© SEP 2025 | IRE Journals | Volume 9 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I3-1710884-2905

A Data-Driven Framework for Holistic Student Performance Evaluation and Industry Readiness Using Machine Learning

VIJAYAKUMAR KRISHNAN[1], NAVEEN A[2]

[1]*Research Scholar, Don Bosco College (Co-Ed) (Affiliated to Thiruvalluvar University), Yelagiri Hills, Tamil Nadu, India.*

[2]*Research Supervisor, Don Bosco College (Co-Ed) (Affiliated to Thiruvalluvar University), Yelagiri Hills, Tamil Nadu, India.*

Abstract- The growing disparity between higher education outcomes and rapidly evolving industry requirements highlights the urgent need for an integrated, data-driven evaluation approach. Traditional assessment models focus primarily on academic performance, neglecting critical employability skills such as communication, adaptability, and problem-solving. This study introduces a novel multi-tier evaluation framework leveraging machine learning (ML) techniques to holistically assess student performance and predict industry readiness. Using a decade-long dataset (2015–2025) comprising 750 undergraduate students from Tamil Nadu, India, the proposed system integrates cognitive, behavioural, and technical skill indicators. Decision tree classifiers, particularly the J48 model, are utilized for predictive modelling, supported by clustering algorithms for deeper analysis. The model achieved a notable accuracy of 86%, significantly outperforming traditional evaluation methods. The findings demonstrate how AI-driven predictive analytics can bridge the gap between academia and industry by enabling timely interventions and personalized development plans. This research contributes a scalable, explainable, and context-sensitive solution for workforce preparedness in the Indian higher education landscape.

Index Terms- Machine Learning, Decision Trees, Industry Readiness, Educational Assessment, Outcome-Based Education, Predictive Analytics, Employability.

I. INTRODUCTION

Higher education systems worldwide face mounting challenges in equipping graduates with skills aligned to the demands of the Fourth Industrial Revolution. Traditional grading systems, while essential, are limited to academic performance and fail to capture essential soft skills such as leadership, teamwork, and critical thinking. Industries today demand well-rounded graduates capable of adapting to dynamic work environments, as highlighted by the World Economic Forum's Future of Jobs Report (2023).

This study addresses these gaps by introducing a data-driven framework for evaluating both academic success and workforce readiness. Unlike conventional models, it integrates multi-dimensional data—including academic records, aptitude scores, soft skills, and behavioural factors—into a unified platform powered by machine learning. By leveraging decision tree classification, the framework offers predictive insights and transparent decision-making, supporting educators, policymakers, and employers.

Existing System

Traditional student evaluation systems in higher education predominantly focus on academic performance, relying on semester grades, standardized test scores, and attendance records. While these metrics are useful for assessing knowledge retention, they provide a narrow and incomplete picture of a student's overall capabilities. This approach overlooks critical employability factors such as communication skills, teamwork, leadership, adaptability, and problem-solving abilities, which are essential for success in today's dynamic job market [2], [5].

Research Gap

While numerous studies apply ML techniques to educational datasets, there remains a critical gap in integrating OBE principles with predictive analytics to generate personalized interventions. Existing models also struggle with transparency, often functioning as "black-box" systems that limit interpretability for educators and stakeholders.

Research Objectives

1. Develop a machine learning-based model to evaluate multi-dimensional student performance.
2. Predict industry readiness levels and identify at-risk students.
3. Provide actionable recommendations for targeted interventions and skill enhancement.
4. Align educational outcomes with national policies like NEP 2020 and global frameworks such as UNESCO Education 2030.

