Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 2, No.2, February : 2025

IMPROVING ONLINE EDUCATION THROUGH MACHINE LEARNING INSIGHTS ON FLEXIBILITY

Dasari Veeraswamy, PhD Scholar, Osmania University, Hyderabad, Telangana, India and Assistant Professor, Department of Computer Science and Engineering, Aditya University, surampalem, Kakinada, Andhra Pradesh. India. dveeraswamy6008@gmail.com
Dr. A V Krishna Prasad, Associate professor, Department of Information Technology, Maturi Venkata Subba Rao Engineering College, Hyderabad, Telangana, India. kpvambati@gmail.com

ABSTRACT:
The rapid evolution of technology has reshaped the educational landscape, fostering the growth of online learning platforms. To address the persistent challenges in adaptability and misclassification accuracy within predictive models, this research aimed to enhance classification performance by proposing a Two-Stage Hybrid Classification Model. The study fills the gap in improving misclassification handling through an innovative framework combining a primary base classifier with an XGBoost-based meta-learner. The methodology involved training the primary classifier on a comprehensive dataset to identify misclassified samples, which were then processed by the XGBoost meta-learner to refine predictions. The novelty lies in integrating the strengths of traditional and advanced machine learning techniques, effectively leveraging XGBoost's iterative learning capabilities. The proposed model achieved an impressive accuracy of 97%, significantly outperforming traditional approaches such as Random Forest (85%) and Decision Tree (83%). These findings underscore the model's potential for applications in fields requiring high predictive precision, such as education and healthcare. This work demonstrates the transformative impact of hybrid machine learning frameworks in tackling complex classification challenges, providing a robust solution for improving reliability and accuracy in real-world applications.
Keywords: Two-Stage Classification, XGBoost, Hybrid Model, Online Education, Predictive Accuracy, Machine Learning.

INTRODUCTION:
The rapid advancement of technology has transformed the educational landscape, enabling the proliferation of online learning platforms [1]. This shift has been particularly significant in recent years, where virtual classrooms have become integral to delivering quality education across diverse demographics. However, the effectiveness of online education is not solely determined by the accessibility of technology but is also influenced by the ability of students to adapt to the online learning environment. Flexibility in adjusting to digital platforms, navigating new tools, and managing self-paced learning are critical factors that determine a student's success in such settings [2]. Addressing these aspects is essential for enhancing the overall efficacy of online education.
Despite its advantages, online education presents a unique set of challenges for both educators and learners [3]. Students often encounter difficulties in adapting to online learning due to a lack of structured environments, varying technological proficiencies, and inconsistent internet access [4]. Furthermore, differences in individual adaptability levels, influenced by factors such as age, prior exposure to technology, and socio-economic conditions, contribute to disparities in learning outcomes. These challenges highlight the need for a deeper understanding of the factors influencing student flexibility and the development of strategies to support learners more effectively.
One of the primary challenges lies in identifying and addressing the diverse adaptability levels among students [5]. Educational institutions and policymakers frequently face limitations in providing personalized support due to a lack of detailed insights into individual learning behaviors [6]. Moreover, traditional methods of assessing adaptability often fail to account for the dynamic and multifaceted nature of online learning environments. This gap emphasizes the need for innovative

approaches that can analyze and predict adaptability levels with greater accuracy. Without such mechanisms, the potential of online education to provide equitable learning opportunities remains underutilized.

The integration of data-driven methodologies to enhance student adaptability further faces obstacles related to data collection, interpretation, and application. Educational datasets often exhibit complexity and variability, requiring sophisticated analytical techniques to extract meaningful patterns [7]. Additionally, ensuring the ethical use of data while maintaining student privacy poses significant challenges. The lack of robust frameworks for implementing such methodologies can hinder the ability of educators to design interventions that improve student engagement and learning outcomes in virtual settings.

Machine learning has emerged as a transformative tool capable of addressing these challenges by providing actionable insights into student adaptability. By leveraging vast datasets, machine learning models can identify patterns and predict factors influencing student flexibility with high precision. These insights enable educators and policymakers to design tailored interventions that foster a more inclusive and supportive online learning environment [8]. Through its ability to process complex data and generate predictive models, machine learning holds the potential to bridge the gap between technological advancements and effective educational practices, ensuring that online education becomes more accessible, equitable, and impactful for learners worldwide.

LITERATURE SURVEY:

N. S. Koti Mani Kumar Tirumanadham et al. [9] introduced a BR2-2T feature selection technique combining Ridge (L2) regularization and Boruta optimization to enhance prediction accuracy. A three-tier ensemble model was developed, integrating Random Forest with Bayesian Optimization, SVM with random search, and Gradient Boosting with PSO for hyperparameter tuning. Techniques like Z-score normalization, SMOTE, and MICE were employed for data standardization, imbalance handling, and error management. The proposed approach achieved a maximum accuracy of 98.74%, outperforming conventional methods. It significantly improved educational outcome predictions and decision-making processes. The study provides insights into leveraging advanced algorithms for student success and educational strategies.

