Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Predicting Students Growth in Academic career using Artificial Intelligence and Machine Learning Techniques

Pranjali Bahalkar[1], Dr Prasadu Peddi[2], Dr. Sanjeev Jain[3]

[1]SJJT University, Rajasthan, Dr. D Y Patil Institute of Technology, Pune, pranjali85bahalkar@gmail.com
[2]SJJT University, Rajasthan, peddiprasad37@gmail.com
[3]Sanjeevan Engineering and Technology Institute, Kolhapur, jain@seti.edu.in

This article presents a critical review of prediction-based AI models on the growth of students in education. The paper presents evidence and elaborates on the predicted impacts of artificial intelligence on the students' results, decision making concerning one's career, and the level of general student satisfaction. It integrates the studies that investigate the usefulness, trustworthiness, and other measures of social responsibility concerning the predictions regarding education and career paths made by artificial intelligent systems. The article presents evidence for the use of AI models for enhancing the achievement of students as well as information that would be of relevance to teachers, scholars, and even policy makers. In educating the readers on the need, pros, and corns of AI to foster student growth and application of AI in education, this paper attempts to do so in a very wide scope of applications. Give a paraphrase of the above paragraph and also change all the sentences to active voice only.
Keywords: Prediction-based AI models, student development, educational outcomes, career guidance, machine learning, student satisfaction, ethical considerations, educational pathways, career trajectories, AI applications in education, student success.

1. Introduction

The 2020 education policy proposed by the Government of India is a paradigm shift in the educational discourse, which advocates more for holistic and multidisciplinary education than rote learning. In this regard, this education reform measures ensures that children aged 3 to 18 years receive quality education that is regional languages inclusive, as much as it is with Hindi and English. Additionally, it speaks of a paradigm where students are evaluated by their abilities to apply knowledge rather than just rote learning[1].

Importantly, the policymakers propose a paradigm shift in education whereby technology is integrated not only in pedagogical methods but also in the curriculum where emphasis is laid on the skills and knowledge that are required. It also provides guidelines for enhancing the quality of education by imposing teacher eligibility tests at all levels and providing in-service training to teachers.

It seeks to create a seamless system of education that cuts across geographical barriers, especially the rural- urban interface and stresses on the need for industrial training of the learners.

For decades' education went many revolutions but still the systems have remained in place. It embraces such rooted change and therefore the ancient 10+2 system will eventually make way for the more modern 5+3+3+4 system.

The changes it will bring include revamping the board exams, introducing a national assessment platform, PARAKH, and a greater prominence of regional languages during initial stages of education. Even so, the policy allows for streamlined entry to postgraduate education and flexible options for mobility between disciplines of study within higher education institutions.

Now looking at these monumental changes another pressing question comes to mind: how do students adapt to the new educational environment in such a way that they are able to make the right choices when it comes to their future profession[1]. In this sense, traditional approaches to career guidance are becoming irrelevant.

2. Overview of AI Principles

The most popular AI models for predicting student achievement are compiled in this area. Furthermore, the standard assessment metrics and validation techniques are examined.

2.1. Overview of AI Models

There are different types of artificial intelligence (AI) models used in predicting student performance. Researchers in student performance evaluation often focus on supervised learning, where the model is trained using input data along with corresponding output data.

Supervised learning AI models can be broadly categorized into classification and regression methods. Classification tasks involve predicting discrete categories or classes, while regression tasks involve predicting continuous, real-valued outputs.

While certain AI methods may be used for particular tasks, others are flexible and, with little modifications, can be used for both classification and regression techniques[2].

