or other educational levels; future research should test the generalizability of these findings across a broader range of institutions, disciplines, and educational levels; the study should be replicated to determine whether the importance of learning-platform activity, quiz performance, and attendance as found here persist over time in the academic outcomes of students identified as at risk; and the longer-term effect of SHAP-informed interventions on the academic outcomes of students identified as at risk should be explored.

References

1. Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1), 3–17.
2. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
3. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). Association for Computing Machinery. <https://doi.org/10.1145/2939672.2939785>
4. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>
5. Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, 46(1–3), 389–422. <https://doi.org/10.1023/A:1012487302797>
6. Hussain, M., Zhu, W., Zhang, W., & Abidi, S. M. R. (2018). Student engagement predictions in an e-learning system and their impact on student course assessment scores. *Computational Intelligence and Neuroscience*, 2018, Article 6347186. <https://doi.org/10.1155/2018/6347186>
7. Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, 97(1–2), 273–324. [https://doi.org/10.1016/S0004-3702\(97\)00043-X](https://doi.org/10.1016/S0004-3702(97)00043-X)

8. Kotsiantis, S. B. (2012). Use of machine learning techniques for educational proposes: A decision support system for forecasting students' grades. *Artificial Intelligence Review*, 37(4), 331-344. <https://doi.org/10.1007/s10462-011-9234-x>
9. Kotsiantis, S. B., Pierrakeas, C. J., & Pintelas, P. E. (2004). Predicting students' performance in distance learning using machine learning techniques. *Applied Artificial Intelligence*, 18(5), 411-426. <https://doi.org/10.1080/08839510490442058>
10. Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1), 56-67. <https://doi.org/10.1038/s42256-019-0138-9>
11. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 4765-4774). Curran Associates.
12. Marbouti, F., Diefes-Dux, H. A., & Madhavan, K. (2016). Models for early prediction of at-risk students in a course using standards-based grading. *Computers & Education*, 103, 1-15. <https://doi.org/10.1016/j.compedu.2016.09.005>
13. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135-1144). Association for Computing Machinery. <https://doi.org/10.1145/2939672.2939778>
14. Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 33(1), 135-146. <https://doi.org/10.1016/j.eswa.2006.04.005>
15. Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, 40(6), 601-618. <https://doi.org/10.1109/TSMCC.2010.2053532>
16. Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, 51(1), 368-384. <https://doi.org/10.1016/j.compedu.2007.05.016>
17. Waheed, H., Hassan, S.-U., Aljohani, N. R., Hardman, J., Alelyani, S., & Nawaz, R. (2020). Predicting academic performance of students from VLE big data using deep learning models. *Computers in Human Behavior*, 104, Article 106189. <https://doi.org/10.1016/j.chb.2019.106189>



2026 by the authors; Journal of *ComputeX - Journal of Emerging Technology & Applied Science*. This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).