Education & Career Success

Tran Dang Huy, Sia Chawla Chawla, Jiaqi Xu

Table of contents

1 Executive summary

(Maximum of 4 sentences)

2 Introduction

(Maximum 10 sentences)

Graduates today face many questions about how their experiences influence their careers. Grades alone may not determine who gets the more job opportunities. This report explores which student experiences are associated with receiving more job offers.

To answer this question, we use a kaggle dataset which contains 5,000 records from recent graduates. It includes information about students' academic background, personal demographic, and career-related outcomes. Rather than testing predefined theories, this project takes an open-ended, pattern-oriented approach. The aim is to explore which types of experiences appear most consistently linked to job outcomes and personal satisfaction.

3 Methodology

(Maximum 300 words; Should include a figure and a table and those must be referenced in the text and have adequate captions)

3.1 Data source and cleaning

The data was collected from Kaggle and initially contains records for 5000 recent university graduates. During our data processing, we found 3.8% (192 rows) of the data have inconsistent gender values, as shown in Figure 1. Sine this is likely to provide misleading information, we decided to exclude them from further analysis.

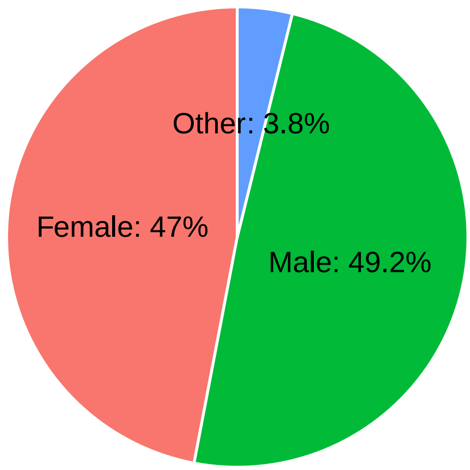

Figure 1: Number of Gender Categories

3.2 Model selection

Table 1 presents the first five records in the dataset and highlights the diversity of input variable types. The dataset includes numerical variables such as `SAT_Score`, categorical fields like `Field_of_Study`, and ordinal variables such as `Networking_Score`. This variety makes it unsuitable for models that require specific data type.

Table 1: First 5 records in filtered dataset

	1	2	3	4	5
Student_ID	S00001	S00003	S00004	S00005	S00006
Age	24	28	25	22	24
Gender	Male	Female	Male	Male	Male
High_School_GPA	3.58	3.42	2.43	2.08	2.40
SAT_Score	1052	1193	1497	1012	1600
University_Ranking	291	715	170	599	631
University_GPA	3.96	2.63	2.81	2.48	3.78
Field_of_Study	Arts	Medicine	Computer Science	Engineering	Law
Internships_Completed	3	4	3	4	2

Table 1: First 5 records in filtered dataset

	1	2	3	4	5
Projects_Completed	7	8	9	6	3
Certifications	2	1	1	4	2
Soft_Skills_Score	9	1	10	10	2
Networking_Score	8	9	6	9	2
Job_Offers	5	0	1	4	1
Starting_Salary	27200	42400	57400	47600	68400
Career_Satisfaction	4	9	7	9	9
Years_to_Promotion	5	3	5	5	2
Current_Job_Level	Entry	Entry	Mid	Entry	Entry
Work_Life_Balance	7	7	5	2	8
Entrepreneurship	No	No	No	No	Yes

Figure 2 shows the distribution of the outcome variable, `Job_Offers`, which represents the number of job offers received. The distribution is relatively balanced across classes, meaning no class dominates the data, and no transformation is required. Furthermore, since `Job_Offers` is ordinal, it emphasis the use of classification model over regression model.