Jewoong Moon et al. [10] developed a predictive model to assess representational flexibility (RF) in autistic adolescents undergoing VR-based cognitive training. The study utilized data from 178 sessions with eight participants, integrating behavioral cues, physiological responses, and interaction logs. Advanced machine learning techniques, including random forest with decision-level data fusion, were applied to analyze this multimodal dataset. The model outperformed single-source approaches, achieving high accuracy in predicting RF development. This research highlights VR's potential to enhance cognitive skills and demonstrates the value of multimodal data fusion in understanding complex cognitive processes.

Tariq Khasawneh et al. [11] introduced a method to interpret subregions of the observation space by fitting simple linear models to subsets rather than the entire space, yielding parsimonious fits. Applied to the French Motor Claims dataset, this approach approximated a black-box predictive model and outperformed alternative surrogate models. The method achieved an ROC AUC score of 0.67, strong Spearman correlation, and no significant median difference compared to the black-box model per Wilcoxon’s test. It also showed a 5% higher intraclass correlation coefficient (ICC) than the next-best surrogate model. The method serves as both a black-box model interpreter and a standalone predictive model, offering superior performance and interpretability.

Chaman Verma et al. [12] identified challenges in hybrid learning, such as internet disconnections, limited technical support, reduced competitiveness, exam cheating risks, and decreased focus and interactivity. They recommended enhanced assistance and safety measures for higher education during pandemics. Random Forest (RF) was used to predict student happiness with hybrid learning features, achieving 88% accuracy and outperforming Logistic Regression (LR), XGBoost, and other classifiers. The RF model's performance was validated using metrics like F1-score, precision, recall, and specificity, confirming its effectiveness.

Ahmed M. Khedr et al. [13] reviewed the role of deep learning (DL) and machine learning (ML) in various aspects of supply chain management (SCM), including supplier selection, production, inventory control, transportation, demand forecasting, and sales estimation. The study highlights strategies to enhance operational efficiency, address current challenges, and explore future research opportunities. It provides a detailed examination of DL and ML integration with SCM and their potential to optimize processes. A comprehensive literature table summarizes existing research, outlining objectives, findings, and areas for improvement. This table offers quick insights into advancements in SCM powered by DL and ML, presenting a clear understanding of the evolving field.

Mohd Javaid et al. [14] examined Industry 4.0's impact on Flexible Manufacturing Systems (FMS) and its role in enhancing performance through advanced technologies. The study emphasizes Industry 4.0's ability to boost flexibility by leveraging virtual infrastructure and cloud services, enabling auto-scaling to match changing resource demands. This adaptability allows production facilities to respond swiftly to market changes, with plant control systems adjusting outputs based on utility rates to reduce costs. Industry 4.0 practices significantly improve production efficiency and flexibility, demonstrating remarkable progress and benefits in recent years.

Teresa M. C. Pereira et al. [15] reviewed flexible cardiac sensing devices for ECG monitoring, focusing on key features like flexibility, durability, biocompatibility, and sensitivity. The study examines fabrication methods and materials used for flexible electrodes and their various applications. It highlights the role of machine learning (ML) in cardiac health monitoring, using techniques like deep learning, support vector machines, and random forest for tasks such as heart disease classification, emotion detection, and biometric recognition. The integration of ML with flexible sensors is emphasized for advancing ECG-based monitoring. The paper concludes with current advancements and future research directions in this domain.

Amira Bourechak et al. [16] investigated the integration of AI and edge computing across diverse domains, highlighting its potential and identifying new opportunities. The combination of AI and edge enhances user experiences in critical areas like the Internet of Vehicles, where delays or inaccuracies can lead to accidents. The review examines eight key applications: smart agriculture, environment, grid, healthcare, industry, education, transportation, and security/privacy. A qualitative comparison emphasizes AI's roles, objectives, and enabling technologies at the network edge. The study also discusses open challenges, future research directions, and perspectives, concluding with insights into the confluence of AI and edge computing.

PROPOSED MODEL:

XGBoost :

The code demonstrates the use of XGBoost (eXtreme Gradient Boosting) for classification tasks. XGBoost is a powerful machine learning algorithm that builds an ensemble of decision trees iteratively, focusing on reducing prediction errors at each step.

The implementation begins with the following steps:

Initialization and Training: An XGBoost classifier is initialized with key hyperparameters:

- **n_estimators=100**: The number of boosting rounds (trees) to train.
- **max_depth=5**: The maximum depth of each tree, controlling model complexity and overfitting.
- **learning_rate=0.1**: The step size shrinkage used to update weights after each tree iteration, ensuring gradual learning.
- **random_state=42**: A seed value for reproducibility.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 2, No.2, February : 2025

The classifier is trained using the fit function on the training dataset X_train and y_train.

