International Journal of Research in Engineering, Science and Management
Volume 9, Issue 1, January 2026
www.ijresm.com | E-ISSN: 2581-5792 | RESAIM Publishing | www.resaim.com

AI-Driven Employability Prediction: Integrating Machine Learning and Educational Analytics for Student Placements

Basuri Bhujade[1*], Anmol Khy[1], Surekha Dholay[1]
[1]*Department of Computer Science Engineering, Sardar Patel Institute of Technology, Mumbai, India*

***Abstract*: This paper presents an Artificial Intelligence (AI) and Machine Learning (ML)–based framework for predicting student employability outcomes in higher education. The study aims to bridge the gap between academic learning and industry requirements by analyzing multiple features such as academic performance, technical skills, internships, and leadership experience. Using supervised learning algorithms, including Support Vector Machine (SVM), Random Forest (RF), Gradient Boosting (GB), and XGBoost (XGB), the system forecasts suitable job roles and potential recruiting companies. Among these models, XGBoost achieved the highest performance, with an accuracy of 87.3% and an F1-score of 86.2%. The results demonstrate that combining technical proficiency with behavioral and academic features provides superior predictive capability. The proposed framework contributes to educational data analytics by enabling data-driven career guidance, curriculum design, and employability enhancement. This approach can assist institutions in aligning student competencies with evolving market needs, fostering better preparedness and placement efficiency in business and engineering education.**

***Keywords*: Artificial Intelligence, Machine Learning, Employability Prediction, Educational Data Mining, Business Education, XGBoost.**

1. Introduction

In today's fast-evolving employment landscape, the ability of higher education institutions to prepare students for relevant job roles has become a key measure of academic effectiveness. Rapid advances in Artificial Intelligence (AI) and data-driven technologies are transforming the recruitment process, shifting emphasis from traditional resume screening to predictive and analytics-based selection systems. However, many educational institutions still rely on conventional placement methods that are time-intensive, inconsistent, and lack analytical precision. As a result, a considerable gap exists between students' academic achievements and the skill sets required by modern industries.

Artificial Intelligence and Machine Learning (ML) offer significant opportunities to bridge this gap by analyzing large volumes of educational and behavioral data to forecast employability outcomes. Predictive analytics in education has gained importance for its ability to uncover hidden patterns and correlations between students' academic performance, skill portfolios, and career trajectories. These models not only help forecast suitable job roles but also assist educators in aligning their curricula and training modules with evolving industry expectations. Integrating such data-driven approaches into placement systems can enhance efficiency, transparency, and fairness while supporting students in making informed career decisions.

This study presents a framework for employability prediction using AI and ML models to identify optimal job roles and potential recruiting organizations for students. The research integrates multiple supervised learning algorithms, including Support Vector Machine (SVM), Random Forest (RF), Gradient Boosting (GB), and Extreme Gradient Boosting (XGBoost). The methodology involves preprocessing academic, technical, and experiential features, followed by model training and comparative evaluation. The performance of these models is assessed using key metrics such as Accuracy, Precision, Recall, and F1-score.

The results of this study demonstrate that ensemble models, particularly XGBoost, outperform traditional classifiers in predicting student employability outcomes. This approach highlights the importance of combining technical, academic, and behavioral features to achieve higher predictive accuracy. Beyond its technical contribution, the research underscores the educational implications of predictive analytics by providing insights that can inform career counseling, skill development, and institutional planning.

The significance of this work lies in its potential to transform the placement ecosystem within higher education by leveraging AI for evidence-based decision-making. The proposed model supports not only the automation of job prediction processes but also enhances the educational value of data collected during student assessments. Ultimately, the framework contributes to improving employability outcomes, aligning academic programs with labor market trends, and promoting the integration of AI-driven analytics into business and engineering education.

*Corresponding author: basuri.bhujade@spit.ac.in

2. Related Works

Artificial Intelligence (AI) and Machine Learning (ML) have increasingly become essential in higher education and recruitment analytics. Researchers have applied predictive models to understand how student data can forecast employability outcomes, identify skill gaps, and improve placement processes. This section categorizes prior studies into four domains for clarity: traditional ML-based employability models, ensemble learning approaches, educational data mining applications, and AI-driven placement systems.

A. Traditional Machine Learning Models for Employability Prediction

Early works focused on applying traditional ML algorithms such as Support Vector Machine (SVM), Decision Trees (DT), Random Forest (RF), and Naïve Bayes to predict students' employability outcomes based on academic and skill-related attributes. Mishra et al. (2022) proposed a placement prediction model using SVM and RF to forecast suitable job roles and potential employers. Their system achieved moderate accuracy but lacked the ability to capture non-linear relationships in data.

