SvedbergOpen
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper
Open Access

A Novel Data Model Framework For Analytic In Online Education

B. Fathima Josepin Prasanna[1] , Dr. T. Balaji[2]

[1]Research Scholar , Madurai Kamaraj University, Madurai, Tamil Nadu, India.
Email: bfathimajosepin@gmail.com
[2]Research Supervisor, Associate Professor, PG Department of Computer Science, Government Arts College, Melur, Tamil Nadu, India
Email: bkmdgacm1976@gmail.com
Corresponding author mail id : bfathimajosepin@gmail.com

ABSTRACT
The rapid digitalization of education has created an urgent need for intelligent, scalable, and adaptable analytical systems. In response to this demand, this research introduces a novel integrated data model framework that effectively bridges demographic insights, device usage patterns, platform preferences, and learner feedback to support data-driven decision-making in online education. Unlike traditional models that focus solely on either engagement metrics or academic performance, the proposed framework offers a holistic understanding of learner behavior by analyzing a dataset comprising 1,000 responses with both structured and semi-structured attributes. At the core of this framework lies the Adaptive Randomized Fusion-layered Artificial NeuroNet (ARF-ANN), a custom deep learning architecture designed to integrate and enhance predictive capabilities. The multi-layered system includes stages for data preprocessing, encoding, and transformation through Python-based feature engineering pipelines. To ensure robustness and generalizability, the research also explores ensemble methods and hybrid modeling strategies. Results reveal significant correlations between profession types, device accessibility, and satisfaction ratings. Performance evaluation using metrics such as F1-score, precision, recall, and accuracy that ranged from 96% to 98% confirms the effectiveness of both ML and DL components. By embedding the innovative ARF-ANN and leveraging hybrid analytics, this framework marks a pioneering step toward real-time, personalized online learning analytics, enabling smarter and more adaptive educational ecosystems.

Keywords: Online Education, Educational Technology, Student Engagement, Adaptive Randomized Fusion layered Artificial NeuroNet, Predictive Modeling, Likert-Scale Analysis, ANN, Data Model Framework.

1. Introduction

Online education, also known as e-learning or digital learning, is one of the most significant developments in the history of the dissemination of human knowledge [1]. Fundamentally, online learning is an outgrowth of the connection between education and technology, offering a way to create and distribute educational materials and academic instruction through technology-based channels, mostly the internet [2]. Online education has continued to evolve, illustrating a larger shift in connection with how learners access content and connect with instructors and fellow students [3]. This shift in educational delivery has moved away from traditional classroom-based learning to an experience that is more robust in options and personalized for learning [4].

It created opportunities for educational content to be brought to learners in engaging formats like videos, interactive simulations, and discussion boards [5]. This change prompted colleges and universities to allow for the introduction of online courses on the global market, a first for higher education. In the 21st century, cloud computing, mobile technologies, and affordable internet access helped to democratize education again by allowing anyone with a device and internet to access meaningful educational resources from anywhere in the world [6].

In online education, a number of principles determine its design and delivery. One of the principles is accessibility, which allows individuals from different geographic, economic, and cultural backgrounds to access the learning resources anytime and anywhere, rather than dragging them to a specific institution at a certain time, as is the case with traditional education [7]. Working professionals, disabled learners, and individuals residing in rural areas are all great beneficiaries of this accessibility to learning [8].

Another core area of online education is learner autonomy. Many online platforms allow students to take ownership of their learning, as students can work at their own pace, review materials, adapt education time to their schedules, and achieve the learning goals in their own way [9]. A more personalized learning experience is more likely to happen; meanwhile, digital platforms often provide a huge range of resources, from text to presentations to multimedia content, to serve multiple modes of learning while enhancing understanding in a variety of forms [10]. Online education offers flexibility, affordability, efficiency, and scalability, making it appealing for adults and learners seeking careers. It supports students in balancing studies with personal and work priorities, often providing them a smarter and less costly way to gain knowledge [11].

The current online education systems lack a unified, scalable data model for effective analytics. Existing frameworks struggle with real-time processing, heterogeneous data integration, and limited personalization. These limitations hinder data-driven insights, reducing the ability to enhance learning outcomes. Hence, this research aims to develop an intelligent and adaptive data model framework that integrates learner demographics, device usage, and feedback to accurately predict engagement and satisfaction, using a hybrid approach combining ARF-ANN, machine learning, and explainable AI techniques. The key contributions are as follows:

- **Hybrid Dataset Integration for Analytics:** Obtained and used a hybrid dataset of learners that included both structured and other semi-structured data in reports from several platforms.
- **Robust Preprocessing and Feature Extraction Pipeline:** Developed a Python-based preprocessing to clean, encode, and extract key features such as satisfaction, engagement, and instructional quality from Likert–scale responses.
- **Development of the ARF-ANN Hybrid Model:** Created the Adaptive Randomized Fusion-layered Artificial NeuroNet (ARF-ANN) model, combining Random Forest decision strength with deep learning capabilities of ANN for accurate learning classification.
- **Real-Time Predictive Learning Analytics:** The complete ARF-ANN system is designed to provide real-time predictive learning analytics of learner satisfaction and engagement, which foster adaptive and personalized environments of online learning.

The introduction discusses the effects of digitalization and the requirement for verified, scalable analytics capabilities in online education, and the related work indicates limitations of existing models in processing semi-structured data and providing real-time insights. The methodology reviews the hybrid dataset used, the techniques for preprocessing and feature extraction, and the ARF-ANN, which hybridizes adaptive Random Forest and deep neural networks. The experimental results demonstrate how effective the framework is at predicting learner satisfaction and engagement. Overall, the research provides a scalable and adaptive approach to real-time analytics in educational settings.

2. Related works

The section discusses applications of AI and learning analytics in online education for improving engagement and outcomes. It identifies gaps in real-time feedback, adaptive personalization, and semi-structured data integration.

The application of Learning Analytics (LA) tools in higher education during the COVID-19 epidemic was examined in [12]. Problems such as students' desire for prompt assistance, time management issues, teachers' lack of pedagogical expertise, and institutions unprepared for digital transformation were all highlighted. LA tools have been used to track, organize, encourage participation, ease evaluation, boost communication, enhance retention, and be easy to use.

To enhance student learning in a collaborative learning environment, research [13] combined learning analytics techniques with an AI performance prediction model. The integrated strategy raised student engagement, enhanced performance, and reinforced satisfaction, according to the investigation, which was carried out in an online engineering course. It offered fundamental implications for future development and advances in AI-driven learning analytics.

To forecast and comprehend online video consumption, investigation [14] introduced a novel video feature architecture that makes use of computer vision and machine learning approaches. The 771 online films from MasterClass and the 1,127 traditional education videos from Crash Course were the two data sets to which the methodology was applied. The framework's guarantee for companies was demonstrated by the analysis, which shows that it correctly forecasted both aggregate video popularity and individual-level customer behavior.

