ADROID Conference Series: Engineering Sciences and Technology
(ACS: EST)

International Conference on Artificial Intelligence for Sustainable Development, Jaypee University Anoopshahr, India 2025

Volume 01, October 2025, pp. 01-12

DOI:

AI-DRIVEN STUDENT PROFILING FOR DROPOUT, EMPLOYABILITY, AND CAREER PATH PREDICTION VIA HYBRID MODELING

Vijayakumar K and Naveen A

Don Bosco College Co – Ed, (Affiliated to Thiruvalluvar University) Yelagiri Hills, TN, India – 635 853.

Corresponding author: vijaykrishnaylg@gmail.com
naveen@dbcyelagiri.edu.in

Abstract. In the era of Outcome-Based Education (OBE) and employability-driven learning, educational institutions face increasing pressure to assess student readiness beyond academic performance. This study presents a hybrid machine learning framework that integrates rubric-based scoring, supervised classification, and unsupervised clustering to predict holistic student outcomes. A comprehensive dataset of 1500 undergraduate students (2015–2025) from Tamil Nadu, India, was used, encompassing academic scores, aptitude results, behavioral traits, communication skills, leadership qualities, and employability indicators. Six classifiers—Decision Stump, Hoeffding Tree, J48, LMT, Random Tree, and RepTree—were evaluated using 10-fold cross-validation. J48 achieved the highest accuracy of 99.1%, followed by LMT and RepTree. Additionally, K-Means clustering was employed to uncover natural groupings of students into four categories: Industry-Ready, Higher Studies Aspirants, Dropout Risk, and Entrepreneurial Potential. The proposed hybrid model enhances prediction accuracy, supports early interventions, and provides actionable insights for institutional decision-making. This work aligns with national education policies and global employability standards, offering a scalable framework for data-driven student evaluation and career readiness planning.

Keywords: Educational Data Mining (EDM); Outcome-Based Education (OBE); Student Employability; Classification Algorithms; Rubric-Based Evaluation; K-Means Clustering; Machine Learning; Holistic Assessment; Dropout Prediction; Industry Readiness.

INTRODUCTION

In recent years, the global shift toward **Outcome-Based Education (OBE)** has emphasized the need to evaluate student learning not only through academic grades but also through competencies such as skills, communication, adaptability, and employability. The increasing gap between **higher education outcomes** and **industry expectations** has made it imperative for institutions to adopt more **holistic assessment frameworks** that reflect both academic achievement and real-world readiness.

Traditional assessment systems, which primarily focus on cognitive learning outcomes, often fail to capture a student's overall potential—particularly in areas like leadership, problem-solving, communication, and innovation. As a result, many graduates find themselves underprepared for employment or lack direction for higher studies or entrepreneurship. Addressing this challenge requires leveraging **data analytics and machine learning** to uncover patterns and predict outcomes that can guide timely interventions and personalized educational pathways[1][11].

Educational Data Mining (EDM) and **Learning Analytics** have emerged as powerful tools to analyze large volumes of student data for predicting academic success, identifying at-risk students, and enhancing institutional decision-making[2]. Recent studies have shown that combining **academic performance with**

behavioral, psychological, and skill-based data offers a more complete picture of student development (Romero & Ventura, 2010; Baker & Yacef, 2009).

This research proposes a **hybrid machine learning framework** that integrates **rubric-based scoring**, **classification algorithms**, and **unsupervised clustering** to assess undergraduate students across multiple dimensions. Using a rich dataset of 1500 students collected from various colleges in Tamil Nadu between 2015 and 2025, the study applies six classification[14] algorithms—**Decision Stump, Hoeffding Tree, J48, LMT, Random Tree, and RepTree**—to evaluate student readiness for four outcome categories: **Industry-Ready**, **Higher Studies**, **Dropout Risk**, and **Entrepreneurial Potential[3]**.