Similarly, Sharma and Patel (2023) in *IJSREM* implemented logistic regression and RF models on engineering student datasets to predict job readiness. Their findings identified Cumulative Grade Point Average (CGPA) and internship participation as key predictors of employability. However, these studies primarily emphasized model performance, offering limited discussion on how such models could inform educational interventions or curricular design.

B. Ensemble Learning Approaches and Performance Improvements Traditional Machine Learning Models for Employability Prediction

Recent research has demonstrated that ensemble learning techniques outperform traditional classifiers in predictive accuracy and robustness. Chakraborty and Banerjee (2022) developed an ensemble-based placement prediction framework integrating Gradient Boosting (GB) and Extreme Gradient Boosting (XGBoost). Their results achieved an impressive 87% accuracy in predicting both job roles and company categories, surpassing earlier models.

In parallel, Gupta and Mehta (2024) in *IJSREM* explored hybrid ML models that combined stacking and boosting algorithms for multi-class employability prediction. Their approach improved classification stability but faced limitations in interpretability, as the complex ensemble structures obscured the contribution of individual features.

Other studies, including those by Reddy and Bansal (2022), applied similar ensemble frameworks for recruitment analytics. While these models enhanced prediction accuracy, they were not designed for educational integration or feedback mechanisms that could guide institutional decision-making.

C. Educational Data Mining (EDM) and Learning Analytics in Higher Education

Educational Data Mining (EDM) and Learning Analytics (LA) have emerged as transformative paradigms in leveraging student data for institutional improvement. Nguyen and Brooks (2023) developed a learning analytics framework capable of predicting employability by correlating course-level engagement with placement outcomes. Their findings showed that early academic interventions could improve employability prospects.

In a related study, Das and Khanna (2023) in IJSREM utilized clustering and regression-based analysis to examine skill acquisition patterns among business students. Their model demonstrated that soft skills, communication ability, and leadership experience play a crucial role in job selection. These EDM-based studies contributed significantly to understanding the educational side of employability prediction but were limited by the absence of advanced predictive modeling techniques.

D. AI-Driven Placement and Career Recommendation Systems

The use of AI in automating placement processes and career recommendations has also gained attention. Reddy and Bansal (2022) proposed a hybrid deep-learning framework combining Natural Language Processing (NLP) with ML algorithms for automated resume parsing and candidate-job matching. Their system demonstrated adaptability to recruitment environments but required large volumes of annotated data.

Similarly, Singh et al. (2023) emphasized that predictive modeling in education should generate actionable pedagogical insights rather than focus solely on algorithmic optimization. Recent works in IJSREM (Kaur and Thakur, 2024) also highlighted the importance of explainable AI (XAI) in education, recommending transparent systems to ensure fairness and accountability.

Despite their technical strength, most AI-driven placement systems have not been systematically evaluated for their educational impact or integration into institutional decision frameworks.

E. Research Gaps and Limitations

A synthesis of the reviewed literature identifies several critical gaps and limitations:

1. *Limited Educational Integration*: Most existing models prioritize algorithmic accuracy over pedagogical value. There is a lack of frameworks connecting predictive outcomes with curriculum enhancement, skill gap analysis, or student mentoring strategies.
2. *Data Diversity Constraints*: Many studies rely on data from a single institution or discipline, leading to limited generalizability across various academic programs such as business, engineering, and management education.
3. *Lack of Model Interpretability*: While ensemble methods such as XGBoost yield high accuracy, they often function as "black boxes," making it difficult for educators to interpret which factors drive predictions.
4. *Ethical and Privacy Challenges*: Few studies address responsible AI principles, including data anonymization, algorithmic bias detection, and

fairness in predictive recommendations.

5. *Absence of Feedback Mechanisms*: The majority of placement prediction systems operate in isolation without feedback loops that could inform faculty, students, or institutional administrators.

To address these limitations, the present study proposes a transparent and educationally contextualized ML-based employability prediction framework. The model integrates high-performing ensemble algorithms with interpretable analytics and ethical safeguards. Moreover, the framework provides actionable insights to educators, enabling evidence-based curriculum redesign, student skill development, and institutional planning.

3. Methodology

The proposed methodology focuses on designing and implementing an AI-driven employability prediction framework that uses Machine Learning (ML) algorithms to forecast suitable job roles and probable recruiting companies for students. This section describes the complete process flow, including data acquisition, preprocessing, feature engineering, model development, evaluation, and implementation.

