

Student Placement Prediction Using Machine Learning

Abstract

This project focuses on predicting student placement outcomes using machine learning techniques. By analyzing academic and demographic data, the study aims to identify patterns influencing placements. Data preprocessing, exploratory analysis, and classification models were applied. Logistic Regression and Random Forest algorithms were used, and their performance was evaluated.

Introduction

Campus placement is an important indicator of a student's academic and professional readiness. Machine learning provides effective tools to analyze student data and predict placement outcomes.

Objectives

- Analyze student placement data
- Clean and preprocess data
- Train machine learning models
- Evaluate model performance

Dataset Description

The dataset consists of student academic records and placement status. Each row represents a student, and the target variable indicates whether the student was placed.

Methodology

The methodology includes data loading, data cleaning, handling missing values, encoding categorical features, splitting data into training and testing sets, and training machine learning models.

Model Implementation

Logistic Regression was used as a baseline model, while Random Forest was used to improve performance. Random Forest achieved higher accuracy due to its ability to capture non-linear relationships.

Results and Discussion

Logistic Regression achieved 50% accuracy, while Random Forest achieved 100% accuracy. The higher accuracy may indicate overfitting, hence results were interpreted carefully.

Conclusion

The project successfully demonstrates a complete machine learning pipeline for placement prediction. Model comparison highlights the importance of preprocessing and algorithm selection.

Future Scope

- Apply cross-validation
- Perform hyperparameter tuning
- Deploy model as a web application

References

1. Scikit-learn Documentation
2. Pandas Documentation