Prediction and Error Calculation: Predictions are made on the test dataset using predict. The model's performance is evaluated using the classification report, which provides metrics like precision, recall, and F1-score. Misclassified samples are identified by comparing y_test with y_pred_stage1.

Residual Analysis: The misclassified samples (instances where the model made incorrect predictions) are extracted from the dataset. These residuals serve as inputs for the next stage of training, allowing the subsequent model to focus on harder-to-predict cases.

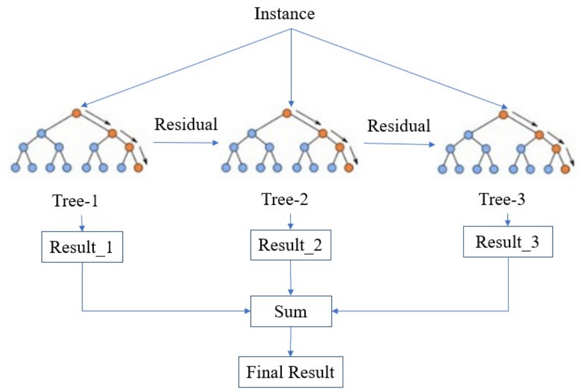

Figure 1: XGBoost Architecture

The figure 1 illustrates the architecture of XGBoost, which operates as follows:

Sequential Tree Construction: Each tree in the ensemble predicts the residual errors from the previous tree, refining the overall predictions iteratively.

Residual Minimization: After each tree, the residuals (differences between actual values and predictions) are calculated and used as input for the next tree. This process ensures that each tree focuses on correcting the errors of its predecessors.

Final Prediction: The predictions from all trees are summed to generate the final result, as shown in the diagram. This cumulative approach is key to reducing errors and achieving high accuracy.

TWO-STAGE HYBRID CLASSIFICATION MODEL

The Two-Stage Hybrid Classification Model is a novel approach designed to improve the predictive performance of classification tasks by integrating the strengths of a primary classifier and a meta-learner. This model is particularly effective in scenarios where the primary classifier's performance is limited by specific misclassifications. By introducing a second stage to focus on these misclassified samples, the model refines its predictions and achieves greater overall accuracy and robustness.

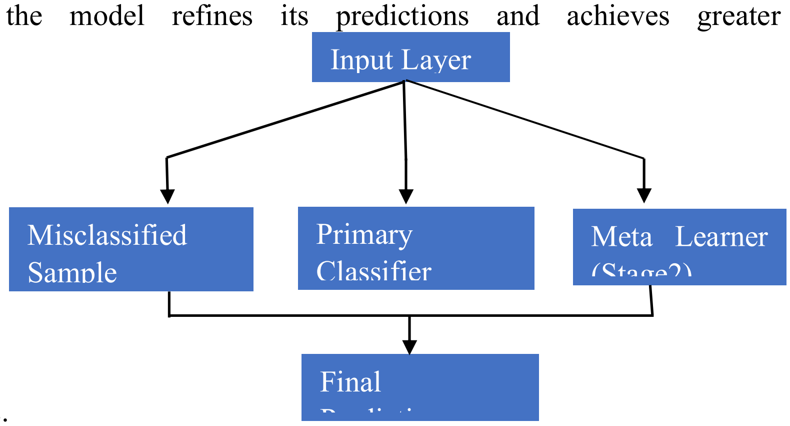

Figure 2: Block Diagram Structure of Two-Stage Hybrid Classification Model

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 2, No.2, February : 2025

The enhanced implementation builds upon the two-stage hybrid classification model, as shown in Figure 2, with the following steps:

Stage 1: Initial Classification: In the first stage, a baseline classifier is trained on the complete dataset to provide initial predictions. This stage establishes a general understanding of the classification problem and identifies the primary trends and patterns in the data. The predictions from this stage, denoted as $\hat{y}_{stage1}$, form the foundation for further refinement.

Identifying Misclassified Samples: After obtaining the predictions, the model evaluates the discrepancies between the predicted labels and the actual labels. A Boolean mask misclassified_mask is generated to highlight instances where the predictions are incorrect ($y_{test} \neq \hat{y}_{stage}$). The indices of these misclassified samples are then extracted and used to isolate the corresponding features and labels from the test set for further processing.

Stage 2: Meta-Learner for Misclassified Samples: The second stage employs an XGBoost classifier as the meta-learner to focus exclusively on the misclassified samples identified in the first stage. By training on this subset, the meta-learner is able to target the weaknesses of the primary classifier and make specialized adjustments to its predictions. The meta-learner's predictions, $\hat{y}_{meta}$, are then used to correct the errors made in Stage 1.