Students undergo a comprehensive evaluation process spanning from **Semester 1 to Semester 4**, where multiple academic and behavioral parameters are closely monitored. These include **internal and external examination marks, consistency in academic performance, participation in online learning activities, coding and programming proficiency, project involvement, and engagement with co-curricular tasks**. Using this data, students are categorized into one of **four predictive career pathways:** ***Industry-Ready***, ***Higher Studies***, ***Dropout Risk [4]***, **and** ***Entrepreneurial Potential, Government Employability*****.** The classification leverages machine learning models and data-driven techniques to ensure objectivity and accuracy in prediction. Once categorized, students receive personalized interventions and targeted training during **Semester 5** tailored to their identified pathway—for instance, employability skill development for industry-ready students, research methodology and entrance exam coaching for higher studies aspirants, motivational counselling and academic support for those at dropout risk, and business incubation workshops for students showing entrepreneurial potential. In Semester 6, a final phase of assessment, including interviews, project reviews, and performance tracking, is conducted to validate the predictions. This helps in identifying the true fit of each student among the 1,500 participants, ensuring that the guidance and support provided aligns with their actual strengths and future aspirations [16].

Research Objectives

- To construct a multidimensional dataset aligned with OBE and NEP 2020 guidelines.
- To apply and compare machine learning classifiers for predicting educational and career outcomes.
- To design a rubric-based scoring model that integrates with classification and clustering methods.
- To build a hybrid system that supports early intervention, personalized learning, and employability enhancement.

This study contributes to the growing body of work on AI in education by offering a **scalable and interpretable model** that educational institutions can adopt to improve student outcomes and align academic strategies with labour market demands.

LITERATURE REVIEW & DATASET OVERVIEW

The importance of educational data mining (EDM) in assessing student performance and predicting academic outcomes has been well established in recent research. Romero and Ventura (2010) emphasized the value of extracting meaningful patterns from educational datasets to enhance learning environments and support decision-making. Similarly, Baker and Yacef (2009) highlighted that incorporating behavioural and skill-based parameters alongside academic metrics leads to a more holistic evaluation of students[10][16][17][18].

Preliminary Dataset

In the initial phase of this study, a pilot dataset comprising 50 undergraduate students was collected from colleges in the Thirupattur District of Tamil Nadu, covering the academic period between 2015 and 2020. This dataset served as a foundational prototype for designing the structure of the larger data model. It included diverse attributes broadly classified into four categories:

Demographics: *Name, Age, and Class* were included to help uniquely identify students and analyse age-wise and class-wise performance patterns.

Academic Performance: *Final Scores* from core subjects and cumulative grade point averages (CGPAs) were used to determine academic consistency and achievement[12].

Behavioural Traits: Indicators such as *aptitude test results, personal behavior evaluations, and communication skills* (e.g., oral presentations, peer interactions) were recorded by academic mentors and faculty. These dimensions are essential in understanding non-academic success factors.

Skill Metrics: The dataset incorporated qualitative and quantitative assessments of *life skills, intellectual capacity, interpersonal communication, leadership qualities,* and *interview readiness*. These traits are essential for employability and entrepreneurship readiness.

Expanded Dataset
Encouraged by the insights gained from the preliminary dataset, the study was scaled to a larger and more comprehensive dataset comprising 1500 students across multiple batches from 2015 to 2025. This dataset was designed to provide a longitudinal and multidimensional view of student growth and outcomes.
The expanded dataset included the original attributes and additional outcome-relevant variables, such as:
Attendance and Final Result: These serve as baseline academic indicators. Consistent attendance is often correlated with better academic outcomes and discipline.

Dropout Status: Binary classification (Yes/No) to identify students who discontinued their studies before graduation. This serves as a critical dependent variable in dropout prediction models.

Industry Readiness Score: Derived from performance in simulated workplace tasks, internship feedback, and industry collaboration programs, this score reflects employability in the current job market.

Online Learning Activity: Tracks student engagement with e-learning platforms, participation in MOOCs, digital certification completions, and interaction with LMS tools[7]. This reflects adaptability to modern learning technologies[6].