Learning analytics to gain valuable insights into students' learning in the context of video-based online learning by addressing an automatic interaction in student-video [15]. 72 college students who were watching films were found to exhibit four distinct behavioral patterns: surfing, social engagement, knowledge seeking, and environment configuration. Passive learners only browsed, whereas active learners regularly engaged in these activities. Higher learning achievement was attained by engaged students. Table 1 demonstrates the related works for online education technology.

Table 1: Summary of online education technology in recent studies

Research No	Aim	Technology Used	Dataset	Findings
[16]	Measured educational video effectiveness	VAS + EDA (Video Assignment System + Engagement Data Analytics)	25 novice EE students' interaction logs	Storytelling enhances learner engagement
[17]	Explored data-driven EdTech	AI + LA (Artificial Intelligence + Learning Analytics)	Industry discourse from EdTech firms	Personalization was hindered by ambiguity
[18]	Analyzed peer learning interaction	IAM + TAA (Interaction Analysis Model + Text Analysis Automation)	Online discussion forum transcripts	Automation boosts analysis efficiency
[19]	Integrated EDM with theory	EDM (Educational Data Mining)	Theoretical integration with classroom examples	EDM supports adaptive personalization
[20]	Research collaborative learning regulation	ENA + CA (Epistemic Network Analysis + Content Analysis)	45 student teachers' online activities	Regulation improves group performance

2.1 Problem statement

Previous research emphasizes the potential of AI and learning analytics in online learning. However, issues remain in terms of semi-structured feedback, real-time adaptability, and providing effective personalized support.

- In research [13], a performance prediction model was applied in an online engineering course for improving engagement and satisfaction. However, the model did not adapt in real-time and did not use semi-structured learner feedback, which limited the personalization possibilities of the model across distinct educational backgrounds.
- The video consumption prediction in [14] leveraged computer vision and machine learning to ascertain the popularity of educational video content. However, although it was beneficial for forecasting educational content, it did not consider learner behavior metrics, which would have made it more useful for learner-centered analytics.

To address these limitations, this research suggests the ARF-ANN model, which brings together structured and semi-structured data that includes learner characteristics and demographics, device usage, and Likert-scale feedback into a single deep learning model. By layering the Random Forest decision processes with advanced ANN processes, the model supports real-time learner classification, personalized satisfaction predictions, and adaptive engagement monitoring within scalable online learning systems.

3. Methodology

The methodology involves examining a hybrid dataset of 1,000 learners, consisting of structured attributes and semi-structured feedback, to evaluate satisfaction and engagement in online learning. The pre-processing phase included a label encoder and composite metrics extraction, such as satisfaction score and engagement index. Other than developing an adaptive Random Forest-ANN model as a new model of ARF-ANN, which combined an adaptive Random Forest and a deep neural network, thereby improving the prediction accuracy and reliability. The overall process for analytics for online education can be found in Figure 1.

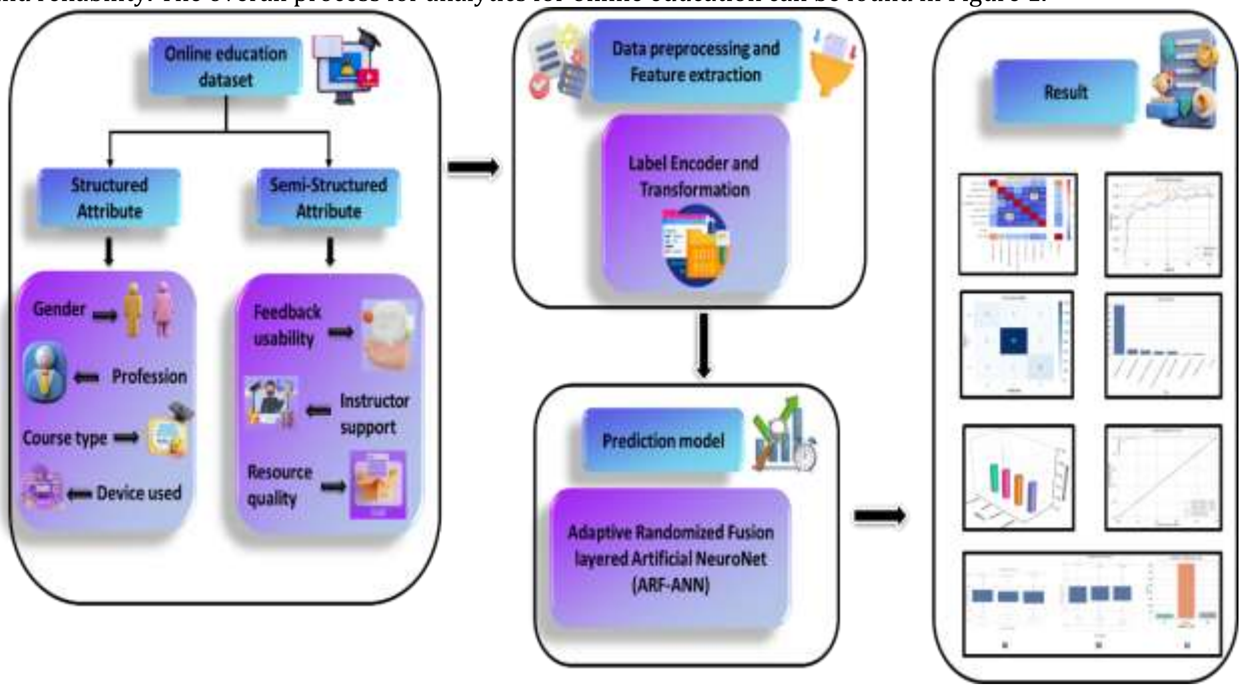


Figure 1: The overall process for analytics for online education

3.1 Dataset

The online education dataset consists of 1,000 responses from learners of different ages, backgrounds, and fields of education. The dataset includes both structured attributes, such as gender, occupation, age category, type of course, and device used, as well as semi-structured attributes, such as feedback on usability, instructor support, course design, and resource quality. Additionally, learners indicated their opinions on 15 Likert-type items assessing their overall experience in online education, interaction, motivation, and fairness of assessment. With this dataset, it is possible to analyze learners' behavior, satisfaction, and engagement in online learning contexts comprehensively. Table 2 shows the sample dataset.

