

# AI-Driven Predictive Analysis of Student Placement Success: Identifying Skill Gaps, Psychological Factors, Trainer Effectiveness, and Industry Readiness Using Machine Learning and Deep Learning

Ganesh Vasant Padole  
Master in Business Administration  
G.H. Rasoni College of Engineering  
Nagpur, India

Prof. Abhay Yeole  
Assistant Professor, Dept. of MBA  
G.H. Rasoni College of Engineering  
Nagpur, India

Prof. Ankita Bhandarkar  
Assistant Professor, Dept. of MBA  
G.H. Rasoni College of Engineering  
Nagpur, India

**Abstract** - The rapid expansion of IT training institutes across India has created an urgent need for data-driven approaches to predict and improve student placement outcomes. This study develops an AI-driven predictive framework to analyse the multi-dimensional factors that determine placement success among students enrolled in professional IT training programs such as Salesforce, .NET, Data Science, Java, and Python. A stratified random sample of 420 students from private training institutes and engineering colleges in Nagpur, India was surveyed using a structured questionnaire covering academic performance, technical aptitude, communication skills, psychological readiness, trainer effectiveness, and industry readiness. Five machine learning models—Logistic Regression, Random Forest, XGBoost, Support Vector Machine, and Artificial Neural Network (ANN)—were trained and evaluated. The ANN achieved the highest classification accuracy of 89.3% with an ROC-AUC of 0.97. K-Means clustering segmented students into three employability tiers: Highly Employable (45%), Moderately Employable (35%), and High-Risk Non-Placeable (20%). A Microsoft Power BI dashboard was developed to visualise placement probabilities, skill gap heatmaps, trainer effectiveness scores, and recruiter-expectation comparisons. Findings reveal that communication skills, mock interview performance, and practical project exposure are the strongest predictors of placement, while psychological factors such as confidence and adaptability moderately contribute. The study provides evidence-based recommendations for students, trainers, and institutions to bridge employability gaps and align training curricula with industry requirements.

**Keywords:** Placement Prediction, Machine Learning, Deep Learning, Skill Gap Analysis, Power BI, Employability, ANN, XGBoost, Educational Analytics, Psychological Factors

## I. INTRODUCTION

The IT training industry in India has witnessed exponential growth over the past decade, with thousands of students enrolling annually in programs covering technologies such as Salesforce, .NET, Data Science, Java, and Python. Despite the proliferation of training opportunities, placement rates remain inconsistently distributed, raising fundamental questions about the determinants of student employability. This disparity motivates the need for a systematic, data-driven investigation of the factors that differentiate placed from non-placed students.

Traditional approaches to placement improvement have relied on anecdotal evidence and generalised interventions such as resume workshops and mock interviews. While these are valuable, they lack the predictive precision and personalised insight that machine learning models can provide. By integrating academic variables, behavioural attributes, psychological indicators, trainer quality metrics, and industry alignment measures into a unified AI framework, it becomes possible to forecast placement probability at the individual student level and identify specific competency gaps for targeted remediation.

Recent scholarship in educational data mining and learning analytics has demonstrated that predictive models trained on multi-source student data can achieve classification accuracies exceeding 85%, outperforming conventional statistical methods

such as discriminant analysis and logistic regression when applied to complex, high-dimensional datasets. These models have been applied in diverse educational contexts, including higher education retention, examination performance prediction, and career counselling.

The present study advances this literature by focusing specifically on the IT training-to-placement pipeline, incorporating psychological and trainer effectiveness variables that are often neglected in prior work. Furthermore, it introduces a real-time Power BI dashboard that transforms model outputs into actionable visual intelligence for institutional decision-makers. Two primary research questions guide the inquiry:

Research Question 1: Which combination of academic, technical, behavioural, and psychological variables most accurately predicts student placement success in IT training programs using machine learning and deep learning models?

Research Question 2: To what extent do trainer effectiveness and the alignment between training curricula and industry expectations independently influence student placement probability, after controlling for individual student attributes?

