Engineering Student Placements

Group 4

Members:

Akif Eltahir
Regina Enrile
Connor Farrell
Mohamed Metwalli
Nicole Young

Goals

Build a Machine Learning model to predict whether a student will be selected for an internship based on several factors mentioned in the following slides.

Reason For Selecting This Topic

Students are in the beginning of their career and it is a challenging time for them to gain the experience needed to put them in the workforce. Understanding the data available from career services in colleges and universities will help these organizations in supporting the students to land in their dream job. Our success in developing such a model will enhance student experience and provide each one with the right guidance to start a career of their choice.

# Description Of Tools Used to Create Final Dashboard

We have chosen to use Tableau to create the final dashboard. This is because Tableau is a powerful interactive visualization tool that helps simplify the dataset into understandable format when presenting.

The clean dataset was loaded into Tableau as a CSV and three visualizations were created based on CGPA, Gender and Stream of the students. This shows us which criterias were mostly required for the student to be placed for an internship or not.

Description of Interactive Elements

1. Our first interactive element is a scatter plot that shows the amount of internship placements based on the applicant's Cumulative Grade Point Average

2. Our second graph shows the amount of applicants that were either placed or not solely by their gender

3. The third visualization shows the number of placements by an applicant's field of study, and then split by gender
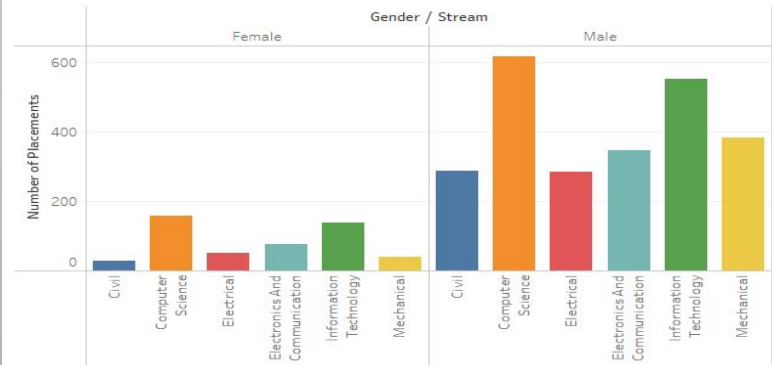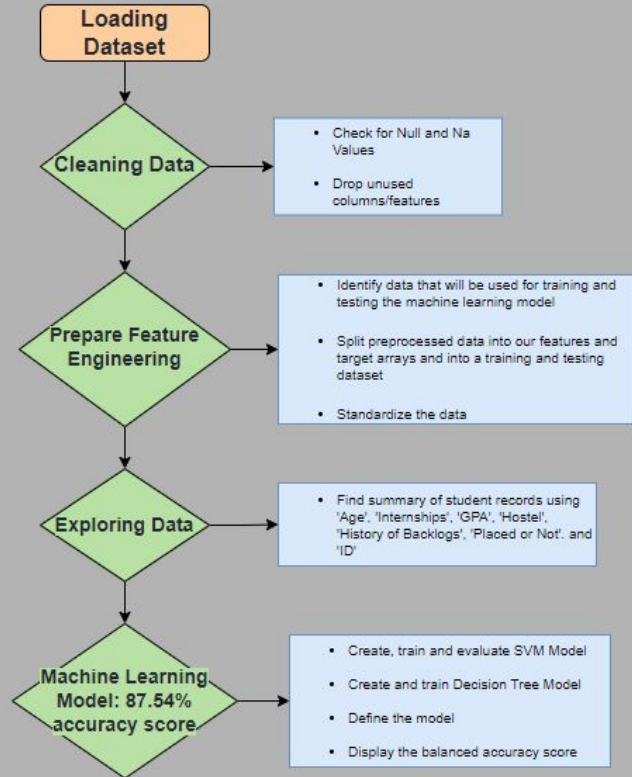

4
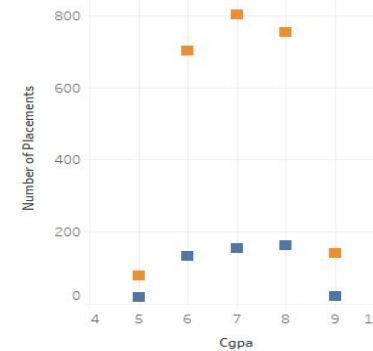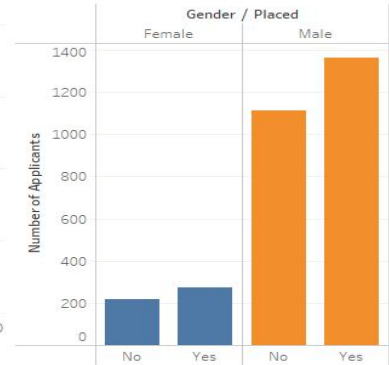
Data Source

The data was sourced from Kaggle from an Engineering Placement dataset which features the following relevant factors:

- Age
- Gender
- Field of Study
- Past Internships
- Cumulative Grade Point Average (CGPA)
- Dwelling Provided
- History of Backlogs

Machine Learning Model

Machine Learning Model Type: Classification model

Utilizing the features mentioned from the dataset, we are going to train a classification model - a subcategory of Supervised Machine Learning – to determine the likelihood of acceptance for future applicants. The following are the steps we will take for data preprocessing.

6

Database Integration

- Create data tables using PostgresSQL
- Join Tables: Student_Information & Student_Predictors (see ERD below)
- Database Host: AWS RDS Service
- Database Server: capstone-project