Interdisciplinary Journal of Virtual Learning in Medical Sciences

Original Article

Using Ensemble Machine Learning and Feature Engineering to Increase the Accuracy of Predicting Learners' Performance in an Online Educational Environment

Seyede Fatemeh Noorani[1]*, Maryam Karimi[2], Zahra Gholijafari[1]

[1]Department of Information Technology and Computer Engineering, Payame Noor University, Tehran, Iran
[2]Department of Computer Sciences, Faculty of Mathematical Sciences, Shahrkord, Iran

ABSTRACT

Background: Online training has gained popularity as an effective teaching method, necessitating diligent monitoring of learner progress and engagement. The challenge of predicting academic performance in online courses is crucial for supporting learners at risk of academic loss. This study aimed to develop a robust model for predicting learners' performance using ensemble machine learning and feature engineering techniques.

Methods: This research employed a classification approach based on the Digital Electronic Education and Design Suite (DEEDS) dataset, which records real-time interactions of learners within an online educational environment. The dataset analyzed in this research included activity logs from 115 undergraduate students majoring in computer engineering who participated in a digital electronics course at the University of Genoa, Italy, between September and December 2015. Various machine learning algorithms, including Random Forest (RF), Adaptive Boosting (AdaBoost), Gradient Boosting (GB), Light Gradient-Boosting Machine (LightGBM), and eXtreme Gradient Boosting (XGBoost), were applied. The study also utilized ensemble learning methods such as Boosting and Stacking to enhance prediction accuracy. Feature engineering techniques were implemented to extract and select relevant features from the dataset, leading to the development of a predictive model.

Results: The proposed model achieved an accuracy of 97.43%, a precision of 96.20%, and an F1-score of 98.06%, indicating an acceptable predictive capability. Notably, the findings revealed that feature selection significantly enhanced performance; in the absence of feature selection, the accuracy dropped to 92.15%. Additionally, ensemble methods like Boosting and Stacking provided a 15% enhancement in prediction accuracy compared to traditional approaches. Overall, the integration of feature engineering and ensemble techniques acceptably optimized the model's ability to predict learners' academic performance in online educational settings.

Conclusion: This research validates the effectiveness of employing ensemble machine learning techniques and feature engineering in predicting learners' academic performance in online education. Future studies should explore additional ensemble methods and incorporate diverse feature types to enhance prediction accuracy.

Keywords: Information Science, Supervised Machine Learning, Educational, Data Mining, Dimensionality Reduction, Computer-Assisted

*Corresponding author:
Seyede Fatemeh Noorani, Department of Information Technology and Computer Engineering, Payame Noor University, Tehran, Iran
Tel: +98 21 22455076
Email: sf.noorani@pnu.ac.ir

Please cite this paper as:
Noorani SF, Karimi M, Gholijafari Z. Using Ensemble Machine Learning and Feature Engineering to Increase the Accuracy of Predicting Learners' Performance in an Online Educational Environment. Interdiscip J Virtual Learn Med Sci. 2024;15(4):369-387.doi:10.30476/ijvlms.2024.101157.1279.

Received: 23-12-2023
Revised: 04-11-2024
Accepted: 21-11-2024

Introduction

Learning analytics involves the measurement, collection, analysis, and reporting of data pertaining to students and their contexts. Its primary objective is to enhance the understanding and optimization of learning processes and the contexts in which they occur (1). This interdisciplinary domain draws upon principles from various fields, including machine learning, artificial intelligence, information retrieval, statistics, and data visualization. Learning analytics can be categorized into four primary types: descriptive, diagnostic, prescriptive, and predictive analytics, each addressing distinct facets of learner data (2, 3). Descriptive analytics summarizes historical data to provide insights into past learner behaviors and engagement levels. In contrast, diagnostic analytics investigates the underlying reasons for performance and engagement issues. Prescriptive analytics offers actionable recommendations to enhance learning outcomes, while predictive analytics employs algorithms to forecast future learners' academic performance based on historical data (2-4).

Predicting learners' academic performance is essential in personalized education, as it provides educators with valuable insights to modify their teaching strategies for students encountering difficulties. This predictive ability enables the anticipation of learners' results on future evaluations, thereby mitigating the likelihood of academic failure and maintaining the integrity of e-learning. Additionally, comprehending students' potential contributes to the formulation of suitable plans for their educational trajectories, fostering a greater awareness of their capabilities (5). This aspect is vital for teachers and administrators, enabling them to monitor student progress and tailor educational programs to maximize learning improvements. Furthermore, it can help mitigate administrative challenges, such as dismissals, which is particularly important in developing countries where fostering a responsible generation is essential for national advancement (6). Therefore, predicting learners' academic performance remains a challenging yet critical area that contributes to ensuring teaching quality and facilitating future development (7).

A prominent tool for predicting learners' academic performance is data mining (8). Data mining has demonstrated significant success across various sectors, including business development, e-commerce, and education. Its applications in education are expanding rapidly (9). This technique identifies patterns within data and uncovers hidden insights in datasets, leading to more effective decision-making (10). A data mining algorithm consists of a sequence of procedures designed to extract practical information and develop classification and prediction models through discovering patterns in datasets (11). Educational planners can leverage data mining to identify critical issues related to improving academic quality and analyze learners' achievements, needs, challenges, and learning habits (9).

Recent studies have utilized five machine learning algorithms—Support Vector Machine (SVM), Random Forest (RF), Naive Bayes (NB), Logistic Regression (LR), and Multi-layer Perceptron (MLP)—to predict learners' academic performance influenced by online class behavior, with RF yielding the highest accuracy (11). In a recent study, researchers analyzed learners' online activities serving as input and exam results as output. The findings revealed that Gradient Boosting (GB), a hybrid machine learning technique, demonstrated superior performance, achieving a correlation coefficient of 0.7558 and the lowest root mean square error of 9.3595 (12).

While initial machine learning methods provided valuable insights into predicting learner performance, their accuracy was frequently constrained by inadequate feature representation. The application of feature engineering markedly improves these models by converting data into more pertinent features, which in turn enhances prediction accuracy and supports better generalization of learners' academic outcomes in online settings (13). Several studies have indicated that performing feature engineering prior to data classification and regression is essential to improving the accuracy of predicting learners' academic performance in online classes (12-14).

Another effective strategy for enhancing prediction accuracy involves employing ensemble machine-learning methods (15). Ensemble learning is a machine learning approach that improves accuracy and robustness in forecasting by combining predictions from various models (16). These algorithms have numerous practical applications for enhancing prediction accuracy (17). By integrating several base machine learning algorithms, ensemble methods improve result accuracy, reduce overfitting, and enhance flexibility across diverse classification challenges (15).

In recent years, researchers have made significant strides in predicting learner efficiency and presented optimal algorithms in this domain. Notably, various studies have utilized the Digital Electronics Education and Design Suite (DEEDS) dataset (10).

DEEDS is a simulation platform designed for e-learning in digital electronics. It offers educational resources via dedicated browsers tailored for students, who are tasked with addressing a range of problems with different levels of difficulty (18).

A pioneering study examined the influence of learners' behavior during tests on their scores, employing complexity matrices and data processing techniques to establish relationships between scores and behaviors. The findings revealed a positive correlation between the complexity matrix and learners' scores, alongside a negative correlation with session difficulty levels (10). Additional research employed machine learning algorithms, mainly Artificial Neural Networks (ANNs) and SVMs, to analyze DEEDS data, focusing on features like average time and keystrokes to predict students grades, demonstrating superior accuracy compared to other methods (19). Findings suggest that ANNs yield the highest accuracy in predicting student performance by leveraging engagement and historical data, while demographic factors exert minimal influence (18). In a recent study, researchers predicted learners' academic performance based on online class behavior using algorithms such as ANN, LR, SVM, NB, and Decision Tree (DT). The results indicated that SVM accurately predicted 94% of learners' academic performance (9).