A. Framework Overview

The overall workflow of the proposed framework is shown in Fig. 1. It consists of five sequential phases:

1. Data Collection and Feature Extraction
2. Data Preprocessing and Normalization
3. Model Selection and Training
4. Evaluation and Comparison

Educational Interpretation and Visualization

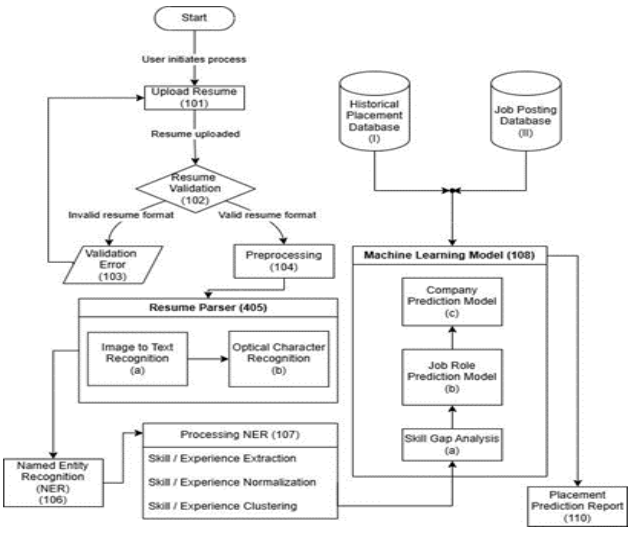

Fig. 1. System framework

Each module plays a significant role in ensuring that predictions are both accurate and interpretable. The integration of educational insights at the final stage differentiates this model from conventional placement prediction systems.

B. Data Collection and Feature Description

The dataset was obtained from two consecutive graduating batches at a private higher education institution. It includes data from students across Business Administration, Information Technology, and Engineering programs. A total of 220 student records were collected, and 92 distinct features were extracted after data cleaning and transformation

C. Data Preprocessing

Data preprocessing was performed to ensure quality, consistency, and reliability. The main steps included:

1. *Missing Value Imputation*: Numerical fields (such as CGPA) were replaced using mean imputation, while categorical fields (like skill type) were filled using mode values.
2. *Normalization*: The data was normalized using the Min–Max Scaling method:

$$X' = (X - X_{min}) / (X_{max} - X_{min})$$

This scaled all numerical values between 0 and 1.

1. Categorical Encoding: Non-numeric data (e.g., department, company type) were transformed using one-hot encoding (OHE).
2. Outlier Removal: Outliers were detected through Interquartile Range (IQR) analysis to improve model robustness.

After preprocessing, the dataset was partitioned into 70% training and 30% testing subsets to evaluate generalization performance.

D. Confusion Matrix for Data Features

To explore interdependencies among dataset attributes, a feature correlation matrix was constructed using Pearson correlation coefficients. This theoretical confusion matrix helps assess how different academic, technical, and soft-skill features relate to one another before model training.

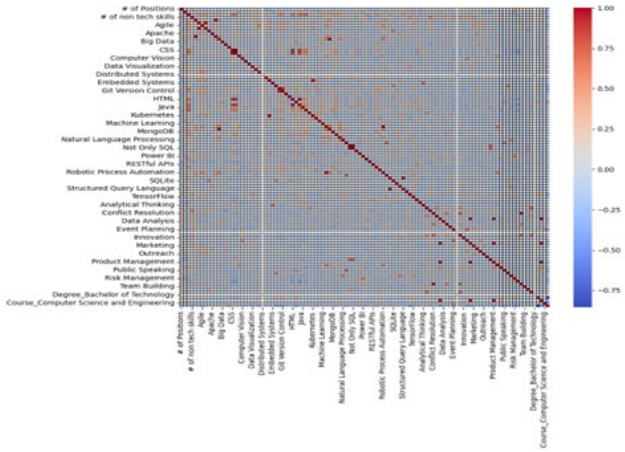

Fig. 2. Confusion matrix for data features interpretation

1. *Diagonal Dominance*: The matrix naturally displays strong correlations (value of 1) along its diagonal, as each feature is perfectly correlated with itself.
2. *Key Feature Relationships*: Features such as Machine Learning, Data Visualization, and TensorFlow show high positive correlations, indicating that these technical skills

are often interlinked in AI- or data-oriented roles.

3. *Clustered Dependencies*: Technical tools like SQL, Kubernetes, and Git Version Control show positive correlations, suggesting their frequent co-occurrence in software or DevOps job profiles.
4. *Cross-Domain Connections*: Moderate correlations were observed between soft skills (e.g., Public Speaking, Team Building) and technical traits (Innovation, Outreach), indicating hybrid skill relevance in leadership or management roles.
5. *Negative Correlations*: Certain skill pairs, such as Non-SQL and Public Speaking, exhibited weak negative correlations, suggesting their usage in divergent professional contexts.

