Cureus Journal of Computer Science

Open Access Original Article

Artificial Intelligence-Based Students' Strengths, Weaknesses, Opportunities, and Challenges (SWOC) Analysis

Received 09/04/2025
Review began 09/14/2025
Review ended 10/27/2025
Published 11/17/2025

DOI:
https://doi.org/10.7759/s44389-025-09909-w

Shraddha Patil [1] , Surabhi Kakade [1] , Laher Patni [1] , Kaivalya Pagrut [1] , Krushna Pawar [1] , Tilak Patel [1]

1. Artificial Intelligence and Data Science, Vishwakarma Institute of Technology, Pune, IND

Corresponding author: Shraddha Patil, shraddha.patil23@vit.edu

Abstract

This research proposes an artificial intelligence-driven framework for thoroughly conducting efficient strengths, weaknesses, opportunities, and challenges (SWOC) analysis of students. A proposed method correctly identifies and categorizes the SWOC of students by integrating rule-based classification with machine learning methods, namely eXtreme Gradient Boosting (XGBoost). A total of 500 responses were collected from students and preprocessed into a dataset. The XGBoost model classified SWOC categories with an accuracy rate of 95.5%. With Streamlit, an easy-to-use web application was developed to host the SWOC analysis tool where users can enter their answers and receive customized SWOC reports. The framework helps teachers and learners understand their SWOC. This leads to better choices and learning experiences that fit their needs.

Categories: AI applications, Data Analysis, Machine Learning (ML)
Keywords: machine learning, xgboost, rule-based system, multi-label classification, student performance

Introduction

Identifying a student's strengths, weaknesses, opportunities, and challenges (SWOC) is imperative in personalized learning, as well as career counseling. Traditional methods are typically opinion- or standard test-based, which cannot discover the unique skill sets and interests of an individual. In this study, an AI-based approach is proposed in an effort to perform comprehensive SWOC analysis of engineering students to overcome this shortcoming. The model uses the hybrid approach that involves the combination of machine learning algorithms, in this case, eXtreme Gradient Boosting (XGBoost) and rule-based classification. The model correctly identifies and classifies independent SWOC components after considering an enormous number of student answers. The task is of immense significance to the XGBoost algorithm, which is of high precision and is able to deal with intricate datasets with ease. The SWOC analysis is done via a simple online application. Both teachers and students can access a perceptive analysis offered by the application. The data can be utilized for creating personal goals and well-informed decisions for future education and career choices. This research addresses a novel approach as it employs a new method for the evaluation of student potential by combining machine learning with rule-based classification. By using AI, the system is more efficient and scalable compared to previous methods. The SWOC analysis provides a clear analysis of the strengths, weaknesses, opportunities, and challenges of each engineering student. This valuable information allows for the optimization of learning outcomes, identification of areas of improvement, and the enabling of individual skill development.

The multi-label feature-aware XGBoost model was proposed by Hanumanthappa and Prakash [1] to solve the problems of class imbalance and multi-label classification. The model provides a strong framework for the estimation of educational data and improves personalized education decision-making by improving prediction accuracy. For personalized guidance, Deshpande et al. [2] highlight the need to identify students' weaknesses and strengths. In the article, the fusion of predictive analysis with continuous feedback can potentially provide more meaningful insights to enhance learning outcomes even if conventional machine learning techniques such as Support Vector Machine, Naive Bayes, and Decision Trees are applied extensively in education data analysis. A paper presented by Abu-Alaish et al. [3] focused on the application of machine learning and data mining techniques such as regression, classification, clustering, and association rules to improve the data analysis while deriving valuable insights. Using Weka, the authors showed that these methods are effective, but they still have limitations and need to be enhanced to better address student satisfaction gaps and improve educational quality assurance systems. In previous research, Khan et al. [4] presented a generalized and interpretable prediction model designed to help the teachers to provide timely interventions for struggling children. By overcoming the limitations of existing localized methods, this model improves the ability to predict and assist students across multiple learning environments and provides greater generalizability. To overcome the limitations of traditional methods for the prediction of academic performance, the hybrid machine learning model was developed by Alshanqiti and Namoun [5]. By combining supervised and unsupervised approaches, the research enhanced the accuracy of prediction and provided deeper insights into the primary major factors that affect students' academic performance. A novel AI-based technique for accessing the performance of students in real-time during online lab experiments was introduced by Abd El-Haleem et al. [6]. The model addresses the need for providing immediate insights while conducting experiments, which is essential for enhancing the learning in virtual environments. Additionally, it enhances the prediction accuracy and allows on time support to students. A study by Karale et al. [7] showed how predictive analytics empowered by AI techniques, like support vector regression, decision trees, and random forests, can estimate student outcomes. The study highlights the need to implement a more comprehensive evaluation of machine learning methods in the academics, especially when students make the transition from high school to college, in order to provide the effective initial support and the resource optimization. The performance of XGBoost as a predictive modeling tool was underscored in a study by Thejovathi and Chandra Sekhara Rao [8], particularly when it involves big data. It added feature extraction methods like Principal Component Analysis and Recursive Feature Elimination, which improve model interpretability while preventing the unnecessary overfitting by highlighting essential prediction factors. Additionally, it utilized regularization and parallel processing to enhance the performance. A study by Maulana et al. [9] questioned the idea that smaller datasets yield high accuracy, which cannot capture the variety of factors that influence student success. The study filled earlier gaps in the literature by comparing different machine learning algorithms on bigger datasets, which enhanced the accuracy of prediction and conducted further analysis of academic outcome.

