Campus Recruitment Prediction Project

1. Executive Summary

This project aimed to develop a machine learning model to predict campus placement outcomes for students. Using a dataset of 10,000 student records with 15 features, we implemented and compared three classification models: Random Forest, Decision Tree, and K-Nearest Neighbours. The Random Forest model achieved the highest accuracy of 93.7%, demonstrating strong potential for predicting student placement success.

2. Introduction

2.1 Background

Campus recruitment is a critical process for both educational institutions and students. Predicting placement outcomes can help in tailoring career development strategies and improving overall placement rates.

2.2 Objectives

- To develop machine learning models for predicting student placement status
- To identify key factors influencing placement outcomes
- To compare the performance of different classification algorithms

3. Methodology

3.1 Dataset Description

- Source: Reputable academic resource
- Size: 10,000 entries, 15 features
- Target Variable: PlacementStatus (Binary: 1 for placed, 0 for not placed)
- Key Features: CGPA, 12th and 10th grade scores, project experience, internships, hackathon participation, technical skills

3.2 Data Preprocessing

- Conducted Exploratory Data Analysis (EDA)
- Applied label encoding for categorical variables
- Performed data splitting (70% training, 30% testing)
- Implemented feature scaling

3.3 Model Development

Three models were developed and tuned:

1. Random Forest
2. Decision Tree
3. K-Nearest Neighbours (KNN)

Hyperparameter optimization was performed using GridSearchCV.

4. Results and Analysis

4.1 Model Performance

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	93.7%	High	High	High
Decision Tree	< 93.7%	Good	Good	Good
KNN	86.9%	Moderate	Moderate	Moderate

4.2 Feature Importance

Top influencing factors:
- Positive: Skills, Internship experience, 10th grade percentage, Workshop participation
- Negative: Academic backlogs

4.3 Ensemble Method

A voting classifier was implemented but did not outperform the individual Random Forest model.

5. Discussion

5.1 Model Interpretation

The Random Forest model demonstrated superior performance, likely due to its ability to handle complex interactions between features and resistance to overfitting.

5.2 Key Insights

- Technical skills and internship experience are crucial for placement success
- Academic consistency (lack of backlogs) significantly impacts placement chances
- The 10th grade percentage remains a relevant factor in placement outcomes

6. Conclusion and Recommendations

6.1 Conclusion

The project successfully developed a predictive model for campus placements, with the Random Forest classifier showing the best performance. The model and insights gained can be valuable tools for educational institutions in guiding student career development.

6.2 Recommendations

1. Implement the Random Forest model as a predictive tool for placement offices
2. Focus on enhancing students' technical skills and encouraging internship participation
3. Provide additional support to students with academic backlogs

6.3 Future Work

- Incorporate additional features such as extracurricular activities and soft skills
- Explore advanced ensemble techniques to potentially improve model performance
- Conduct longitudinal studies to validate the model's predictive power over time