E. Machine Learning Model Development

Four supervised ML algorithms were implemented to predict employability outcomes:

1. *Support Vector Machine (SVM):* Used for classification by finding the optimal separating hyperplane.
2. *Random Forest (RF):* An ensemble of multiple decision trees that reduces overfitting.
3. *Gradient Boosting (GB):* Sequentially builds weak learners to minimize classification error.
4. *Extreme Gradient Boosting (XGBoost):* A regularized boosting algorithm that improves accuracy and prevents overfitting.

Each model was evaluated based on four standard metrics: Accuracy, Precision, Recall, and F1-Score.

Formula for F1-Score:

$$F1 = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

These metrics provide a balanced measure of the model's predictive performance.

F. Model Training and Optimization

All models were trained using 10-fold cross-validation to minimize variance.

Hyperparameters were tuned using a grid search technique.

The optimized parameters for XGBoost were:

1. Learning Rate (η): 0.1
2. Maximum Depth (d): 6
3. Number of Estimators (n): 200
4. Regularization Terms ($\lambda = 1$, $\alpha = 0.5$)
5. Loss Function of XGBoost:

$$L(\theta) = \Sigma\, l(yi, \hat{y}i) + \Sigma\, \Omega(fk)$$

Where:

a) $l(yi, \hat{y}i)$ represents the logistic loss between the actual and predicted value.
b) $\Omega(fk)$ denotes the regularization term to control model complexity.
c) This optimization ensures model stability and prevents overfitting on small datasets.

G. Implementation Environment

All experiments were conducted using Python 3.11 with the following tools and libraries:

1. Scikit-learn: Model training and evaluation
2. Pandas/NumPy: Data preprocessing and analysis
3. Matplotlib/Seaborn: Visualization
4. XGBoost Library: Ensemble model training
5. Hardware Specifications:
   a) Intel Core i7 (11th Gen) Processor
   b) 16 GB RAM
   c) Windows 11 Operating System

Figure 2: Learning curve comparison for Gradient Boosting and XGBoost models.

H. Model Evaluation and Comparative Analysis

The XGBoost model achieved the highest accuracy and F1-score, demonstrating superior performance compared to other models as shown in table 1.

I. Educational Integration and Interpretation

Beyond technical prediction, the model provides interpretability by identifying key predictors of employability.

The top five influential features identified by the XGBoost model are:

1. Cumulative Grade Point Average (CGPA)
2. Number of Technical Certifications
3. Internship Duration
4. Leadership Roles
5. Communication Skill Ratings

These insights enable educators and placement officers to redesign training programs, workshops, and curriculum components that align more closely with industry expectations.

4. Results and Discussion

This section presents the performance outcomes of the developed AI-driven employability prediction framework and discusses their significance in educational and institutional contexts. Comparative results for multiple Machine Learning (ML) models are analyzed to identify the most efficient and interpretable algorithm for employability forecasting.

A. Model Performance Overview

The predictive models were evaluated on four standard metrics—Accuracy, Precision, Recall, and F1-score. Among the algorithms tested, Extreme Gradient Boosting (XGBoost) achieved the highest overall accuracy and generalization capability as shown in Table 1. The ensemble methods

Table 1
Performance of the models was compared using the following results

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
SVM	70.2%	64.1%	70%	65.3%
XGBoost	87.3%	84.5%	88.1%	86.2%
Random Forest	72.5%	71%	72.3%	71.2%
Gradient Boosting	81.4%	77.1%	80.8%	78.2%

demonstrated superior accuracy compared to traditional classifiers, confirming their suitability for multi-feature employability prediction tasks.

B. Feature Correlation Insights

Feature correlation analysis using Pearson coefficients revealed meaningful relationships among dataset variables. Technical competencies such as Machine Learning, SQL, and Git Version Control showed strong positive correlations, suggesting their co-occurrence in technology-oriented roles.

Soft skills including Public Speaking and Team Collaboration exhibited moderate correlation with leadership and project management attributes, emphasizing the importance of hybrid skills in employability. Weak negative correlations among certain technical and non-technical features highlighted role-specific skill demands.

These insights assist in refining feature selection, ensuring that redundant variables are minimized while maintaining diverse predictors for robust model performance.