A developed prediction model extracted 86 unique statistical features categorized by activity type, timing, and peripheral activity count, with the RF classifier achieving the highest classification accuracy of 97.4% (20). This study explored different algorithm combinations, such as FCM-MLP and FCM-RF, presenting detailed outcomes for each approach (20). Various machine learning techniques, including Fuzzy C-means, MLP, LR, and RF, were evaluated to predict students' performance. A recent investigation utilizing the DEEDS dataset implemented five machine learning algorithms, including SVM, RF, LR, MLP, and NB, to predict learners' efficiency based on behavior, with RF achieving the best accuracy of 97.4% (21). Previous research emphasized the necessity of predicting learners' academic performance while addressing the lack of effective unbalanced data processing mechanisms that accurately capture learners' characteristics and progress. In a recently proposed method aimed at tackling the challenges associated with unbalanced data while improving the predictive performance for final course grades, Open University Learning Analytics (OULAD) and Semi-Supervised Learning and Progressive Distillation (SPD) datasets were used, outperforming Convolutional Neural Network (CNN), SVR, RF, Categorical Boosting (CatBoost), and eXtreme Gradient Boosting (XGBoost) models (19).

Despite the variety of existing methods for predicting learner performance, incorporating feature engineering and ensemble machine learning techniques can acceptably enhance results. Feature engineering optimizes input data by highlighting critical patterns, while ensemble techniques combine multiple models to improve robustness and accuracy. By leveraging these approaches, superior predictive performance can be achieved, showcasing their effectiveness in this domain.

The current research is significant for two main reasons: First, it utilizes ensemble learning algorithms, including RF, GB, XGBoost, Light Gradient-Boosting Machine (LightGBM), Adaptive Boosting (AdaBoost), and Stacking. Second, it employs feature engineering, which is anticipated to enhance performance accuracy compared to previous studies. Consequently, this study seeks to investigate the potential for achieving greater accuracy in predicting learners' academic effectiveness via the use of ensemble machine learning techniques and feature engineering.

Methods

Study Design and Setting

In the present research, the predictive analytics method was applied, and ensemble machine learning and feature engineering were used to predict learners' performance. The study is based on the prepared DEEDS dataset, which is accessible from https://github.com/ZhenghaoXiao32/educational-process-mining.

Initially, pre-processing was performed to eliminate missing data and prepare the dataset for the subsequent data mining procedures. Subsequently, feature engineering techniques, encompassing both feature extraction and selection, were implemented. Following this, an ensemble machine learning approach was employed to predict learners' performance. Finally, the results were assessed. Figure 1 illustrates the steps of implementing the proposed model for current research.

Participants and Sampling

The DEEDS dataset was used to capture real-time in-class learners' interactions and

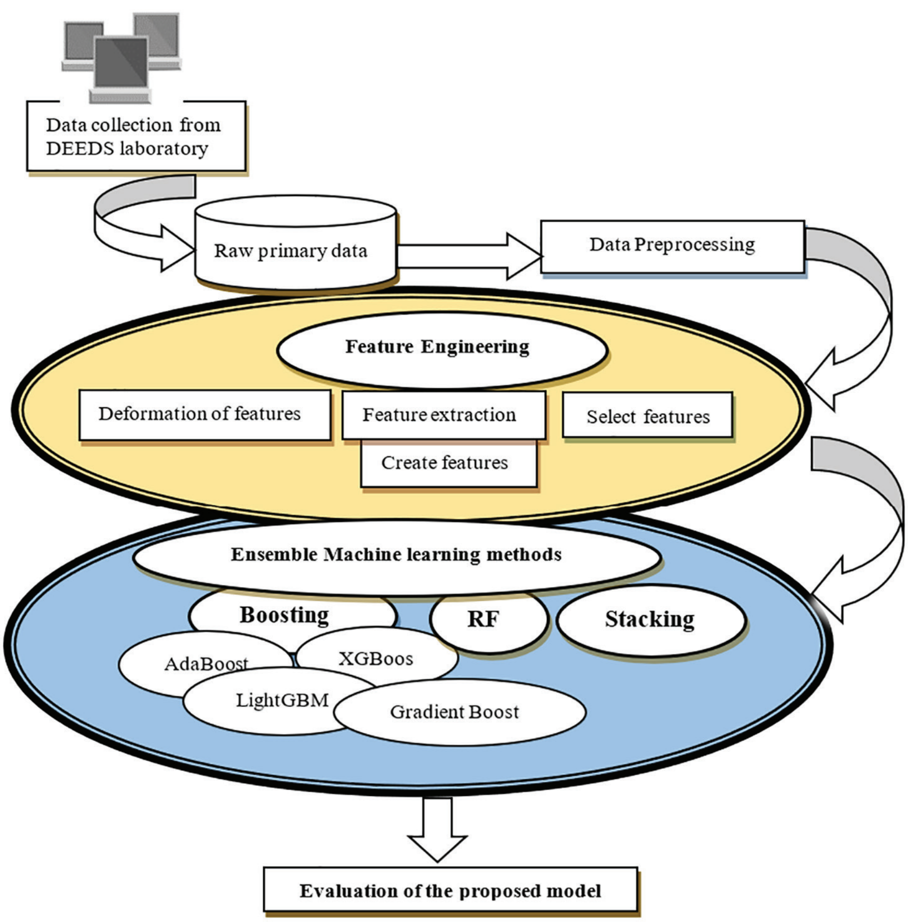

Figure 1: The process of implementing the proposed model. *RF: Random Forest; LightGBM: Light Gradient-Boosting Machine; XGBOOST: eXtreme Gradient Boosting; AdaBoost: Adaptive Boosting

behaviors engaged in a registered technology-enhanced learning platform. The dataset analyzed in this research included activity logs from 115 undergraduate students majoring in computer engineering who participated in a digital electronics course at the University of Genoa, Italy, between September and December 2015. In this paper, the whole dataset was used, and only seven learners who had not attended any of the sessions were removed during the pre-processing stage.

Tools/Instruments

The Visual Basic for Application (VBA) coding modules in Excel were used in the feature extraction phase, and the Python language in Visual Studio, commonly known as the VS code environment, was used to program ensemble machine learning.

Validity and Reliability - In the field of machine learning, when a model is proposed, the values derived from the confusion matrix are utilized to evaluate the accuracy and reliability of the proposed model. In a binary case where we have positive and negative classes, the confusion matrix comprises the following values: True Positive (TP), which represents the count of positive instances accurately identified by the model. True Negative (TN), which represents the count of negative cases correctly identified by the model. False Positive (FP), which represents the count of negative instances wrongly identified as positive by the model. False Negative (FN), which represents the count of positive instances wrongly identified as negative by the model.

Based on TP, TN, FP, and FN, four criteria are defined to evaluate the model:

Accuracy: It describes how closely measurements align with the actual value of the quantity being assessed.

Precision: It can be described as the ratio of accurate predictions for the positive class in relation to the overall count of positive predictions generated.

Recall: also known as sensitivity, serves as a metric to evaluate how frequently a machine learning model accurately recognizes positive instances compared to the total number of genuine positive samples in the dataset.

