Placement Prediction Project Report

1. Introduction

This project aims to analyze student placement data and develop a machine learning model to predict whether a student will be placed based on academic and non-academic factors. The dataset used contains student records, including academic scores, employability test scores, and placement status.

2. Data Overview

The dataset consists of multiple features such as:

1. **Academic Scores:**SSC percentage, HSC percentage, Degree percentage
2. **Work Experience:** Whether the student has prior work experience
3. **Specialization:**Field of study
4. **Employability Test Scores**:Scores on tests and assessments for job readiness
5. **Placement Status:**Target variable about whether a student will be placed

3. Exploratory Data Analysis (EDA)

EDA was undertaken to gain insight into the dataset. The steps included:

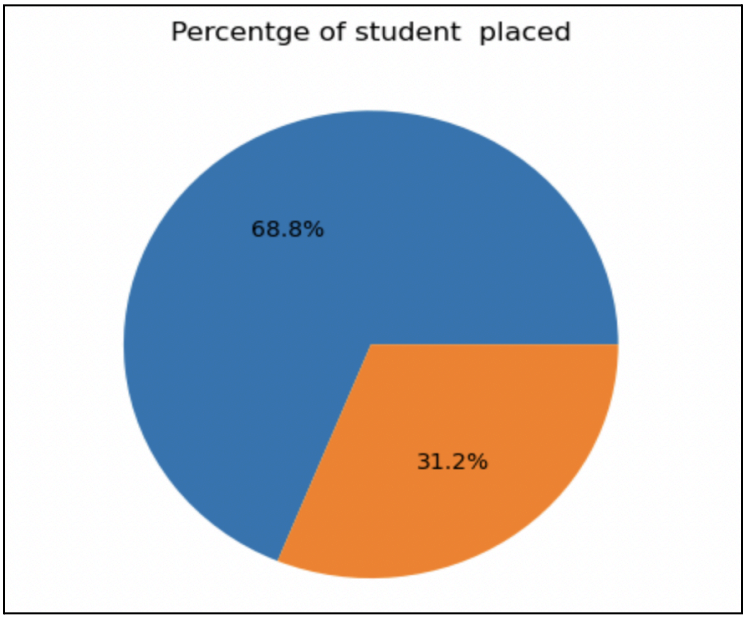
- Checking the level of missing values, handling inconsistencies
- Analysis of distribution of key features
- Plotting a graph between marks secured in the exam, employability scores, and the placement status

Main Findings

- Greater the academic scores, greater is the possibility of being placed
- Greater work experience has increased chances of being placed
- Specific specializations show greater percentages of placement

3.1 Distribution of Placement Status

- **Plotted Pie Chart**: Shows the proportion of placed vs. not placed students.



- **Observation**: Majority of students are placed, but a significant percentage remains unplaced.

3.2 Academic Performance and Placement

- **Scatter Plots & Box Plots** were used to analyze:
    - 10th %, 12th %, Degree %, and MBA % vs. Placement Status.
    - Higher scores generally increase placement probability, but some students with good scores remain unplaced.

3.3 Effect of Work Experience

- **Bar Plot** shows that students with work experience have a higher placement rate.

3.4 Specialization & Placement

- **MBA Specialization** (HR vs. Marketing/Finance) influences placement rates.

4. Data Preprocessing

During preprocessing for the dataset, for model training purposes, the following were carried out:

1. **Handling Missing Values:** Missing values have been spotted and filled up appropriately.
2. **Encoding Categorical Variables:**Converting categorical variables such as specialization into numerical format.
3. **Feature Scaling:**Standardizing numerical values where necessary.
4. **Splitting Data:**The dataset was split into training and testing sets (e.g., 80%-20%).

5. Training Machine Learning Models

Multiple machine learning models were trained to predict placement status:

- Logistic Regression
- Decision Tree Classifier
- Random Forest Classifier
- Support Vector Machine (SVM)
- K-Nearest Neighbors (KNN)

Model Evaluation:

The performance of each model was assessed using:

- **Accuracy Score:** Measures the percentage of correct predictions.
- **Confusion Matrix:** Graphic illustration of the number of correct and incorrect predictions.
- **Precision & Recall:** Assess how correctly the model classifies.
- **F1 Score:** A balanced metric between precision and recall.

Model Performance Metrics

Model	Accuray	Precision	Recall	F1-Score
Logistic Regression	84%	0.85	0.94	0.89

Decision Tree	88%	0.91	0.94	0.92
Random Forest	81%	0.83	0.94	0.88
SVM	77%	0.78	0.94	0.85
KNN	77%	0.78	0.94	0.85

Best Performing Model:
Decision Tree: The model with the highest accuracy (88%).
Feature importance analysis showed strong predictability power by employability test scores and degree percentage.

6. Conclusion & Findings

- For higher SSC, HSC and Degree percentages better placement opportunities occur.
- Strong placement results show that work experience improves significantly
- Employability test scores represent an important predictability factor towards the placement.
- Machine learning models, mostly ensemble methods in this case-Random Forest-classify as effective in predicting their output.

7. Recommendations & Future Work

- Gather extra features like internship experience, projects, and extracurricular activities to enhance the accuracy of the model.
- Deep learning techniques to improve predictive performance
- Deploy the model on a cloud platform (for example, AWS) for real-time placement prediction.

This report provides an end-to-end overview of the placement prediction project from data analysis to model evaluation and offers valuable insights into factors that influence student placements.