Table 2: Sample dataset

S.No	Gender	Choose your stand/profession	The type of course you are attending through online education	If Others, please specify	Mention your field of study	Select the age group to which you belong
1	Female	Working Professional	Other certificate courses		Commerce	20 - 25
2	Female	Full-time Student	MOOC courses		Engineering	25 - 30
3	Male	Working Professional	Other certificate courses	No	Economics	25 - 30
4	Female	Full-time Student	Degree programs	No	Commerce	20 - 25
5	Male	Part-time Student	Degree programs		Computer Science	25 - 30
6	Male	Part-time Student	MOOC courses		Engineering	25 - 30

7	Male	Part-time Student	Other certificate courses		Engineering	25 - 30
8	Male	Full-time Student	Degree programs	-	Biology	15 - 20
9	Female	Full-time Student	MOOC courses		Engineering	25 - 30

The collective data set serves as a foundation for further exploration of learner behavior in online education, which has currently been input into the next step for preprocessing and feature extraction.

3.2 Data pre-processing and Feature extraction

Pre-processing began with loading the online education dataset via the Pandas package and cleaning the column names, where whitespace and newlines were stripped. The last 15 Likert-scale columns were identified and renamed as Q1 to Q15 for consistency. Categorical columns such as gender, profession, and platform used were assigned using LabelEncoder to convert textual data into numerical data. Likert responses were indexed on a 1–5 scale for quantitative assessment. Learners were classified as High, Moderate, or Low satisfaction groups, and lastly, the clean dataset was saved. Table 3 shows the pre-processed table outputs.

Table 3: Pre-processed outputs for online education dataset

S.No	Gender	Choose your stand/ profession	The type of course you are attending through online education	If Others, please specify	Mention your field of study	Select the age group to which you belong
1	0	2	Other certificate courses	0	Commerce	20 - 25
2	0	0	MOOC courses	0	Engineering	25 - 30
3	1	2	Other certificate courses	1	Economics	25 - 30
4	0	0	Degree programs	1	Commerce	20 - 25
5	1	1	Degree programs	0	Computer Science	25 - 30
6	1	1	MOOC courses	0	Engineering	25 - 30
7	1	1	Other certificate courses	0	Engineering	25 - 30
8	1	0	Degree programs	2	Biology	15 - 20
9	0	0	MOOC courses	0	Engineering	25 - 30

The feature extraction phase begins when the Satisfaction Score is established as the mean of fifteen Likert scale responses. The Engagement Score is derived from Q2, Q6, Q10, and Q11 to represent learner engagement. The Resource Effectiveness Index is derived from Q7, Q8, and Q14, which represent the utility of the learning resources. The Communication Effectiveness score is derived from Q9, Q11, and Q13, and the Instruction Quality score is extracted from Q4, Q5, and Q12. Q14 is input for the Practical Learning Experience, and Q15 is the Confidence Score representing a learning outcome. Additional features include binary values for device and learner type. The final target label is encoded with 1 through 6 to represent satisfaction levels. Table 4 shows the feature extraction table outputs.

Table 4: Feature extraction outputs for online education dataset

Satisfaction_Score	Engagement_Score	Resource_Effectiveness	Communication_Effectiveness	Instruction_Quality	Practical_Experience	Confidence_Score	Is_Youth	Target_Label
3.533333	4.5	2.666667	4.333333	3.666667	1	1	0	2
3.266667	4.25	2.666667	4	3.333333	3	3	0	1
3.666667	4	4.333333	3.666667	3	3	5	0	2

2.6	2.25	2	2.666667	4	3	1	0	1
2.933333	3.5	3.333333	2.333333	2.666667	1	2	0	1

The pre-processed and feature extraction data set is presently organized for modeling. The resulting dataset is then input into the proposed ARF-ANN model for predictive analysis and learner classification.

3.3 Adaptive Randomized Fusion-layered Artificial NeuroNet (ARF-ANN)

The ARF-ANN is a hybrid model created for improvement in predictive analytics in online education that combines the decision strength of Random Forest with the depth of learning associated with artificial neural networks. This model is designed to address the variability and non-linearity of online learners' behaviour informed by known factors. The ARF-ANN aims to provide not only real-time classification but also to classify online learners in personalized interventions to increase learner satisfaction, engagement, and performance.

The ARF component modifies the classical RF model with a dynamic node-splitting method. Rather than using fixed node-splitting criteria, the model integrates a linear function of Information Gain (from ID3) and Gini Index (from CART) using Equation (1):

$$G = \min_{\alpha,\beta\in Q}[\alpha.Gini(C,b) - \beta.Gain(C,b) \quad where\ \alpha + \beta = 1 \tag{1}$$

The ANN section in the ARF-ANN model recognizes complex nonlinear relationships within online learner data. The model operates on features such as demographics, feedback scores, and engagement insufficiency across several layers, including randomization and adaptive normalization. The layers allow for generalization over noisy or incomplete datasets and increased robustness in predictions. The ANN maps learned features to satisfaction and engagement outcomes for the purpose of enhancing learning analytics in real-time education systems.

$$\widetilde{x_{pi}} = x_{pi} + \epsilon_i, \quad \epsilon_i \backsim N(0,\sigma^2) \tag{2}$$

The combination of ARF and ANN allows the ARF-ANN model to benefit from both interpretable tree-based decisions and the high level of abstraction from the neural network. ARF-ANN is well suited to online educational systems where the level of noise, variability, and semi-structured input data can be observed.

3.3.1 Advanced Random Forest

The Random Forest algorithm is highly recognized for its high classification accuracy, robustness against outliers and noise, and low degree of overfitting. To facilitate data-driven decisions in online education, this research proposes a novel method to enhance the ARF algorithm to improve the decision process of the internal trees using adaptive mechanisms for node splitting.

Within online education contexts, learner behaviors are influenced by multiple variables ranging from demographics, device usage, platform of choice, and their feedback patterns. It can be asserted that fixed node splitting strategies could produce Random Forest models with inconsistent decision tree structures and ultimately, consistency in prediction performance.

The proposed hybrid technique applies a separate criterion for splitting and evaluates multiple splitting criteria, such as Information Gain (ID3) and Gini Index (CART), in combination. This allows the algorithm to identify the most informative attribute for each decision node, thus improving accuracy and generalization across various learner datasets.