Materials And Methods

Dataset

Students were subjected to structured questions about their SWOC using a Google Form to generate the dataset. The domains in the questionnaire focused on Academic Performance, Skills and Abilities, Personal Attributes, and Goals and Aspirations. To maintain sufficient data for training and generalization of a model, the dataset was gradually incremented from 100 students to 300 and then to 500 students.

Data Preprocessing

The initial responses were preprocessed to make sure that they were consistent. Preprocessing involved the following actions:

• Handling Missing Data: Unresolved or missing responses were removed.

• Data Transformation: After all the responses were entered into SWOC categories, using multi-hot encoding, the categories were transformed into binary vectors. Thus, each student's response is represented as a vector, indicating the presence of a specific SWOC characteristic.

• Feature Engineering: Key features related to the academic and career life of students were extracted from the responses and used as input features for the model.

SWOC classification using rule-based system

Students' answers were initially organized into SWOC categories through a rule-based system. Using a collection of pre-specified rules, the system was able to find specific keywords and patterns in the text responses. The correct SWOC categories were then used to sort the responses according to these patterns. A predefined set of SWOC categories (Strengths, Weaknesses, Opportunities, Challenges) was first created. Rules were applied to map each student’s response into these categories. Then the resulting categorical outputs from the rule-based stage are added to the dataset as new features.

For example:

• Strengths were assigned if a student demonstrated consistently high ratings in technical subjects.

Rule Example: If Understanding of Mathematics, Programming, Web/App Development, and Data Structures & Algorithms ⩾3 (on a 5-point scale), and the student reported proficiency in at least one programming language, then the student was labeled as having the strength “Strong understanding of core technical subjects.”

• Weaknesses were identified if the student reported difficulties in decision-making or goal-oriented behavior.

Rule Example: If Decision-making, Confidence in achieving goals, and Handling unexpected changes ⩽2, then the student was labeled with the weakness “Struggles with decision-making and problem-solving.”

However, certain limitations were associated with the process, as it was unable to handle delicate language and its sensitivity to keyword matching. To enhance classification, the enriched dataset (original features + rule-based categorical features) was transformed into numerical vectors and used as input for XGBoost.

The overall workflow of the methodology is illustrated in Figure 1.

