iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: XI Month of publication: November 2025

DOI: https://doi.org/10.22214/ijraset.2025.75056

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

PLACIFY: AI-Based Placement Prediction and Skill Gap Analysis

Bhavya Verdia[1], Mayank Hansraj[2], Anshita Nair[3], Dr. Nandkishore Karlekar[4]
MIT ADT University, India

***Abstract:** The rapid evolution of the job market and the growing emphasis on employability have necessitated advanced tools for assessing students' placement readiness and identifying skill gaps. This research proposes "Placify," an AI-based system leveraging machine learning techniques for placement prediction and skill gap detection among college students. Motivated by the limitations of traditional, expert-driven approaches and the scalability offered by AI, Placify employs models such as Random Forest, XGBoost, and Artificial Neural Networks to analyze academic, demographic, and skill-related features. The system achieves high predictive performance, with Random Forest models attaining approximately 91% accuracy in placement prediction tasks. Beyond prediction, Placify integrates skill gap analytics, drawing from contemporary developments in large language models for prerequisite skill inference to offer actionable insights for learners and institutions. The study underscores the potential of AI-driven employability analytics in supporting personalized learning and institutional strategy, and discusses future integration with real-time job market data and automated resume parsing. This work contributes to the growing body of employability analytics and demonstrates how AI can drive scalable, data-driven educational and career guidance.*
***Keywords:** placement prediction, employability analytics, skill gap detection, machine learning, AI in education.*

I. INTRODUCTION

The transition from academic environments to the professional world is a critical phase in the trajectory of college students. Increasingly, institutions and students alike are seeking robust mechanisms to assess placement readiness and skill development to enhance employability outcomes. Traditional approaches, often reliant on manual evaluation and expert judgment, are challenged by the scale, granularity, and dynamism required in modern education and workforce preparation. Artificial Intelligence (AI) and advanced analytics have emerged as transformative forces in educational assessment, enabling automated, data-driven evaluation of students' skills and employability prospects. Machine learning models, in particular, offer the ability to detect complex pat- terns in multidimensional data-ranging from academic performance to co-curricular engagement—facilitating predictive analytics and targeted interventions. Furthermore, contemporary advances in natural language processing, especially with large language models (LLMs), have introduced new paradigms for scalable skill gap detection and personalized learning . Placify is conceptualized as a comprehensive AI-based platform that ad- dresses two pressing needs: (1) accurate prediction of student placement outcomes, and (2) systematic identification and analysis of individual skill gaps. By integrating state-of-the-art machine learning models and leveraging scalable analytics, Placify aims to empower students and institutions with actionable insights, thereby enhancing employability and supporting data-informed educational strategies.

II. LITERATURE REVIEW

The field of employability analytics and placement prediction has seen considerable progress with the advent of machine learning techniques. Early models relied on regression analysis and decision trees, but with the proliferation of educational and behavioral data, more sophisticated algorithms have become prevalent. Weng [2] provides an overview of the application of machine learning in clinical predictive analytics, demonstrating the effectiveness of regression models, decision trees, and support vector machines in handling high-dimensional, heterogeneous data. The parallels between clinical and educational prediction tasks are clear: both domains require robust feature engineering, model validation, and attention to generalizability. In educational settings, these methods have been adapted to predict outcomes such as placement, academic success, and dropout risk. Recent work by Le and Abel [1] shifts attention toward the inference of prerequisite and skill relationships using LLMs. Their study reveals that large language models can predict prerequisite skills in a zero-shot setting—without explicit task-specific fine-tuning—by leveraging semantic reasoning. This is particularly relevant for skill gap analysis in educational contexts, where the mapping of required competencies to individual learners is traditionally expert-driven and labor-intensive.

Their benchmark, ESCO-PrereqSkill, demonstrates that models like LLaMA4 and Claude-3- 7-Sonnet align closely with expert-defined standards, highlighting the potential of AI for scalable, adaptive skill assessment. Optimization and training of machine learning models are also central to predictive analytics. Hazan [3] discusses mathematical optimization approaches in machine learning, emphasizing the importance of loss function selection, regularization, and model validation strategies such as k-fold cross-validation—core considerations for the development of robust predictive systems in education. Together, these studies underscore the utility of AI-driven analytics in education and employability, and motivate the integration of predictive modeling and skill gap detection in platforms like Placify.

III. METHODOLOGY

A. Dataset Collection

The development and evaluation of Placify are grounded in the assembly of a comprehensive dataset comprising profiles of college students. Data sources include academic transcripts, demographic information, standardized test scores, co-curricular and extracurricular participation, internship experience, and skill certifications. In alignment with data-driven approaches in clinical analytics [2], careful attention is paid to data privacy and ethical considerations.