To facilitate large-scale learning datasets and real-time data analysis, the ARF algorithm is implemented using the Apache Spark MLlib. Spark is a distributed processing system that streamlines working with large semi-structured datasets, like those obtained from an online learning system. The adaptive evaluation of the node begins with splitting metrics computed for each attribute b on dataset C are illustrated in (3) and (4). Subsequently, a selection strategy based on linear combination is proposed to determine the best split based on a set of metrics for each node.

$$Gain(C,b) = Ent(C) - \sum_{u=1}^{U}\frac{|C^u|}{|C|}Ent(C^u) \tag{3}$$

$$Gini(C,b) = \sum_{u=1}^{U}\frac{|C^u|}{C}Gini(C^u) \tag{4}$$

Where C^u denotes that every sample in the C with a value of b^u on the attribute b is contained in the u branch node, as shown in Equations (5) and (6).

$$Ent(C) = -\sum_{l=1}^{|z|} o_l \log_2 o_l \tag{5}$$

$$Gini(C) = \sum_{l=1 l'\neq l}^{|z|} olol' = 1 - \sum_{l=1}^{|z|} ol^2 \tag{6}$$

The combination node splitting formula and the adaptive parameter selection procedure are as follows in Equation (7), as the goal of node splitting is to increase the integrity of the data set following the division.

$$G = \min_{\alpha,\beta \in Q} E\{C,b\} = \alpha Gini(C,b) - \beta Gain(C,b) \begin{cases} \alpha+\beta = 1 \\ 0 \le \alpha,\beta \le 1 \end{cases} \quad (7)$$

In this case, α *and* β stand for the attribute splitting weight coefficient. In the meantime, G has a very low value. The procedure of adaptive parameter selection is used to obtain the ideal parameters for the combination. This indicates that to enhance the classification effect, ID3 and CART are both the best node splitting criteria. The experiment's performance is measured by the classification error rate and accuracy rate. Equation (8) defines the sample C's categorization error rate.

$$F(e;C) = \frac{1}{n}\sum_{j=1}^{n} JJ(e(w_j) \neq z_j) \quad (8)$$

Equation (9) is used to define the accuracy rate.

$$acc(e;C)\frac{1}{n}\sum_{j=1}^{n} JJ\big(e(w_j) = z_j\big) = 1 - F(e;C) \quad (9)$$

The addition of adaptive learning mechanisms inside the Random Forest model and implementation in a scalable distributed environment adds significant value to learner classification, engagement prediction, and satisfaction analysis in online educational systems.

Artificial NeuroNet (ANN)

Artificial Neural Networks (ANNs) are exceptionally useful for solving nonlinear, complex problems that often cannot be adequately modelled with traditional quantitative models or alongside statistical approaches. ANNs offer considerable advantages in modeling learner behavior, predicting satisfaction, and classifying engagement in the online education context, even with noisy or incomplete data as inputs. This research utilized a customized ANN architecture for adaptive data-driven decisions in online learning contexts with education analytic data. The proposed architecture has multiple layers, each with layers and features that aid the deepening pattern recognition processes and predictive outcomes.

Input layer

The input layer acts as an initial buffer for cleaned features obtained from the learner dataset. These features include characteristics, such as demographic information, devices used, platforms used, and feedback from the learner. All the information fed through each input node represents individual properties or characteristics of online learning behavior. The completed feature set for the learner is represented as in Equation (10):

$$x_p = [x_{p1}, x_{p2}, \ldots . x_{pn}] \quad (10)$$

This vector x_p represents the input feature set for the p^{th} learner. Each x_{pi} is a normalized or encoded feature such as age, type of device, session duration, or rating of feedback. These variable values will be passed to the next layer for further transformation.

Randomization Layer 1

This layer adds controlled randomness to the input space, allowing the model to be more resilient to variability, noise, and uncertainty in the learner data. By adding a small perturbation to the input features, the network learns to better across the many learning contexts the agent experiences and to reduce overfitting to specific patterns. It modifies each feature as depicted in Equation (11):

$$\widetilde{x_{pi}} = x_{pi} + \in_i, \quad \in_i \backsim N(0,\sigma^2) \quad (11)$$

Every input feature x_{pi} is disturbed by a small random noise $\in_i$, which is drawn from a normal distribution with a mean of 0 and variance σ^2. This allows the network to be more generalizable toward noisy and irregular learner data.

Fusion Layer 1

The first fusion layer combines the randomized features with a nonlinear transformation that permits the network to learn complex interrelationships and hidden dependencies among the data. The first fusion layer is to combine multiple input dimensions into one representation of learner behavior. The output is computed as in Equation (12):

$$z_{pj}^{(1)} = f^{(1)}\Big(\sum_{i=1}^{n} w_{ij}^{(1)}\widetilde{x_{pi}} + b_j^{(1)}\Big) \quad (12)$$

$z_{pj}^{(1)}$ denotes Fusion layer 1 output, $f^{(1)}$ represents the nonlinear activation function, $w_{ij}^{(1)}$ is the connection weight value, $\widetilde{x_{pi}}$ stands for randomized input feature, $b_j^{(1)}$ is the bias for neuron, i refers to the input feature index, j is the fusion layer index, p denotes the pattern or learner, and n indicates the total input features.

Adaptive Randomization Layer

This layer enhances adaptability by adjusting the parameters of the internal transformation based on the distribution of the features. The network employs the distributions listed in the layers below to adaptively recalibrate its understanding of learner inputs, particularly in evolving educational environments. The transformation is defined as in Equation (13):

$$\widehat{Z_{pj}} = z_{pj}^{(1)} \cdot \gamma_j + \beta_j \quad (13)$$

The output $\widehat{Z_{pj}}$is formed by performing a feature-wise affine transformation, with γ_j and β_j being learnable scale and shift parameters. This process is inspired by batch normalization and enables the network to normalize the internal activations adaptively.

Fusion Layer 2

The second fusion layer further focuses the feature representations by fusing the two previous layers' outputs. This processed hierarchical fusion captures the higher-order dependencies within the fused space, giving the network a well-rounded idea of the learning efficacy and satisfaction progression of the learner, which is computed as in Equation (14).

$$z_{pk}^{(2)} = f^{(2)}\left(\sum_{j=1}^{m} w_{jk}^{(2)} \widetilde{z_{pj}} + b_k^{(2)}\right) \quad (14)$$

Adaptive Randomization Layer

A secondary adaptive randomization stage has been included to accommodate adaptive changes before final classification. This additional layer supports the adaptability of the model, furthering its resilience to outliers or unusual behavior, which is often found in online education datasets, as illustrated in Equation (15).

$$\widehat{Z_{pk}} = z_{pk}^{(2)} \cdot \gamma_k' + \beta_k' \quad (15)$$

Output layer

The output layer provides the final predictions of the network. These predictions may include learner engagement classification, satisfaction level prediction, and performance outcome prediction. Each output node weighs on a specific decision variable relevant to personalized online learning analytics.

Predicting learner engagement and satisfaction in online education with accuracy is the goal of the suggested ARF-ANN model. This hybrid approach is scalable and flexible enough to be used in a real-world online learning environment, allowing for individualized insights without depending on visual inputs.

4. Results and Discussion

The ARF-ANN framework utilized Python and accompanying libraries (Pandas, NumPy, Scikit-learn, and PyTorch), which enable scalable, data-driven learning analytics. The flexibility of Python allowed for natural integration of pre-processing and adaptive Random Forest iterations and other deep neural network components. By integrating both structured and semi-structured educational data, the system produces personalized real-time predictions of learner satisfaction and engagement to facilitate adaptive interventions in online educational contexts.