In supervised learning artificial intelligence, input data is fed into the model for classification or clustering. Common AI models for predicting student success and the tasks that go along with them are shown in the following table:

Model	Task	Reason
Random Forest	Classification and Regression	Random Forest proved highly effective for both regression and classification tasks, giving robustness with good performance and relatively less parameter tuning as is often the case with some other models. It copes well with large data and complex interactions, which is beneficial when considering the diverse nature of educational data.
Long Short-Term Memory Neural Networks	Classification and Regression	LSTMs are well suited to time-series and sequential data, which is very common in educational applications-for example, analyzing the performance of students over time. Their ability to remember long-term dependencies makes them particularly valuable for modeling how students' progress and develop.
Support Vector Machine	Classification	SVMs are strong for classification tasks, mainly suitable for small to medium-size datasets. SVMs can be used in very high-dimensional spaces and are very versatile with a number of different kernel functions. SVMs can prove very handy for classifying students under categories or predicting outcomes based on various features.
Extreme Gradient Boosting (XGBoost)	Classification and Regression	XGBoost is high performing with great efficiency, although it is applied to both classification and regression tasks. It fits missing values very well and can be tailored to a great extent. Suitable for detailed and accurate predictions of the student performance and development as it captures complex patterns in the data.

2.2. Evaluation and Validation Strategies

Because each model produces different output, both regression and classification tasks in the student performance prediction require different assessment criteria. Nevertheless, despite this difference, the validation procedures for the two problem areas use comparable strategies, especially when it comes to dividing the data into training and testing sets[3].

2.2.1. Evaluation Metrics for Classification Tasks

Accuracy is a widely used parameter in educations, learning of learners or students performance prediction by classification, where the objective is to classify students into classes like Low/Medium/High or Pass/Fail[4]. The percentage of properly identified samples relative to the total number of samples in the test set is known as accuracy. When working with unbalanced datasets, where one class may have much more or less samples than others, accuracy is limited and may impact how well a model performs. The formula for accuracy is shown as:

$$\text{Accuracy} = \frac{\mathrm{TP+TN}}{\mathrm{FP+FN+TP+TN}}$$

TP, FP, TN, and FN stand for true +positive, false -positive, true -negative, and false -negative values, respectively.

Particularly when working with two-class unbalanced data, the Receiver Operating Characteristics Area under Curve (ROC AUC) is often used to overcome uncertainties brought on by imbalanced data. The model's performance is assessed using the false positive rate and the area under the recall (sensitivity) curve.

Another important metric is the F1 score, which is the harmonic mean of accuracy and recall. It is often used to binary and multiclass problems with imbalanced data and is calculated using:

$$\mathrm{FIScore} = \frac{2.\,\mathrm{TP}}{2.\,\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}$$

For classification tasks, recall (sensitivity), specificity, and accuracy are additional assessment criteria. These metrics, which include the following formulae, assess the model's capacity to identify certain output classes:

Recall = $\frac{\mathrm{TP}}{\mathrm{TP+FN}}$

Specificity = $\frac{\mathrm{TN}}{\mathrm{TN+FP}}$

Precision = $\frac{\mathrm{TP}}{\mathrm{TP+FP}}$

Together, these metrics provide information on how well classification models perform, especially when managing unbalanced datasets and differentiating between output classes[5].

2.2.2 Evaluation Metrics for Regression Tasks

When predicting student performance, regression studies forecast raw data such as quiz scores, final grades, total marks, and CGPA. Model assessment is accomplished by comparing the anticipated and actual values since regression issues deal with real-valued data rather than preset classifications[6].

MSE, or mean squared error: The average squared difference between expected and actual values is measured by MSE. Because of squaring, it gives outliers more weight, which might lead to an underestimation of mistakes brought on by acceptable deviations. The formula for MSE is:

$$\mathrm{MSE} = \frac{1}{2}\sum_{\mathrm{i}=1}^{\mathrm{N}} (\mathrm{y_i} - \bar{\mathrm{y}})^{\,2}$$

Here, N represents the number of samples, $\mathrm{y_i}$ is the observed value, and y is the predicted value.

Mean Absolute Error (MAE): This statistic ignores the direction of errors and calculates the average absolute difference between expected and actual values. When compared to MSE, it yields more reliable findings. The formula for MAE is:

$$\mathrm{MAE} = \frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}} |\mathrm{y_i} - \bar{\mathrm{y}}|$$

Determination Coefficient The percentage of variation in the observed data that can be predicted from the projected values is shown by the R_2 score. Higher numbers on the scale of 0 to 1 indicate a better match between the model and the data. The formula for the R_2 score is:

$$R_2\text{score} = 1 - \frac{\sum(y_{1-\bar{y}_i})}{\sum(y_i - \widehat{y_i})}$$

Here, $\bar{y}_1$ represents the mean value of the observed data, and $\hat{y}_i$ represents the mean value of the predicted data.