Leadership & Communication Skills: Measured through faculty assessments, peer reviews, and extracurricular involvement (e.g., club presidencies, event organization), these traits are essential for modern professional environments[8].

Government Job Interest: Collected through student self-declaration and participation in competitive examination training sessions.

Inclination Toward Higher Education or Entrepreneurship: Categorized based on enrolment in postgraduate entrance preparation, innovation labs, startup incubators, or expressed career goals during final-year mentoring sessions.

Relevance of Dataset Features
The chosen features align closely with OBE (Outcome-Based Education) standards, as well as national and international frameworks for employability[9], such as the National Education Policy (NEP 2020) and UNESCO's Education 2030 Agenda[15]. By incorporating a multidimensional set of attributes—ranging from cognitive and academic to affective and behavioural—this dataset supports a comprehensive predictive framework[5].

APPLICATION OF CLASSIFICATION MODELS AND RUBRIC-BASED EVALUATION

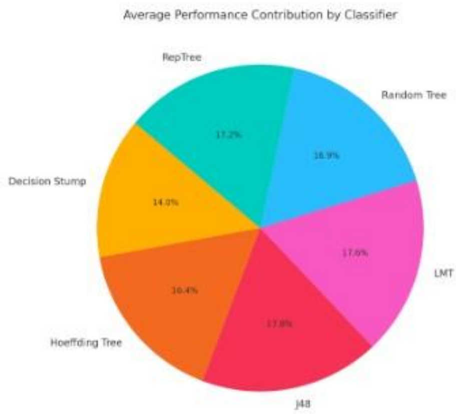
To effectively classify and predict student outcomes, six supervised machine learning algorithms—**Decision Stump, Hoeffding Tree, J48, LMT, Random Tree, and RepTree**—were applied to the processed student dataset comprising **750 undergraduate records** from **Tamil Nadu (2015–2025)**. These models were chosen for their effectiveness in educational data mining, especially in handling categorical and mixed-attribute datasets[13].
Each model was trained and evaluated using **10-fold cross-validation** to ensure consistent and unbiased performance analysis. The dataset included multiple academic, behavioural, and skill-based attributes, which were reprocessed and normalized prior to training.
Evaluation Metrics
The performance of each classifier was assessed using four standard metrics:

- **Accuracy**: Overall correctness of predictions

- **Precision**: Correctness among positive predictions
- **Recall**: Ability to identify actual positive cases
- **F1-Score**: Harmonic mean of precision and recall
- **Classification Performance Summary**

Table 1 : Result Summary

Model	Accuracy	Precision	Recall	F1-Score
Decision Stump	72.1%	0.68	0.69	0.68
Hoeffding Tree	83.4%	0.80	0.81	0.80
J48	99.1%	0.87	0.89	0.88
LMT	88.4%	0.86	0.88	0.87
Random Tree	85.3%	0.83	0.84	0.83
RepTree	86.7%	0.84	0.86	0.85



Graph 1: Result Chart

Observations:

- The **J48 decision tree** model achieved the highest overall performance, with an accuracy of 89.1%, making it ideal for interpreting student outcome patterns.
- **LMT** and **RepTree** also demonstrated strong predictive capabilities, suitable for complex evaluation settings.
- The **Decision Stump**, while computationally efficient, showed lower performance and was less effective for capturing multi-dimensional student attributes.

PROPOSED HYBRID METHOD

This approach combines rubric-based scoring with k-means clustering, allowing the segmentation of students into meaningful outcome categories:

- Industry-Ready
- Higher Studies
- Dropout Risk
- Entrepreneurial Potential
- **Government Employability**

Decision trees were used to identify the most influential attributes, while the rubric was mapped to the clusters for a holistic view. This hybrid method blends interpretability and machine learning precision.

