Presenter: Nicole

Good evening everyone, we are group 4 and our project is called Engineering Student Placements. I’ll first start off by introducing our 5 members, we have Akif, Regina, Connor, Mohamed and my name is Nicole. And we’re excited to show what we have been working on for the past couple of weeks. And we’re excited to show what we have been working on for the past couple of weeks.

New slide

Presenter: Nicole

First off, I will start by speaking on our Goals and the Reason For Selecting This Topic

**Goals:**

Our goal was to 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:**

The reason for selecting this topic is because students are in the beginning of their career, and I’m sure some of you can relate, it’s a challenging time for them to gain the experiences 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. It will also give the students the added incentive to do well in their remaining exams when they see what is required to be placed for an internship. Our success in developing such a model will enhance student experiences and provide each one with the right guidance to start a career of their choice.  
New slide

Presenter: Nicole

**Data Source**

The dataset we used was sourced from Kaggle from an Engineering Placement dataset. The dataset was fairly clean so there wasn’t much work needed to clean it up. But we all agreed that the fields included on the dataset were all relevant factors when being placed for an internship. And it features relevant factors like:

* Age
* Gender
* Field of Study
* Past Internships
* Cumulative Grade Point Average (CGPA)
* Whether Dwelling was Provided
* AND History of Backlogs

**And now I’ll pass it onto Akif who will talk about the questions we wanted to answer**

New slide

Presenter: Akif

* Questions slide:
* Using the datasets, we will highlight a few questions we sought to answer through our MLM which will be discussed later.
* First, Are students more likely to be accepted into a placement if they have a higher GPA?,
* Question 2 is; Are males or females more likely to be accepted into a placement?
* and lastly, question 3 is, Are students more likely to receive a placement based on the stream they are in?

New slide

Presenter: Akif

Database slide:

First, to answer these questions, we will have to create a database to host all our data. Using the data files we retrieved from Kaggle, we created tables in PostGresSQL (mention titles and point to ERD) and imported the CSV files into pgAdmin. Next, we joined (student\_information and student predictors, on ID) the tables to create a data table that will contain all the information for the database. We used AWS RDS Service to host the database on the server capstone project.

Next, we integrated our database into the MLM to perform accuracy testing and exploratory data analysis.

New slide

Presenter: Mohammed

ML Model Slide:

As the data is imported to the notebook. EDA was performed, the presence of null values, and duplicates were checked. Then, plots were created to see if there is any trends or insights. Later, the data was preprocessed. It was hot encoded, then standardized and the training set was 80% of the dataset and it was randomized. Finally, three ML models were developed, SVM, Decision Trees, and a deep learning model. The results were similar to each other and the average was about 86% accuracy.

New slide

Presenter: Connor

New slide

We can see here that males have received many more placements than females. But this also shows that the amount of male applicants is much higher than the amount of female applicants. This could possibly be due to there being more males in fields that are more popular for internships

Since 55% of male applicants are placed and 56% of female applicants are placed, we can see that both genders are essentially equally likely to be accepted into an internship. The following factors are more indicative of the likelihood of placement.

New slide

From this scatter plot, we can see that the placement of males based on their cumulative grade point average. is a fairly standard bell curve, where as the placement data of females shows a negative skew. It would seem appropriate for this to be a more linear plot, with the most amount of people receiving internships being in the range of 8 and 9 CGPA. But there are most likely fewer people achieving these spectacular grades so the pool of applicants in this range would be smaller.

We can see for both genders, the placement of people with a CGPA of five is quite low. Since there is such a large jump between a CGPA of five and six, we can conclude that applicants will most likely be placed if they have a CGPA of 6 or above.

New slide

From this bar graph we can see that computer science and information technology majors from both genders had the highest rate of placement. Civil and electrical engineering have approximately the same number of placements.

We can conclude that regardless of gender, it is most likely you will be placed into an internship if your engineering focus is either Computer Science or Information Technology.

New slide

We also wanted to include a recommendation for future analysis and that is to use a dataset with more fields and a more inclusive gender field as they only provided the male and female genders.

This is the end of our presentation. Thank you everyone for listening, and we will open it up for Q&A.