Combining Predictions: The final predictions $\hat{y}_{final}$ are generated by combining the outputs of both stages. For correctly classified samples from Stage 1, the initial predictions are retained. For misclassified samples, the predictions from the meta-learner are used to update the final output:

$$\hat{y}_{final}[\mathrm{i}] = \begin{cases} \hat{y}_{meta}[\mathrm{i}], & if\ i\ \in\ misclassified\_indices \\ \hat{y}_{stag}\ [\mathrm{i}], & otherwise. \end{cases}$$

Evaluation: The performance of the combined model is evaluated using standard classification metrics such as precision, recall, F1-score, and overall accuracy. These metrics provide a comprehensive assessment of the model's effectiveness in handling both the general dataset and the targeted misclassifications.

$$\hat{\boldsymbol{y}} = \begin{cases} \hat{y}_{Stage}\ , & if\ correctly\ classified \\ \hat{y}_{Stage2}, & if\ misclassified\ in\ Stage1 \end{cases}$$

Where:

- $\hat{y}_{Stage1}$: is the prediction from the first classifier.
- $\hat{y}_{Stage}$: is the correction from the meta-learner.

ALGORITHM WITH MATHEMATICAL EQUATIONS:

The algorithm for Proposed Model involves the following steps:

Algorithm: Two-Stage Classification with Meta-Learner

Step 1: Identify Misclassified Samples

1. Compute a misclassification mask:

$$M[i] = \begin{cases} 1, & if\, y_{test}[i] \neq y_{pred,stag}\ \ [i] \\ 0, & otherwise \end{cases}$$

$$M = \{i | \hat{y}_i \neq y_i\}$$

Where:
M: represents the set of indices of misclassified samples.
$\hat{y}_i$: is the predicted label.
y_i: is the actual label.

2. Extract the indices of misclassified samples:

$$Indices_{misclassified} = i | M[i] = 1$$

3. Extract the features ($X_{misclassified}$) and true labels($y_{misclassified}$):

$$X_{misclassified} = X_{test}\left[Indices_{misclassified}\right]$$

$$y_{misclassified} = y_{test}[Indices_{misclassified}]$$

Step 2: Train the Meta-Learner

1. Define the meta-learner f_{meta} which is an XGBoost classifier with the following parameters:

- Number of estimators ($n_{estimators}$=50).
- Maximum depth (d_{max}= 3).
- Learning rate (η=0.1).

2. Train the meta-learner on the misclassified samples:

$$f_{meta} \leftarrow Train(f_{meta}, X_{misclassified}, y_{misclassified})$$

Step 3: Meta-Learner Predictions

1. Use the meta-learner to predict labels for the misclassified samples

$$y_{meta\_pred} = f_{meta}(X_{misclassified})$$

Step 4: Final Prediction

1. Initialize final predictions as Stage 1 predictions:

$$y_{final} = y_{pred,stage1}$$

2. Update the predictions for misclassified samples:

$$y_{final}[Indices_{misclassified}] = y_{meta\_pred}$$

Step 5: Evaluate the Combined Model

1. Use a classification evaluation metric (e.g., precision, recall, F1-score):

$$Evaluate(y_{test}, y_{final})$$

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

$$Recall = \frac{TP}{TP + FP}$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Where:

TP: True Positives
FP: False Positives
FP: False Negatives

The Proposed Model introduces a novel two-stage classification framework designed to address the limitations of single-stage predictive systems by enhancing accuracy and focusing on misclassified samples. In the first stage, a primary classifier generates predictions ($y_{pred,stage1}$) based on the input features (X_{test}). Misclassified instances are identified by comparing these predictions to the true labels (y_{test}), using a boolean mask to isolate difficult cases for further refinement. These misclassified samples are extracted ($X_{misclassified}, y_{misclassified}$) and serve as input for a second-stage meta-learner, an XGBoost classifier. This meta-learner is specifically trained to correct errors made in the initial stage, leveraging its ability to handle complex data structures and relationships.

The key contribution of this framework lies in combining the strengths of both stages to form an integrated hybrid model. After training the meta-learner, its predictions (y_{meta_pred})are used to update the final predictions (y_{final}) for misclassified cases, resulting in a robust and adaptive model. This approach ensures that the misclassified samples, which are typically harder to predict, receive focused attention, thereby improving overall classification performance. The model's effectiveness is evaluated using comprehensive metrics, including precision, recall, and F1-score, demonstrating its capability to deliver superior accuracy and reliability compared to single-stage classification methods. By addressing the challenges of traditional models, this proposed two-stage framework showcases its potential to significantly enhance predictive accuracy in complex and imbalanced datasets.