F1-score: this measure specifies a harmonic mean of recall and precision (22).

In this research, all four mentioned metrics were utilized to assess the proposed model. Table 1 shows the calculation of the mentioned criteria based on confusion matrix parameters.

Data Collection

DEEDS is an advanced simulation platform for e-learning in digital electronics. It includes three design tools: Digital Circuit Simulator (Deeds-DcS), Finite State Machine Simulator (Deeds-FsM), and Micro Computer Emulator (Deeds-McE) to train digital electronics. At the time of composing this document, the latest released version of the software was 2.50.200, published on February 18, 2022. This version is available for installation and use on both Windows and Mac operating systems. Additional information is available from www.digitalelectronicsdeeds.com.

Table 1: Evaluation criteria for the proposed model based on confusion matrix parameters

Criteria	Formula
Accuracy	$\frac{(TP+TN)}{(TP+TN+FP+FN)}$
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$
F1-score	$\frac{2TP}{2TP+FP+FN}$

*TP: True Positive; TN: True Negative; FP: False Positive; FN: False Negative

In this study, the DEEDS' log file was used, which is available from https://github.com/ZhenghaoXiao32/educational-process-mining.

As previously noted, the DEEDS dataset encompasses 115 first-year engineering students who interacted with the system to study digital electronics throughout six instructional sessions. Throughout this instructional period, a detailed collection of log data was gathered, capturing various aspects of learners' activity, including time spent on tasks, mouse clicks, and keystrokes. In total, the dataset consists of 230,318 individual records. At the conclusion of each session, students were tasked with completing an assignment, and their corresponding grades were documented in the intermediate grade dataset.

The number of assignments varies across the six sessions. Sessions 1, 3, and 5 each have four assignments, while sessions 2 and 6 contain six assignments. Session 4 has five assignments. At the end of the semester, the final exam grades for all participants were recorded in the final grades dataset. Table 2 provides a detailed description of the features of this dataset.

Out of the 115 students reported in the dataset, seven did not attend any of the sessions; the information for these seven students was excluded from the analysis of the dataset. Table 3 provides a selection of entries from the DEEDS database, featuring details from the first session of two students.

Figure 2 shows the distribution of DEEDS entries by session. The majority of logged data (23%) belong to session 6, and the fewest belong to session 3.

Based on the DEED's log file, there are 15 activities. Figure 3 indicates the activities and their frequency in the DEEDS dataset. Some activities lacked adequate representation. For instance, "Text Editor no Exercise: 0.02%", "FsM related: 0.1%", and "Deeds no exercise: 0.2%". The highest frequency is associated with the activity "Text Editor Exercise," which accounts for 16.33%, followed closely by "Deeds Exercise" at 15.92%. The "Text Editor Exercise" involves the process of students documenting the results of their work using a text editor, such as Word, for eventual submission to the instructor. In contrast, the "Deeds Exercise" pertains to students engaging in specific tasks within the Deeds simulator.

Data Analysis

Data pre-processing is the first stage of most machine learning algorithms, which includes transforming the data or removing duplicate, noisy, unrelated, or incomplete data from the raw data. In this research, to prepare the data, the following operations performed on DEEDS:

The incomplete information was removed. Session 1 was excluded because no score was recorded for it, and as the efficiency prediction relied on the score, deleting this session was necessary.

Table 2: Features of the DEEDS dataset

Feature Name	Description
Session	• The number of lab instructional sessions ranges from 1 to 6
Student-ID	• The identification number of students varies from 1 to 115
Exercise	• The identification number of the exercise a student is engaged in from 1 to 6 for each session
Activity	• The form of activity is categorized from 1 to 15
Start_Time	• The start date and time of a specific action in form of dd.mm.yyyy hh:mm:ss
End_Time	• The end date and time of a specific task in form of dd.mm.yyyy hh:mm:ss
Idle_time	• The idle time between the begin and end time of a task in milliseconds
Mouse_wheel	• The quantity of mouse wheel usage executed while performing a task
Mouse_wheel_click	• The total count of mouse wheel clicks executed while performing a task
Mouse_click_left	• The total count of left mouse clicks executed while performing a task
Mouse_click_right	• Total number of right mouse clicks executed while performing a task
Mouse_movement	• The distance traveled by the mouse throughout the task
Keystrokes	• The total count of keystrokes executed while performing a task

Table 3: Entries from the DEEDS dataset featuring the first session of two participants

Session	Student-ID	Exercise	Activity	Start_Time	End_Time	Idle time	Mouse_wheel	Mouse_Wheel_click	Mouse_click_left	Mouse_click_right	Mouse_movement	Keystrokes
1	1	Es_1_1	Study_Es_1_1	2.10.2014 11:27:18	2.10.2014 11:27:45	165188	0	0	4	0	715	0
		Es_1_1	Deeds_Es_1_1	2.10.2014 11:27:46	2.10.2014 11:27:49	234	0	0	2	0	214	0
		Es_1_1	Deeds_Es_1_1	2.10.2014 11:27:51	2.10.2014 11:33:57	11510470	0	0	230	54	16970	7
		Es_1_2	Deeds_Es_1_2	2.10.2014 12:6:34	2.10.2014 12:6:54	6868	0	0	23	2	1410	0
		Es_1_2	Deeds_Es_1_2	2.10.2014 12:6:56	2.10.2014 12:9:46	1138015	0	0	128	18	9375	26
		Es_1_2	Study_Es_1_2	2.10.2014 12:9:47	2.10.2014 12:9:50	206	1	0	2	0	242	0
	2	Es_1_2	Deeds_Es_1_2	2.10.2014 12:22:51	2.10.2014 12:23:3	31312	0	0	0	0	0	0
		Es_1_2	TextEditor_Es_1_2	2.10.2014 12:23:17	2.10.2014 12:24:20	293130	0	0	0	0	0	118
		Es_1_2	Deeds_Es_1_2	2.10.2014 12:24:21	2.10.2014 12:24:25	47	0	0	0	0	0	0
		Es_1_2	TextEditor_Es_1_2	2.10.2014 12:24:26	2.10.2014 12:25:41	343476	0	0	0	0	0	50
		Es_1_2	Deeds_Es_1_2	2.10.2014 12:25:42	2.10.2014 12:26:33	338434	0	0	0	0	0	0

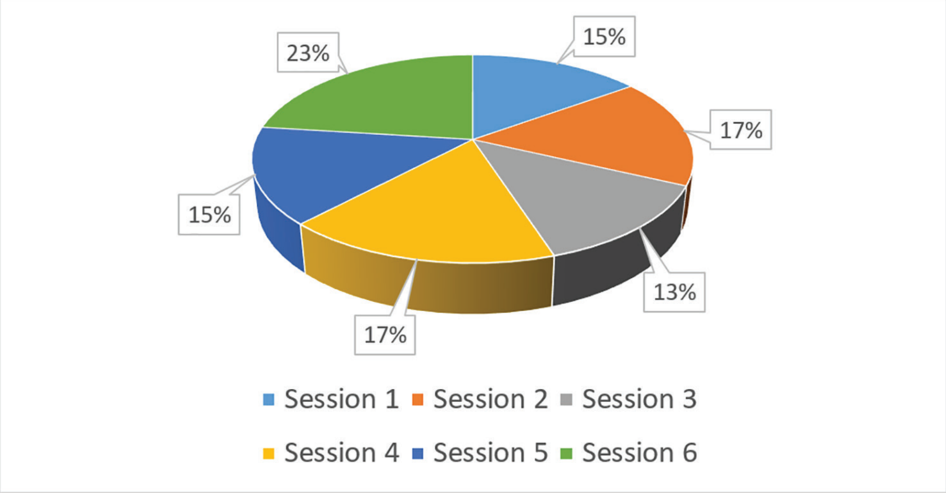

Figure 2: Dataset distribution across sessions (20)

At this stage, less than 1% of the six different activity types were deleted, and 9 other activities remained to be used in data analysis.