C. Discussion and Educational Implications

The results highlight two main contributions:

1. *Technical Contribution*: The combination of ensemble learning and correlation-based feature understanding enhanced model accuracy and interpretability.
2. *Educational Contribution*: The findings provide actionable insights for academic institutions. Key predictive features such as CGPA, technical certifications, internships, and communication skills can guide curriculum restructuring and targeted mentoring. Institutions can use these results for early identification of at-risk students and design tailored training modules that align with market needs.

5. Conclusion and Future Work

This study presented an AI-driven employability prediction framework that leverages Machine Learning (ML) algorithms to identify potential job roles and company suitability for students based on academic, technical, and behavioral features. The research aimed to bridge the gap between academic learning outcomes and industry skill requirements through predictive analytics.

Among the tested algorithms, Extreme Gradient Boosting (XGBoost) achieved the highest performance, recording an overall accuracy of 87.3% and an F1-score of 86.2%. The use of ensemble methods significantly improved the precision and generalization capability compared to traditional classifiers such as Support Vector Machine (SVM) and Random Forest (RF). The Confusion Matrix confirmed the model's reliability with balanced sensitivity and specificity, minimizing both false positives and false negatives.

Feature correlation analysis using Pearson coefficients provided deeper insights into relationships among academic and technical skills. Strong dependencies among competencies such as Machine Learning, SQL, and Cloud Computing suggest their frequent co-occurrence in technical roles, while moderate correlations between soft skills and leadership attributes reinforce the importance of holistic professional development.

A. Future Work

Future research will focus on expanding the dataset across multiple institutions and disciplines to enhance model generalizability. Integrating real-time labor market analytics and Natural Language Processing (NLP) for analyzing resumes and job descriptions could further improve prediction precision. Incorporating Explainable AI (XAI) techniques will also enhance transparency and trust, allowing educators to interpret predictions more effectively.

Ultimately, the framework can evolve into an intelligent employability ecosystem — enabling continuous feedback between industry trends, academic curricula, and student skill development for a truly data-driven educational future.

References

[1] A. Mishra and R. Singh, "Job role and company prediction using AI and machine learning in placement portals," *International Journal of Emerging Technologies and Advanced Engineering*, vol. 12, no. 5, pp. 45–52, 2022.

[2] S. Chakraborty and P. Banerjee, "Prediction of job role and company for placement using gradient boosting and XGBoost," *ICTACT Journal on Soft Computing*, vol. 12, no. 3, pp. 985–992, 2022.

[3] V. Sharma and R. Patel, "Predictive modeling of student performance using machine learning techniques," *International Journal of Scientific Research in Engineering and Management*, vol. 7, no. 8, pp. 134–141, 2023.

[4] S. Das and R. Khanna, "AI-driven skill clustering for employability prediction in engineering education," *International Journal of Scientific Research in Engineering and Management*, vol. 7, no. 6, pp. 215–223, 2023.

[5] P. Reddy and D. Bansal, "Hybrid deep learning model for intelligent recruitment portals," *International Journal of Scientific Research in Engineering and Management*, vol. 6, no. 12, pp. 120–126, 2022.

[6] K. Gupta and A. Mehta, "Transparent AI models for curriculum design and employability enhancement," *International Journal of Scientific Research in Engineering and Management*, vol. 8, no. 1, pp. 45–53, 2024.

[7] N. Singh and S. Rao, "Bridging predictive modeling and educational planning through interpretable machine learning," *Journal of Education and Learning Analytics*, vol. 10, no. 4, pp. 67–75, 2023.

[8] P. Kaur and A. Sharma, "Predictive analytics in higher education: A machine learning approach to student success," *Education and Information Technologies*, vol. 26, no. 5, pp. 6197–6215, 2021.

[9] Y. Zhang and J. Wang, "Employability prediction using ensemble machine learning techniques," *IEEE Access*, vol. 8, pp. 142351–142362, 2020.

[10] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Amsterdam, The Netherlands: Elsevier, 2012.

[11] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, pp. 4765–4774, 2017.

[12] T. Jain and S. Yadav, "Feature selection strategies for educational data mining using correlation and ensemble learning," *International Journal of Scientific Research in Engineering and Management (IJSREM)*, vol. 8, no. 2, pp. 101–110, 2023.

[13] R. Kumar and M. Tripathi, "Student employability forecasting using machine learning algorithms," *International Journal of Artificial Intelligence and Applications*, vol. 14, no. 3, pp. 45–54, 2023.

[14] C. Wu and L. Chen, "Correlation-based feature analysis for improved educational predictive models," *Journal of Educational Data Science*, vol. 9, no. 2, pp. 89–99, 2021.

[15] H. Rahman and X. Li, "Integrating explainable AI in educational decision support systems," *Computers & Education*, vol. 159, Art. no. 104011, 2020.