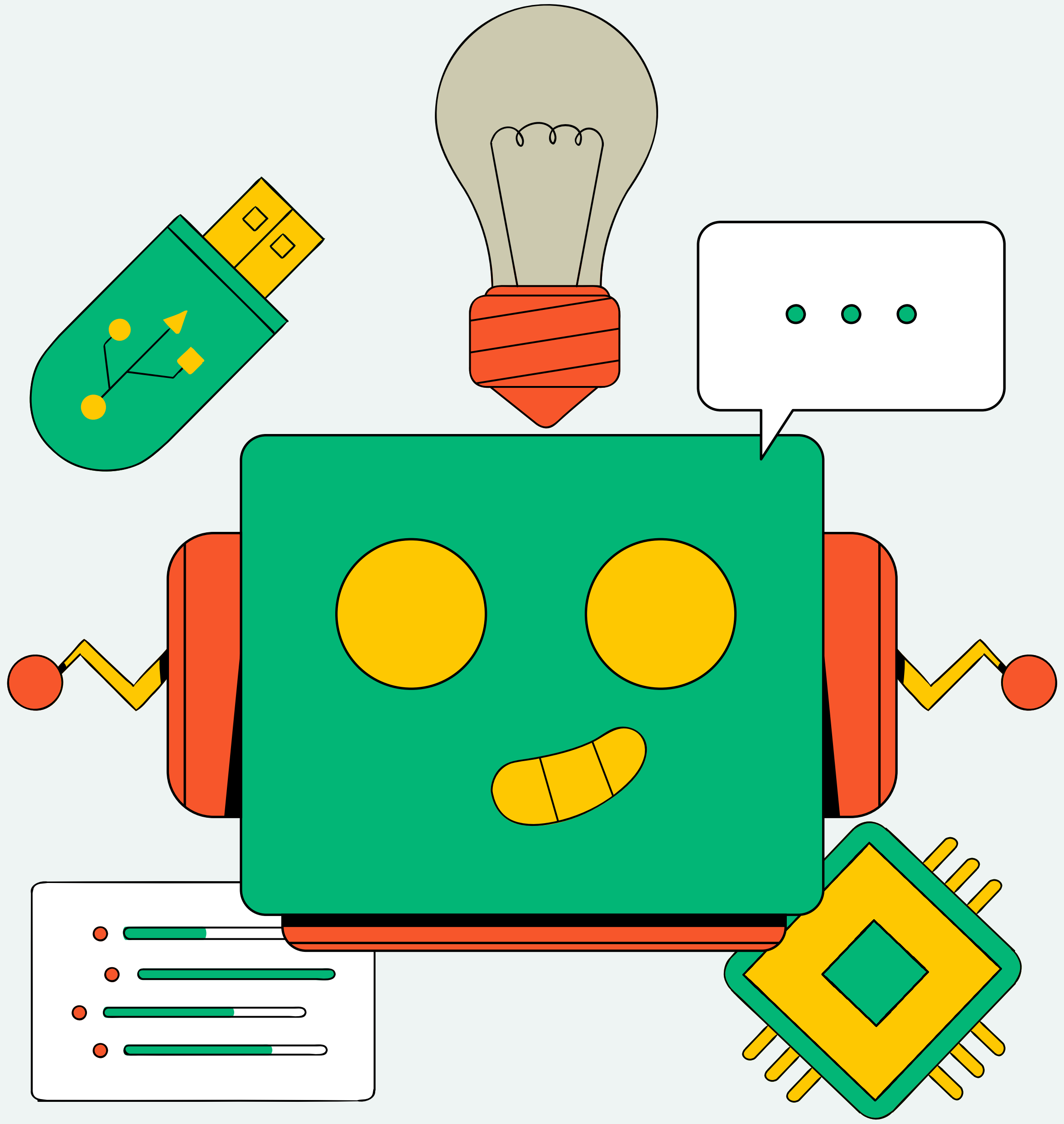
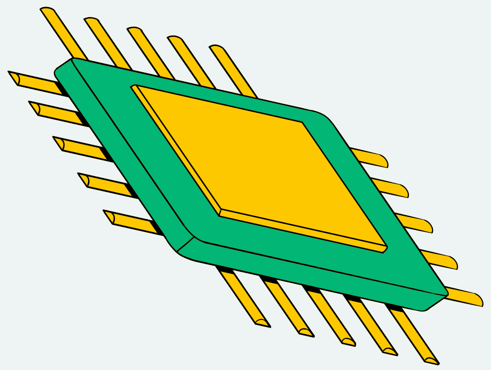