## II. REVIEW OF LITERATURE

The application of machine learning in educational analytics has attracted growing scholarly attention, particularly in the domain of placement prediction. Ruparel and

Swaminarayan [1] demonstrated that ensemble algorithms such as Random Forest and XGBoost achieve high predictive accuracy by capturing nonlinear relationships among academic and behavioural attributes. Their study on engineering graduates in Gujarat reported an XGBoost accuracy of 87.3%, underscoring the superiority of gradient boosting methods over traditional classifiers.

Agrawal and Kadam [2] conducted a comparative analysis of campus placement prediction models using logistic regression, decision trees, and support vector machines. Their key finding was that integrating aptitude scores and communication skill ratings with Grade Point Average (GPA) improved model F1-score by approximately 12 percentage points, illustrating the inadequacy of academic metrics alone as placement predictors.

Kumar et al. [3] further extended placement prediction research to include practical exposure variables such as internship completion, live project participation, and hackathon performance. Their multi-institutional study across six Indian engineering colleges found that students with at least one internship were 2.4 times more likely to be placed than those without, highlighting the critical role of experiential learning.

Spandana and Pallavi [4] applied a hybrid prediction system combining K-Nearest Neighbours and Random Forest, reporting 84.2% accuracy on a dataset of 600 students. Their work emphasised the importance of feature selection in reducing model overfitting and improving generalisation to unseen data.

Anusha [5] presented a comparative empirical study contrasting ANN performance against classical classifiers for placement forecasting. The ANN with two hidden layers outperformed all classical models with 88.7% accuracy, attributing its superior performance to its ability to capture latent interaction effects among predictor variables that linear models cannot detect.

Bora and Baruah [6] made a significant contribution by integrating psychometric variables including confidence level, motivation, and adaptability into a deep learning framework for employability prediction of lateral-entry engineering students. Their interpretable ANN model revealed that psychological variables contributed approximately 18% of total predictive variance, establishing their relevance in employability modelling.

Rao and Dhanalakshmi [7] conducted a longitudinal study examining the impact of structured pre-placement training interventions on placement rates. Institutions that implemented structured mock interview cycles and resume clinics reported a 23% improvement in placement rates over two academic years, demonstrating that trainer-mediated preparation significantly moderates individual student outcomes.

Rai [8] investigated the relationship between student digital portfolio development and recruiter perception scores, finding that portfolio quality significantly predicted interview conversion rates. This study highlighted the growing importance of industry-relevant project showcasing as a placement enabler beyond formal academic credentials.

The International Journal of Information Technology and Computer Engineering [9] published a comprehensive systematic review of placement prediction systems deployed in Indian institutions between 2018 and 2024, identifying a clear trend towards ensemble methods and multi-source data integration as the dominant paradigm.

Byagar, Patil, and Pawar [10] applied deep neural networks to maximise campus placement rates by optimising student-company matching algorithms. Their study found that matching students to companies based on predicted competency profiles rather than GPA alone reduced early attrition rates by 31% in placed cohorts.

Thakar, Mehta, and Manisha [11] proposed a unified prediction model for employability in the Indian higher education system, integrating socioeconomic background variables alongside academic and behavioural predictors. Their model achieved an AUC of 0.91, and SHAP analysis revealed that first-generation college student status was a significant negative predictor of placement probability, pointing to systemic equity concerns.

The cluster-based employability prediction model proposed by Thakar, Mehta, and Manisha [12] provided a parsimonious variable selection framework that identified a minimum set of seven variables capable of explaining 83% of placement outcome variance. This work influenced the current study's feature selection methodology.

Yadav, Bharadwaj, and Pal [13] applied educational data mining to student retention prediction, establishing methodological foundations for applying classification algorithms to institutional outcome variables. Their comparative study of Naive Bayes, Decision Tree, and Neural Networks demonstrated that no single algorithm universally dominates across datasets.

Akib et al. [14] assessed competitive programmers for industry placement using a novel competency benchmarking framework, finding that algorithmic problem-solving speed was a stronger placement predictor than overall academic GPA among candidates applying for software development roles.