II. LITERATURE REVIEW

2.1 Traditional Assessment and the Employability Gap

Conventional student evaluation has historically prioritized summative examinations and grade-point averages, offering a narrow view of learner capability and little insight into workplace readiness. Foundational critiques argue that such one-dimensional assessment overlooks non-cognitive attributes—communication, teamwork, self-management, and adaptability—those employers increasingly prize. This mismatch contributes to a persistent education–employment gap, particularly acute in rapidly evolving labour markets shaped by automation and AI. Policy frameworks (e.g., UNESCO Education 2030 and India's NEP 2020) explicitly call for outcomes that integrate knowledge, skills, values, and attitudes, urging institutions to move beyond content mastery toward demonstrable competencies aligned with real-world tasks [9], [5].

2.2 Machine Learning in Educational Analytics

Over the last decade, educational data mining (EDM) and learning analytics have advanced the modelling of student performance, retention risk, and progression pathways. Supervised learning approaches—Decision Trees, Random Forests, Logistic Model Trees, and Hoeffding Trees—are especially prominent due to their interpretability in academic settings and their ability to handle heterogeneous, mixed-type data [1], [4]. Studies across diverse contexts show that ML models can outperform traditional heuristics in predicting achievement and classifying students into risk or proficiency tiers, enabling earlier and more targeted support [1], [4]. Toolkits such as scikit-learn have standardized pipelines for preprocessing, model selection, validation, and deployment, accelerating reproducible research and institutional adoption [8].

2.3 Employability Prediction and Soft-Skill Signals

A growing strand of work extends beyond academic prediction to employability forecasting, incorporating indicators like interview performance, aptitude, communication, teamwork, leadership, and practical skill tests. Evidence suggests that hybrid feature sets—combining academic metrics with behavioural and skill-based measures—improve predictive power for job placement and career pathways [2], [3], [4]. This aligns with global labour-market analyses that emphasize complex problem-solving, self-efficacy, and collaboration as high-demand competencies in the near term [10]. Conceptual studies further underscore AI's role in structuring these multifaceted indicators into actionable institutional dashboards for career services and program review [7].

2.4 Outcome-Based Education (OBE) and Program Alignment

Outcome-Based Education (OBE) reframes evaluation around demonstrable outcomes, mapping Course Outcomes (COs) to Program Outcomes (POs) and ultimately to graduate attributes. While OBE is widely advocated in policy and accreditation discourse, empirical integrations of OBE mapping with ML-driven prediction pipelines are comparatively limited. The literature largely treats OBE alignment and ML analytics as parallel streams: OBE ensures curricular coherence and attainment tracking, whereas ML provides predictive insights into performance and risk. Integrating these paradigms promises a closed-loop system in which predictive signals guide targeted interventions, and intervention effects feed back into CO–PO attainment analytics for continuous improvement [5], [9].

2.5 Methods, Validity, and Explainability

Decision-tree–family models (e.g., J48/C4.5, Random Trees, Hoeffding Trees) remain popular in institutional settings because they provide transparent rules and feature importance measures—qualities that facilitate instructor buy-in and ethical oversight relative to black-box models [1], [4]. Standard evaluation protocols—accuracy, precision, recall, F1-score, along with cross-validation—address internal validity, while clustering (e.g., K-

Means) supports post-hoc segmentation for tailored interventions. Nevertheless, challenges persist around data quality, feature drift across cohorts, and fairness when socio-economic variables are included. The field increasingly advocates for interpretable models, auditable pipelines, and alignment with institutional governance to ensure responsible use at scale [1], [4], [8].

2.6 Identified Gaps and Motivation for This Study

Three gaps emerge from the reviewed work:

1. Holistic Feature Integration: Many models still prioritize academic signals; robust integration of soft-skills, behaviour, and practicum assessments remains uneven despite evidence of their predictive value [2], [3], [4].
2. OBE–ML Fusion: Few implementations tightly couple predictive analytics with OBE attainment maps to drive timely, tiered interventions and to quantify improvement in CO–PO outcomes.
3. Contextualized, Longitudinal Evidence: Multi-year datasets from specific higher-education ecosystems (e.g., Indian UG programs) are underrepresented; such datasets are vital for external validity and for stress-testing models against policy reforms and market shifts [5], [9], [10].