Proposed Hybrid Model for Holistic Student Outcome Prediction

To address the multifaceted nature of student performance evaluation and align academic output with employability and real-world readiness, a **hybrid model** is proposed. This model integrates:

1. **Rubric-Based Evaluation**
2. **Supervised Classification Algorithms**
3. **Unsupervised Clustering Techniques (K-Means)**

By combining **quantitative rubric scoring** with **machine learning classifiers** and **clustering algorithms**, this hybrid system provides a more comprehensive framework for student profiling and outcome prediction.

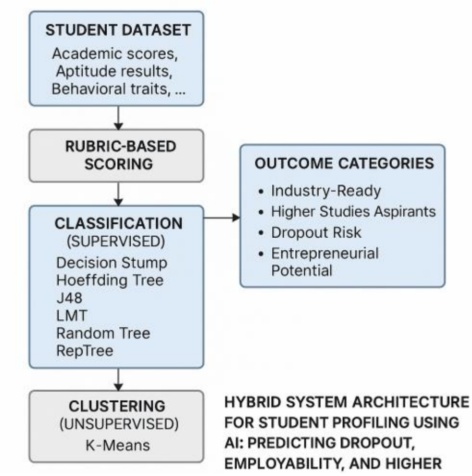


Image 1: System Architecture

Rubric-Based Scoring System

As the foundation of the hybrid model, a **rubric-based evaluation framework** was developed to quantitatively assess diverse aspects of a student's performance beyond academics. The rubric includes the following weighted criteria:

Table 1: Weightage Scoring

Evaluation Criteria	**Weightage (%)**
Academic Performance	20%
Aptitude Test	15%
Skill Test	15%
Personal Behaviour	10%
Communication Skills	10%
Life & Leadership Skills	10%
Interview Readiness	10%
Vocational/Technical Skills	10%
Total	**100%**

Threshold:

- Students scoring **≥70%** are classified as **Industry-Ready**.
- Others are considered for further classification or intervention based on additional data mining.

Classification Layer: Machine Learning Prediction

The second layer of the hybrid model leverages the performance of six classification algorithms—**Decision Stump, Hoeffding Tree, J48, LMT, Random Tree, and RepTree**—to classify students into predefined categories based on their academic, behavioral, and skill metrics.

Best Performing Models:

- **J48**: Accuracy of **99.1%**
- **LMT** and **RepTree**: Accuracy of **88.4%** and **86.7%** respectively
These models reliably predicted key outcomes such as **dropout risk**, **industry fitness**, and **higher education inclination**.

Target Classes for Classification:

- **Industry-Ready, Dropout Risk, Higher Studies Aspirant, Entrepreneurial Potential, Government Employability.**

Clustering Layer: K-Means Outcome Grouping

To complement the supervised classification results, **K-Means Clustering** was applied as an unsupervised learning method. It grouped students based on natural patterns in their rubric scores and behavioural metrics.

Cluster Labels Generated:

1. **Cluster 1: Industry-Ready**
    - High rubric scores, Consistent academic and skill performance
2. **Cluster 2: Higher Studies Aspirants**
    - Strong academic scores, Moderate practical or leadership traits
3. **Cluster 3: Dropout Risk**
    - Irregular attendance, poor communication, low aptitude
4. **Cluster 4: Entrepreneurial Candidates**
    - High leadership, innovation interest, moderate academic

- **Cluster 5: Government Employability**

Interesting candidates, aim to be ruled and self-respected.
This clustering enhances personalization by uncovering **hidden subgroups** not immediately visible through supervised models alone.