In this step, the data with mismatches were removed. One is a mismatch between Session and Exercise, and the other is a mismatch between Exercise and Activity.

The start and end times for each exercise in the initial dataset were recorded in days, hours, minutes, and seconds. These values were converted into seconds for analytical purposes. Additionally, the idle time in the initial dataset, originally measured in milliseconds, was also transformed into seconds.

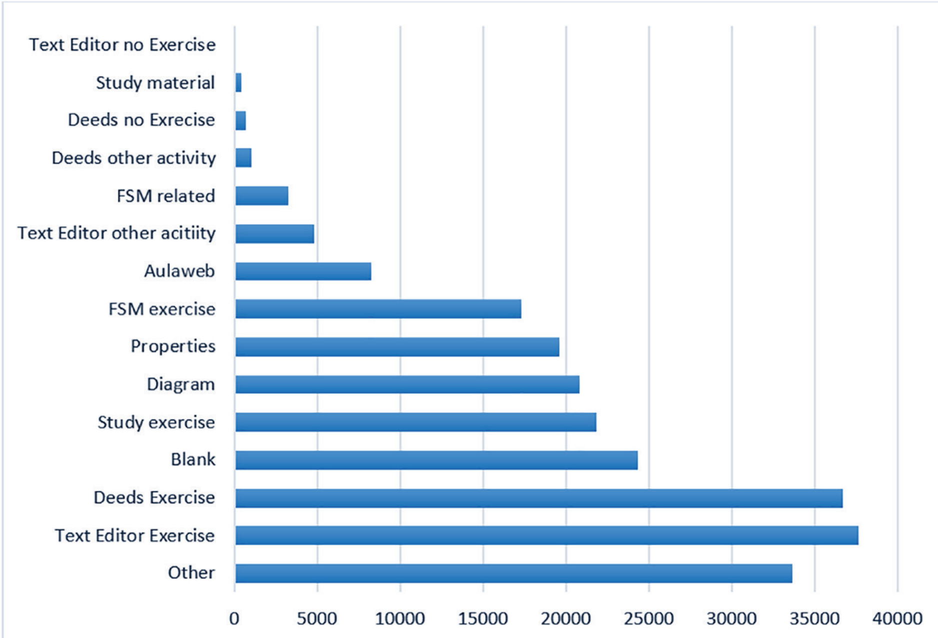

Figure 3: The proportional frequency of activities in the DEEDS dataset (20)

Feature Extraction and Selection

Feature extraction involves converting raw data into numerical features that can be analyzed while keeping the information from the original dataset. It is necessary to change the representation of pre-processed data to be used as their inputs (15). This type of change is called feature extraction, which is considered a key issue in the area of machine learning research. In the field of education, there are different types of features, including 1) Features of the learner's previous performances, 2) Demographic characteristics of the learner, 3) The characteristics related to the interactive behavior of the learner in the class, 4) Personality traits to better explain the learner's abilities, and 5) The characteristics of the educational institution to further explain the educational methods and strategies (10).

In this study, a total of 124 features were extracted from the initial dataset, which was then compiled for each learner to create new records for subsequent analysis. These features are detailed in Table 4 and encompass:

Features related to activities: Initially, nine frequent activities (Aula web, Blank, Deeds, Diagram, FsM, Other, Properties, Study, and Text Editor) were selected. Each exercise was calculated and recorded separately. Since we had six exercises, the number of features to represent the number of activities in each exercise was 6 × 9 = 54. Then, the total number of activities for each session and learner was calculated, which was nine features. The sum of all activities in each exercise was then calculated, which added six separate features to our dataset. Finally, the sum of all activities performed in all exercises is stored as one feature. So, the total number of features in this stage was 54+9+6+1=70. Seven empty features were removed. Thus, the total number of features in this step reached 63.

Features related to time: In the initial dataset, the time spent on each activity was recorded separately in terms of seconds. In this step, the total time spent on each exercise was computed, which included six features. The total idle time in each exercise was also recorded, which provided the second six features. So, the total number of time-related features was 6+6=12.

Features related to mouse and keyboard data: Within the DEEDS dataset, data pertaining to Mouse wheel click, Mouse click (left, right), Mouse movement, Mouse wheel, and Keystrokes have also been recorded.

Table 4: The extracted features from the initial dataset

Feature	Explanation	Feature	Explanation	Feature	Explanation	Feature	Explanation
F1	Number of activity 1 in exercise 1	F32	Number of activity 5 in exercise 4	F63	Sum of activity 9 in session	F94	Number of Mouse wheel in exercise 4
F2	Number of activity 2 in exercise 1	F33	Number of activity 6 in exercise 4	F64	Time of exercise 1	F95	Number of Mouse wheel click in exercise 4
F3	Number of activity 3 in exercise 1	F34	Number of activity7 in exercise4	F65	Time of exercise 2	F96	Number of Mouse click left in exercise 4
F4	Number of activity 4 in exercise 1	F35	Number of activity 8 in exercise 4	F66	Time of exercise 3	F97	Number of Mouse click right in exercise 4
F5	Number of activity 5 in exercise 1	F36	Number of activity 9 in exercise 4	F67	Time of exercise 4	F98	Number of Mouse movement in exercise 4
F6	Number of activity 6 in exercise 1	F37	Number of activity 1 in exercise 5	F68	Time of exercise 5	F99	Number of Key strokes in exercise 4
F7	Number of activity 7 in exercise1	F38	Number of activity 2 in exercise 5	F69	Time of exercise 6	F100	Number of Mouse wheel in exercise 5
F8	Number of activity 8 in exercise 1	F39	Number of activity 3 in exercise 5	F70	Idle Time of exercise 1	F101	Number of Mouse wheel click in exercise 5
F9	Number of activity 9 in exercise 1	F40	Number of activity 4 in exercise 5	F71	Idle Time of exercise 2	F102	Number of Mouse click left in exercise 5
F10	Number of activity 1 in exercise 2	F41	Number of activity 5 in exercise 5	F72	Idle Time of exercise 3	F103	Number of Mouse click right in exercise 5
F11	Number of activity 2 in exercise 2	F42	Number of activity 6 in exercise 5	F73	Idle Time of exercise 4	F104	Number of Mouse movement in exercise 5
F12	Number of activity 3 in exercise 2	F43	Number of activity 7 in exercise 5	F74	Idle Time of exercise 5	F105	Number of Key strokes in exercise 5
F13	Number of activity 4 in exercise 2	F44	Number of activity 8 in exercise 5	F75	Idle Time of exercise 6	F106	Number of Mouse wheel in exercise 6
F14	Number of activity 5 in exercise 2	F45	Number of activity 9 in exercise 5	F76	Number of Mouse wheel in exercise 1	F107	Number of Mouse wheel click in exercise 6
F15	Number of activity 6 in exercise 2	F46	Number of activity 1 in exercise 6	F77	Number of Mouse wheel click in exercise 1	F108	Number of Mouse click left in exercise 6
F16	Number of activity 7 in exercise 2	F47	Number of activity 2 in exercise 6	F78	Number of Mouse click left in exercise 1	F109	Number of Mouse click right in exercise 6
F17	Number of activity 8 in exercise 2	F48	Number of activity3 in exercise6	F79	Number of Mouse click right in exercise1	F110	Number of Mouse movement in exercise 6