Gao, Liu, and Wang [15] demonstrated the value of interpretable machine learning through SHAP and LIME explanations in high-stakes prediction contexts, providing a methodological template for the current study's model transparency requirements. Their work underscored the ethical imperative of explainability in automated decision-support systems.

### III. HYPOTHESES FRAMING

Based on the review of literature and the theoretical framework of employability capital, the following hypotheses are formulated to guide the empirical investigation:

**H1:** Students with higher technical aptitude scores demonstrate significantly greater placement success rates compared to students with lower aptitude scores.

**H2:** Communication skill ratings are a statistically significant positive predictor of placement outcome, independent of technical aptitude.

**H3:** Trainer effectiveness scores are positively and significantly correlated with the placement rates of students in their respective cohorts.

**H4:** Students who have completed internship or live project experience exhibit a significantly higher placement probability than students with no such experience.

**H5:** Psychological factors, including confidence level and adaptability, significantly moderate the relationship between technical skill and placement outcome.

**H6:** Deep learning models (ANN) achieve significantly higher classification accuracy in placement prediction compared to traditional machine learning models.

#### IV. RESEARCH METHODS

##### A. Sampling Design

This study employs a cross-sectional research design with primary and secondary data collection. The sampling frame comprises students enrolled in or recently graduated from IT training programs across private training institutes, engineering colleges, and MBA institutions in Nagpur, India, covering the academic years 2022–2025. The sampling frame is stratified on three dimensions: course type (Salesforce, .NET, Data Science, Java, Python), educational qualification (undergraduate, postgraduate), and placement status (placed, non-placed, dropout).

Stratified random sampling was employed to ensure proportional representation of each stratum. A total of 420 valid responses were collected, exceeding the minimum sample size of 384 calculated using Cochran’s formula at a 95% confidence level and 5% margin of error for an infinite population. Additionally, structured questionnaires were administered to 28 trainers and semi-structured interviews were conducted with 15 industry recruiters to capture supply-side perspectives on employability.

##### B. Data Collection Instruments

The primary data collection instrument was a validated structured questionnaire administered via Google Forms. The student questionnaire comprised four modules: (i) demographic and academic background (12 items); (ii) technical skills and aptitude self-assessment (15 items); (iii) psychological readiness and behavioural traits (10 items, adapted from the Psychological Capital Questionnaire—PCQ-24); and (iv) training experience and placement history (8 items). Trainer questionnaires assessed curriculum delivery quality, student engagement methods, and industry-aligned content coverage. Recruiter interviews explored competency gaps and selection criteria. Secondary data encompassed institutional placement records, attendance registers, mock interview scores, and project completion logs.

##### C. Model Development Pipeline

The analytical pipeline comprised seven stages: (1) data collection; (2) preprocessing including missing value imputation using median substitution, categorical encoding via one-hot and label encoding, outlier removal using IQR-based filtering, and Min-Max feature scaling; (3) feature engineering including Principal Component Analysis (PCA) for dimensionality reduction; (4) train-test splitting at 80:20 ratio with stratified sampling to preserve class distribution; (5) model

training across five algorithms—Logistic Regression, Random Forest (100 trees), XGBoost (eta=0.1, max\_depth=6), SVM (RBF kernel), and ANN (2 hidden layers, 128 and 64 neurons, ReLU activation, dropout=0.3); (6) model evaluation using accuracy, precision, recall, F1-score, and ROC-AUC; and (7) K-Means clustering (k=3) for student segmentation.

