Predicting qualified employees for promotion using Classification model

Banan Alhethlool

Randa Mohammed

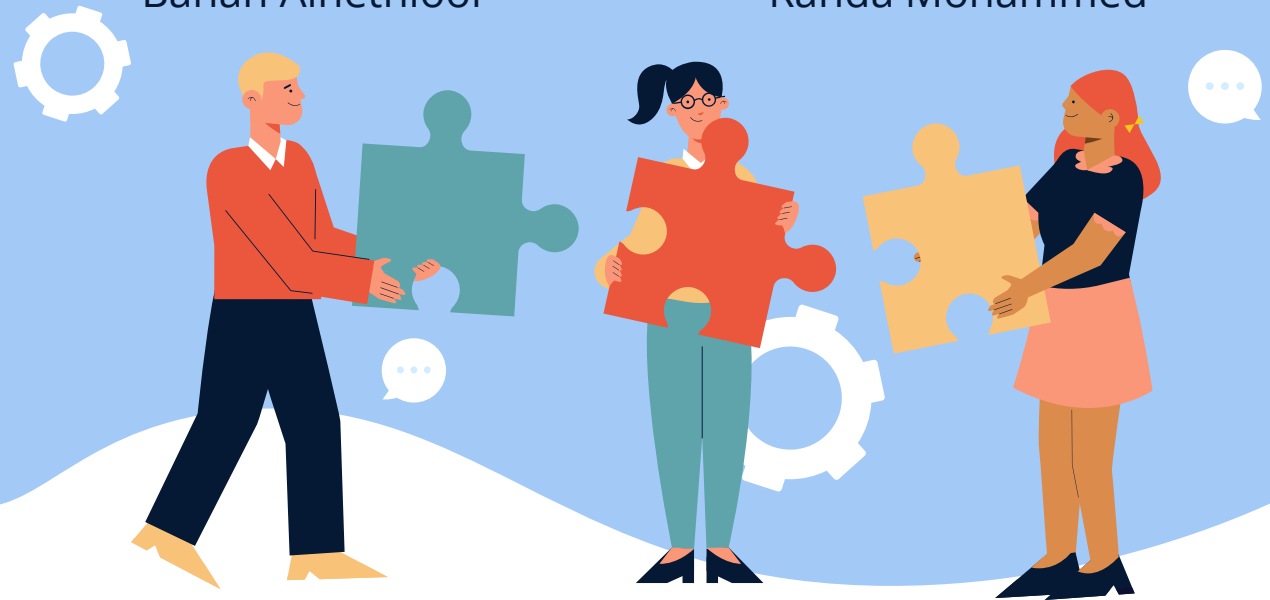


TABLE OF CONTENTS

Introduction

Dataset

Original Dataset

13 Columns

54808 Rows

After Pre-processing

17 Columns

52399 Rows

Approach & Methodology

01 EDA

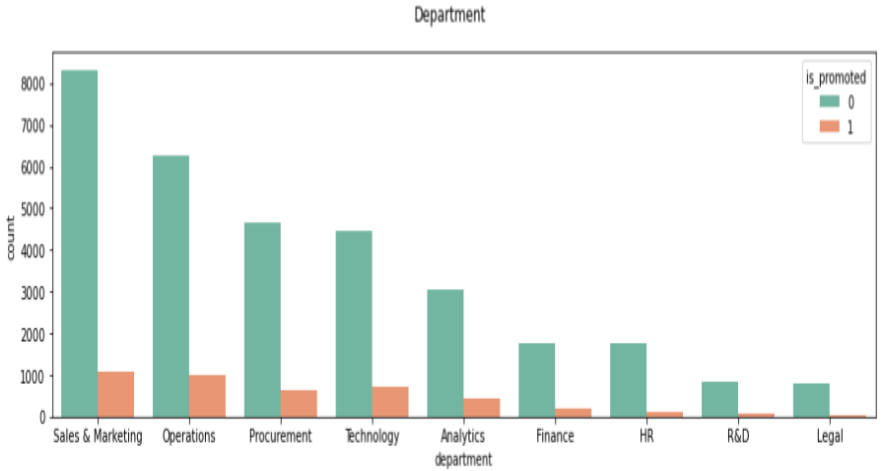
02 Data Preparation

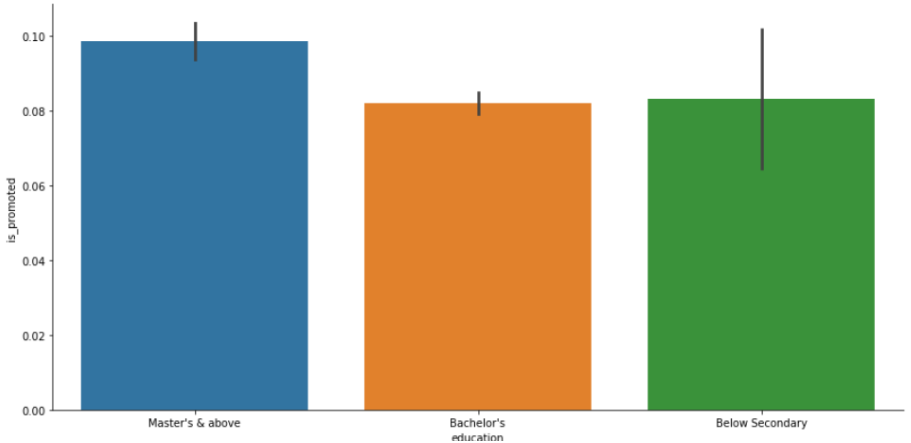
03 Balance Data

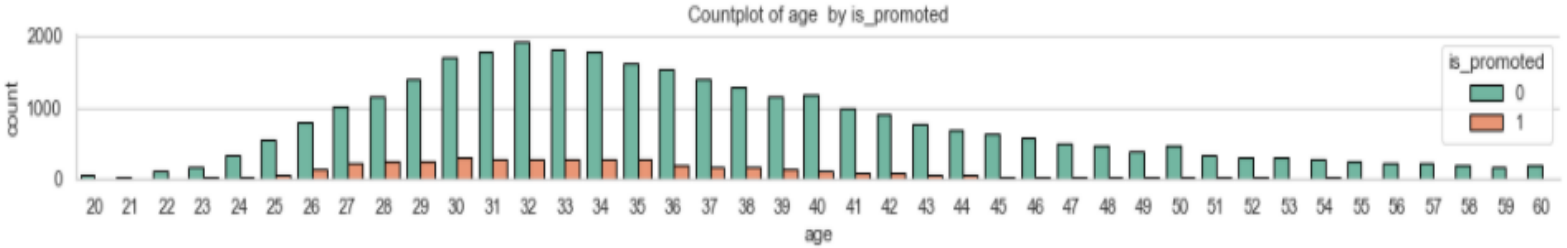
04 Classification Algorithms

05 Best model

EDA


Department
is_promoted
0
1
count
8000
7000
6000
5000
4000
3000
2000
1000
0
Sales & Marketing