4.1 Metrics for evaluating the effectiveness of the proposed method

The performance of the ARF-ANN model was assessed in terms of F1-score, recall, accuracy, and precision to evaluate the model's ability to adapt to different learner behaviors and deliver personalized feedback on satisfaction and engagement in online learning.

- Accuracy provides insight into how well the model accurately classified learners into satisfaction levels in an online learning environment. It indicates the overall success of the model in predicting learners' outcomes.
- Precision is the number of students that were accurately predicted to be highly or moderately satisfied. Precision helps the model to have few false alarms when it predicts satisfied or engaged students.

- Recall refers to how well the model can detect all the truly satisfied or engaged students from the data. It is important in online learning not to leave out any truly satisfied students in predictions.
- F1-score synthesizes precision and recall to provide a balanced measure of the model's capacity to deliver accurate and complete feedback. It is especially valuable in online learning when satisfaction levels can be unbalanced.

4.2 Performance evaluation of the proposed ARF-ANN model

This confusion matrix demonstrates how well the model predicted the satisfaction level of learners, categorized into three classes: 0 (Low), 1 (Moderate), and 2 (High). Most predictions were correct for class 1, with 163 out of 164 correct predictions, indicating strong model performance concerning moderately satisfied learners. Class 2 is also very good, with 22 correct predictions, and class 0 had 10 correct predictions, though with some misclassifications. There were very few incorrect predictions per class, indicating that minimal confusion was produced between the classifications. Overall, this model produced high accuracy and strong classification ability. Figure 2 shows the confusion matrix for three classes.

Figure 2: Confusion matrix for predicting the satisfaction level of learners

The connection between the extracted features and the final target label in the online education dataset is shown graphically in Figure 3. A strong positive correlation of 0.74 is identified between satisfaction score and target label, along with moderate correlations with engagement score (0.41), resource effectiveness (0.35), and communication effectiveness (0.34). This suggests that satisfaction is a large part of predicting learner outcomes. Weak correlations are identified with confidence score (0.19) and practical experience (0.15), which suggests a limited role in prediction. Specifically, instruction quality and youth show negligible correlation. This information helps to identify which features most strongly influence satisfaction and performance in an online education environment.

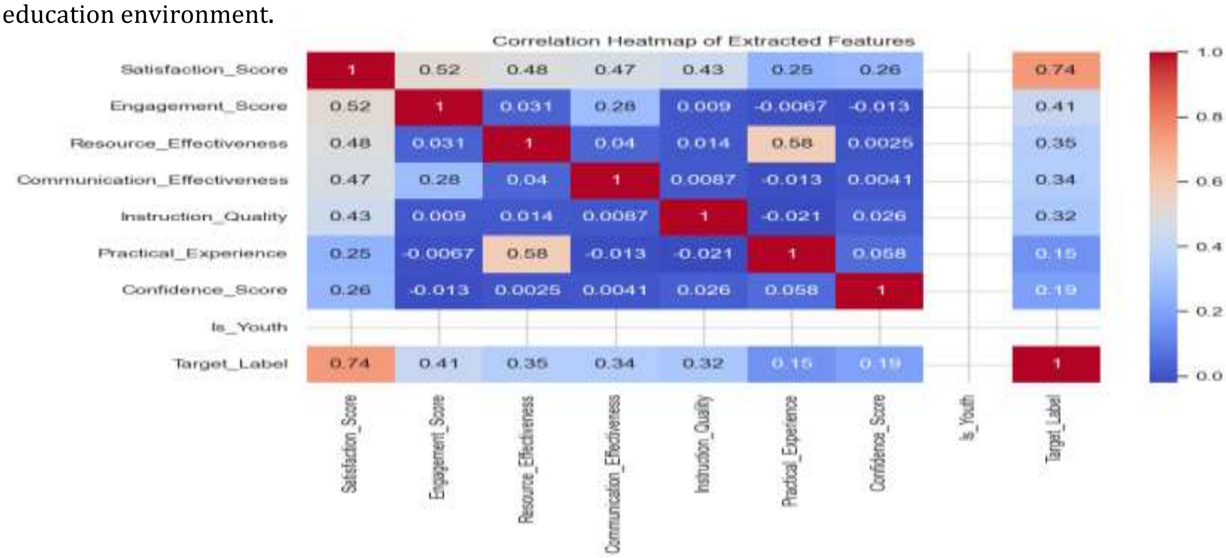


Figure 3: Correlation heat map of extracted features

The feature importance graph in Figure 4 shows which criteria have the biggest impact on predicting the target label in the context of online learning. The satisfaction score was found to be the most contributory factor, contributing to the prediction model by over 70%. The other factors, engagement score, resource effectiveness, and communication effectiveness, are also of importance, but of lesser importance. Overall, this graph shows that learner satisfaction is the primary driver while predicting success or satisfaction levels in online learning environments.

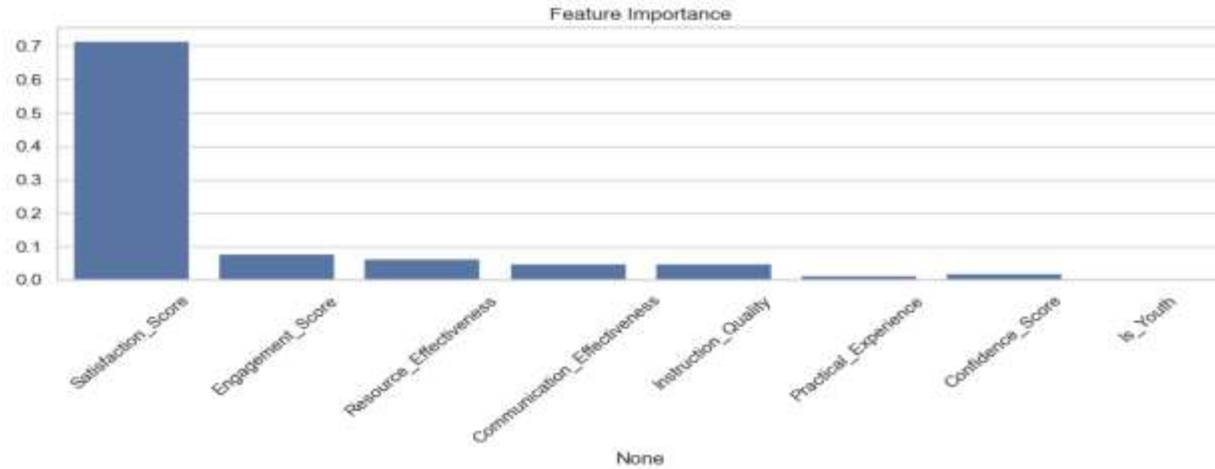


Figure 4: Feature importance graph for predicting the target label

The reliability of the ARF-ANN model's training and validation throughout 43 epochs is shown in Figure 5. Both curves increase quickly during the initial phases, indicating effective learning. The validation accuracy remains consistently higher than the training accuracy and peaks at 98%, indicating good generalization and minimal overfitting. The training accuracy increases gradually and eventually stabilizes at 94%. Overall, the model performs well and has a reliable performance in classifying learner satisfaction.