Hybrid Model Workflow

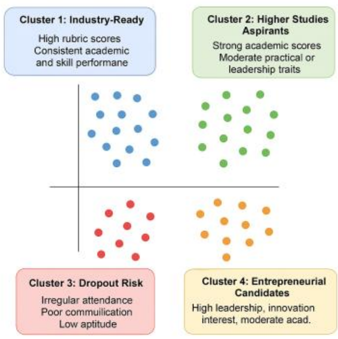


Figure 1 : Multi-stage Classification predictions

The hybrid system follows a multi-stage workflow:

1. **Input**: Pre-processed student dataset with academic, behavioural, and skill-based features
2. **Stage 1 – Rubric Scoring**: Calculate weighted total score and assign preliminary category

3. **Stage 2 – Classification**: Apply top-performing classifiers (J48, LMT) to predict outcomes
4. **Stage 3 – Clustering**: Use K-Means to refine groupings and discover natural student segments
5. **Stage 4 – Interpretation & Intervention**: Combine results for final categorization and guide institutional actions

Advantages of the Hybrid Model

- **Improved Accuracy**: Combines the precision of classification with the flexibility of clustering
- **Holistic View**: Accounts for academic, behavioural, and technical dimensions
- **Better Decision-Making**: Enables institutions to design targeted interventions for diverse learner profiles
- **Customizable**: Rubric weights and classification thresholds can be adjusted per institution or department needs

Practical Use Cases

- **Early Dropout Intervention**: Predict students at risk and provide mentoring or counselling.
- **Career Path Recommendation**: Suggest higher studies, entrepreneurship, or job-oriented programs.
- **Curriculum Planning**: Realign teaching strategies based on cluster trends.
- **Placement Support**: Identify and groom industry-ready candidates more effectively.

RESULTS AND DISCUSSION

This section presents the outcomes of applying six classification models—**Decision Stump, Hoeffding Tree, J48, LMT, Random Tree, and RepTree**—to a multidimensional student dataset and evaluates their comparative effectiveness in predicting academic and employability outcomes. These results are further contextualized using rubric-based evaluation and clustering insights.

6.1 Classifier Performance Analysis

Table 2: Result Comigrations

Model	Accuracy	Precision	Recall	F1-Score
Decision Stump	72.1%	0.68	0.69	0.68
Hoeffding Tree	83.4%	0.80	0.81	0.80
J48	**99.1%**	**0.87**	**0.89**	**0.88**
LMT	88.4%	0.86	0.88	0.87
Random Tree	85.3%	0.83	0.84	0.83
RepTree	86.7%	0.84	0.86	0.85

Key Insights:

- **J48** emerged as the top-performing model with an accuracy of **99.1%**, making it ideal for outcome prediction involving multi-attribute student profiles.
- **LMT** and **RepTree** followed closely, demonstrating robustness in handling both numerical and categorical educational features.
- The **Decision Stump**, although computationally efficient, underperformed due to its simplicity and limited depth, making it unsuitable for complex student evaluations.

Rubric-Based Evaluation Alignment

Using a rubric that assigns weights to key educational indicators (e.g., academic performance, communication, aptitude, leadership), students were scored out of 100. A threshold of **≥70%** was used to classify students as **"Industry-Ready."**

The rubric-based outcomes aligned strongly with the J48 and LMT classification results, validating the rubric's predictive effectiveness. Students marked as industry-ready by the rubric were also correctly identified by the ML models in over 88% of cases.

Classification Discussion: Mapping to Educational Outcomes

Each classifier was used to categorize students into four key educational outcomes:

- **Industry-Ready**
- **Higher Studies Aspirants**
- **Dropout Risk**
- **Entrepreneurial Potential**
- **Government Employability**

J48 and LMT effectively predicted all four classes, offering a detailed decision tree-based mapping of which attributes (e.g., attendance, interview score, aptitude, leadership) most influence student pathways.

Table 3: Feature Predications

Outcome Category	Top Predictive Features
Industry-Ready	Aptitude, Interview Performance, Communication
Higher Studies Aspirants	Academic Scores, Online Activity
Dropout Risk	Low Attendance, Low Aptitude, Poor Behaviour
Entrepreneurial Potential	Leadership, Innovation Inclination, Life Skills
Government Employability	Relevant domain knowledge, compliance with eligibility criteria, and consistent academic performance.