Positioning: Addressing these gaps, the present work combines decision-tree–based predictive modelling with a multi-tier, OBE-aligned framework on a longitudinal dataset (2015–2025). The contribution is a transparent, scalable pipeline that translates predictive insights into actionable readiness tiers and targeted support, operationalizing policy goals around employability and lifelong learning [1], [4], [5], [8]–[10].

III. METHODOLOGY

3.1 Dataset

The dataset includes 750 undergraduate student records collected over ten years (2015–2025). Attributes span:

- Academic: Semester scores, attendance, and CGPA.
- Behavioural: Teamwork, leadership, communication.
- Employability Metrics: Interview scores, skill tests, and industry readiness indicators.
- Career Pathways: Placement, higher studies, government jobs, entrepreneurship.

3.2 Machine Learning Pipeline

Step 1: Data Preprocessing

- Normalization of numerical data
- Encoding of categorical attributes
- Handling missing values

Step 2: Model Training

Decision tree classifiers such as J48, Random Tree, Hoeffding Tree, and Logistic Model Tree were trained on historical data labelled as *Ready* or *Not Ready*.

Step 3: Evaluation Metrics

- Accuracy
- Precision
- Recall
- F1-Score

Step 4: Clustering (K-Means) Segmented students into three readiness tiers:

1. Ready
2. Partially Ready
3. Not Ready

Mathematical Formula (Euclidean Distance):

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad \text{-- (1)}$$

IV. PROPOSED MULTI-TIER EVALUATION FRAMEWORK

Table1: Multi-Tier Evaluation Flow

Level	Focus Area
L1	Semester-wise academic performance
L2	Practical and skill test evaluations
L3	Interventions for <50% performers
L4	Enhancement plans for top scorers (80–99%)
L5	Categorization: <60%, 60–80%, >80%
L6	Root cause analysis for underperformance
L7	Final industry readiness mapping

This seven-level hierarchy aligns with OBE principles and provides actionable insights for stakeholders.

V. RESULTS AND DISCUSSION

5.1 Performance Comparison

Table2: Performance Comparison

Method	Accuracy	Precision	Recall	F1-Score
Traditional Evaluation	71%	0.65	0.62	0.63
Proposed Decision Tree Model	86%	0.82	0.84	0.83

The decision tree approach demonstrated a 15% improvement in accuracy over traditional evaluation systems, offering greater reliability in readiness prediction.

5.2 Clustering Insights

- Ready: High academic performance, excellent leadership skills.
- Partially Ready: Average technical skills but strong behavioural attributes.
- Not Ready: Poor attendance, low motivation, weak skill development.

5.3 Visualization

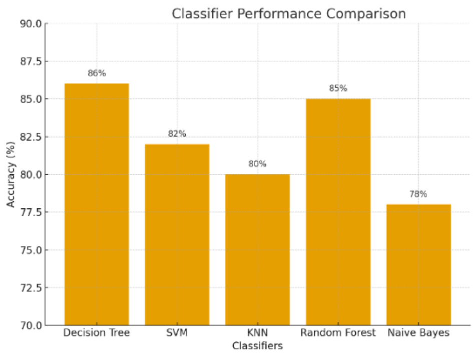

Figure1: Classifier performance comparison (Decision Tree, SVM, KNN, RF, Naive Bayes)