Feature	Explanation	Feature	Explanation	Feature	Explanation	Feature	Explanation
F18	Number of activity 9 in exercise 2	F49	Number of activity 4 in exercise 6	F80	Number of Mouse movement in exercise1	F111	Number of Key strokes in exercise 6
F19	Number of activity 1 in exercise 3	F50	Number of activity 5 in exercise 6	F81	Number of Key strokes in exercise 1	F112	Sum of Mouse and Keyboard data in exercise 1
F20	Number of activity 2 in exercise 3	F51	Number of activity 6 in exercise 6	F82	Number of Mouse wheel in exercise 2	F113	Sum of Mouse and Keyboard data in exercise 2
F21	Number of activity 3 in exercise 3	F52	Number of activity 7 in exercise 6	F83	Number of Mouse wheel click in exercise 2	F114	Sum of Mouse and Keyboard data in exercise 3
F22	Number of activity 4 in exercise 3	F53	Number of activity 8 in exercise 6	F84	Number of Mouse click left in exercise2	F115	Sum of Mouse and Keyboard data in exercise 4
F23	Number of activity 5 in exercise 3	F54	Number of activity 9 in exercise 6	F85	Number of Mouse click right in exercise2	F116	Sum of Mouse and Keyboard data in exercise 5
F24	Number of activity 6 in exercise 3	F55	Sum of activity 1 in session	F86	Number of Mouse movement in exercise 2	F117	Sum of Mouse and Keyboard data in exercise 6
F25	Number of activity 7 in exercise 3	F56	Sum of activity 2 in session	F87	Number of Key strokes in exercise 2	F118	Number of Mouse wheel in exercise 6
F26	Number of activity 8 in exercise 3	F57	Sum of activity 3 in session	F88	Number of Mouse wheel in exercise 3	F119	Number of Mouse wheel click in exercise
F27	Number of activity 9 in exercise 3	F58	Sum of activity 4 in session	F89	Number of Mouse wheel click in exercise 3	F120	Number of Mouse click left in exercise
F28	Number of activity 1 in exercise 4	F59	Sum of activity 5 in session	F90	Number of Mouse click left in exercise 3	F121	Number of Mouse click right in exercise
F29	Number of activity 2 in exercise 4	F60	Sum of activity 6 in session	F91	Number of Mouse click right in exercise3	F122	Number of Mouse movement in exercise
F30	Number of activity 3 in exercise 4	F61	Sum of activity 7 in session	F92	Number of Mouse movement in exercise 3	F123	Number of Key strokes in exercise
F31	Number of activity 4 in exercise 4	F62	Sum of activity 8 in session	F93	Number of Key strokes in exercise 3	F124	Number of Mouse wheel in exercise

Features were developed to represent these activities across six distinct exercises, resulting in a total of 36 individual features. Furthermore, the sum of these activities was calculated for each exercise, yielding an additional six features. The overall sum of each activity across all exercises contributed six more features to the feature vector. Lastly, the total sum of all activities across all exercises was considered as the final feature. Consequently, the total number of side activity features was calculated to be 1+6+ 6+36=49.

Feature selection is one of the stages of feature engineering and is significant in increasing the accuracy of prediction in machine learning algorithms. The feature selection techniques can be categorized into three main types: 1) filter methods, 2) wrapper methods, and 3) built-in methods. For instance, one study employed an online feature selection method known as Alpha, successfully selecting 14 out of 30 features (23). In another investigation, Waikato Environment for Knowledge Analysis (WEKA) was utilized to apply three distinct methods—Gain Ratio, Chi-Square, and Info Gain—for the purpose of feature selection (12). Similarly, Shannon's entropy method was implemented to reduce the feature set and identify the most effective features within the model (21).

In this study, we employed the Select K Best and Chi-Square methods, both of which are classified as wrapper methods. The Select K Best method selects the K features with the maximum scores. In classification tasks utilizing Select K Best, Chi-Square is typically employed as the scoring function. The objective of using these methods is to identify a combination of features that maximizes prediction accuracy in the selected model. Given that the optimal value of K for achieving the highest accuracy was not predetermined in this research, the algorithm was executed for various K values ranging from 1 to 124.

Ensemble Machine Learning Algorithms

An ensemble machine learning method is a set of classifiers that are combined in different ways. Extensive research has focused on developing ensemble methods, with many resulting models demonstrating superior accuracy compared to the individual models they comprise (19). The ideal ensemble method consists of several basic high-precision methods that exhibit substantial diversity. When the individual base models produce different errors, the overall error can be minimized. Conversely, if the models are too similar, their combination typically yields no significant improvement, as the outcomes remain unchanged (18). In the following, Boosting, Stacking, and Random Forest (RF) are introduced as highly effective examples of ensemble methods.

Random Forest method: The bagging method is derived from the two bootstrap and aggregating words. It is one of the simple ensemble methods with good results. This method trains various basic models on randomly selected subsets of data with placement, and the maximum votes of the trained basic models are combined to make the final decision.

The RF method is a special version of bagging that combines unpruned classifiers and regression trees. The algorithm that is usually used as a decision tree is Classification and Regression Trees (CART). Like bagging, RF also uses samples with placement. In this method, the trees grow to the maximum depth and each tree performs an independent classification (18).

Boosting ensemble method: Boosting is an ensemble technique that enhances the performance of weak learners by combining them. Most boosting algorithms iteratively train weak classifiers based on their performance relative to a specific distribution, ultimately contributing to a robust final classifier. In this process, misclassified samples are assigned greater weight, while correctly classified samples receive reduced weight. As a result, subsequent weak learners focus more on the cases in which the previous poor learners classified them incorrectly. Among the most widely used boosting algorithms are AdaBoost, Gradient Boosting (GB), LightGBM, and XGBoost.

Stacking ensemble method: Stacking is an improved form of voting technique. It simultaneously runs multiple models on the data and combines the results using a meta-model to build the final model. Unlike Boosting, first-level models are trained in parallel, and their predictions are utilized for the training where the meta-level of the model can be better trained. Unlike Bagging and Boosting, in a stacking ensemble method, different types of models can be combined. The main steps are as follows:

1. Divide the training set into two separate sets.
2. In the first step, use several basic algorithms.
3. In the second step, test the results of the first step.
4. Train the higher-level learner using the predictions from step 3 as input and the correct answers as output.

The initial three steps of the proposed methodology align with the principles of cross-validation; however, rather than employing a winner-take-all approach, the underlying learning machines may be combined non-linearly. In this research, the RF and LightGBM methods were used for the basic models, while LR served as the Meta model.

To achieve data classification, this research implemented six widely recognized hybrid machine learning models to identify the most accurate one. These models were RF, AdaBoost, GB, LightGBM, XGBoost, and Stacking.

Two methods for training and evaluation were employed to mitigate the risk of overfitting the models. In the initial approach, 80% of the data was randomly selected for training, while the remaining 20% was designated for testing. This method is referred to as the 80%-20% split. The random selection and train/test process were repeated a thousand times, and the average of each evaluation criterion across these 1000 iterations was reported.

In the second method, 10-fold cross-validation was utilized. In K-Fold Cross Validation, the dataset is generally divided into k subsets, referred to as folds. The model was trained on all subsets except one (k-1), which was set aside for evaluation. This process was repeated k times, with a different subset reserved for testing during each iteration.