EXPERIMENTAL RESULTS:

In this subsection, we provide a detailed analysis of the results obtained from the proposed approach during the ongoing simulations. The dataset utilized for these simulations was sourced from the Student Flexibility in Online Learning [17]. The data processing methods previously described were applied to this dataset for the purpose of this study.

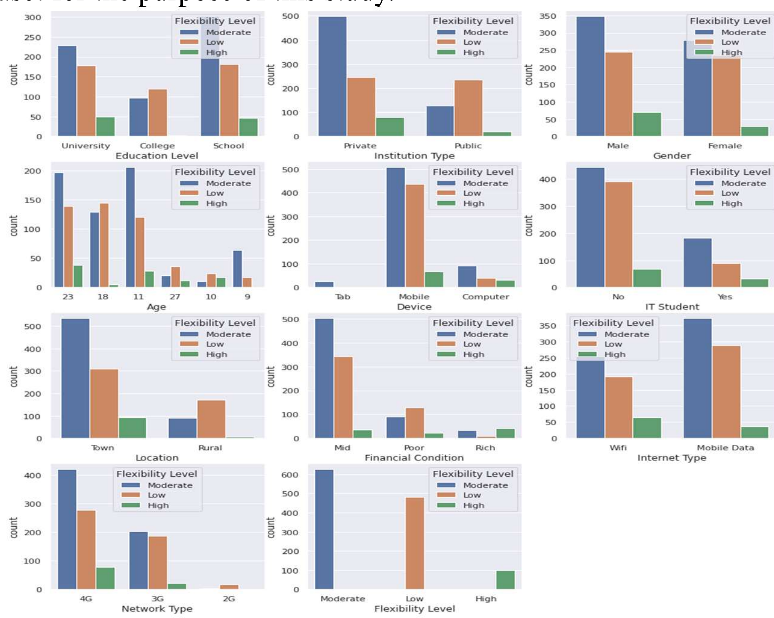

Figure 3: Flexibility Level Distribution Across Various Categories

Education Level: This chart shows the distribution of flexibility levels (Low, Moderate, and High) across different education levels: University, College, and School. University students exhibit the highest moderate flexibility, while school students show a notable presence of low flexibility. High flexibility remains less common across all groups.

Institution Type: The comparison between Private and Public institutions highlights that private institution students dominate the moderate flexibility category. In contrast, public institutions have a relatively higher proportion of students with low flexibility, indicating disparities in adaptability between the two types of institutions.

Gender: Male participants exhibit a significantly higher count in the moderate flexibility category compared to females. However, low flexibility is nearly evenly distributed between genders, suggesting that gender influences adaptability primarily in higher flexibility levels.

Age: Age categories (e.g., 23, 18, 11, etc.) reveal a consistent trend where moderate flexibility is most prevalent across all age groups. However, younger participants (e.g., age 10 and 9) are predominantly in the low flexibility category, highlighting a potential correlation between age and adaptability.

Device: This chart compares flexibility levels based on device usage (Tab, Mobile, and Computer). Mobile device users exhibit the highest moderate flexibility, likely due to its widespread usage and accessibility, whereas computer users are more evenly distributed across flexibility levels.

IT Student: The classification of participants as IT students or not demonstrates that non-IT students exhibit higher moderate flexibility. However, IT students have a greater proportion of high flexibility, potentially due to their technical expertise.

Location: Urban and rural settings significantly affect flexibility levels. Town-based participants dominate the moderate flexibility category, while rural participants are more concentrated in the low flexibility level, reflecting differences in access to resources and technological familiarity.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 2, No.2, February : 2025

Financial Condition: This chart shows flexibility levels for participants with Mid, Poor, and Rich financial conditions. Participants from mid-level financial conditions exhibit the highest moderate flexibility, whereas those with poor financial conditions dominate the low flexibility category.

Internet Type: The comparison between WiFi and Mobile Data users highlights that WiFi users have higher moderate flexibility. Mobile data users are relatively concentrated in the low flexibility category, suggesting an influence of internet stability on adaptability.

Network Type: Flexibility levels are distributed across 4G, 3G, and 2G network types. Participants with 4G access dominate the moderate flexibility category, whereas those using 3G and 2G networks are predominantly in the low flexibility level, reflecting the importance of network quality for adaptability.

Overall Flexibility Level: This chart summarizes the overall distribution of flexibility levels. Moderate flexibility dominates, followed by low flexibility, while high flexibility remains rare, highlighting the need for strategies to improve adaptability across various groups.

Each subsection provides valuable insights into how demographic and technological factors influence adaptability in different scenarios. This detailed analysis helps identify areas that require intervention to enhance flexibility levels across groups.