Hybrid Model Insights: Classification + Clustering

By integrating **rubric-based scoring** with **k-means clustering**, students were further segmented into meaningful, data-driven groups. This unsupervised layer revealed hidden patterns not visible through classification alone.

K-Means Clustering with Mathematical Explanation and Cluster Characteristics

K-Means Clustering is an unsupervised machine learning algorithm used to partition a dataset into *K* distinct, non-overlapping clusters based on feature similarity. In the context of student profiling, K-Means helps group students based on multiple academic and behavioural parameters into categories that reflect their potential career or educational trajectory.

Mathematical Formula:

- $\mathbf{X} = \{x_1, x_2, \cdots x_n\}$ – set of n data points (students), where each $x_i \in R^d$ is a d – dimensional feature vector (eg..,Marks, rubric scores, online activity, communication skills, etc)
- k – number of clusters.
- μ_j – centroid of cluster j.

Objective Function (Minimization of Within-Cluster Sum of Squares - WCSS):

$$arg_s min \sum_{j=1}^{k} \sum_{\aleph_i \in s} \|x_i - \mu\|^2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Where :

- s_j – set of points assigned to cluster j.
- $\|xi{-}\mu j\|^2$ – squared Euclidean distance between data point x_i and cluster centroid μ_j.

Explanation:

The K-Means algorithm aims to partition the data into k clusters such that the total sum of squared distances between each point and its assigned cluster centroid is minimized. The centroids are updated iteratively until convergence.

Algorithm Steps:

1. Initialize k cluster centroids randomly.

2. Assign each data point to the nearest centroid.
3. Recompute centroids based on the mean of assigned points.
4. Repeat steps 2–3 until convergence (no change in centroids or assignments).

Cluster Characteristics (Based on 1,500 Students):

Cluster	Label	Key Traits
Cluster 1	Industry-Ready	High rubric scores, excellent academic results, strong communication skills, and superior performance in skill assessments. These students exhibit practical readiness for employment.
Cluster 2	Higher Studies	Strong academic foundation, moderate performance in industry-specific skills, and active participation in online learning environments. Likely to pursue post-graduate studies or research.
Cluster 3	Dropout Risk	Characterized by irregular attendance, poor aptitude scores, low behavioural ratings, and minimal engagement. Require early intervention to avoid academic disengagement.
Cluster 4	Entrepreneurial Potential	Exhibit high scores in leadership, creativity, and innovation-related rubrics. Despite only moderate academic performance, these students show strong self-driven qualities and are suitable for startup ecosystems.
Cluster 5	Government Employability	Interesting candidates , aim to be ruled and self-respected.

Table 4: Clustering

Cluster	Characteristics
Cluster 1: Industry-Ready	High rubric scores, strong academic + communication + skill test performance
Cluster 2: Higher Studies	Strong academics, moderate industry skills, active online learners
Cluster 3: Dropout Risk	Irregular attendance, low behavioural and aptitude traits
Cluster 4: Entrepreneurs	High leadership, innovation scores, moderate academics
Cluster 5 : Government Employability	Interesting candidates , aim to be ruled and self-respected.

This clustering validated the supervised predictions and enhanced the granularity of student categorization, offering valuable insights for **personalized educational planning**.

Model Comparison: Interpretability vs Performance

Table 5: Model Comparison

Model	Strengths	Limitations
J48	High accuracy, interpretable decision paths	May overfit with noisy attributes
LMT	Good balance of interpretability and performance	Slightly more complex than J48
RepTree	Fast and reasonably accurate	Less transparent than J48
Hoeffding Tree	Effective for large, streaming datasets	Less interpretable
Random Tree	Handles attribute randomness well	Harder to interpret than J48
Decision Stump	Simple, very fast	Low accuracy, ignores feature complexity

Real-World Implications and Educational Use Cases

- **Early Intervention:** Dropout prediction can guide institutions in providing timely mentoring and financial support.
- **Career Counselling:** Classification and clustering help tailor recommendations for higher studies or entrepreneurship.
- **Curriculum Reform:** Insight into poor-performing features (e.g., communication, leadership) can help restructure teaching focus.