# CLASSIFYING JOB PLACEMENT AND SALARY PREDICTION USING MACHINE LEARNING

İLAYDA AKINET

CAN KANKILIÇ

AHMET YUSUF DEMİR



# PRESENTATION OUTLINE

- Introduction
- Algorithms and Motivation
- Experimental Scheme
- Results
- Improvements

# INTRODUCTION

University students often face concerns about job placement and salary prospects after graduation.
Using data from 700 individuals, this model can support students in planning their careers, guide recruiters in hiring decisions, and help universities improve their programs.

This project addresses these concerns by developing a machine learning model to:

- Predicts if a person will get a job based on their personal and academic information.
- Estimates the salary for those who are successfully placed.

# ALGORITHMS AND MOTIVATION

**Job Placement Prediction:**

- We selected the **K-Nearest Neighbors (KNN)** algorithm due to its strong performance in binary classification problems. KNN effectively groups similar data points based on proximity, making it ideal for predicting whether an individual will secure a job.

**Salary Prediction:**

- Initially, **Linear Regression** was chosen for its simplicity and suitability for continuous data prediction. However, when better accuracy was desired, we moved to **Random Forest Regression**, which accounts for non-linear relationships and reduces overfitting.

# EXPERIMENTAL SCHEME

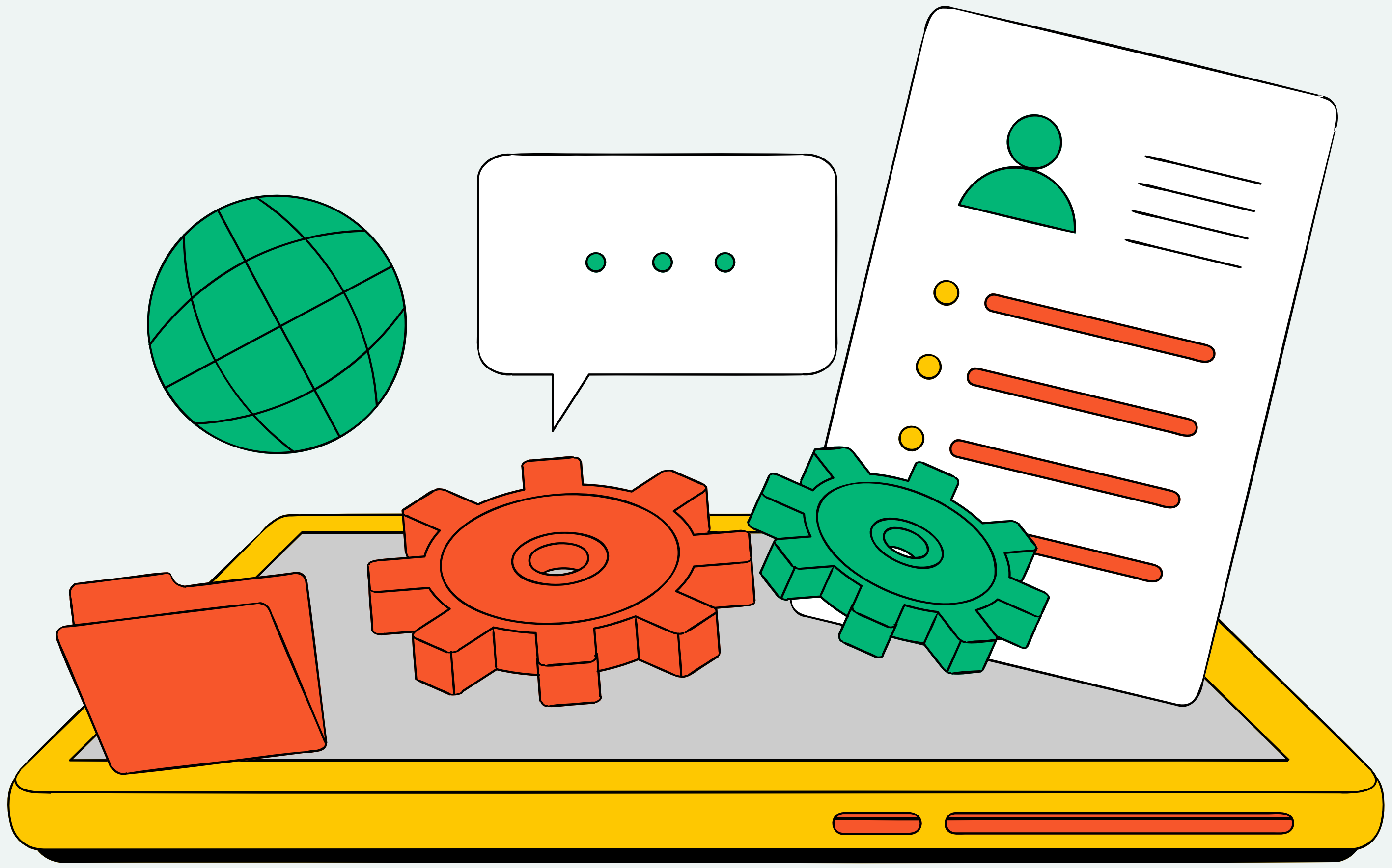
Dataset Preparation

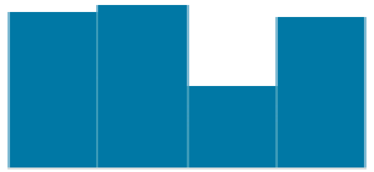
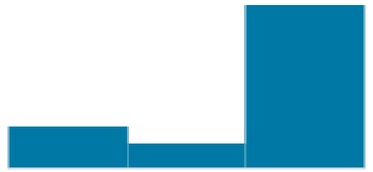
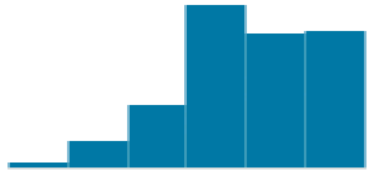
KNN with 560 training points and 140 test points.

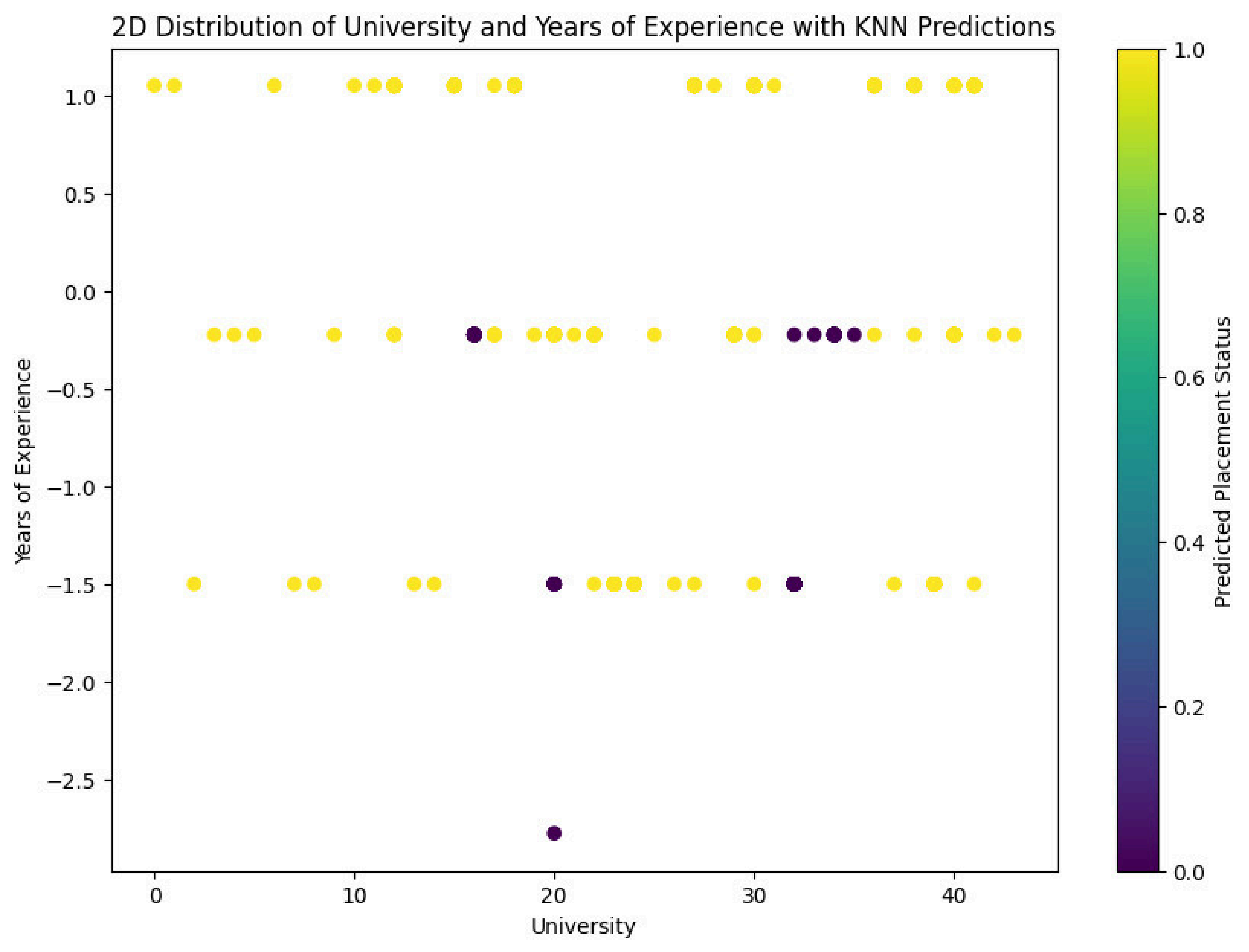
Linear Regression with 412 training and 148 test points.