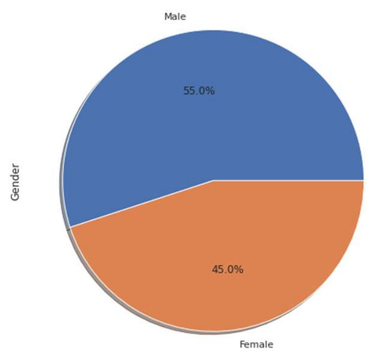

Figure 4: Gender Distribution

Figure 4 illustrates the gender distribution of the participants using a pie chart. The chart reveals that 55% of the participants are male, while the remaining 45% are female. This nearly balanced distribution provides a comprehensive representation of both genders, ensuring that insights derived from the data are inclusive and account for gender-based variations in adaptability, preferences, or behaviors in the study. Such gender balance is critical for analyzing flexibility levels across various dimensions and identifying trends or disparities influenced by gender.

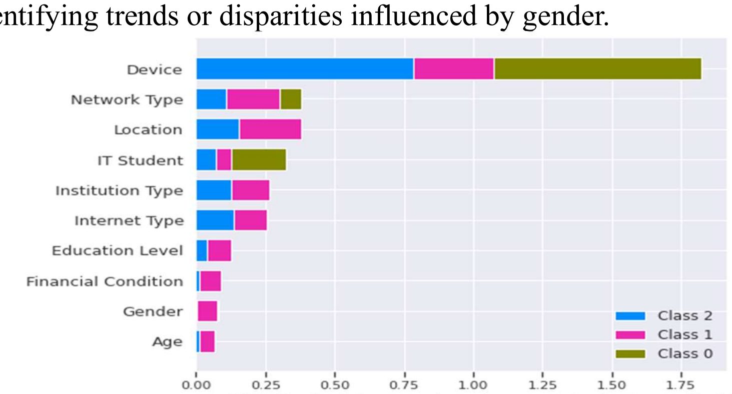

Figure 5: SHAP Value Analysis: Dominant Features Influencing Model Prediction

Figure 5 illustrates the feature importance for a multi-class classification model using SHAP values, focusing on the average impact of each feature on the model's output for Classes 0, 1, and 2. The feature "Device" emerges as the most significant contributor, particularly for Class 0, followed by

"Network Type" and "Location," which show varying levels of influence across the three classes. Other features, such as "IT Student," "Institution Type," and "Internet Type," contribute to the classification decisions with lower magnitude. The color-coded segments within each bar highlight the respective contributions to each class, enabling a deeper understanding of how each feature drives predictions and their relative importance in the model's decision-making.

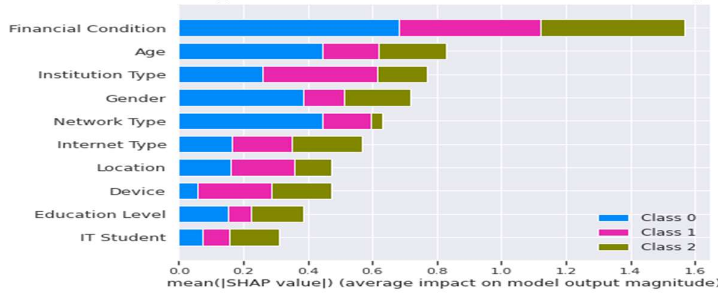

Figure 6: SHAP Value Analysis: Feature Importance Across Classes

Figure 6 provides a visual representation of the feature importance in a multi-class classification model using SHAP (SHapley Additive exPlanations) values. Each bar represents the average impact of a feature on the model's output across three classes: Class 0, Class 1, and Class 2. "Financial Condition" has the highest influence across all classes, followed by "Age," "Institution Type," and "Gender," with varying contributions to different classes. Features such as "IT Student" and "Education Level" have relatively lower impacts. The color-coded segmentation indicates the specific contribution of each feature to the three classes, helping identify key drivers for classification and their relative importance. This analysis aids in understanding the model's decision-making process.

$$\phi_i = E_{x'} \sim D[f(x')|x'] - E_{x'} \sim D[f(x')]$$

Where:

- ϕ_i: is the SHAP value of feature i.
- $E_{x'} \sim D[f(x')]$: represents the expectation of the model output.

Table 1: Classification Report

	Precision	Recall	F1-Score
0.0	1.00	0.98	0.99
1.0	0.97	0.93	0.95
2.0	0.93	0.99	0.96
Total Accuracy	0.97		

The classification report summarizes the performance of a model across three classes (0.0, 1.0, and 2.0). For class 0.0, the model achieved near-perfect performance with a precision of 1.00, recall of 0.98, and an F1-score of 0.99. For class 1.0, the precision is 0.97, recall is 0.93, and the F1-score is 0.95, indicating slightly lower performance but still strong results. For class 2.0, the model achieved a precision of 0.93, a high recall of 0.99, and an F1-score of 0.96, showing robust detection. Overall, the model's total accuracy is 0.97, demonstrating high overall effectiveness in classifying the data.