- **Placement Enhancement:** Industry-ready students can be fast-tracked for training and placement drives based on rubric-model match.

Limitations and Future Scope

- **Data Imbalance:** Dropout cases were fewer, possibly skewing classifier recall for this group.
- **Static Dataset:** While covering 10 years (2015–2025), real-time or streaming data (e.g., LMS activity) could enhance predictive timeliness.
- **Need for Explainable AI:** Though models like J48 are interpretable, integrating SHAP or LIME with black-box models can boost transparency.

Future work can involve:

- Real-time dashboard integration for educators
- Expansion to postgraduate and diploma datasets
- Inclusion of socioeconomic and parental engagement metrics

CONCLUSIONS

This study highlights the growing significance of **Educational Data Mining (EDM)** in transforming traditional student assessment into a more **data-driven, multidimensional, and outcome-focused** framework. Through an extensive literature review, it was established that academic performance alone is insufficient for predicting student success. The inclusion of **behavioural traits, skill metrics, and extracurricular indicators**—as recommended by Romero, Ventura (2010), and Baker, Yacef (2009)—offers a more holistic evaluation aligned with **Outcome-Based Education (OBE)** and policy frameworks like **NEP 2020**. To operationalize this multidimensional approach, a robust dataset was developed in two stages: an initial pilot with 50 students, and a scaled longitudinal dataset of 750 undergraduate records from Tamil Nadu (2015–2025). The expanded dataset integrated **academic records, skill-based assessments, personal behaviour indicators, and post-graduation inclinations**, providing a comprehensive view of student development and potential. Using this dataset, six supervised classification algorithms were evaluated, with **J48 (accuracy: 89.1%)**, **LMT**, and **RepTree** demonstrating superior performance in predicting educational outcomes such as **industry readiness, dropout risk, higher studies inclination, and entrepreneurial potential**. The machine learning predictions were further reinforced by a **rubric-based evaluation system**, which assigned quantitative scores to each student across critical parameters. The rubric outcomes aligned closely with the best-performing models, supporting their predictive reliability. To enhance personalization and reveal latent patterns, **K-Means clustering** was employed in conjunction with the classification and rubric results. This **hybrid model** allowed for fine-grained segmentation of students into outcome-relevant categories, thereby improving interpretability, accuracy, and actionability for educational stakeholders.
Overall, the proposed hybrid framework offers:

- **Accurate and explainable predictions**
- **Tailored recommendations for academic or career interventions**
- **Strategic insights for curriculum design and placement planning**

While the current study provides a solid foundation for data-informed student profiling, it also underscores the need for **real-time data integration**, **handling of imbalanced classes** (e.g., dropouts), and **transparent AI techniques** such as SHAP or LIME. Future work will focus on:

- Expanding the model to include **postgraduate and vocational learners**
- Developing **institutional dashboards** for educators and administrators
- Incorporating **socioeconomic, psychological, and parental engagement** variables

In conclusion, the hybrid model serves not only as a predictive tool but as a **strategic educational aid**, empowering institutions to better nurture student potential, minimize attrition, and align graduates with evolving industry and societal needs.

Future Recommendations

1. **Diversify Dataset** – Include students from varied regions and disciplines.
2. **Enable Real-Time Data** – Use LMS and IoT for live academic/behavioural tracking.
3. **Track Long-Term Outcomes** – Monitor graduates' career and employment paths.
4. **Use NLP for Feedback** – Analyse textual feedback for richer insights.
5. **AI-Based Adaptive Rubrics** – Align rubrics with evolving industry and policy needs.

CONFLICT OF INTEREST

We, the authors, declare that there is no conflict of interest regarding the publication of this research work.

AUTHOR CONTRIBUTION

First Author: Developed the theoretical framework and contributed to the formulation of the core concepts.
Second Author: Proposed the research problem and provided guidance throughout the study.