2.2.3. Validation Methods

Numerous techniques are used to verify AI models. Despite the notable distinctions among these approaches, research has examined hold-out, cross-validation, and data selection (data mining) techniques to verify models throughout their convergence[7]. Data mining approaches assist in identifying instances or attributes, allowing for direct data selection for training. When data mining approaches do not choose training data, cross-validation turns out to be the most effective method[8].

Even little modifications to the training data might have a significant effect on the results. In contrast, the hold-out strategy employs separate dataset divisions (e.g., 70% training, 30% testing) for training and testing separately[9]. The true efficacy is shown by cross-validation, where all data are utilized for both training and testing. By dividing the dataset into k parts and using each portion again throughout the training and testing phases, cross-validation yields more accurate model performance outcomes.

3. Results of Systematic Literature Review

This section presents the methodology and results of the systematic literature review.

3.1 Methodology

3.1.1 Search Strategy

After duplicate studies were eliminated, the titles and abstracts were reviewed to identify relevant research based on the preset inclusion and exclusion criteria. As part of the screening procedure, each author marked articles that satisfied the exclusion criteria. After then, the remaining research were evaluated further. For a graphic representation of the research selection procedure, see Figure 1, which displays the approach's whole flow chart.

3.1.2. Inclusion and Exclusion Criteria

Inclusion Criteria:

- Research papers published in scientific peer-reviewed journals.
- Studies implementing or proposing machine learning and/or deep learning models to predict student performance across all levels of education.
- The research being reported in English.

Exclusion Criteria:

- Studies focusing on student performance without implementing machine learning models.
- Review studies, abstracts, commentaries, book chapters, or editorials.

3.1.3. Selection Procedure and Data Extraction

After duplicate studies were eliminated, the titles and abstracts were reviewed to identify relevant research based on the preset inclusion and exclusion criteria. As part of the screening procedure, each author marked articles that satisfied the exclusion criteria. After then, the remaining research were evaluated further. For a graphic representation of the research selection procedure, see Figure 1, which displays the approach's whole flow chart.

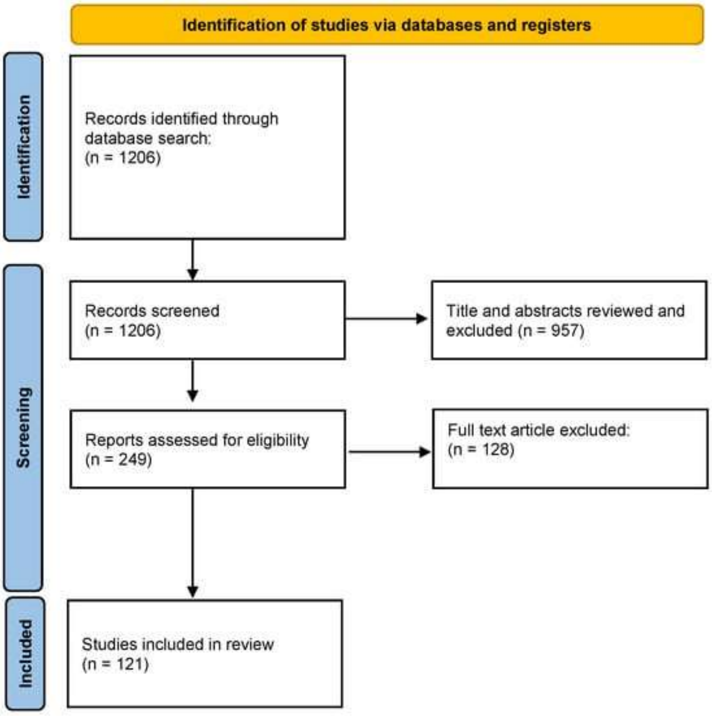

Figure 1. Study of Selection procedure

Data extraction was conducted based on five main objectives: determining the study's aim—whether it focused on regression or classification, identifying educational expectations addressed by the study, examining the methodology, including validation and evaluation techniques used, specifying the datasets utilized, and listing the machine learning models considered.