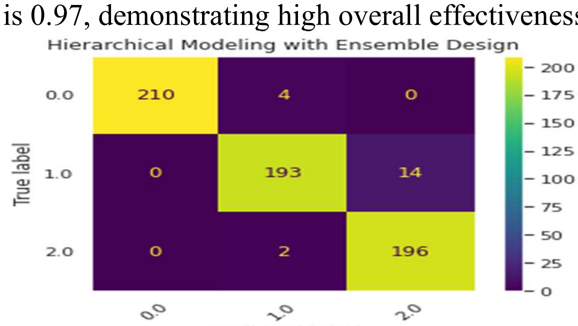

Figure 7: Confusion Matrix

The confusion matrix, shown in Figure 7, evaluates the performance of a multi-class classification model by summarizing the counts of correct and incorrect predictions for each class. Rows represent the actual classes, while columns correspond to the predicted classes. This matrix is crucial for assessing model efficiency and identifying misclassification patterns.

In this matrix:

- The top-left cell (210) represents the true positives (TP) for class 0.0, where predictions correctly match the actual class.
- The first row's off-diagonal cell (4) represents false positives (FP) for class 0.0, showing cases incorrectly predicted as class 1.0.
- The second row's diagonal cell (193) indicates the true positives for class 1.0, while the off-diagonal cell (14) in the same row represents instances misclassified as class 2.0.
- For class 2.0, the bottom row shows 196 true positives (TP) and 2 false negatives (FN) classified as class 1.0. Overall, this confusion matrix highlights the model's effectiveness, with a strong concentration of predictions along the diagonal, indicating accurate classification for all three classes.

Table 2: Comparative Analysis

Methods	Accuracy
Linear Regression [18]	0.63
Support Vector Classifier [19]	0.73
KNeighbours Classifier [20]	0.80
Decision Tree Classifier [21]	0.83
Random Forest Classifier [22]	0.85
Proposed Model	**0.97**

Table 2 presents a comparative analysis of various classification methods based on their accuracy. Traditional models like Linear Regression and Support Vector Classifier (SVC) achieved accuracies of 0.63 and 0.73, respectively, while more advanced algorithms such as KNeighbors Classifier, Decision Tree Classifier, and Random Forest Classifier demonstrated improved performances with accuracies of 0.80, 0.83, and 0.85. However, the proposed Two-Stage Hybrid Classification Model, which combines a base classifier with an XGBoost meta-learner, significantly outperformed all other methods, achieving an impressive accuracy of 0.97. This highlights the effectiveness of the hybrid approach in leveraging the strengths of both the base model and XGBoost to address misclassifications and optimize overall performance.

CONCLUSION:

The research concludes that the proposed Two-Stage Hybrid Classification Model, which integrates a primary classifier with an XGBoost-based meta-learner, significantly enhances classification accuracy, achieving an impressive 97% compared to traditional models like Random Forest (85%) and Decision Tree (83%). The model's key components include an initial stage for general classification, identification of misclassified samples, and a second stage where the meta-learner targets these errors for refinement. The novelty lies in its ability to address weaknesses in conventional single-stage classifiers by leveraging a focused and adaptive learning mechanism. The results demonstrate that this hybrid approach effectively reduces misclassification, especially in complex datasets, making it a robust solution for practical applications like education, healthcare, and fraud detection. Within the broader field, this methodology highlights the potential of combining traditional and advanced machine learning techniques to improve model reliability and predictive performance. This work underscores the importance of innovative hybrid frameworks in advancing classification tasks and their practical implications in diverse domains.

REFERENCES:

[1] Rahimi, Ramy A., and Grace S. Oh. "Rethinking the role of educators in the 21st century: navigating globalization, technology, and pandemics." Journal of Marketing Analytics (2024): 1-16.

[2] Song, Chuanxiang, Seong-Yoon Shin, and Kwang-Seong Shin. "Implementing the Dynamic Feedback-Driven Learning Optimization Framework: A Machine Learning Approach to Personalize Educational Pathways." Applied Sciences 14, no. 2 (2024): 916.

[3] Hossain, Muhammad, Morris Anglin, Ayazullah Safi, Tajbir Ahmed, and Salim Khan. "Adapting to the Digital Age: An Evaluation of Online Learning Strategies in Public Health and Social Care Education." Education Research International 2024, no. 1 (2024): 5079882.

[4] Sohel, Md Salman, Md Khaled Sifullah, Babul Hossain, Md Alamgir Hossain, Md Fouad Hossain Sarker, Israfil Hossain, Mohamed Emran Hossain, and Md Rafiz Uddin. "E-learning experience of indigenous rural communities in the face of COVID-19 crisis in Chittagong hills tracts region, Bangladesh: a qualitative investigation." International Journal of Community Well-Being (2024): 1-24.