# DATASET PREPARATION

Attributes like ID, Name, and Degree were removed as they provided no predictive value. ("In our dataset, all recorded degrees were 'Bachelor's'.")
Strings were encoded to numerical values for machine learning compatibility.



| gender<br>Gender of individual | age<br>Age of individuals | stream<br>Field of specialization | college_name<br>Name of the university | placement_status<br>indicated individual placed or not | salary<br>Salary earned upon placement | gpa<br>GPA of individual | years_of_experie...<br>Years of work experience |
|---|---|---|---|---|---|---|---|
| Female 52%<br>Male 48% | <br>23 26 | Computer Science 31%<br>Information Tech... 22%<br>Other (334) 48% | University of Calif... 6%<br>University of Mich... 6%<br>Other (614) 88% | Placed 81%<br>Not Placed 19% | <br>0 68.0k | <br>3.4 3.9 | <br>1 3 |
| Male | 25 | Computer Science | Harvard University | Placed | 60000 | 3.7 | 2 |
| Female | 24 | Electrical Engineering | Massachusetts Institute of Technology | Placed | 65000 | 3.6 | 1 |
| Male | 26 | Mechanical Engineering | Stanford University | Placed | 58000 | 3.8 | 3 |
| Female | 23 | Information Technology | Yale University | Not Placed | 0 | 3.5 | 2 |
| Male | 24 | Computer Science | Princeton University | Placed | 62000 | 3.9 | 2 |
| Female | 25 | Electronics and Communication | Columbia University | Placed | 63000 | 3.7 | 1 |
| Male | 26 | Information Technology | California Institute of Technology | Placed | 59000 | 3.8 | 3 |
| Female | 24 | Computer Science | University of Chicago | Not Placed | 0 | 3.6 | 2 |
| Male | 25 | Electrical Engineering | University of Pennsylvania | Placed | 64000 | 3.7 | 2 |
| Female | 23 | Mechanical Engineering | Northwestern University | Placed | 57000 | 3.5 | 1 |
| Male | 24 | Computer Science | Duke University | Placed | 61000 | 3.9 | 2 |
| Female | 25 | Electronics and Communication | Johns Hopkins University | Not Placed | 0 | 3.8 | 3 |
| Male | 26 | Information Technology | University of California--Berkeley | Placed | 63000 | 3.7 | 2 |
| Female | 24 | Computer Science | University of Michigan--Ann Arbor | Placed | 64000 | 3.6 | 1 |
| Male | 23 | Electrical Engineering | University of California--Los Angeles | Placed | 66000 | 3.8 | 3 |