In this review, a total of papers were considered for further analysis. These studies, which include a variety of educational levels, distinct aims, assessment and validation approaches, and machine learning models, were selected to fully meet all of the study's goals. An overview of the features of the chosen studies for more research is given in Table 2.

Author/Publication Year	Objective	Type	Target Level	Primary ML Method	Primary Evaluation Metric(s)	Validation Technique
Jie Yang et al. (2021)	Predict pre-course student performance	Classification	C	Discriminable Multi-Label Attribute Selection (DMAS)	Accuracy	Not mentioned
Zawqari et al. (2022)	Predict student academic performance in online courses	Classification	C	Feature selection with various algorithms	Accuracy, Precision, Recall, F1-Score	Not mentioned
Nalindren Naicker et al. (2020)	Investigate linear SVM for predicting student performance	Classification	C	Linear SVM	Accuracy, Precision, Recall (possibly)	Not mentioned
Safira et al. (2022)	Combine feature selection and multiclass classification for predicting student performance	Classification	C	Genetically Optimized Feature Selection with Multiclass Classification	Accuracy, potentially others	Not mentioned
Yahia Baashar et al. (2022)	Use ANNs for predicting student performance	Classification	C	Artificial Neural Networks (ANNs)	Accuracy, other metrics not mentioned	Not mentioned
Sujan Poudyal et al. (2022)	Propose a novel architecture combining 2D CNNs with other techniques for predicting student performance	Classification	C	2D CNNs	Accuracy, other metrics not mentioned	Not mentioned
Kaiming et al. (2015)	Explain residual learning and ResNet architecture	N/A	N/A	N/A	Accuracy, ROC AUC, Recall, Specificity, Precision	N/A
Ian J. et al. (2019)	Explain Generative Adversarial Networks (GANs)	N/A	N/A	GAN	MSE	-
M. Riki Apriyadi et al (2022)	Hybrid PSO-GA for feature selection in SVR for student performance prediction	Prediction	C	SVR	Accuracy	Not mentioned
Deepti Aggarwale et al. (2019)	Analyze machine learning algorithms for predictive analytics	N/A	N/A	N/A	N/A	N/A
Phauk Sokkhey et al. (2020)	Feature selection for predicting academic performance	Prediction	C	Various Feature Selection Techniques	Accuracy	Not mentioned
Prasanalakshmi Balaji et al. (2021)	Machine learning techniques for predicting academic performance	N/A	N/A	Various Machine Learning Techniques	N/A	N/A
AYA NABIL et al. (2021)	Deep neural network for predicting academic performance	Prediction	C	Deep Neural Network	Accuracy, Other metrics not mentioned	Not mentioned
Wilson et al. (2021)	Attribute selection and ensembles for predicting student performance in ITS	Prediction	C	Attribute Selection Techniques, Ensemble Models	Accuracy	Not mentioned
Sebastianus Radhya et al. (2022)	Machine learning for predicting student performance	N/A	N/A	N/A	N/A	N/A
Maria Koutina et al. (2011)	Machine learning for predicting postgraduate students' performance	Prediction	HE	Decision trees, logistic regression, support vector machines, neural	N/A	N/A

				networks		
Roberto Bertolini et al. (2021)	Feature selection and cross-validation for student performance prediction	Prediction	C	Feature selection techniques, cross-validation	Accuracy, Generalizability	Not mentioned
Md. Ahsan Arif et al. (2021)	Enhance accuracy and efficiency of prediction	Classification	HE	Decision Trees, Support Vector Machines (possible)	Accuracy (possible), other metrics not mentioned	Not mentioned
Phauk Sokkhey et al. (2020)	Develop methods for educational data mining to predict underachieving students	N/A	N/A	Data Mining	N/A	Prediction model comparison
Raza Hasan et al. (2020)	Use video learning analytics to predict student success	N/A	HE	Classification algorithms, Clustering approaches (possible)	Accuracy, other metrics not mentioned	Not mentioned
Gomathy Suganya Ramaswami et al. (2020)	Identify relevant features and build a model to predict student performance	Classification	HE	Filter, Wrapper, or Embedded methods (possible), Decision Trees, Support Vector Machines, Logistic Regression (possible)	Accuracy, Precision, Recall, F1-Score (possible)	Not mentioned
Teo Susnjak et al. (2021)	Use machine learning to predict student success using LMS data	Classification	HE	Machine learning methods (unspecified)	Not mentioned	Not mentioned