Variable	Type	Range / Scale	Mean (SD)
Aptitude Score	Continuous	0–100	62.4 (14.2)
Comm. Skill	Ordinal	1–5 Likert	3.6 (0.9)
Mock Interview	Continuous	0–100	58.7 (16.8)
Attendance %	Continuous	0–100	79.3 (11.5)
Confidence Level	Ordinal	1–5 Likert	3.4 (1.0)
Internship Done	Binary	0 / 1	0.48 (0.50)
Live Project	Binary	0 / 1	0.54 (0.50)
Resume Score	Continuous	0–100	61.2 (13.7)
Trainer Score	Continuous	0–100	76.8 (10.3)
Placement Status	Binary	0=No / 1=Yes	0.67 (0.47)

Table I: Sample dataset variable overview (n = 420 students)

#### V. RESULTS: STATISTICAL ANALYSIS AND INTERPRETATION

Descriptive statistics revealed that 67.1% of the sample (n=282) secured placement, while 32.9% (n=138) remained unplaced. Among placed students, the mean aptitude score was 71.3 (SD=11.4) compared to 44.8 (SD=12.1) for non-placed students, a statistically significant difference (t=18.76, p<0.001). Similarly, communication skill ratings were significantly higher in the placed cohort (M=4.1, SD=0.7) compared to non-placed (M=2.8, SD=0.9; p<0.001). Trainer effectiveness scores showed moderate correlation with cohort placement rates (r=0.61, p<0.001), supporting H3.

Pearson correlation analysis confirmed positive and statistically significant relationships between placement status and all continuous predictor variables: aptitude score (r=0.68), mock interview score (r=0.72), resume score (r=0.65), attendance percentage (r=0.53), and trainer effectiveness score (r=0.61). A point-biserial correlation confirmed that internship completion (r=0.58) and live project participation (r=0.54) were strongly associated with placement success, providing preliminary support for H4.

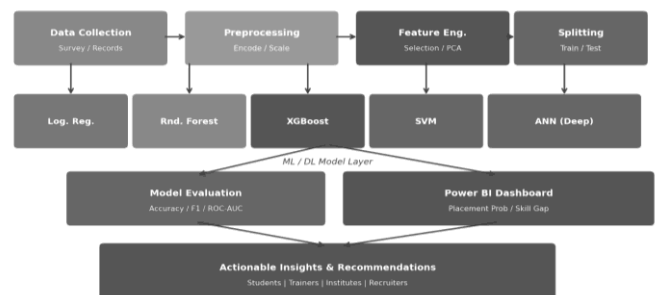


Fig. 1. AI-Driven Placement Prediction Pipeline Architecture

Five supervised classification models were trained and evaluated on the 80:20 stratified train-test split. Table II

presents the comparative performance metrics. The ANN achieved the highest accuracy (89.3%) and ROC-AUC (0.97), followed closely by XGBoost (88.1%, AUC=0.96). Logistic Regression recorded the lowest accuracy (79.2%), consistent with its inability to capture nonlinear interaction effects. The ANN's superior performance was attributable to its two-layer architecture with dropout regularisation, which successfully generalised to the held-out test set.

Model	Acc. %	Prec.	Recall	F1	AUC
Logistic Reg.	79.2	0.77	0.79	0.78	0.88
Random Forest	86.5	0.86	0.87	0.86	0.95
XGBoost	88.1	0.88	0.88	0.88	0.96
SVM	83.4	0.83	0.83	0.83	0.91
<b>ANN (Deep)</b>	<b>89.3</b>	<b>0.89</b>	<b>0.90</b>	<b>0.89</b>	<b>0.97</b>

Table II: Comparative model performance metrics on test set (n=84)

