Report: IDENTIFYING TRENDS AND PATTERNS IN CAMPUS PLACEMENT PREDICTION

Date: 27/08/2023
Team: TARS
Manaswi B Patil (lead)
Rohit Thakur

ABSTRACT:

The Campus Placement process is a critical aspect of higher education, as it bridges the gap between academic and the corporate world. Predicting Campus Placements can offer valuable insights to both Students and Institutions, aiding in better decision-making and resource allocation. This project aims to identify trends and patterns in Campus Placement prediction using data analysis and machine learning techniques. By analyzing historical placement data and student information, this study explores the factors that significantly influence placement outcomes and develops a predictive model.

INTRODUCTION:

The objective of this project is to uncover valuable insights from campus placement data and student attributes, ultimately enabling accurate placement predictions. This information can assist students in making informed career choices and help institutions improve their placement strategies.

METHODOLOGY:

- Data Collection: The project begins by gathering the comprehensive dataset related to Campus Placements. This dataset includes information about students' academic performance, technical skills, professional skills and personal details. The data also encompass placement outcomes such as whether a student was placed or not.

- Data Preprocessing: Before analysis, the data undergoes preprocessing to handle missing values, outliers, and irrelevant features. Categorical variables are encoded, and continuous variables are scaled appropriately. Also new features are added by doing some feature engineering like the domain interest and average of scores. This ensures that the data is suitable for analysis and modelling.

- Exploratory Data Analysis: EDA involves visualizing the data to identify trends, patterns, and correlations. Heatmaps, scatter plots, and histograms help uncover relationships between attributes and outcomes. EDA helps pinpoint potential factors influencing placements, such as academic performance, and specialization.

- Feature Selection: Through techniques like correlation analysis and feature importance from machine learning algorithms, the most influential features are selected. This helps build a more focused and effective predictive model.

- Modeling and Evaluation: Different Machine Learning algorithms. such as Decision trees, Random Forest Classifier, Support Vector Classifier, K-Nearest Neighbours , are applied to create a predictive model. Random Forest Classifier turned out to be the best fit. The model is trained on a subset of data and evaluated using appropriate metrics like accuracy score, classification report. Clustering algorithms are also applied on data. Kmodes clustering is used to cluster students based on their skill sets. And K-means clustering is applied on the numerical columns to group students based on their scores, skills and interests.

- Model Deployment: Deployment of the prediction and clustering models in the IBM Cloud through IBM Watson Studio.

RESULTS:

- The EDA phase revealed that CGPA and the skills, particularly technical skills, strongly correlates with placement outcomes.
- The Verbal and Logical Reasoning scores had very weak positive correlation with the placement outcomes.
- Specific specialization subjects such as, Data analytics, Quality Assurance and Artificial Intelligence are associated with highest placement rates.
- The developed machine learning model, based on the selected features, exhibited promising predictive accuracy.
- The clustering model rightly segregates the student based on certain features and also displays which cluster the Student belongs.

CONCLUSION:

The project successfully identified significant trends and patterns in campus placement prediction. The insights gained can guide students in enhancing their employability by focusing on academic performance and gaining relevant experience. Institutions can utilize the findings to tailor their placement strategies, allocate resources efficiently, and enhance collaboration with industries.

FUTURE DIRECTIONS:

Future enhancements to this project include:

- Incorporating more diverse data sources, such as alumni feedback and industry trends, to improve model accuracy.

- Exploring advanced machine learning techniques, such as ensemble methods, neural networks for even better predictions.

In conclusion, this project provides valuable insights into campus placement prediction trends and patterns. By leveraging data analysis and machine learning, students and institutions can make informed decisions that positively impact placement outcomes.