Data Collection:

This is the first step where different data sources, which would be of prime importance for a comprehensive analysis, are brought together. For instance, this would include student demographics, academic records, and interaction logs from any learning management system or other digital platforms. Institutional data privacy rules, therefore, are always followed whereby methodologies employed include survey, interview, and automated tools in order to capture nuanced insights about their engagement.

Data Preprocessing:

The raw data is scrubbed in detail to eliminate any duplicate ones and also missing values before being standardized. Unstructured data, such as text comments, may undergo other processes, such as normalization and natural language processing, in order to clean up the data before analyzing to ensure that the integrity of data is maintained and consistent.

Data Analysis:

Descriptive and inferential statistical analyses will reveal trends and underlying relationships hidden in the data. Regression, classification, and clustering are examples of machine learning approaches that assist predict future outcomes by identifying the factors that influence student participation. Results are presented to stakeholders using visual aids like heat maps and infographics.

Intervention Design:

Use the findings from data analysis to design appropriate interventions and learning strategies for the right needs of individual students. Adaptations to instructional material, on top of early warning systems, shall be constructed to basically service the needs of individual learners. This could encourage proactive support systems with regard to improving engagement and academic success.

Implementation and Evaluation:

Within existing educational structures, AI-powered analytic solutions are being applied. This guarantees integration and ease of use. Ongoing monitoring and evaluations address the different interventions over a period of time which is then complemented by more advanced research designs such as randomized controlled trials or quasi-experimental designs.

Ethical Considerations:

There are ethical principles embedded in the system that ensures appropriate usage of student data, confidentiality as well as making decisions with the assistance of algorithms without discrimination. The issues of bias and fairness risk mitigation are critical in the responsible implementation of any AI related to student welfare and trust.

Objectives

Assessment of Analytical Interventions Incorporated into AI on Student Engagement in Tertiary Institutions. This article sets out to evaluate the use of AI analytics in any tertiary institution and examine its effects on the involvement of students. The paper seeks to explore how effective artificial intelligence systems are in recognizing various characteristics, inclinations, and educational approaches of learners.

Examine how personalisation affects student engagement: This seeks to ascertain if using AI-based analytics to create individualised learning experiences helps students become more engaged. This will have to do with looking at the creation and use of customised interventions, support services, and adaptable instructional materials.

The ethical and privacy ramifications of AI-driven analytics in higher education will be assessed by this study. It addresses concerns about algorithmic bias, data privacy, and equality to ensure that AI-driven analytics projects are implemented appropriately, ethically, and fairly.

Significance

This research will contribute toward advancement in educational practices by providing insights into an effective strategy for the use of AI analytics in support of student engagement and improved retention in higher education. The studies inform policy decisions and represent necessary features toward creating inclusive learning environments that make progress in educational innovation much less challenging.

4. Challenges and Opportunities

1. Challenges

Challenges are related to data privacy regulation complexity, reductions in algorithmic bias, and technical and institutional adoption.

2. Opportunities:

AI-based analytics hold opportunities for personalized learning experience, intervention strategies early on, prediction of student success, and efficiencies in educational operations.

Results

Higher education institutions gained important insights on student engagement and academic results thanks to AI-driven analytics. Links regarding the measures for academic success, retention rates, and student involvement were found via the examination of student data. By recognising trends in students' behaviour, preferences, and learning styles, the algorithms were able to adjust support plans and interventions appropriately. Accurately predicting the pupils' results was made possible in large part by predictive analytics. By using AI algorithms, early warning systems were able to identify pupils who were at danger and alert them in advance of any potential mistakes. All of these contributed to higher retention rates and, as a result, increased student success potential. AI-driven analytics might be used to create adaptive learning experiences based on the requirements of each individual learner.