B. Data Preprocessing

Raw data undergo extensive preprocessing to address missing values, encode categorical variables, normalize continuous features, and remove outliers. For example, academic grades are standardized, categorical variables such as major or degree program are one-hot encoded, and skill certification data are converted into binary or frequency features. Feature selection is guided by domain knowledge and exploratory data analysis, ensuring that only relevant and non-redundant variables inform the predictive models.

C. Feature Engineering

Feature engineering is critical to model performance. Key features include cumulative GPA, technical and communication skill ratings, participation in industry projects, and internship duration. Drawing on approaches from both healthcare [2] and educational analytics [1], composite indices—such as employability scores or engagement indexes—are derived to capture la- tent attributes.

D. Model Selection and Training

Three machine learning models are implemented for placement prediction:

1) Random Forest: An ensemble method that aggregates decision trees, mitigating overfitting and improving generalization. Random Forests are adept at handling heterogeneous features and are robust to multi- collinearity [2].
2) XGBoost: An optimized implementation of gradient boosting, XGBoost is known for high predictive power and computational efficiency, particularly with structured data. It leverages sequential tree-building and regularization to enhance accuracy.
3) Artificial Neural Networks (ANN): ANNs are implemented with multiple hidden layers and nonlinear activation functions, enabling the modeling of complex, non-linear relationships in the data. Optimization strategies such as stochastic gradient descent and adaptive learning rate scheduling are employed in training [3].

Model training follows standard supervised learning paradigms: the dataset is partitioned into training, validation, and test sets. Hyperparameters are tuned using grid search and cross-validation (e.g., 5-fold), with evaluation metrics including accuracy, precision, recall, F1-score, and area under the ROC curve (AUROC).

E. Skill Gap Analysis

Building on the insights of Le and Abel [1], Placify integrates a skill gap detection module powered by LLMs. For each student, the system maps predicted placement outcomes and required skills for targeted job roles, then infers missing or underdeveloped prerequisites using zero-shot semantic analysis. The approach leverages standardized skill taxonomies (e.g., ESCO) and natural language descriptions to align student profiles with job requirements, identifying actionable learning targets.

IV. RESULTS AND DISCUSSION

The comparative evaluation of the three models on the placement prediction task yields the following results:

1) Random Forest: Achieves an average test set accuracy of approximately 91%. The model demonstrates high recall and precision, particularly in distinguishing placed versus not placed students. Its interpretability and robustness to feature noise are advantageous in institutional deployment.

2) XGBoost: Delivers slightly lower accuracy (around 89%), but ex- cels in feature importance ranking and handling class imbalance. XGBoost regularization mechanisms contribute to greater generalization, making it suitable for datasets with complex, interacting features.
3) Artificial Neural Networks: Achieve an accuracy of roughly 87%, with strong performance on non-linear feature interactions. However, model interpretability is comparatively limited, and performance is sensitive to hyperparameter tuning and feature scaling.