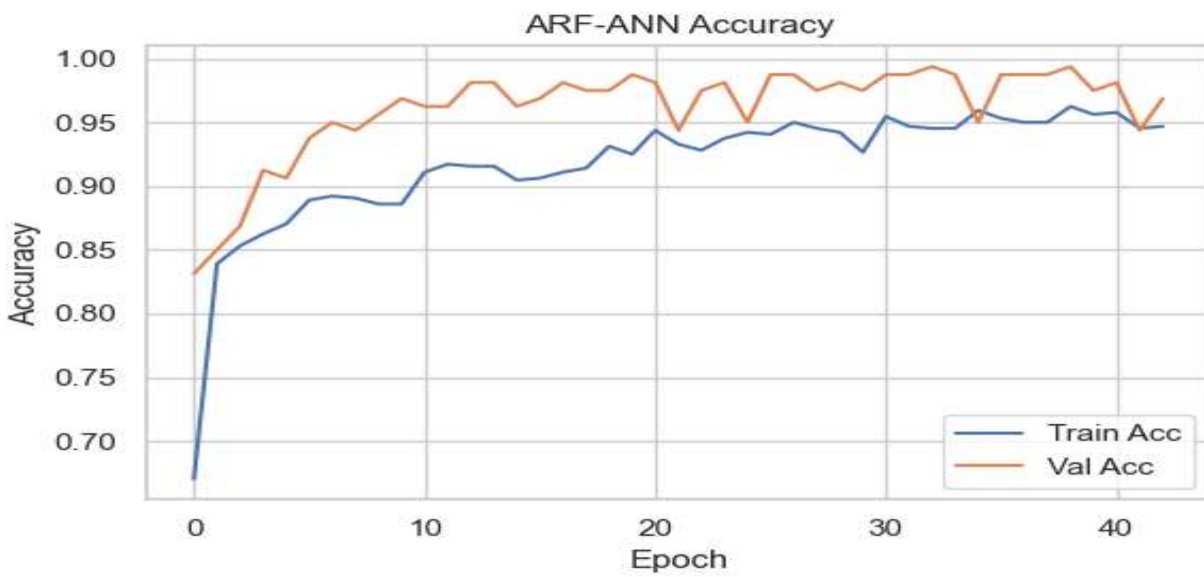


Figure 5: ARF-ANN Training and Validation Accuracy Curve

The ROC (Receiver Operating Characteristic) curve depicts the performance of a multi-class classification model for predicting learner satisfaction levels, as shown in Figure 6. This curve depicts the true positive rate against the false positive rate for each class. Class 0 and Class 2 had perfect predictions, as shown with an AUC (Area Under Curve) of 1.00, while Class 1 performed exceptionally well with an AUC of 0.99. These high values for AUC demonstrate that the model is highly accurate and are able to distinguish satisfaction levels effectively.

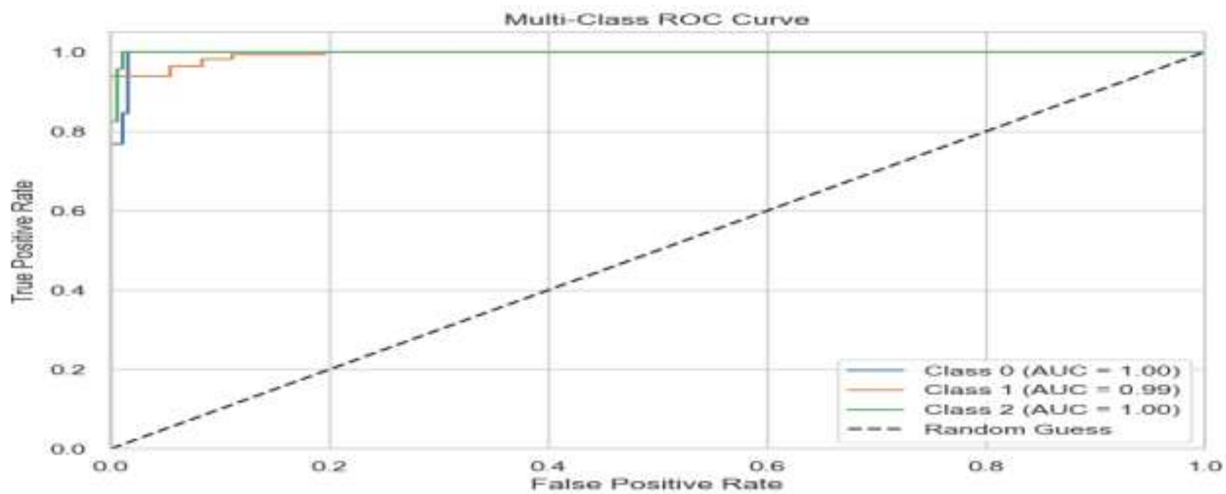
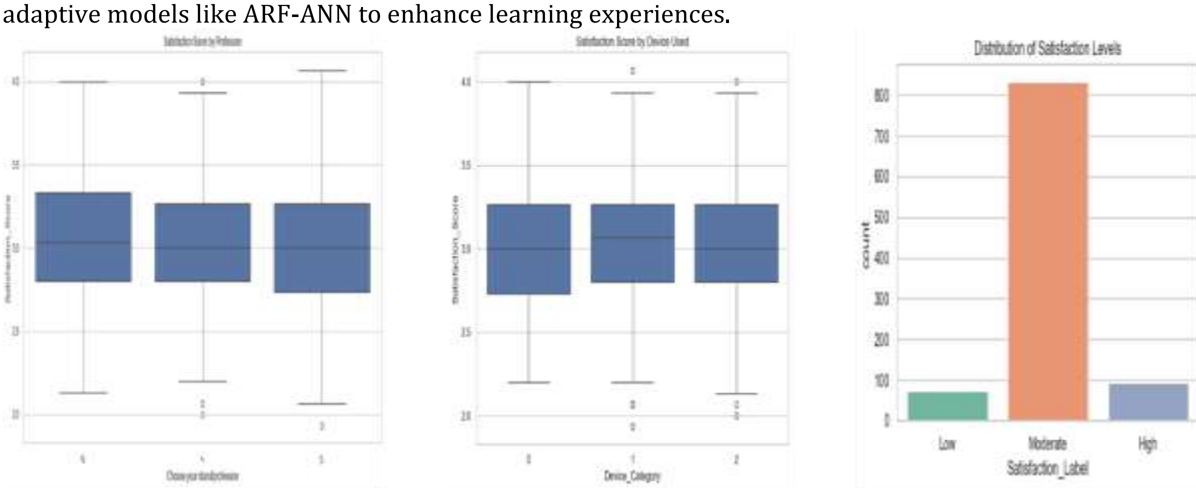


