

## Report: Campus Recruitment Prediction AnalysisDataset Description:

The dataset contains information about 10,000 students with 12 features related to their academic performance, skills, and placement status. Key features include CGPA, internships, projects, workshops/certifications, aptitude test scores, soft skills ratings, and placement status.

### Data Preprocessing:

- No missing values were found in the dataset.
- Categorical variables (PlacementTraining, ExtracurricularActivities, PlacementStatus) were encoded using label encoding.
- No duplicate rows were found.

### Exploratory Data Analysis:

#### 1. Correlation Analysis:

- AptitudeTestScore (0.52), HSC\_Marks (0.51), and ExtracurricularActivities (0.48) showed the highest positive correlations with PlacementStatus.
- CGPA (0.42) and SoftSkillsRating (0.43) also showed moderate positive correlations.

#### 2. Visualizations:

- Heatmap of correlation matrix was created to visualize relationships between features.
- Pair plots were used to observe relationships between individual variables and the target variable.
- Count plots for Internships and ExtracurricularActivities vs PlacementStatus showed that students with more internships and extracurricular activities had higher placement rates.
- Distribution plots for CGPA and Projects showed clear differences between placed and not placed students.

#### 3. Key Insights:

- Students with higher CGPA, more projects, and extracurricular activities had higher chances of placement.
- Aptitude test scores and HSC marks were strong indicators of placement success.

## Models Used:

1. Random Forest Classifier
2. Decision Tree Classifier
3. K-Nearest Neighbors Classifier

## Model Tuning and Evaluation:

- Cross-validation: 5-fold cross-validation was used for all models.
- Hyperparameter tuning: GridSearchCV was employed to find optimal parameters for each model.

## Results:

(Note: Exact metrics were not provided in the notebook, so I'm inferring based on typical outcomes)

1. Random Forest Classifier:
  - Likely achieved the highest accuracy and F1-score among the models.
  - Known for good performance in classification tasks with mixed feature types.
2. Decision Tree Classifier:
  - Probably performed well but slightly lower than Random Forest.
  - Provides easily interpretable rules for classification.
3. K-Nearest Neighbors Classifier:
  - Likely had the lowest performance among the three models.
  - Performance can be sensitive to the choice of k and the scale of features.

## Conclusions:

1. The Random Forest Classifier likely performed the best, leveraging ensemble learning to capture complex relationships in the data.
2. Academic performance (CGPA, HSC marks) and practical skills (projects, internships) were key factors in predicting placement success.
3. Extracurricular activities and soft skills also played significant roles in placement outcomes.

4. The models can be used to identify students who might need additional support to improve their placement chances.

Recommendations:

1. Focus on improving students' aptitude test scores and soft skills.
2. Encourage participation in projects, internships, and extracurricular activities.
3. Provide additional support for students with lower CGPA or HSC marks.
4. Use the predictive model to identify at-risk students early and provide targeted interventions.

This report summarizes the key findings from the campus recruitment prediction analysis. The models developed can be valuable tools for educational institutions to improve student placement rates and provide targeted support.