5. Discussion

Furthermore, the ability of AI-powered analytics to revolutionize the education sector particularly in enhancing students' active participation and attainment cannot be overlooked. Enhanced by advanced technologies, it is possible to monitor and analyze students' behavioral traits, preferences, and learning tendencies, and such learning will be individualized. This enables some tracking of the at-risk students so that preventive measures can be put in place before they start posing problems. In other words, improved student outcomes will be possible through adaptive learning experiences as AI-enabled analytics will ensure that educational resources and support services are well tailored to meet individual students' requirements. In connection with this are issues of moral concern and privacy which shall be dealt with head on. Data collection and usage needs to be open to scrutiny, responsible, and equitably done in order to reduce the biases that are unfortunately associated with fair education policies.

6. Conclusion

Implementation of AI-driven analytics: It will be an excellent opportunity for any college or university to transform the lives of their students, increase involvement and as a result achieve academic achievement. Large data sets and cutting-edge technology will ensure that students get more assistance from institutions' actions and tactics. Predictive analytics is also used by institutions, which facilitates effective measures aimed at assisting retention and success of students while adaptive learning technologies improve the level of engagement and learning outcomes. Hence, it is important that ethical sociological issues on data protection and algorithm bias are dealt with seriously in order to enhance equity and justice in teaching methods.

References

1. M. Yağcı, "Educational data mining: prediction of students' academic performance using machine learning algorithms," Smart Learn. Environ., vol. 9, no. 1, 2022, doi: 10.1186/s40561-022-00192-z.
2. S. B. Kotsiantis, "Use of machine learning techniques for educational proposes: a decision support system for forecasting students' grades," Artif. Intell. Rev., vol. 37, no. 4, pp. 331–344, 2012, doi: 10.1007/s10462-011-9234-x.
3. P. Guleria and M. Sood, Explainable AI and machine learning: performance evaluation and explainability of classifiers on educational data mining inspired career counseling, vol. 28, no. 1. Springer US, 2023. doi: 10.1007/s10639-022-11221-2.
4. S. Kour, R. Kumar, and M. Gupta, "Analysis of student performance using Machine Learning Algorithms," in 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), 2021, pp. 1395–1403. doi: 10.1109/ICIRCA51532.2021.9544935.
5. M. F. Musso, C. F. R. Hernández, and E. C. Cascallar, "Predicting Key Educational Outcomes in Academic Trajectories," High. Educ., vol. 80, no. 5, pp. 875–894, 2020.
6. Y. Wang, L. Yang, J. Wu, Z. Song, and L. Shi, "Mining Campus Big Data: Prediction of Career Choice Using Interpretable Machine Learning Method," Mathematics, vol. 10, no. 8, 2022, doi: 10.3390/math10081289.
7. J. L. Rastrollo-Guerrero, J. A. Gómez-Pulido, and A. Durán-Domínguez, "Analyzing and predicting students' performance by means of machine learning: A review," Appl. Sci., vol. 10, no. 3, 2020, doi: 10.3390/app10031042.
8. F. Cruz-Jesus et al., "Using artificial intelligence methods to assess academic achievement in public high schools of a European Union country," Heliyon, vol. 6, no. 6, p. e04081, 2020, doi: 10.1016/j.heliyon.2020.e04081.
9. D. Buenaño-Fernández, D. Gil, and S. Luján-Mora, "Application of machine learning in predicting performance for computer engineering students: A case study," Sustain., vol. 11, no. 10, pp. 1–18, 2019, doi: 10.3390/su11102833.
10. S. Kotsiantis, C. Pierrakeas, and P. Pintelas, "Predicting students' performance in distance learning using machine learning techniques," Appl. Artif. Intell., vol. 18, no. 5, pp. 411–426, 2004, doi: 10.1080/08839510490442058.
11. Zunzarrao Pankaj Deore, Pranjali Bahalkar, "Multicamera Object Detection and Tracking System and Path Prediction Using AI and Machine Learning for Enhanced Surveillance", ISSN: 1092-910X Vol 28 No. 2 (2025). DOI: https://doi.org/10.52783/anvi.v28.1915