Figure 6: ROC curve for multi-class classification model

Learner satisfaction levels in online learning are clearly displayed in Figure 7. Plot (a) shows that satisfaction scores are generally similar across different professions, such as full-time student, part-time student, and working professional, with medians around 3.0; however, some variability and small group outliers exist. Plot (b) compares satisfaction, taking into account the use of different devices, finding similar satisfaction levels regardless of device type, suggesting minimal influence of the device type. In both cases, a few extreme cases appear to be related to some personal technical or contextual issues. Plot (c) gives an overview of satisfaction levels that showed most learners to be moderately satisfied, with only a small number of the learners reporting low and high levels of satisfaction, allowing the potential for adjustment and improvement. These visualizations help identify where engagement and quality can be optimized. Together, the graphs validate the need for adaptive models like ARF-ANN to enhance learning experiences.



(a) Satisfaction score by profession (b) Satisfaction score by device used (c) Distribution of satisfaction levels

Figure 7: Learner satisfaction levels in online learning (a) profession, (b) Device used, and (c) distribution of satisfaction levels

Performance measures of the new ARF-ANN model reveal its strong efficiency in classifying student satisfaction in online learning environments. The model achieved a level of prediction accuracy of 97.50%, demonstrating an exceptional overall prediction ability. A precision of 97.54% confirms the model's consistency in precise classification of satisfied students, while the recall value of 97.50% shows its performance ability in identifying the majority of real cases. The 97.41% F1-score metric captures an equally weighted precision and recall performance that verifies the robustness and flexibility of the model in working with online learner data. Figure 8 shows the overall performance of the proposed ARF-ANN model.

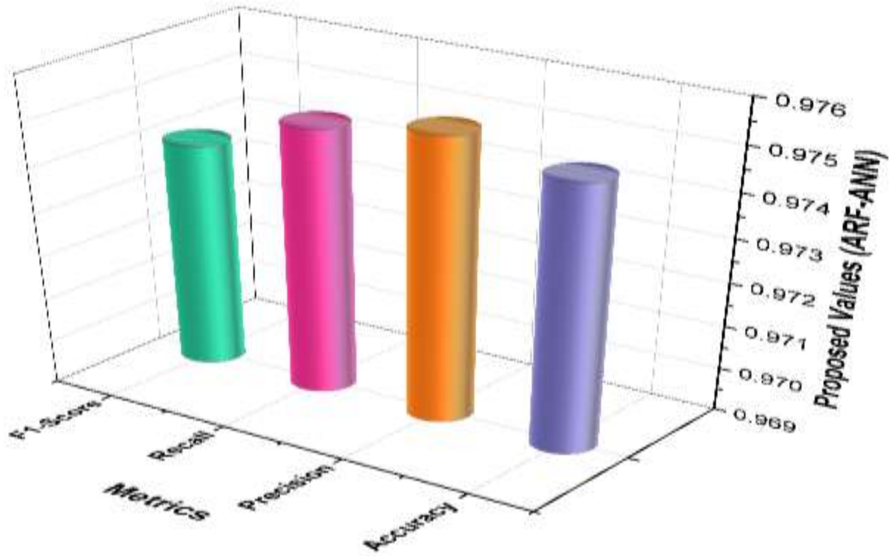


Figure 8: The overall performance of the proposed ARF-ANN model

The proposed ARF-ANN model aims to accurately predict learner satisfaction and engagement in online education. This hybrid approach enables personalized insights without relying on visual inputs, making it scalable and adaptable to a real-world online learning environment.

4.3 Discussion

Online learning is a digital mode of education that allows learners to access instructional content and engage with educators remotely through internet–based platforms. The purpose of [21] was to investigate the devices through which various forms of peer feedback shape learning burnout in online learning environments. Deep learning models, such as Bidirectional Encoder Representations from Transformers (BERT), Bi-directional Long Short-Term Memory (Bi-LSTM), and BERT-Bi-LSTM, were employed to annotate peer comments with cognition and affective content supported by multiple linear regression (MLR) analysis. However, limitations include the relatively small sample size (116 participants) and the potential subjectivity in interpreting feedback sentiment. Additionally, model performance may vary across different learner populations and platforms. The present research [22] seeks to promote student engagement in online learning through tracking real-time emotions during virtual lessons. It utilizes innovative deep models combining facial landmark localization, emotion recognition, and survey weights to establish a Mean Engagement Score (MES). But limitations relied on camera quality, lighting, and students' willingness to have their cameras on. Moreover, emotional interpretation can be subjective from person to person and therefore might not be accurate.

To overcome these limitations, this research suggests the ARF-ANN model based on structured learner data and semi-structured feedback to forecast engagement and satisfaction without the need for facial analysis or subjective interpretive emotion. In contrast to previous visual-based or shallow sample-sized strategies, ARF-ANN benefits from multiple behavior and feedback indicators from a rich dataset. Its adaptive Random Forest-neural network blend ensures robust performance regardless of learner profiles. This enables reliable, large-scale, and camera-independent measurement of involvement. Thus, ARF-ANN offers a more inclusive and reliable solution for enhancing online learning experiences.

5. Conclusion

Online education has quickly become a common mode of learning with flexibility, accessibility, and choices globally for learners. The proposed Adaptive Randomized Fusion-layered Artificial NeuroNet (ARF-ANN) architecture will facilitate accurate predictions of learner satisfaction and engagement. By analyzing a comprehensive dataset of 1,000 responses, integrating structured and semi-structured characteristics, a complete general view of the learner was produced. The multi-layered structure can be developed using Python-based feature engineering and hybrid modeling approaches, and can be flexible, scalable, and very robust. The evaluation metrics, F1-score (0.9741), precision (0.9754), recall (0.9750), and accuracy (0.9750), validate the effectiveness of the ARF-ANN model, while important correlations between learner occupation, access to device use, and learner satisfaction highlight key influencers in online learning. Overall, this research contributes a compelling, real-time, and personalized analytics framework to improve decision making and learner engagement in digital education ecosystems. The research is limited because it relies on self-reported data, which may introduce bias, as well as the data set size, which may not capture all international learner behavior variability.