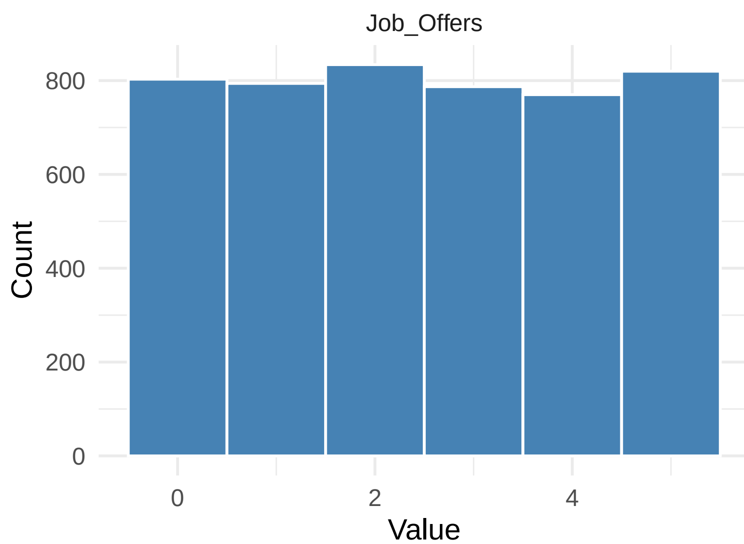

Figure 2: Distributions of Outcome Variable

Given these characteristics, random forest is an appropriate modelling choice. It is a flexible and robust algorithm that can naturally handle both numerical and categorical inputs without extensive preprocessing. In addition, it captures non-linear relationships and performs well even when input variables vary in scale or measurement type.

3.3 Accesing associations.

After fitting the random forest model, we use permutation importance to rank all input variables (Figure 3). This method measures how much the model's accuracy drops when the values of one variable are randomly shuffled. A larger drop means the variable was more important for the model's predictions. This allows us to assess which factors the model relies on most when estimating job offers.

The top 5 predictors were selected based on their importance scores (Figure 3) and further explored using boxplot with trend lines for the direction of association (**?@fig-trend**).

4 Results

(Maximum 200 words. Should include either a figure or a table.)

4.1 Associated factors for receiving job offers.

Figure 3 ranks predictors by their importance in the random forest model. Students' practical experiences (Completed Internships and Academic Projects), academic background (High School GPA and Field of Study), and interpersonal networking skills are more associated with the number of job offers received. In contrast, factors such as University Ranking and Soft Skills Score appear to have minimal influence.

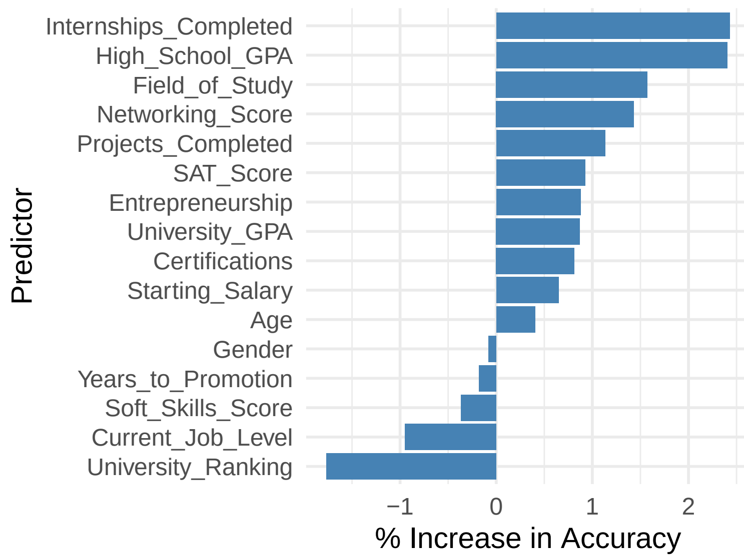

Figure 3: Importance of Predictors for Number of Job Offers Received

4.2 Association between important factors and job offers.

5 Discussion, conclusion and recommendations

6 Reference section

(Include at least 1 reference.)