This study examined the significance of various features in the application of Select K Best and Chi-Square feature selection techniques. Ultimately, the accuracy of the model developed in this research was compared to that of existing predictive methods utilized by other studies (7, 9, 16), all of which were examined using the DEEDS dataset.

Ethics - To ensure the privacy of participants, as indicated in Table 3, the DEEDS dataset does not include any personal identifiers, such as names or student numbers; instead, individuals are assigned a unique identification number beginning with 1. This dataset was sourced from [https://github.com/ZhenghaoXiao32/educational-process-mining], where the practice of sharing datasets is well-established within the data mining community. Publicly available datasets on reputable platforms are widely recognized as valuable resources for analysis using diverse algorithms. Such practices not only advance the field of research but also enhance transparency and foster collaboration among scholars. Moreover, the use of these datasets allows for comparative analyses, thereby enabling rigorous evaluation and validation of findings across various studies. Consequently, utilizing this dataset is in accordance with ethical standards in data analysis and promotes the responsible utilization of shared resources in the discipline.

Results

In this research, we examined the impact of feature engineering alongside ensemble learning methods in increasing the accuracy of learners' performance prediction. To achieve this, we initially employed 80% of the data for training and the other 20% (briefly as 80%-20%) and then 10-fold cross-validation without feature selection, utilizing ensemble algorithms (RF, GB, LightGBM, XGBoost). Subsequently, we applied feature selection in conjunction with the ensemble algorithms.

Initially, the entire DEEDS dataset, which comprised 124 features, was utilized. The dataset was partitioned into 80% for training and 20% for testing (80%-20%). Subsequently, ensemble machine learning algorithms were applied. To enhance the reliability of the results, this procedure was repeated 1000 times. The mean values for accuracy, precision, recall, and F1-score are summarized in Table 5.

The findings revealed that the Stacking and AdaBoost models achieved the highest levels of accuracy. Following these, XGBoost and LightGBM ranked as the next most accurate models regarding predictive performance. The Stacking model exhibited the highest precision, with AdaBoost and XGBoost closely trailing. Additionally, the AdaBoost model exhibited superior performance in terms of recall and F1-Score.

In the subsequent step, a 10-fold cross-validation approach was applied. The findings are presented in Table 6.

Based on the results, XGBoost achieved the highest accuracy, followed by RF and then the Stacking model in the subsequent rank. The GB model demonstrated the highest level of precision, with the XGBoost model coming in second. In terms of recall, the Stacking model was the most efficient, followed by the RF model. Additionally, RF obtained the highest F1-Score, ranking just above the XGBoost model.

In the final step, feature selection was applied prior to the implementation of machine learning algorithms. Two feature selection methods were utilized: Select K best and Chi-Square based method. Based on the applied methodologies, 102 out of 124 features were identified as optimal, yielding the highest accuracy. Subsequently, we conducted ensemble machine learning 1,000 times using these 102 selected features, employing an 80% training and 20% testing split. Table 7 shows the average of evaluation criteria.

The results presented in Table 7 indicate that feature selection enhanced predictive performance, with XGBoost achieving the highest precision, recall, and F1-Score among the evaluated ensemble learning algorithms.

Finally, The findings from the current study were compared to the results of

Table 5: The average efficiency of the proposed models with 80%-20% split

Model	Accuracy	Precision	Recall	F1-Score
RF	0.9085	0.9197	0.9431	0.9308
GB	0.8932	0.9264	0.9100	0.9176
LightGBM	0.9163	0.9305	0.9429	0.9363
XGBoost	0.9189	0.9353	0.9418	0.9381
AdaBoost	0.9469	0.9373	0.9860	0.9610
Stacking	0.9469	0.9484	0.9731	0.9606

* RF: Random Forest; GB: Gradient Boosting; LightGBM: Light Gradient-Boosting Machine; XGBOOST: eXtreme Gradient Boosting; AdaBoost: Adaptive Boosting

Table 6: The average efficiency of the proposed model with 10-fold cross-validation

Model	Accuracy	Precision	Recall	F1-Score
RF	0.9183	0.9124	0.9442	0.9268
GB	0.9008	0.9219	0.8886	0.8997
LightGBM	0.8817	0.8785	0.9040	0.8900
XGBoost	0.9200	0.9160	09365	0.9251
AdaBoost	0.9043	0.9150	0.9047	0.9081
Stacking	0.9131	0.8993	0.9494	0.9220

* RF: Random Forest; GB: Gradient Boosting; LightGBM: Light Gradient-Boosting Machine; XGBOOST: eXtreme Gradient Boosting; AdaBoost: Adaptive Boosting

Table 7: The effectiveness of the suggested approach, joined with the features selection

Model	Accuracy	Precision	Recall	F1-Score
RF	0.9652	0.9500	1.0000	0.9743
GB	0.9656	0.9610	0.9763	0.9673
LightGBM	0.9740	0.9620	0.9868	0.9806
XGBoost	0.9743	0.9620	1.0000	0.9806
AdaBoost	0.9740	0.9740	0.9868	0.9803
Stacking	0.9652	0.9500	1.0000	0.9743

* RF: Random Forest; GB: Gradient Boosting; LightGBM: Light Gradient-Boosting Machine; XGBOOST: eXtreme Gradient Boosting; AdaBoost: Adaptive Boosting

Table 8: Accuracy, precision, recall, and F1-Score of the current study compared to previous studies

Studies \ Metrics	Accuracy	Precision	Recall	F1-Score
Maksud and colleagues' study (12)	0.9400	0.9500	0.9800	0.9600
Brahim's study (21)	0.9700	0.9700	0.9700	0.9700
Hussain and colleagues' study (23)	0.8000	0.8300	0.9200	0.8700
Current study's Proposed method	0.9743	0.9620	1.0000	0.9806

previous research (12, 21, 23), all of which employed the DEEDS dataset. Given that the proposed method in the current study demonstrated its highest performance with XGBoost, a comparison was explicitly conducted focusing on the outcomes produced by XGBoost. All methods utilized an 80%-20% split, which was executed multiple times to calculate an average. The results indicate that our method, which combined ensemble machine learning with feature engineering, outperforms the others. The precision and accuracy metrics have shown acceptable improvement compared to other studies (12, 23). Furthermore, the recall and F1-score metrics have improved markedly in comparison to all three previous studies (12, 21, 23). Table 8 presents a comparative analysis between the proposed method and prior studies based on the evaluated metrics.