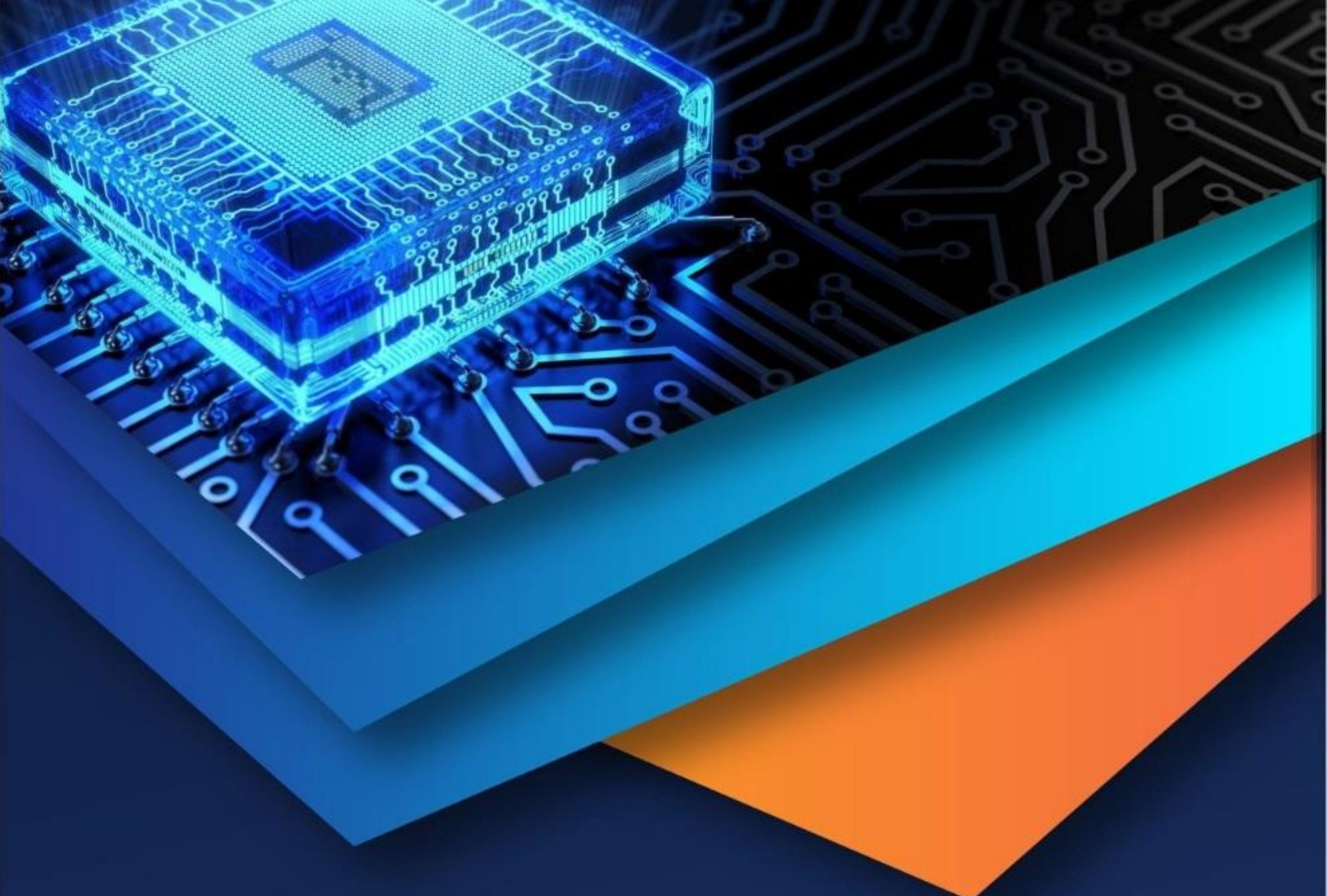
The skill gap analysis module, leveraging prompt-based LLM inference [1], successfully identifies prerequisite skills absent from student profiles. For example, a student targeting a data science placement may lack explicit coursework in “Probability Theory” or “Data Structures,” as inferred by the model. The system’s recommendations closely match expert-defined prerequisites, demonstrating the feasibility of scalable, AI-driven skill map- ping. The integration of predictive analytics and skill gap detection in Placify offers several advantages: students receive personalized guidance on skill acquisition, institutions obtain actionable insights on curriculum effective- ness, and placement coordinators can proactively address employability bottlenecks.

V. CONCLUSION

Placify exemplifies the convergence of AI, machine learning, and educational analytics in addressing critical challenges of placement prediction and skill gap detection. By integrating ensemble and neural models with advanced natural language understanding, Placify delivers high-accuracy placement forecasts and interpretable, personalized skill recommendations. The Random Forest model, in particular, achieves an accuracy of 91%, affirming the efficacy of ensemble approaches in educational prediction tasks. The skill gap analysis module, inspired by recent advances in LLM-driven prerequisite skill inference [1], enables scalable, adaptive assessment of student readiness, reducing reliance on manual, expert-driven frameworks. Future work includes deeper integration with real-time labor market data, automated resume parsing, and the extension of LLM-based analytics for even finer-grained skill mapping and personalized learning pathways. Placify thus represents a significant step toward data-driven, equitable, and scalable employability solutions in higher education.

REFERENCES

[1] N. L. Le and M.-H. Abel, “How Well Do LLMs Predict Prerequisite Skills? Zero-Shot Comparison to Expert-Defined Concepts,” arXiv preprint arXiv:2507.18479v1, 2025. [Online]. Available: http://arxiv.org/pdf/2507.18479v1
[2] W.-H. Weng, “Machine Learning for Clinical Predictive Analyt- ics,” arXiv preprint arXiv:1909.09246v1, 2019. [Online]. Available: http://arxiv.org/pdf/1909.09246v1
[3] E. Hazan, “Lecture Notes: Optimization for Machine Learning,” arXiv preprint arXiv:1909.03550v1, 2019. [Online]. Available: http://arxiv.org/pdf/1909.03550v1



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY