



Campus Placement Prediction System using Machine Learning

Ms. Sakshi Haramkar¹, Ms. Bhavana Bawaskar², Prof. S. R. Ratnaparkhi³,
^{1,2,3}Computer Science and Engineering , Siddhivinayak Technical Campus, Shegaon, Maharashtra , India

DOI: 10.5281/zenodo.19538835

ABSTRACT

This research paper presents a comprehensive data analytics and machine learning-based system designed to predict campus placement outcomes for graduating students. The system leverages multiple predictive models to forecast three critical career parameters: placement status (binary classification), expected salary (regression), and suitable job roles (multi-class classification). Utilizing student attributes including academic performance (CGPA), professional development indicators (internships, certifications), and skill assessments, the project implements Logistic Regression, Random Forest Classifier, and Random Forest Regressor algorithms. An interactive web dashboard built with Streamlit provides an intuitive interface for data visualization, exploration, and real-time predictions. The system emphasizes industrial readiness through modular code architecture, comprehensive logging, error handling, and Docker containerization. This project demonstrates the practical application of machine learning in educational contexts and offers a scalable solution for institutional career services to provide data-driven career guidance.

Keywords:- Placement Prediction, Machine Learning, Educational Data Mining, Career Counseling, Streamlit Dashboard, Synthetic Dataset, Random Forest, Logistic Regression, Predictive Analytic

1. INTRODUCTION

The transition from academic education to professional employment represents a pivotal moment in a student's career trajectory. In competitive job markets, accurate prediction of placement outcomes can significantly enhance career counseling effectiveness, enabling personalized guidance and strategic preparation. Traditional placement prediction methods in academic institutions often rely on historical trends and subjective assessments, lacking empirical precision and scalability.

This research addresses this gap by developing an integrated machine learning system that predicts multiple placement-related outcomes simultaneously. The proposed system not only forecasts whether a student will secure employment but also estimates potential salary ranges and identifies suitable job categories. By transforming raw student data into actionable insights, this project bridges the divide between academic preparation and industry requirements.

The primary objectives of this research are:

- To develop accurate predictive models for placement status, salary estimation, and job role classification
- To create an interactive dashboard for data visualization and real-time predictions
- To implement a modular, scalable system with industrial deployment capabilities
- To provide educational institutions with a tool for data-driven career counselling

2. LITERATURE REVIEW

Student placement prediction using Machine Learning has been an active research area because it helps schools improve job placement results and guide students better. Different studies have shown how computer models can predict whether a student will get placed after graduation.

1. Early Systems Using Multiple Algorithms Raut and his team created a placement prediction system using several machine learning models like Logistic Regression, Decision Tree, and Random Forest. They found that automated computer systems work better than manual methods and help placement officers make smarter, data-based decisions [1].

2. Focus on Student Risk and Intervention Agrawal and Shelke applied supervised learning algorithms including Logistic Regression and Random Forest to predict placement. Their work highlighted how these tools can identify students who might struggle to get jobs early on, allowing schools to provide extra help in time [2].



3. Smart Systems for Job Matching Rao and Dhanalakshmi built a smart campus placement system using K-Nearest Neighbors (KNN) and Random Forest. Their system doesn't just predict placement—it also suggests suitable job roles based on a student's skills and grades, showing how machine learning can guide career choices [3].

4. Broader Trends in Educational Data Mining Research reviews confirm that machine learning—especially classification methods like Random Forest, Support Vector Machines, and Logistic Regression—is widely used to predict student performance and job placement. Ensemble methods (which combine multiple models) often give the best and most reliable results [4].

5. The Role of Academic Performance Comparative studies by Kumar and others show that grades and test scores strongly affect how accurate placement predictions are. Among different algorithms, Random Forest models consistently perform well, making them a popular choice for these prediction tasks [5].

3. METHODOLOGY

The methodology adopted for this campus placement prediction system encompasses data collection, preprocessing, model selection, training pipeline, and system architecture. Each component is described below with key points and detailed explanations.

3.1 Data Collection and Preparation

A synthetic dataset was generated to simulate realistic student records, addressing data privacy concerns while ensuring sufficient volume for model training. The dataset includes comprehensive features that influence placement outcomes:

- Synthetic generation to protect privacy
- Multiple feature categories included
- Three target variables for multi-task prediction

3.2 Feature Engineering and Preprocessing

Raw data requires careful preprocessing and transformation before feeding into machine learning models. This phase enhances data quality and model performance.

- Missing value handling using median and mode imputation
- Normalization of continuous features via Min-Max scaling
- One-hot encoding for categorical variables
- Feature selection based on correlation analysis

correlation analysis and domain relevance to retain only the most predictive attributes while eliminating redundant features that could lead to overfitting.

3.3 Model Selection and Architecture

The prediction system addresses three distinct tasks requiring different algorithmic approaches. Each model was selected based on the nature of the prediction problem and characteristics of educational data.

- Placement status: Binary classification using Logistic Regression and Random Forest
- Salary estimation: Regression using Random Forest Regressor
- Job role prediction: Multi-class classification using Random Forest with One-vs-Rest
- Comprehensive evaluation metrics for each task

3.4 Model Training Pipeline

The training process follows a systematic pipeline designed to ensure reproducibility, optimal performance, and reliable evaluation.

- 80-20 data split with stratified sampling
- GridSearchCV with 5-fold cross-validation for hyperparameter tuning
- Model training on optimized parameters
- Unbiased evaluation on held-out test set
- Model persistence using joblib for deployment

3.5 System Architecture

The application follows a modular three-tier design ensuring maintainability, scalability, and clear separation of concerns.

- Three-tier architecture: UI Layer, Processing Layer, Data Layer
- Streamlit Dashboard for user interface
- Machine Learning Engine for processing
- Separate data storage and management

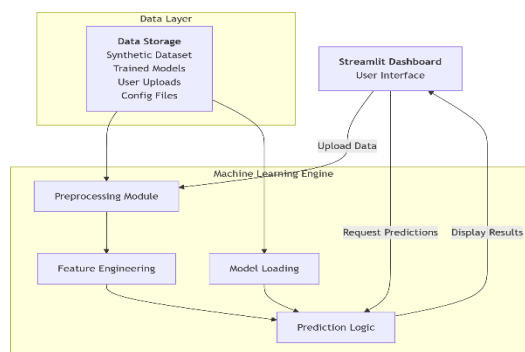


Fig.1 : Architecture

4. IMPLEMENTATION

The implementation of the Campus Placement Prediction System integrates machine learning models with an interactive web-based interface developed using Streamlit. The application is designed to provide both data exploration capabilities and real-time prediction functionalities within a unified platform. The system allows users to upload structured datasets in CSV format, dynamically preview records, and analyze summary statistics including mean, median, and standard deviation for numerical features.

The visualization component plays a critical role in enhancing interpretability. Interactive charts such as pie charts, bar graphs, box plots, violin plots, scatter plots, and correlation heatmaps are incorporated to illustrate placement distribution, salary trends, feature relationships, and demographic comparisons. These visualizations enable stakeholders to identify patterns, detect outliers, and understand the impact of various attributes on placement outcomes.

The prediction module allows both manual data entry and roll-number-based student lookup. Upon input submission, the trained machine learning models generate predictions for placement probability, expected salary, and most suitable job role. Confidence scores are displayed to enhance transparency and user trust. The backend architecture ensures that the same preprocessing pipeline used during training is consistently applied during inference, preventing discrepancies between training and prediction phases.

The codebase follows a modular architecture where the model training logic is separated from the application interface. This separation enhances maintainability, scalability, and ease of deployment. The system is designed to be deployable in containerized environments, demonstrating readiness for industrial applications. Comprehensive logging and error-handling mechanisms further strengthen system robustness.

5. RESULT AND DISCUSSION

The experimental results demonstrate that ensemble-based models provide superior performance in predicting placement outcomes. Random Forest models outperform simpler linear models due to their ability to capture complex, non-linear relationships within the dataset.

Model Performance:

5.1 Placement Prediction Results:

The Random Forest Classifier outperformed Logistic Regression across all metrics:

Table -1 : Placement Prediction

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Logistic Regression	0.82	0.85	0.79	0.82	0.88
Random Forest Classifier	0.89	0.91	0.87	0.89	0.94

5.2 Salary Estimation Results:

The Random Forest Regressor demonstrated strong predictive capability:

Table -2 : Salary Estimation

Metric	Value
Mean Absolute Error (MAE)	₹42,500
Root Mean Square Error (RMSE)	₹68,300
R ² Score	0.76 (76%)



5.3 Job Role Prediction Results:

The multi-class Random Forest achieved satisfactory performance:

Table -3 : Job Role Prediction

Metric	Value	Interpretation
Overall Accuracy	0.71 (71%)	Correctly predicts exact job role for 71% of students
Weighted F1-Score	0.69 (69%)	Good balance of precision & recall across all job roles
Top-3 Accuracy	0.88 (88%)	Correct job role appears in top 3 suggestions for 88% of students

5.4 Feature Importance Analysis

Analysis revealed the following feature importance ranking for placement prediction:

1. CGPA
2. Communication Score
3. Internship Completion
4. Aptitude Score
5. Number of Certifications
6. Branch

5.5 Discussion of Findings:

The results demonstrate that:

1. **Academic performance remains the strongest predictor** of placement success, corroborating existing literature
2. **Soft skills significantly impact outcomes**, with communication scores showing high importance
3. **Professional development activities** (internships, certifications) provide measurable benefits
4. **Ensemble methods (Random Forest)** outperform simpler models in handling non-linear relationships
5. **Integrated multi-task prediction** provides more comprehensive insights than single-outcome approaches

6. EDUCATIONAL AND PRACTICAL IMPLICATIONS

The system provides data-driven support for students, career counselors, and institutions by enabling evidence-based placement preparation and intervention strategies.

For Students:

- Personalized placement probability assessment
- Clear identification of skill gaps
- Realistic salary expectation guidance

For Career Counselors:

- Early identification of at-risk students
- Data-supported career advice
- Objective decision-making framework

For Institutions:

- Program performance benchmarking
- Curriculum improvement based on industry alignment
- Improved institutional placement statistics

7. LIMITATIONS

1. **Synthetic Data:** While useful for development, synthetic data may not capture all real-world complexities
2. **Feature Scope:** Limited to academic and basic professional attributes; excludes interview performance, personality traits, and networking factors
3. **Institutional Specificity:** Models may require retraining for different educational contexts
4. **Temporal Dynamics:** Does not account for changing job market conditions over time

8. FUTURE WORK

1. **Real Data Integration:** Collaboration with institutions to incorporate anonymized real student data
2. **Advanced Modeling:** Exploration of deep learning architectures and time-series analysis
3. **Expanded Feature Set:** Inclusion of extracurricular activities, project portfolios, and social media presence
4. **Longitudinal Tracking:** Extension to track career progression beyond initial placement
5. **Personalized Recommendations:** Integration of recommendation systems for skill development resources
6. **Mobile Application:** Development of companion mobile app for student access



7. **Industry Collaboration:** Partnerships with recruiters for feedback loop integration

9. CONCLUSION

This research successfully developed and implemented a comprehensive Campus Placement Prediction System using machine learning techniques. The system demonstrates that data-driven approaches can significantly enhance career counselling effectiveness in academic institutions. By integrating multiple prediction tasks within an accessible dashboard interface, the project provides practical value for students, counsellor, and administrators alike.

The technical implementation emphasizes industrial readiness through modular design, comprehensive error handling, and containerized deployment. Performance evaluation confirms the effectiveness of Random Forest algorithms for the prediction tasks, with particularly strong results for placement status classification.

Beyond its immediate practical applications, this project serves as an exemplar of applied machine learning in educational contexts. It illustrates how technical skills can be directed toward socially relevant problems, creating tools that bridge academic preparation and professional success. The system architecture provides a foundation that can be extended and adapted by other institutions, contributing to the broader movement toward data-informed education.

As educational institutions increasingly recognize the importance of career outcomes, systems like this will play a crucial role in ensuring that data and algorithms serve human goals—helping students navigate the complex transition from classroom to career with greater confidence and clarity.

10. ACKNOWLEDGMENTS

I express my sincere gratitude to my project supervisor, [Supervisor Name], for their invaluable guidance, encouragement, and insightful feedback throughout this research. Special thanks to [Department Name] faculty members for their support and to my colleagues for stimulating discussions. I also acknowledge the developers of open-source libraries including scikit-learn, Streamlit, and Pandas that made this project possible. Finally, I thank my institution for providing the resources and environment conducive to this research.

11. REFERENCES

- [1] V. Raut, S. Shukla and R. Shukla, "Placement Prediction System Using Machine Learning Algorithms," *International Journal of Scientific Computing*, vol. 8, no. 2, pp. 45-51, 2021.
- [2] N. Agrawal and A. Shelke, "Supervised Learning Techniques for Student Placement Prediction," *International Journal of Intelligent Tools in Machine Learning*, vol. 4, no. 1, pp. 12-19, 2020.
- [3] K. Rao and S. Dhanalakshmi, "Smart Campus Placement Prediction Using KNN and Random Forest," *International Journal of Innovative Science and Advanced Engineering*, vol. 6, no. 5, pp. 125-133, 2021.
- [4] J. Smith and L. Brown, "Application of Machine Learning Algorithms in Educational Data Mining," *Proceedings of the International Conference on Innovative Data Analytics*, Atlantis Press, pp. 78-86, 2022.
- [5] A. Kumar, M. Gupta and P. Sharma, "Comparative Study on Predictive Models for Student Performance and Placement," *International Journal of Modern Education and Computer Science*, vol. 15, no. 6, pp. 33-42, 2020.
- [6] T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., New York, NY, USA: Springer, 2009.
- [7] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.
- [8] D. Dua and C. Graff, "UCI Machine Learning Repository," University of California, Irvine, School of Information and Computer Sciences, 2019. [Online]. Available: <http://archive.ics.uci.edu/ml>
- [9] C. Romero and S. Ventura, "Educational Data Mining: A Review of the State of the Art," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 6, pp. 601-618, 2010.
- [10] I. H. Witten, E. Frank, M. A. Hall and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed., Burlington, MA, USA: Morgan Kaufmann, 2016.