In a detailed comparative analysis, this study evaluated and contrasted the optimal results reported in the existing literature, the peak performance of the proposed method that employed the ensemble learning algorithm without feature selection, and the

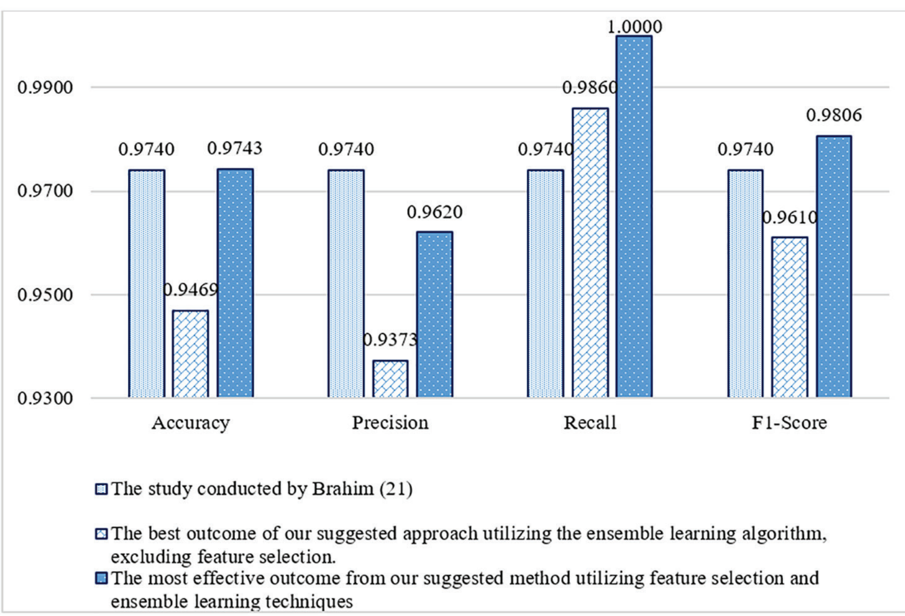

Figure 4: A Comparison of Brahim's research findings (21) with the current study

enhanced outcomes achieved by the method presented in this paper, which integrated both feature selection and the ensemble learning algorithm. Figure 4 demonstrates a comparison between the findings of Brahim's research (21) and the approach proposed in the current study.

As shown in columns 2 to 4 of Table 5, the previous research utilizing the DEEDS achieved the highest evaluation values, with Accuracy=0.9740, Precision=0.9740, Recall=0.9740, and F1-Score=0.9740. In our proposed method, detailed in Tables 5 and 6, AdaBoost produced the best results when using ensemble learning without feature selection. Furthermore, the most favorable outcomes were obtained by combining feature selection with ensemble algorithms, especially when employing the XGBoost algorithm. As illustrated in Figure 4, integrating feature selection with ensemble algorithms leads to higher evaluation results.

Discussion

In this study, we employed an ensemble learning method integrated with feature selection to model the DEEDs dataset effectively. Initially, we conducted modeling without feature selection, utilizing both an 80%-20% data split and 10-fold cross-validation techniques. Following this, we applied feature selection methods, which were instrumental in refining the dataset before constructing the ensemble learning model. The results indicated that our proposed method, which synergistically combined feature selection with ensemble learning, achieved significantly higher accuracy, recall, and F1-Score compared to prior studies (12, 21, 23) that utilized the same dataset. Based on the results obtained in this research, the combination of XGBoost and feature selection achieved an accuracy of 0.9742, a recall of 1, and an F1-Score of 0.9806. These findings surpass those achieved through methods lacking feature selection, as well as the results documented in studies employing the DEEDS dataset. This indicates that feature selection not only streamlines the modeling process but also enhances data quality by reducing noise and mitigating the risks of overfitting.

In the study conducted by Baig and colleagues (20), an integrated approach utilizing machine learning algorithms for predicting students' performance levels was explored. Although the results indicated a reasonable level of accuracy, our focused approach on ensemble algorithms combined with robust feature selection has yielded superior predictive performance. Specifically, the optimized feature selection techniques employed in this research effectively minimized noise and clarified the predictive pathways, thus enhancing the overall model accuracy.

Another study investigated academic performance prediction linked to synchronous online interactive learning behaviors, primarily relied on a single algorithm and overlooked the critical aspect of feature selection (8). In contrast, our application of multiple feature selection methods allowed for the identification of significant and relevant characteristics, which substantially improved prediction accuracy. This highlights the importance of a comprehensive feature selection strategy in developing effective predictive models.

Additionally, the research conducted by Wang and colleagues introduced a comprehensive system for predicting Student Academic Performance, focusing on advanced algorithms (5). This framework, named ProbSAP, effectively tackles the challenges posed by imbalanced datasets. By utilizing collaborative data processing and XGBoost-enhanced prediction, ProbSAP significantly improves prediction accuracy, achieving up to an 84.76% reduction in mean absolute error compared to traditional methods. However, our findings indicate that the combined use of diverse algorithms along with effective feature selection can lead to substantial improvements in model performance. Therefore, the application of varied and practical feature selection techniques in our study has contributed to increased model reliability when compared to the methodologies employed in similar studies.

Ultimately, the results of this research underscore the potential of utilizing hybrid algorithms and feature selection in tandem to enhance predictions of learner performance. This study not only achieved significant improvements in predictive accuracy but also serves as a foundational framework for future research aimed at further developing and optimizing predictive methods in educational contexts.

The theoretical basis for combining feature selection with ensemble learning is rooted in the premise that identifying the most relevant features can lead to more effective model generalization (24). Previous studies have indicated that ensemble methods, designed to leverage the strengths of multiple models through the aggregation of their predictions, generally outperform single classifiers (16, 25, 26). This characteristic makes ensemble methods particularly suitable for complex datasets like DEEDs, where the relationships between features and outcomes are often intricate and non-linear.

These findings have profound practical implications. Educational institutions can significantly enhance their predictive capabilities regarding students' performance by adopting a methodology that prioritizes meaningful feature selection followed by robust ensemble modeling. This approach enables more accurate identification of at-risk students and facilitates tailored interventions, ultimately promoting improved academic outcomes.

Limitations and Suggestions

While the findings of this study demonstrate significant advancements compared to similar research, it is essential to acknowledge certain limitations. Implementing the proposed algorithms and feature engineering processes can be relatively time-consuming. Although previous studies did not report on execution times, our research found that the time spent on feature selection was acceptable. Future research could explore alternative feature selection methods that may reduce execution time without compromising accuracy.

Moreover, the features investigated in this study were limited to behavioral and real-time characteristics of online classes. Incorporating additional information, such as demographic characteristics or comprehensive historical data, could lead to even more accurate predictions. Expanding the feature set in future studies may provide deeper insights into the factors influencing student performance, thereby enhancing the efficacy of predictive models. A further promising area for future investigation is the use of other ensemble machine learning, such as the multi-level stacking method, that can potentially yield improved accuracy and robustness in predictions.

Conclusion

This study investigated the effectiveness of using ensemble learning methods combined with feature selection to predict learners' academic performance in online education. The results indicate that this approach significantly enhances prediction accuracy compared to previous research. These findings highlight the practical benefits for educational institutions aiming to improve their ability to forecast students' performance. By focusing on meaningful feature selection and employing robust ensemble modeling, educators can more effectively identify at-risk students and implement targeted interventions. This research has opened several avenues for future investigation. For instance, exploring the impact of different feature selection methods, such as recursive feature elimination or variance thresholding, on ensemble performance could yield valuable insights. Additionally, investigating the integration of other data types, such as demographic information or personality traits, could further enhance predictive accuracy. This ongoing investigation will contribute to the development of more effective predictive frameworks in education.

Abbreviations

AdaBoost: Adaptive Boosting
ANN: Artificial Neural Network

CART: Classification and Regression Trees
Cat Boost: Categorical Boosting
DEEDS: Digital Electronics Education and Design Suite
DT: Decision Tree
FCM: Fuzzy C-means
GB: Gradient Boosting
JRIP: Repeated Incremental Pruning to Produce Error Reduction
LightGBM: Light Gradient-Boosting Machine
LR: Logistic Regression
LSTM: Long Short-Term Memory
MLP: Multi-Layer Perceptron
NB: Naive Bayes
OULAD: Open University Learning Analytics
RF: Random Forest
SPD: Semi-supervised Learning and Progressive Distillation
SVM: Support Vector Machine
VBA: Visual Basic for Application
WEKA: Waikato Environment for Knowledge Analysis
XGBOOST: eXtreme Gradient Boosting

Acknowledgments

Not applicable.

Authors' Contribution

ZG wrote the Python code and analyzed the results. SFN and MK drafted and finalized the manuscript. All authors reviewed and approved the final version. They take full responsibility for the content and writing of this article.

Conflict of Interest

The authors declare that they have no conflicts of interest to disclose.

Ethical Considerations

This research was carried out in compliance with the local standards set by the Deputy of Research and Technology at Payame Noor University (PNU) of Sharyar, Iran, under the identification code S/7652. To protect participant confidentiality, the DEEDS dataset omits any personal identifiers, such as names or student ID numbers; instead, each individual is assigned a unique identification number starting from 1. Datasets that are publicly accessible on reputable platforms are widely acknowledged as essential resources for analysis employing various algorithms. The use of these datasets adheres to ethical principles in data analysis and fosters the responsible use of shared resources within the field.

Funding/Support

This research did not receive any outside funding or support.

Availability of Data and Materials

The primary data of this study were derived from the following resources available in the public domain: https://github.com/ZhenghaoXiao32/educational-process-mining.

References

1 Siemens G, editor Learning analytics: envisioning a research discipline and a domain of practice. Proceedings of the 2nd international conference on learning analytics and knowledge; 2012.
2 Chatti MA, Dyckhoff AL, Schroeder U, Thüs H. A reference model for learning analytics. International journal of Technology Enhanced learning. 2012;4(5-6):318-31. doi: https://doi.org/10.1504/IJTEL.2012.051815.
3 Qiu F, Zhang G, Sheng X, Jiang L, Zhu L, Xiang Q, et al. Predicting students' performance in e-learning using learning process and behaviour data. Sci Rep. 2022;12(1):453. doi: 10.1038/s41598-021-03867-8. PubMed PMID: 35013396; PubMed Central PMCID: PMCPMC8748729.
4 Banihashem SK, Aliabadi K, Pourroostaei Ardakani S, Delaver A, Nili Ahmadabadi M. Learning analytics: A systematic literature review. Interdisciplinary Journal of Virtual Learning in Medical Sciences. 2018;9(2). doi: https://doi.org/10.5812/ijvlms.63024.

5 Wang X, Zhao Y, Li C, Ren P. ProbSAP: A comprehensive and high-performance system for student academic performance prediction. Pattern Recognition. 2023;137:109309. doi: https://doi.org/10.1109/ISBAST.2013.50.
6 Liu T, Li S. Performance Prediction for Higher Education Students Using Deep Learning. 2022.
7 Yin C, Tang D, Zhang F, Tang Q, Feng Y, He Z. Students learning performance prediction based on feature extraction algorithm and attention-based bidirectional gated recurrent unit network. Plos one. 2023;18(10):e0286156. doi: https://doi.org/10.1371/journal.pone.0286156. PubMed PMID: 37878591; PubMed Central PMCID: PMC10599562.
8 Liang G, Jiang C, Ping Q, Jiang X. Academic performance prediction associated with synchronous online interactive learning behaviors based on the machine learning approach. Interactive Learning Environments. 2023:1-16. doi: https://doi.org/10.1080/10494820.2023.2167836.
9 Batool S, Rashid J, Nisar MW, Kim J, Kwon H-Y, Hussain A. Educational data mining to predict students' academic performance: A survey study. Education and Information Technologies. 2023;28(1):905-71. doi: https://doi.org/10.1007/s10639-022-11152-y.
10 Forero-Corba W, Bennasar FN. Techniques and applications of Machine Learning and Artificial Intelligence in education: a systematic review. RIED-Revista Iberoamericana de Educación a Distancia. 2024;27(1). doi: https://doi.org/10.5944/ried.27.1.37491.
11 Sarker S, Paul MK, Thasin STH, Hasan MAM. Analyzing students' academic performance using educational data mining. Computers and Education: Artificial Intelligence. 2024;7:100263. doi: https://doi.org/10.1016/j.caeai.2024.100263.
12 Maksud M, Ahmad N. Machine learning approaches to digital learning performance analysis. International Journal of Computing and Digital Systems. 2020;10:2-9.
13 Vahdat M, Oneto L, Anguita D, Funk M, Rauterberg M, editors. A learning analytics approach to correlate the academic achievements of students with interaction data from an educational simulator. Design for Teaching and Learning in a Networked World: 10th European Conference on Technology Enhanced Learning, EC-TEL 2015, Toledo, Spain, September 15-18, 2015, Proceedings 10; 2015: Springer.
14 Kou G, Xu Y, Peng Y, Shen F, Chen Y, Chang K, et al. Bankruptcy prediction for SMEs using transactional data and two-stage multiobjective feature selection. Decision Support Systems. 2021;140:113429. doi: https://doi.org/10.1016/j.dss.2020.113429.
15 Zhang Y, Liu J, Shen W. A review of ensemble learning algorithms used in remote sensing applications. Applied Sciences. 2022;12(17):8654. doi: https://doi.org/10.3390/app12178654.
16 Rane N, Choudhary SP, Rane J. Ensemble deep learning and machine learning: applications, opportunities, challenges, and future directions. Studies in Medical and Health Sciences. 2024;1(2):18-41. doi: https://doi.org/10.48185/smhs.v1i2.1225.
17 Ganaie MA, et al. . Recent advances in ensemble machine learning: A comprehensive review. Artificial Intelligence Review. 2023;56(2):1031-65.
18 Tomasevic N, Gvozdenovic N, Vranes S. An overview and comparison of supervised data mining techniques for student exam performance prediction. Computers & education. 2020;143:103676. doi: https://doi.org/10.1016/j.compedu.2019.103676.
19 Wan S, Yang H, editors. Comparison among methods of ensemble learning. 2013 International Symposium on Biometrics and Security Technologies; 2013: IEEE.
20 Baig MA, Shaikh SA, Khatri KK, Shaikh MA, Khan MZ, Rauf MA. Prediction

of Students Performance Level Using Integrated Approach of ML Algorithms. Int J Emerg Technol Learn. 2023;18(1):216-34. doi: https://doi.org/10.3991/ijet.v18i01.35339.
21 Brahim GB. Predicting student performance from online engagement activities using novel statistical features. Arabian Journal for Science and Engineering. 2022;47(8):10225-43. doi: https://doi.org/10.1007/s13369-021-06548-w. PubMed PMID: 35070634; PubMed Central PMCID: PMC8762194.
22 Tan P-N, Steinbach M, Kumar V. Introduction to data mining Pearson Education India. Indian Nursing Council: New Delhi, India. 2018.
23 Hussain M, Zhu W, Zhang W, Abidi SMR, Ali S. Using machine learning to predict student difficulties from learning session data. Artificial Intelligence Review. 2019;52:381-407. doi: https://doi.org/10.1007/s10462-018-9620-8.
24 Bolón-Canedo V, Alonso-Betanzos A. Ensembles for feature selection: A review and future trends. Information Fusion. 2018;52. doi: https://doi.org/10.1016/j.inffus.2018.11.008.
25 Mienye ID, Sun Y. A survey of ensemble learning: Concepts, algorithms, applications, and prospects. IEEE Access. 2022;10:99129-49. doi: https://doi.org/10.1109/ACCESS.2022.3207287.
26 Ajibade S-SM, Bahiah Binti Ahmad N, Mariyam Shamsuddin S, editors. Educational data mining: enhancement of student performance model using ensemble methods. IOP conference series: materials science and engineering; 2019: IOP Publishing.