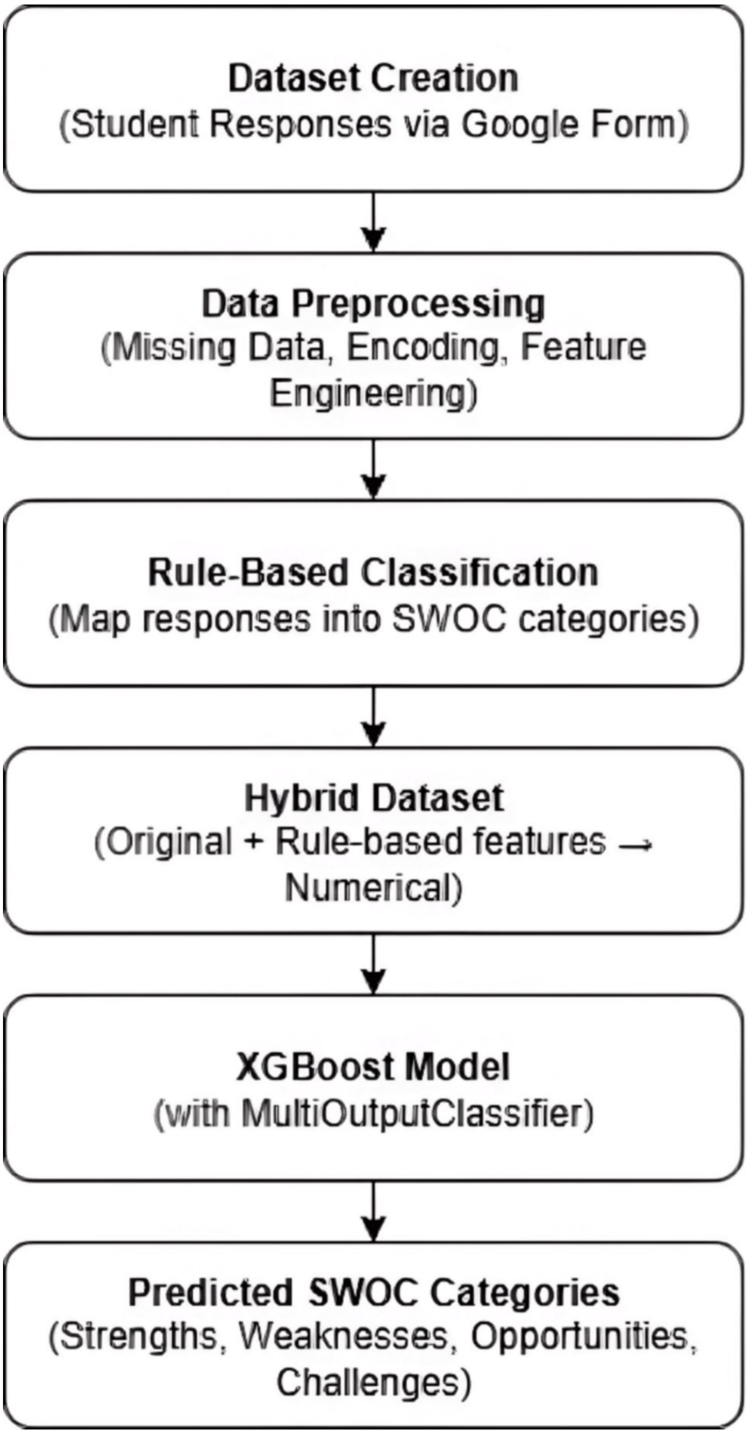

FIGURE 1: Workflow of the proposed methodology

SWOC = Strengths, Weaknesses, Opportunities, and Challenges

Model selection and training with XGBoost

In multi-label classification, a case may be assigned to multiple classes at once, unlike standard single-label classification. Each student's data has been associated with multiple SWOC categories in this framework. For instances like identifying the general strengths and weaknesses of students, where a class label is not unique, multi-label classification has to be used. Because of its ability to handle structured data and appropriateness for multi-label classification, XGBoost was selected. The model was able to learn the input features and SWOC categories' relationships as a result of this transformation. XGBoost is a powerful and scalable machine learning algorithm built on top of gradient boosting frameworks. It expands decision trees sequentially, making the errors of the previous trees correct.

The XGBoost model for SWOC analysis employs range and binary-based features to effectively identify varied student characteristics. As the problem is inherently multilabel, Scikit learn's MultiOutputClassifier was used to encapsulate the XGBoost model so that it can predict more than one label simultaneously (Strengths, Weaknesses, Opportunities, and Challenges). The model can be described as follows:

XGBoost Model Structure

XGBoost algorithm is particularly designed to work with mixed data types effectively by splitting features during building decision trees. Trees are sequentially grown, with the objective of each tree to decrease the residual error of the prior ensemble. L1 and L2 regularization is employed to control model complexity and improve generalization, especially for range-based features. The model provided several SWOC categories simultaneously for every student. The model generated a probability vector for every category, i.e., the probability that the student belonged to that category. The probabilities were thresholded to provide binary output (1 for presence, and 0 for absence). The model was optimized with the learning rate 0.1 to facilitate a gradual convergence upon training.

Training Process

The data were divided into training (80%) and testing (20%) sets. Feature normalization was performed for balancing the contribution of range-based and binary features. The model was trained on the basis of XGBoost to optimize a custom loss function that includes log loss for binary features and mean squared error for range-based features as a solution to the multi-label classification problem.

The complete dataset of student responses used for training and evaluation can be found in Appendix A.

Results And Discussion

The SWOC analysis system for the student is capable of analyzing and classifying student answers into opportunities, challenges, strengths, and weaknesses. The framework thrived on repeated refinement and application of a hybrid approach combining rule-based systems and machine learning.

Experimental setup

The experiments were conducted to evaluate the effectiveness of the proposed hybrid SWOC classification model. The setup was designed to ensure reproducibility and fair assessment of performance.

• Environment: The model was implemented in Python with the Scikit-learn and XGBoost libraries. For deployment and user interaction, Streamlit was used.

• Hardware/Software: All experiments were carried out on a system with an Intel i5 processor, 16 GB RAM, and Windows OS.

• Dataset: Responses were collected from 500 students through a structured questionnaire.

• Dataset Splitting: The data was divided into training (80%) and testing (20%) subsets.

• Evaluation Protocol: Performance was measured using accuracy, precision, recall, F1-score, and hamming loss. Results were averaged across the test set for each SWOC category.

Procedure

1. Collected responses were preprocessed and encoded.

2. Rule-based SWOC categorization was added as categorical features.

3. The enriched dataset was trained using XGBoost, which was wrapped in Scikit-learn's MultiOutputClassifier.

4. Model performance was measured on the test subset, and category-wise metrics were recorded.

Model performance

The model's performance has been checked at different levels of dataset growth and optimization. The model was first checked with a 100-student dataset and recorded a general accuracy of 62.5%. The dataset was grown to 300 students and 500 students, and the accuracy of the model was 93% and 95.5%, respectively. This incremental enhancement demonstrates the most significant contribution of dataset size and diversity to model generalizability. The third hybrid rule-based system-XGBoost model achieved category-specific accuracies of 93% for strengths, 90% for weaknesses, 99% for opportunities, and 100% for challenges. The accuracy of the results is a testament to the power of the hybrid approach to accurately classify student attributes and its excellent performance in classifying opportunities and challenges, a sign of its potential to provide students with useful feedback.

Evaluation criteria

For model validation, precision, recall, F1-score, and hamming loss measures were recorded for every class of SWOC. The model's stability and reliability are assured through high values of the measures for every class. The following Table *1* presents the formula for each of the measures and the respective figures for each SWOC category.

Category	Precision	Recall	F1-Score	Hamming Loss
Strengths	1	0.97	0.99	0.03
Weaknesses	1	0.99	1	0.01
Opportunities	1	1	1	0
Challenges	1	1	1	0

TABLE 1: Precision, Recall, and F1-Score for each SWOC category

SWOC = Strengths, Weaknesses, Opportunities, and Challenges

The table consolidates the model's evaluation measures as four groups: strengths, weaknesses, opportunities, and challenges. The model is exactly accurate (1) for all groups without any false positives. The recall is near-perfect for Strengths (0.97) and Weaknesses (0.99), and perfect (1) for Opportunities and Challenges with high sensitivity. The F1-score is perfect for all groups. The Hamming Loss is low for Strengths (0.03) and Weaknesses (0.01), but zero for Opportunities and Challenges, with perfect performance on multi-label classification tasks.

Visualization and analysis

Figure *2* shows the pie charts used to represent the prevalence of different technical skills. The charts showed skills differences, the areas where the students are predominantly lacking. The teachers can use these numbers to redesign the curriculum and implement special programs to close skill gaps and improve the technical proficiency of the students.

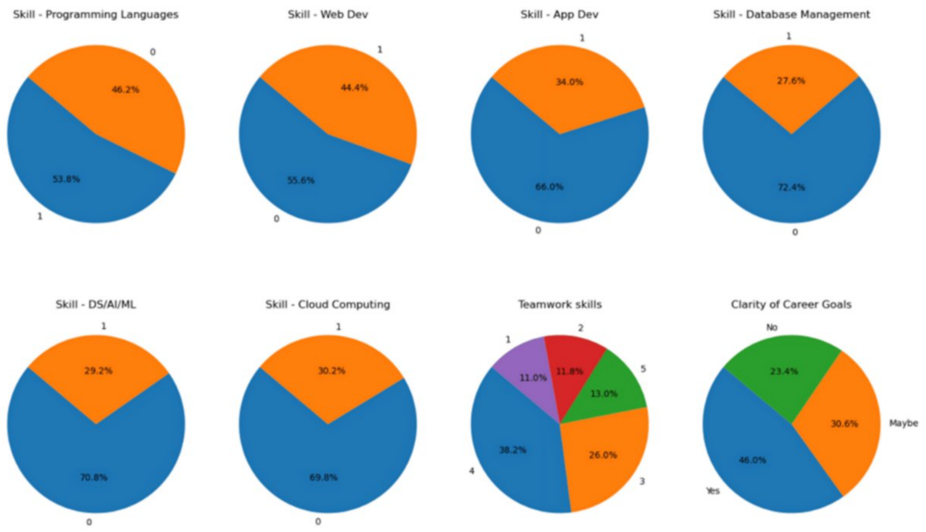

FIGURE 2: Skill distribution analysis of students

Deployment and utilization

Streamlit software is able to capture the functionality of the hybrid model with an easy-to-use interface through which students are able to input their responses and receive a complete SWOC report. The gap between technicality and usability is bridged through the interactive platform, and this makes the system more usable. Figure 3 shows the output interface of Streamlit application used for SWOC analysis.

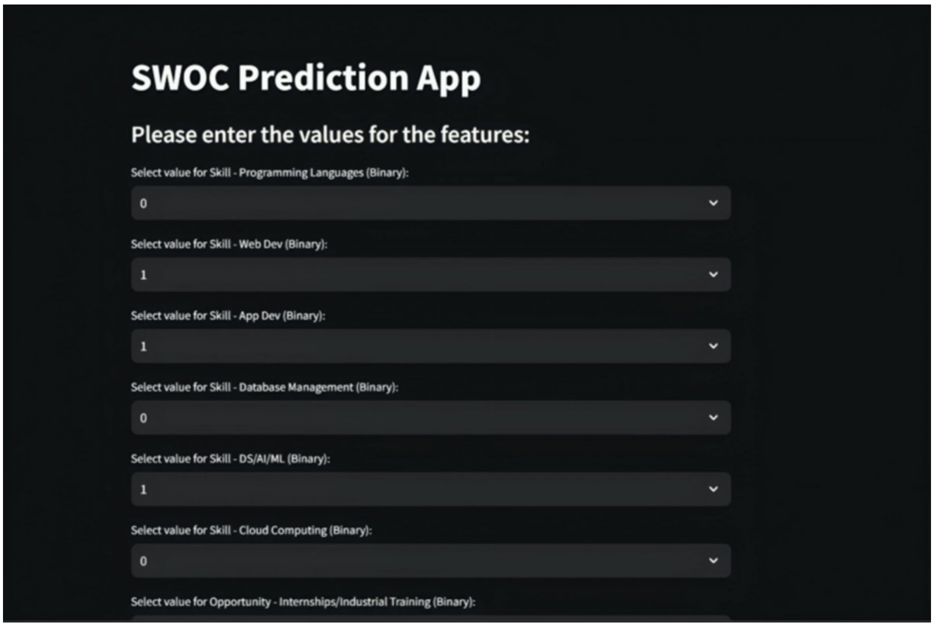

FIGURE 3: Streamlit application interface for SWOC analysis

SWOC = Strengths, Weaknesses, Opportunities, and Challenges

Experimental analysis

The experimental findings highlights several significant insights about the proposed SWOC classification approach.

The hybrid design that combined rule-based features and machine learning performed well. The rule-based stage improved the dataset with category-specific signals, which strengthened the ability of the XGBoost model. This explains the particularly strong performance in classifying Opportunities (99%) and Challenges (100%), categories that frequently contained distinct keywords and phrases.

The results highlights the comparative difficulty of categorizing Strengths (93%) and Weaknesses (90%). These categories are inherently complex, with overlapping vocabulary and slight linguistic variations, which makes them less adaptable to strict keyword-based rules and more dependent on contextual interpretation. This implies that integration of advanced natural language processing methods like contextual embeddings may further improve accuracy.