Future scope

Future research may integrate real-time data on behaviors (for example, keystroke dynamics, eye tracking) and expand data sets across multiple regions because the variety in time supports enhanced generalizability and reliability of the model. Furthermore, this framework could be applied effectively through an e-learning platform, education technology companies, corporate training, and higher education to personalize learning analytics and engagement.

References

1. Matthew, U.O., Kazaure, J.S. and Okafor, N.U., 2021. Contemporary development in E-Learning education, cloud computing technology & Internet of Things. EAI Endorsed Trans. Cloud Syst., 7(20), p.e3. https://doi.org/10.4108/eai.31-3-2021.169173
2. Gandolfi, E., Ferdig, R.E. and Kratcoski, A., 2021. A new educational normal: an intersectionality-led exploration of education, learning technologies, and diversity during COVID-19. Technology in society, 66, p.101637. https://doi.org/10.1016/j.techsoc.2021.101637
3. Peimani, N. and Kamalipour, H., 2021. Online education and the COVID-19 outbreak: A case study of online teaching during lockdown. Education Sciences, 11(2), p.72. https://doi.org/10.3390/educsci11020072
4. Roa González, J., Sánchez Sánchez, N., Seoane Pujol, I. and Díaz Palencia, J.L., 2025. Challenges and perspectives in the evolution of distance and online education towards higher technological environments. Cogent Education, 12(1), p.2447168. https://doi.org/10.1080/2331186X.2024.2447168
5. Pregowska, A., Masztalerz, K., Garlińska, M. and Osial, M., 2021. A worldwide journey through distance education—from the post office to virtual, augmented, and mixed realities, and education during the COVID-19 pandemic. Education Sciences, 11(3), p.118. https://doi.org/10.3390/educsci11030118
6. Carabregu-Vokshi, M., Ogruk-Maz, G., Yildirim, S., Dedaj, B. and Zeqiri, A., 2024. 21st century digital skills of higher education students during Covid-19—is it possible to enhance digital skills of higher education students through E-Learning?. Education and Information Technologies, 29(1), pp.103-137. https://doi.org/10.1007/s10639-023-12232-3
7. El-Sabagh, H.A., 2021. Adaptive e-learning environment based on learning styles and its impact on the development of students' engagement. International Journal of Educational Technology in Higher Education, 18(1), p.53. https://doi.org/10.1186/s41239-021-00289-4
8. Sundeen, T.H. and Kalos, M., 2022. Rural educational leader perceptions of online learning for students with and without disabilities before and during the COVID-19 pandemic. Theory & Practice in Rural Education, 12(2), pp.105-128. https://doi.org/10.3776/tpre.2022.v12n2p105-128
9. Ludwig, C. and Tassinari, M.G., 2023. Foreign language learner autonomy in online learning environments: the teachers' perspectives. Innovation in Language Learning and Teaching, 17(2), pp.217-234. https://doi.org/10.1080/17501229.2021.2012476
10. Kem, D., 2022. Personalised and adaptive learning: Emerging learning platforms in the era of digital and smart learning. International Journal of Social Science and Human Research, 5(2), pp.385-391. https://doi.org/10.47191/jsshr/v5-i2-01
11. Ismail, S. and Ling, Z., 2025. Digital Learning: A Solution for More Inclusive and Affordable Education. International Journal of Education and Digital Learning (IJEDL), 3(4), pp.191-200. https://doi.org/10.47353/ijedl.v3i4.260
12. Celik, I., Gedrimiene, E., Silvola, A. and Muukkonen, H., 2023. Response of learning analytics to the online education challenges during the pandemic: Opportunities and key examples in higher education. Policy Futures in Education, 21(4), pp.387-404. https://doi.org/10.1177/14782103221078401
13. Ouyang, F., Wu, M., Zheng, L., Zhang, L. and Jiao, P., 2023. Integration of artificial intelligence performance prediction and learning analytics to improve student learning in an online engineering course. International Journal of Educational Technology in Higher Education, 20(1), p.4. https://doi.org/10.1186/s41239-022-00372-4

14. Zhou, M., Chen, G.H., Ferreira, P. and Smith, M.D., 2021. Consumer behavior in the online classroom: Using video analytics and machine learning to understand the consumption of video courseware. Journal of Marketing Research, 58(6), pp.1079-1100. https://doi.org/10.1177/00222437211042013
15. Yoon, M., Lee, J. and Jo, I.H., 2021. Video learning analytics: Investigating behavioral patterns and learner clusters in video-based online learning. The internet and higher education, 50, p.100806. https://doi.org/10.1016/j.iheduc.2021.100806
16. Chouhan, R., 2021, December. Effective interactive video assignments and rewatch analytics for online flipped classrooms. In 2021 IEEE 1st International Conference on Advanced Learning Technologies on Education & Research (ICALTER) (pp. 1-4). IEEE. https://doi.org/10.1109/ICALTER54105.2021.9675132
17. Alam, A. and Mohanty, A., 2022, November. Business models, business strategies, and innovations in EdTech companies: integration of learning analytics and artificial intelligence in higher education. In 2022 IEEE 6th Conference on Information and Communication Technology (CICT) (pp. 1-6). IEEE. https://doi.org/10.1109/CICT56698.2022.9997887
18. Megli, A.C., Fallad-Mendoza, D., Etsitty-Dorame, M., Desiderio, J., Chen, Y., Sanchez, D., Flor, N. and Gunawardena, C.N., 2024. Using social learning analytic methods to examine the social construction of knowledge in online discussions. American Journal of Distance Education, 38(1), pp.65-80. https://doi.org/10.1080/08923647.2023.2192597
19. Alam, A., 2023, May. Improving learning outcomes through predictive analytics: Enhancing teaching and learning with educational data mining. In 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 249-257). IEEE. https://doi.org/10.1109/ICICCS56967.2023.10142392
20. Zhang, S., Chen, J., Wen, Y., Chen, H., Gao, Q. and Wang, Q., 2021. Capturing regulatory patterns in online collaborative learning: A network analytic approach. International Journal of Computer-Supported Collaborative Learning, 16(1), pp.37-66. https://doi.org/10.1007/s11412-021-09339-5
21. Huang, C., Tu, Y., Han, Z., Jiang, F., Wu, F. and Jiang, Y., 2023. Examining the relationship between peer feedback classified by deep learning and online learning burnout. Computers & Education, 207, p.104910. https://doi.org/10.1016/j.compedu.2023.104910
22. Bhardwaj, P., Gupta, P.K., Panwar, H., Siddiqui, M.K., Morales-Menendez, R. and Bhaik, A., 2021. Application of deep learning on student engagement in e-learning environments. Computers & Electrical Engineering, 93, p.107277. https://doi.org/10.1016/j.compeleceng.2021.107277