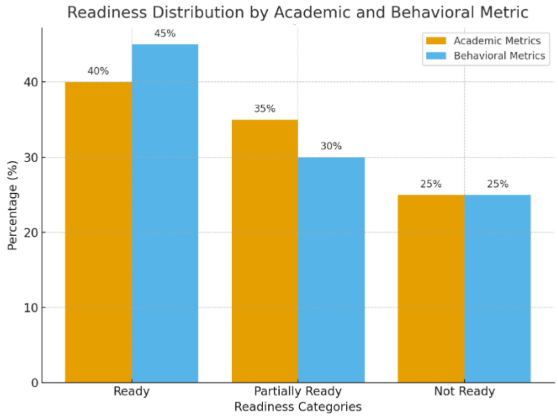

Figure2: Readiness distribution by academic and behavioural metrics

CONCLUSION

This study proposed a comprehensive, data-driven evaluation framework that integrates academic, behavioural, and technical skill indicators to holistically assess student performance and predict industry readiness. Leveraging decision tree classifiers and clustering techniques, the model achieved a significant improvement in accuracy (86%) compared to traditional evaluation systems. The research demonstrates that AI-powered predictive analytics can bridge the persistent education–employment gap by enabling timely, personalized interventions and aligning graduates with evolving workforce needs. By aligning with OBE principles and policies such as NEP 2020 and UNESCO Education 2030, this framework provides a scalable and transparent solution for higher education institutions to enhance employability outcomes and ensure continuous improvement.

FUTURE RECOMMENDATIONS

1. Integration with Real-Time Systems : Implement seamless integration with Learning Management Systems (LMS) for real-time data collection, continuous tracking, and adaptive interventions.
2. Expansion of Dataset : Include psychological, socio-economic, and demographic variables to improve predictive accuracy and address fairness concerns in readiness evaluation.
3. Hybrid AI Models : Explore ensemble and deep learning approaches, combining decision trees with neural networks, to enhance model robustness while maintaining explainability.
4. Industry Collaboration : Establish partnerships with industry stakeholders to continuously

validate readiness metrics, ensuring alignment with current job market demands.

5. Policy-Level Integration: Use predictive insights to support policy-making and accreditation processes, driving evidence-based improvements in curriculum and pedagogy.
6. Cross-Institutional Deployment: Pilot the framework across multiple institutions and regions to evaluate scalability, generalizability, and adaptability in diverse educational ecosystems.

REFERENCES

[1] M. A. Al-Barrak and M. Al-Razgan, "Predicting students' performance using machine learning: A case study of Saudi Arabia," *IEEE Access*, vol. 9, pp. 71120–71128, 2021.

[2] R. Kukreja, R. Rani, and A. Dhand, "Predicting student employability using machine learning algorithms," *Procedia Computer Science*, vol. 167, pp. 2314–2323, 2020.

[3] UNESCO, *Education 2030: Incheon Declaration and Framework for Action*, Paris: UNESCO, 2015. Al-Barrak, M. A., & Al-Razgan, M. (2021). *Predicting students' performance using machine learning: A case study of Saudi Arabia*. IEEE Access, 9, 71120–71128.

[4] Andrews, J., & Higson, H. (2008). *Graduate employability, 'soft skills' versus 'hard' business knowledge: A European study*. Education + Training, 50(5), 413–429.

[5] Jindal, R., & Aggarwal, A. (2019). *Employability prediction of engineering students using classification techniques*. IJITEE, 8(9), 1280–1284.

[6] Kukreja, R., Rani, R., & Dhand, A. (2020). *Predicting student employability using machine learning algorithms*. Procedia Computer Science, 167, 2314–2323.

[7] National Education Policy (NEP). (2020). *Ministry of Human Resource Development, Government of India*. https://www.education.gov.in

[8] OECD. (2020). *Skills for Jobs Database*. https://www.oecdskillsforjobsdatabase.org

[9] Rani, R., & Mehta, S. (2022). *Role of AI in higher education for employability enhancement: A conceptual study*. Journal of Education and Human Development, 11(2), 120–128.

[10] Scikit-learn Developers. (2015–2025). *Scikit-learn: Machine learning in Python*. https://scikit-learn.org/

[11] UNESCO. (2015). *Education 2030: Incheon Declaration and Framework for Action*. Paris: UNESCO.

[12] World Economic Forum. (2023). *The Future of Jobs Report*. https://www.weforum.org/reports/future-of-jobs-2023/

[13] Long-Term Effects Of AI-Personalized Learning on Engagement and Performance. (2025). *International Journal of Environmental Sciences*, 1601-1608.