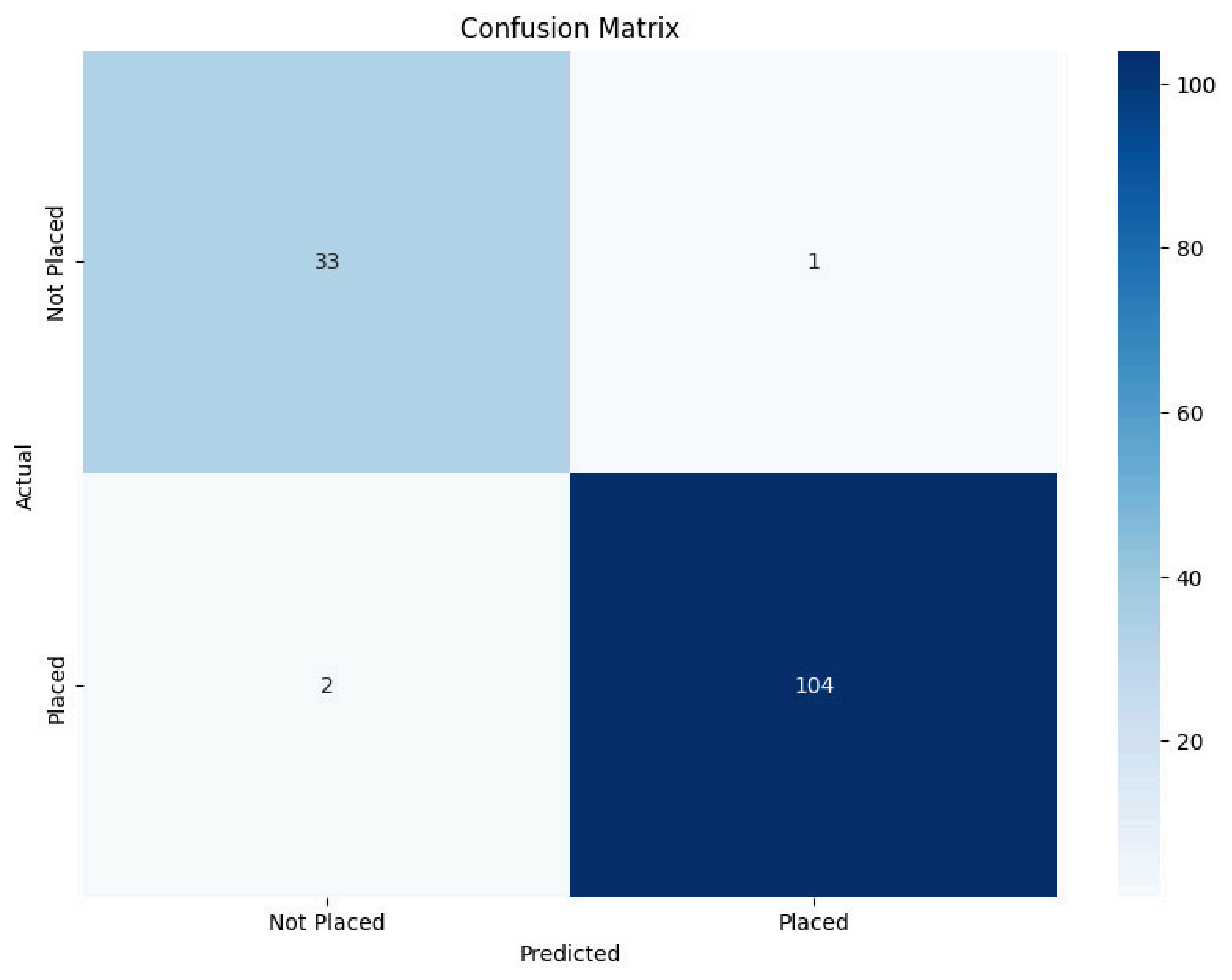

2D Distribution of University and Years of Experience with KNN Predictions
Years of Experience
University
Predicted Placement Status
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
0
10
20
30
40
0.0
0.2
0.4
0.6
0.8
1.0