[5] Ramachandran, Aishwarya, Meg Schwellnus, Derek Gladwin, Ryan Derby-Talbot, and Naoko Ellis. "Cultivating educational adaptability through collaborative transdisciplinary learning spaces." Discover Education 3, no. 1 (2024): 2.

[6] Zamiri, Majid, and Ali Esmaeili. "Strategies, methods, and supports for developing skills within learning communities: A systematic review of the literature." Administrative Sciences 14, no. 9 (2024): 231.

[7] Ikegwu, Anayo Chukwu, Henry Friday Nweke, and Chioma Virginia Anikwe. "Recent trends in computational intelligence for educational big data analysis." Iran Journal of Computer Science 7, no. 1 (2024): 103-129.

[8] Al-Gerafi, Mohammed AM, Shankha Shubhra Goswami, Sushil Kumar Sahoo, Raman Kumar, Vladimir Simic, Nebojsa Bacanin, Quadri Noorulhasan Naveed, and Ayodele Lasisi. "Promoting inclusivity in education amid the post-COVID-19 challenges: An interval-valued fuzzy model for pedagogy method selection." The International Journal of Management Education 22, no. 3 (2024): 101018.

[9] Tirumanadham, NS Koti Mani Kumar. "Improving predictive performance in e-learning through hybrid 2-tier feature selection and hyper parameter-optimized 3-tier ensemble modeling." International Journal of Information Technology (2024): 1-28.

[10] Moon, Jewoong, Fengfeng Ke, Zlatko Sokolikj, and Shayok Chakraborty. "Applying multimodal data fusion to track autistic adolescents' representational flexibility development during virtual reality-based training." Computers & Education: X Reality 4 (2024): 100063.

[11] Khasawneh, Tariq, and Mohammad Azzeh. "Polishing the black box: flexible model-based partitioning surrogate models for interpretable machine learning model." International Journal of Data Science and Analytics (2024): 1-29.

[12] Verma, Chaman, Zoltán Illés, and Deepak Kumar. "An investigation of novel features for predicting student happiness in hybrid learning platforms–An exploration using experiments on trace data." International Journal of Information Management Data Insights 4, no. 1 (2024): 100219.

[13] Khedr, Ahmed M. "Enhancing supply chain management with deep learning and machine learning techniques: A review." Journal of Open Innovation: Technology, Market, and Complexity (2024): 100379.

[14] Javaid, Mohd, Abid Haleem, Ravi Pratap Singh, and Rajiv Suman. "Enabling flexible manufacturing system (FMS) through the applications of industry 4.0 technologies." Internet of Things and Cyber-Physical Systems 2 (2022): 49-62.

[15] Pereira, Teresa MC, Raquel Sebastiao, Raquel C. Conceição, and Vitor Sencadas. "A Review on Intelligent Systems for ECG Analysis: from Flexible Sensing Technology to Machine Learning." IEEE Journal of Biomedical and Health Informatics (2024).

[16] Bourechak, Amira, Ouarda Zedadra, Mohamed Nadjib Kouahla, Antonio Guerrieri, Hamid Seridi, and Giancarlo Fortino. "At the confluence of artificial intelligence and edge computing in iot-based applications: A review and new perspectives." Sensors 23, no. 3 (2023): 1639.

[17] Shariful, "Student Flexibility in Online Learning Dataset," Kaggle, 2025. [Online]. Available:https://www.kaggle.com/datasets/shariful07/student-flexibility-in-online-learning/data. [Accessed: Jan. 22, 2025].

[18] Leem, Subeen, Jisong Oh, Jihoon Moon, Mucheol Kim, and Seungmin Rho. "Enhancing multistep-ahead bike-sharing demand prediction with a two-stage online learning-based time-series model: insight from Seoul." The Journal of Supercomputing 80, no. 3 (2024): 4049-4082.

[19] Pawar, Mousmi, and Jyotshna Dongardive. "Improving Student Performance Prediction Through Feature Selection: Insights from 'Offee'Assessment Data during the Covid-19 Pandemic." Procedia Computer Science 235 (2024): 3075-3084.

[20] Farhood, Helia, Ibrahim Joudah, Amin Beheshti, and Samuel Muller. "Evaluating and enhancing artificial intelligence models for predicting student learning outcomes." In Informatics, vol. 11, no. 3, p. 46. MDPI, 2024.

[21] Hussain, Tahir, Lasheng Yu, Muhammad Asim, Afaq Ahmed, and Mudasir Ahmad Wani. "Enhancing E-Learning Adaptability with Automated Learning Style Identification and Sentiment Analysis: A Hybrid Deep Learning Approach for Smart Education." Information 15, no. 5 (2024): 277.

[22] Han, Hui, and Silvana Trimi. "Analysis of cloud computing-based education platforms using unsupervised random forest." Education and Information Technologies (2024): 1-28.