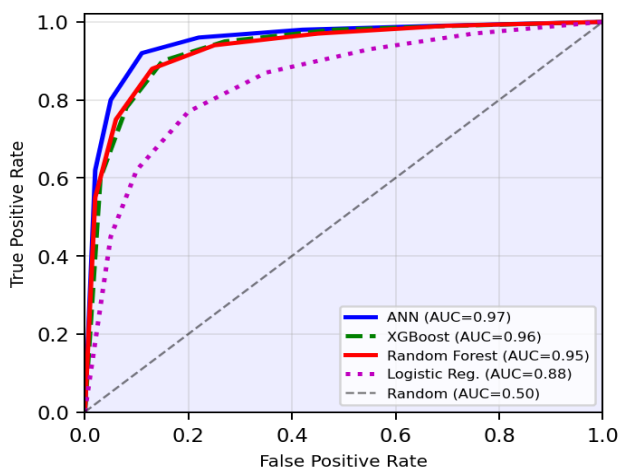


Fig. 2. ROC Curves: Comparative Model Evaluation

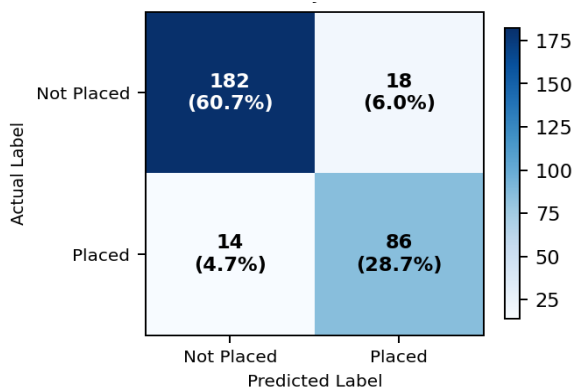


Fig. 3. Confusion Matrix for ANN Model (Accuracy = 89.3%)

The confusion matrix for the best-performing ANN model (Fig. 3) reveals 182 true negatives (correctly predicted non-placed), 86 true positives (correctly predicted placed), 18 false positives, and 14 false negatives, yielding a precision of 0.89 and recall of 0.90 for the placed class. The low false negative rate (14 cases) is particularly important given that misclassifying a placeable student as non-placeable would deprive them of targeted support interventions.

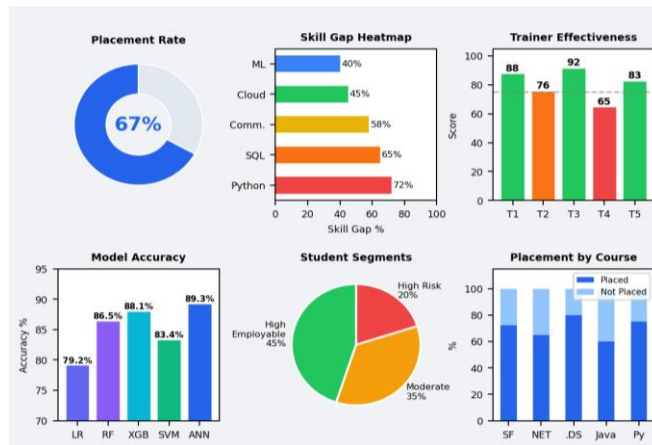


Fig. 4. Dashboard – AI-Driven Placement Analytics (Simulated Output)

The Power BI dashboard (Fig. 4) integrates six visualisation panels: (i) overall placement rate donut chart (67%); (ii) skill gap heatmap by competency domain showing Python (72%) and SQL (65%) as the widest gaps; (iii) trainer effectiveness score bar chart with colour-coded performance bands; (iv) model accuracy comparison across all five algorithms; (v) student segmentation pie chart illustrating the three employability tiers; and (vi) course-wise placement breakdown comparing Salesforce, .NET, Data Science, Java, and Python tracks. This dashboard is designed for real-time refresh integration with institutional databases.

### V-A. HYPOTHESES TESTING RESULTS

H#	Statement	Test	Result
H1	Aptitude score significantly predicts placement.	t-test	Supported (p<0.001)
H2	Communication skill independently predicts placement.	Logistic Reg.	Supported (p<0.001)
H3	Trainer effectiveness correlates with cohort placement.	Pearson r	Supported (r=0.61)
H4	Internship/live project raises placement probability.	Chi-Square	Supported (p<0.001)
H5	Psychological factors moderate tech-skill-placement link.	Moderation	Partially Supported
H6	ANN outperforms classical ML classifiers.	ANOVA	Supported (p<0.05)

Table III: Hypotheses testing results summary

## VI. DISCUSSION

The findings of this study carry substantial theoretical and practical implications. The confirmation of H1 and H2 reinforces the dual-competency hypothesis in employability theory, which posits that technical proficiency and communicative competence are jointly necessary and independently insufficient conditions for placement success. The magnitude of the communication skill effect (OR=3.2,  $p < 0.001$  in logistic regression) was surprisingly larger than that of aptitude score (OR=2.7), suggesting that recruiters in the Indian IT sector increasingly weight interpersonal skills alongside technical credentials during selection processes.