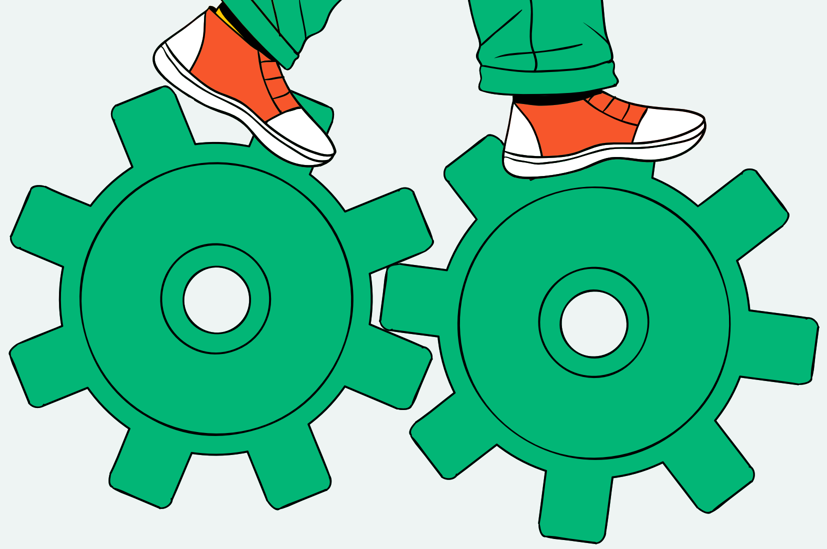
# JOB PLACEMENT PREDICTION USING KNN

Accuracy: 0.98

Classification Report:

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.94 | 0.97 | 0.96 | 34 |
| 1 | 0.99 | 0.98 | 0.99 | 106 |
| accuracy | | | 0.98 | 140 |
| macro avg | 0.97 | 0.98 | 0.97 | 140 |
| weighted avg | 0.98 | 0.98 | 0.98 | 140 |





# SALARY PREDICTION USING LINEAR REGRESSION

Linear Regression: RMSE ≈ 1885, indicating predictions were off by approximately $1885.

```
Mean Squared Error: 3556664.37
R^2 Score: 0.30
     Predicted Salary  Real Salary
622             64066        60000
88              65640        67000
164             64863        64000
492             65534        65000
664             66226        66000
..                ...          ...
247             65805        68000
94              63625        66000
308             65524        66000
518             62615        63000
443             65661        66000
```

```
Mean Squared Error: 674383.02
R^2 Score: 0.87
     Predicted Salary  Real Salary
622             59970        60000
88              66970        67000
164             64000        64000
492             65000        65000
664             65720        66000
..                ...          ...
247             67820        68000
94              66000        66000
308             66011        66000
518             63000        63000
443             66023        66000
```

# RANDOM FOREST FOR BETTER RESULTS

Random Forest Regression improved accuracy, significantly reducing error margins. With Random Forest, the Root Mean Squared Error (RMSE) was reduced to 821.21, and the $R^2$ score increased to 0.87.