The model's reliability is validated by the evaluation metrics (precision, recall, F1-score, and hamming loss). The nearly ideal scores across categories demonstrate robustness, while the low hamming loss indicates good performance with multi-label predictions. Overall, the experiments confirm the efficiency of the hybrid methodology, its relevance for educational analytics, and potential for real-world application in student self-assessment and planning of curriculum.

Conclusions

Through meticulous preprocessing of data, multi-hot encoding, and repeated model training, the student SWOC analysis framework using AI was able to implement a hybrid technique by combining the rule-based system and XGBoost in order to determine students' strengths, weaknesses, opportunities, and challenges. The system reported high precision in SWOC category classification, and the overall accuracy was 95.5%. The use of the Streamlit app for deployment ensured a user-friendly interface for users to access individualized SWOC reports, making the tool accessible and initiating real-time analysis. The findings point towards a promising tool for students and teachers alike to attain actionable insights and streamline learning and teaching.

Implementation of more advanced machine learning algorithms, including deep learning methods, to enhance the system's accuracy and responsiveness, especially in big and heterogenous data, is one such direction in which the future can be searched for the AI-based student SWOC analysis system and its expansion. Another direction of development with immense significance is the integration of personalized recommendations based on the SWOC analysis of the student, which can give learning content, career guidance, or training programs for the learning of particular skills tailored to the areas of strength and weakness of individual students, enhancing the value of the system for both teachers and students.

Appendices

Appendix A: Customized Dataset

The complete, customized dataset used in this study can be accessed via this link: https://drive.google.com/file/d/136IlOV3UngmQkRGd-DoAb8oBJOUT065j/view?usp=sharing

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Shraddha Patil, Kaivalya Pagrut, Tilak Patel, Surabhi Kakade

Drafting of the manuscript: Shraddha Patil, Laher Patni, Kaivalya Pagrut, Tilak Patel

Acquisition, analysis, or interpretation of data: Laher Patni, Krushna Pawar

Critical review of the manuscript for important intellectual content: Krushna Pawar, Surabhi Kakade

Supervision: Surabhi Kakade

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Hanumanthappa S, Prakash C: Multi-label feature aware XGBoost model for student performance assessment

using behavior data in online learning environment. IAES International Journal of Artificial Intelligence (IJ-AI). 2024, 13:4537-4543. 10.11591/ijai.v13.i4.pp4537-4543

2. Deshpande LA, Rege PR, Bewoor MS, Mahajan P: Revolutionizing the education industry: Swot analysis and predictive modeling approach to enhance students' educational proficiency. Educational Administration: Theory and Practice. 2024, 30:3758-3765. 10.53555/kuey.v30i5.3530
3. Abu-Alaish A, Jaradat G, Al-Shugran M, Alodat I: Automating SWOT analysis using machine learning methods. International Journal of Advances in Soft Computing and its Applications. 2021, 13:140-150.
4. Khan I, Ahmad AR, Jabeur N, Mahdi MN: An artificial intelligence approach to monitor student performance and devise preventive measures. Smart Learning Environments. 2021, 8:17. 10.1186/s40561-021-00161-y
5. Alshanqiti A, Namoun A: Predicting student performance and its influential factors using hybrid regression and multi-label classification. IEEE Access. 2020, 8:203827-203844. 10.1109/access.2020.3036572
6. Abd El-Haleem AM, Eid MM, Elmesalawy MM, Hosny HA: A generic AI-based technique for assessing student performance in conducting online virtual and remote controlled laboratories. IEEE Access. 2022, 10:128046-128065. 10.1109/access.2022.3227505
7. Karale A, Narlawar A, Bhujbal B, Bharit S: Student performance prediction using AI and ML. International Journal for Research in Applied Science and Engineering Technology. 2022, 10:1644-1650. 10.22214/ijraset.2022.44032
8. Thejovathi M, Chandra Sekhara Rao MVP: Evaluating the performance of XGBoost and gradient boost models with feature extraction in FMCG demand forecasting: A feature-enriched comparative study. Journal of Theoretical and Applied Information Technology. 2024, 102:4158-4165.
9. Maulana A, Idroes GM, Kemala P, et al.: Leveraging artificial intelligence to predict student performance: A comparative machine learning approach. Journal of Educational Management and Learning. 2023, 1:64-75. 10.60084/jeml.v1i2.132