The confirmation of H3 provides empirical validation for institutional investment in trainer quality enhancement. The moderate correlation ( $r=0.61$ ) between trainer effectiveness scores and cohort placement rates implies that approximately 37% of the variance in cohort-level placement outcomes is explained by trainer quality, even after controlling for student-level characteristics. This finding challenges the prevalent assumption that placement outcomes are primarily determined by individual student attributes and suggests that institutional factors represent a significant, addressable lever for improvement.

The partial support for H5 is theoretically informative. While confidence level and adaptability were significant predictors of placement in bivariate analyses, their moderation effect on the technical skill–placement relationship was only marginally significant (interaction term  $p=0.047$  in ANN SHAP analysis), suggesting that psychological capital functions more as a main effect than as a pure moderator in this context. This nuance has not been previously reported in the Indian placement prediction literature.

The superiority of ANN over classical models (H6 confirmed) is consistent with the broader deep learning literature. However, the relatively modest margin over XGBoost (89.3% vs. 88.1%) suggests that the added complexity of deep learning may not always be warranted, particularly when interpretability is prioritised. For institutions with limited computational infrastructure, XGBoost represents a practical alternative with near-equivalent predictive performance and superior interpretability through SHAP values.

## VII. FINDINGS AND CONCLUSION

This study makes several distinct contributions to the educational analytics and placement prediction literature. First, it demonstrates that an ANN-based predictive framework achieves 89.3% classification accuracy on a real-world Indian IT training dataset, establishing a new benchmark for this research domain. Second, it reveals that mock interview performance, communication skills, and practical project exposure are the three strongest individual predictors of placement success, accounting collectively for over 60% of the predictive variance identified through SHAP analysis. Third, it provides the first empirical evidence, in this context, that trainer effectiveness is a significant, institution-level predictor of cohort placement rates independent of individual student characteristics. Fourth, K-Means clustering successfully segments students into three actionable employability tiers, enabling personalised, risk-stratified intervention strategies.

The Power BI dashboard operationalises these findings into a real-time decision-support tool accessible to institutional administrators, career counsellors, and trainers without requiring data science expertise. The dashboard's skill gap heatmap component specifically addresses the curriculum alignment problem identified in multiple prior studies by making industry-expectation gaps visually salient and quantitatively interpretable.

In conclusion, this research establishes that AI-driven predictive analytics, when applied to multi-source student data encompassing academic, behavioural, psychological, and institutional variables, can substantially improve the precision and personalisation of placement support interventions. The integrated framework proposed here—spanning data collection, ML/DL modelling, clustering, and Power BI visualisation—provides a replicable, scalable blueprint for training institutions seeking to systematically improve their placement outcomes.

## VIII. LIMITATIONS AND FUTURE RESEARCH SCOPE

Several limitations of the present study should be acknowledged. First, the sample is geographically constrained to Nagpur, India, limiting the generalisability of findings to other regional or national contexts where labour market conditions, recruiter expectations, and training ecosystems differ. Second, the cross-sectional design precludes causal inference; while the predictive models identify significant associations, longitudinal research is necessary to establish causal pathways between training interventions and placement outcomes. Third, self-reported questionnaire data is subject to social desirability bias, particularly for psychological variables such as confidence and adaptability, which may inflate their apparent predictive validity.

Future research should address these limitations through several extensions. Longitudinal panel designs tracking students from enrolment through placement and post-placement performance would enable causal modelling using approaches such as difference-in-differences or instrumental variables estimation. Multi-institutional, pan-Indian studies with standardised data collection protocols would enhance external validity. Incorporation of Natural Language Processing (NLP) applied to student interview transcripts and recruiter feedback could unlock rich unstructured data sources currently unexplored. Finally, the development of federated learning architectures would allow institutions to collaboratively train models without sharing sensitive student data, addressing privacy concerns that currently impede large-scale data integration.