# "GENDER", AND "STREAM" ATTRIBUTES REMOVED

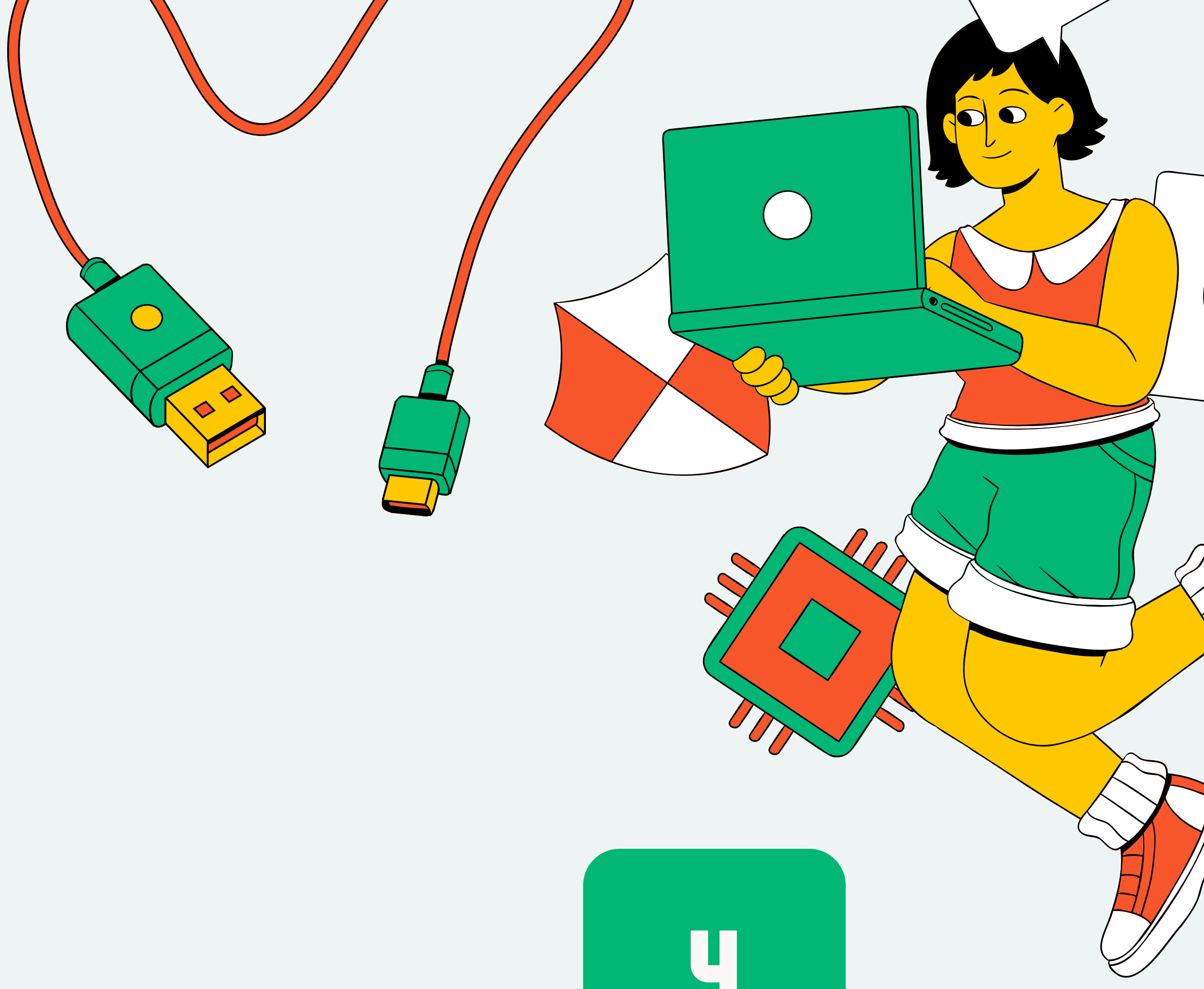
```
Accuracy: 0.99

Classification Report:
              precision    recall  f1-score   support

           0       1.00      0.97      0.99        34
           1       0.99      1.00      1.00       106

    accuracy                           0.99       140
   macro avg       1.00      0.99      0.99       140
weighted avg       0.99      0.99      0.99       140
```

```
Root Mean Squared Error: 826.92
R^2 Score: 0.87
     Predicted Salary  Real Salary
622             60000        60000
88              66674        67000
164             64000        64000
492             65000        65000
664             65893        66000
..                ...          ...
247             67870        68000
94              66000        66000
308             66000        66000
518             62838        63000
443             65893        66000
```



# IMPROVEMENTS

1

## FEATURE ENGINEERING

Enhance the dataset by including additional factors such as industry trends, regional salary variations, and economic indicators to improve the predictive power of the model.

2

## ALGORITHM TUNING

Perform hyperparameter optimization for KNN and Random Forest algorithms to maximize accuracy and minimize errors. Techniques like grid search or random search can be applied.

3

## CROSS-VALIDATION

Implement k-fold cross-validation to ensure the reliability and generalizability of the model's performance across different data splits.

4

## REAL-WORLD VALIDATION

Test the model on real-world HR or recruitment data to assess its practical applicability and identify potential gaps for refinement.

# THANK YOU FOR YOUR ATTENTION