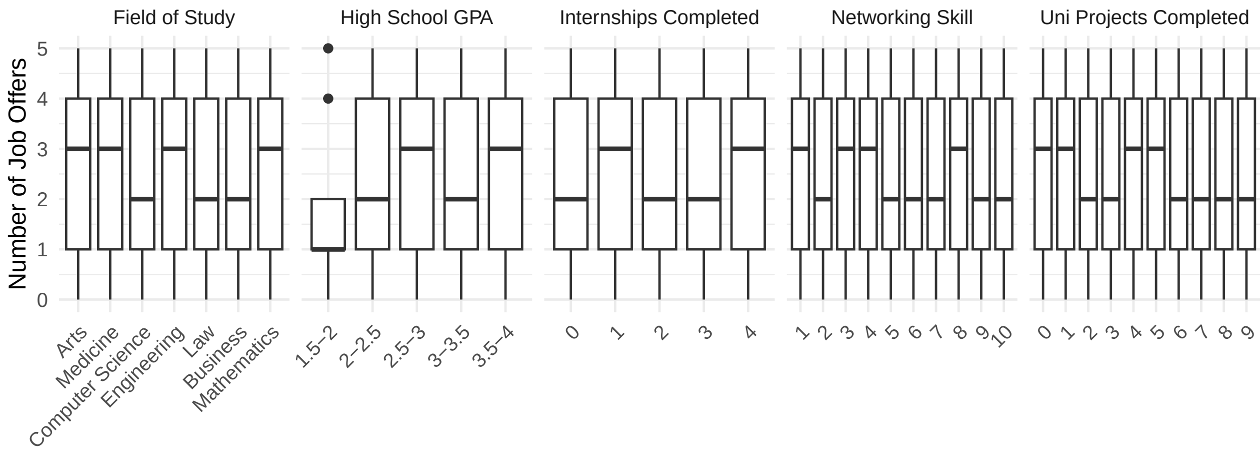

Figure 4: Distribution of job offers across student categories

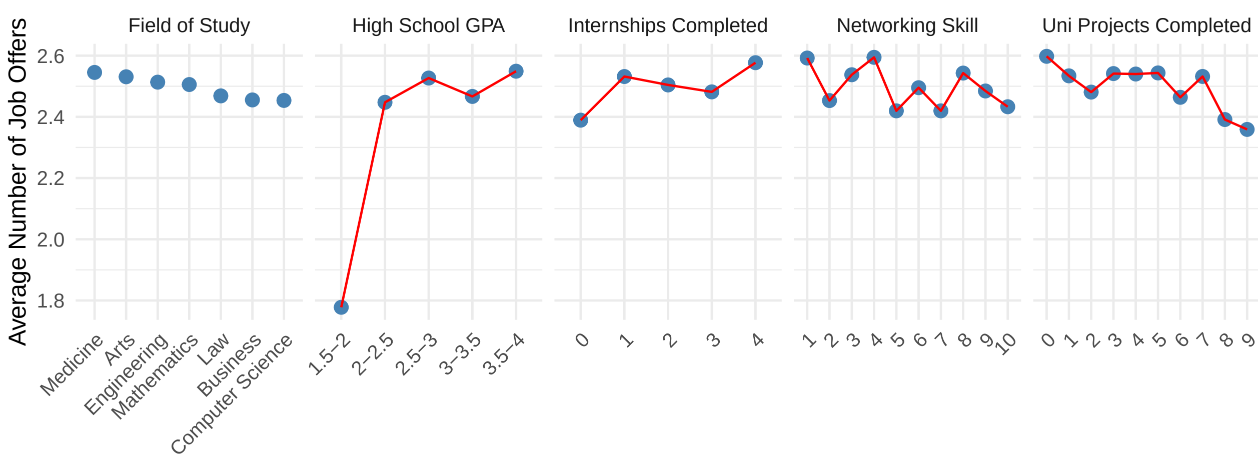

Figure 5: Average number of job offers by predictor category