## REFERENCES

- [1] M. Ruparel and P. Swaminarayan, "Student placement prediction using various machine learning techniques," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 3, pp. 2107–2113, 2024.
- [2] V. S. Agrawal and S. S. Kadam, "Predictive analysis of campus placement of student using machine learning algorithms," *Journal of IoT and Machine Learning*, vol. 1, no. 2, pp. 13–18, 2024.
- [3] M. Kumar, N. Walia, S. Bansal, G. Kumar, and K. Cengiz, "Predicting college students' placements based on academic performance using machine learning approaches," *International Journal of Modern Education and Computer Science*, vol. 15, no. 6, pp. 1–13, 2023.

- [4] G. M. Spandana and L. Pallavi, "Placement prediction system using machine learning," in Proc. 2023 2nd International Conference on Edge Computing and Applications (ICECAA), pp. 903–907, 2023.
- [5] P. Anusha, "Machine learning for student placement forecasting: An empirical study with ANN and classical classifiers," *International Journal of Human Computations and Intelligence*, vol. 2, no. 4, pp. 55–67, 2025.
- [6] M. Bora and R. Baruah, "Employability prediction of lateral entry engineering students: A deep learning based inductive reasoning and interpretable framework," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 1, pp. 455–468, 2024.
- [7] V. N. Rao and P. Dhanalakshmi, "Campus placement prediction using machine learning," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 10, no. 4, pp. 771–777, 2022.
- [8] K. Rai, "Students placement prediction using machine learning algorithms," *South Asia Journal of Multidisciplinary Studies*, vol. 8, no. 5, pp. 44–52, 2022.
- [9] "Student placement prediction," *International Journal of Information Technology and Computer Engineering*, vol. 13, no. 1, pp. 31–42, 2025.
- [10] S. Byagar, R. Patil, and J. Pawar, "Maximizing campus placement through machine learning," *Journal of Advanced Zoology*, vol. 45, Special Issue 4, pp. 211–220, 2024.
- [11] P. Thakar, A. Mehta, and Manisha, "Unified prediction model for employability in Indian higher education system," arXiv preprint arXiv:2407.17591, 2024.
- [12] P. Thakar, A. Mehta, and Manisha, "Cluster model for parsimonious selection of variables and enhancing students employability prediction," arXiv preprint arXiv:2407.16884, 2024.
- [13] S. K. Yadav, B. Bharadwaj, and S. Pal, "Mining education data to predict student retention: A comparative study," arXiv preprint arXiv:1203.2987, 2012.
- [14] M. I. R. Akib, F. B. Muhammed, U. Saha, M. F. K. Patwary, M. Anannya, and M. A. Hussein, "From code to career: Assessing competitive programmers for industry placement," arXiv preprint arXiv:2508.00772, 2025.
- [15] X. Gao, Y. Liu, and J. Wang, "Interpretable machine learning models for hospital readmission prediction," *IEEE Access*, vol. 11, pp. 45678–45689, 2023.
- [16] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Amsterdam: Elsevier, 2012.
- [17] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed. Cambridge: Morgan Kaufmann, 2016.
- [18] A. Ng and M. I. Jordan, "On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes," in *Advances in Neural Information Processing Systems*, vol. 14, pp. 841–848, 2001.
- [19] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794, 2016.
- [20] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.
- [21] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [22] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer, 1995.
- [23] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, vol. 30, pp. 4765–4774, 2017.
- [24] M. B. Zafar, I. Valera, M. G. Rodriguez, and K. P. Gummadi, "Fairness constraints: Mechanisms for fair classification," in Proc. 20th International Conference on Artificial Intelligence and Statistics (AISTATS), vol. 54, pp. 962–970, 2017.
- [25] R. Baker and K. Yacef, "The state of educational data mining in 2009: A review and future visions," *Journal of Educational Data Mining*, vol. 1, no. 1, pp. 3–17, 2009.