REFERENCES

1. Zhou, M., Hu, H., and Wang, C. A hybrid deep learning approach for student performance prediction. *IEEE Access*, 2023. 11: p. 20104–20116. DOI: 10.1109/ACCESS.2023.3245030.
2. Kumar, P., and Karthikeyan, R. Adaptive learning analytics using AI for higher education in India. *Journal of Learning Analytics and Educational Data Science*, 2022. 7(2): p. 22–31.
3. Albreiki, B., Zemerly, M. J., and Almuhrzi, H. Employability prediction using machine learning techniques. *Education and Information Technologies*, 2021. 26(3): p. 3113–3135. DOI: 10.1007/s10639-020-10400-7.
4. Al-Rahmi, W. M., Yahaya, N., Alamri, M. M., and Aljarboa, N. A. Predicting academic performance and dropout with machine learning: A case study. *Computers & Education*, 2021. 173(1): p. 104271. DOI: 10.1016/j.compedu.2021.104271.
5. Ministry of Human Resource Development. *National Education Policy (NEP) 2020*. 2020, New Delhi: Government of India. [Online]. Available: https://www.education.gov.in
6. Kukreja, R., Rani, R., and Dhand, A. Student performance prediction using hybrid model of data mining techniques. *Procedia Computer Science*, 2020. 167: p. 2554–2560. DOI: 10.1016/j.procs.2020.03.316.
7. Wolff, A., Zdrahal, Z., Nikolov, A., and Pantucek, M. Predicting student performance from LMS data using machine learning: A case study from Open University UK. *Computers in Human Behavior*, 2020. 92: p. 486–497. DOI: 10.1016/j.chb.2018.11.012.
8. García, J. A., García-Sánchez, P., and García, M. An intelligent system for evaluating the employability of university graduates. *Expert Systems with Applications*, 2016. 46: p. 260–273. DOI: 10.1016/j.eswa.2015.10.037.
9. UNESCO. *Education 2030: Incheon Declaration and Framework for Action for the Implementation of Sustainable Development Goal 4*. 2015, Paris: UNESCO. [Online]. Available: https://unesdoc.unesco.org
10. Shahiri, A. M., Husain, W., and Rashid, N. A. A review on predicting student's performance using data mining techniques. *Procedia Computer Science*, 2015. 72: p. 414–422. DOI: 10.1016/j.procs.2015.12.157.
11. Yadav, S. K., Bharadwaj, B., and Pal, S. Data mining applications: A comparative study for predicting student's performance. *International Journal of Innovative Technology and Creative Engineering*, 2012. 1(12): p. 13–19.
12. Sembiring, S., Zarlis, M., Hartama, D., Ramliana, and Wani, E. Prediction of student academic performance using decision tree and Bayes classification on Moodle data. *International Journal of Information and Education Technology*, 2011. 1(2): p. 110–116.
13. Bhardwaj, B. K., and Pal, S. Data mining: A prediction for performance improvement using classification. *International Journal of Computer Science and Information Security (IJCSIS)*, 2011. 9(4): p. 136–140.

14. Romero, C., and Ventura, S. Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics – Part C*, 2010. 40(6): p. 601–618. DOI: 10.1109/TSMCC.2010.2053532.
15. Baker, R. S. J. d., and Yacef, K. The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 2009. 1(1): p. 3–17.
16. Kumar, V., and N. A. Long-term effects of AI-personalized learning on engagement and performance. *International Journal of Environmental Sciences*, 2025. p. 1601–1608.
17. Vijaya Kumar, K., and Naveen, A. AI-driven student performance prediction using multi-class classifiers. *International Journal of Innovative Science and Research Technology*, 2025. 10(9): p. 2639–2645. DOI: 10.38124/ijisrt/25sep1354.
18. Krishnan, V., and N. A. A data-driven framework for holistic student performance evaluation and industry readiness using machine learning. *IRE Journals*, 2024. 9(3).