Operations
Procurement
Technology
Analytics
Finance
HR
R&D
Legal
department


is_promoted
0.10
0.08
0.06
0.04
0.02
0.00
Master's & above
Bachelor's
Below Secondary
education


Countplot of age by is_promoted
is_promoted
0
1
count
2000
1000
0
20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
age

Data Preparation

Feature Engineering:

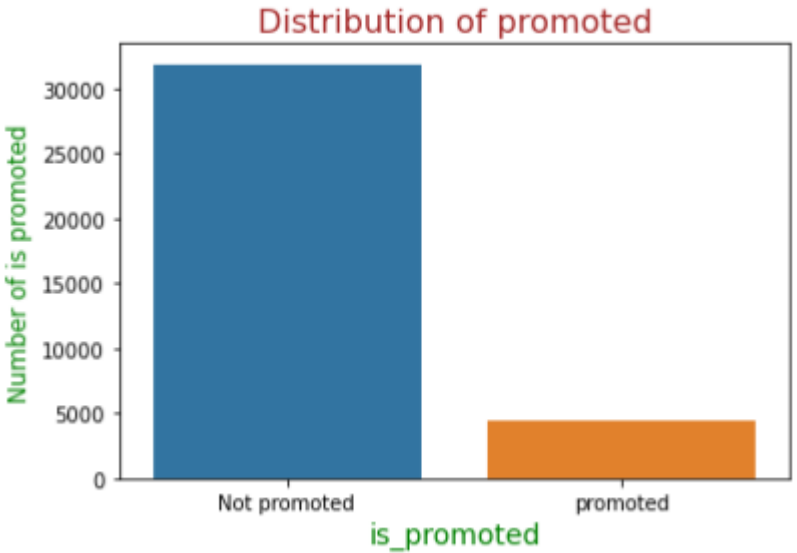
- Encoding (dummy values)
- Scaling (Standardization)

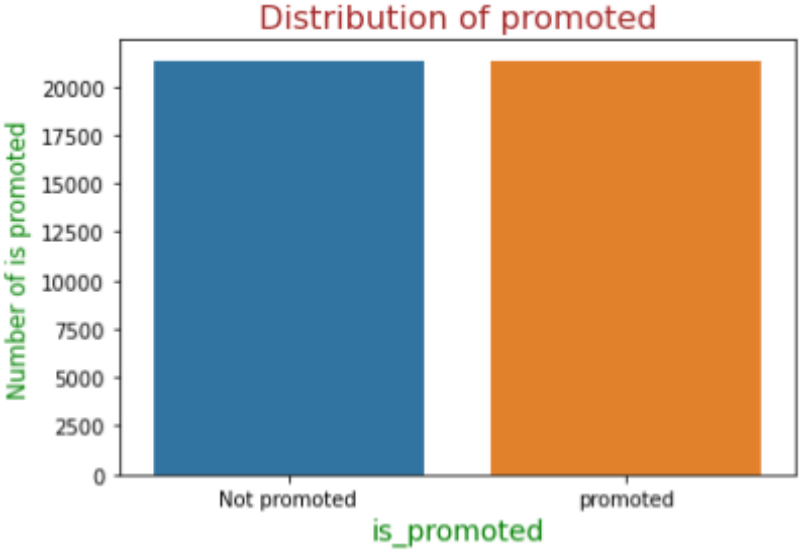
Feature Selection:

- Drop some columns: recruitment_channel, region, gender and employee_id

Balance Data

- **SMOTE**

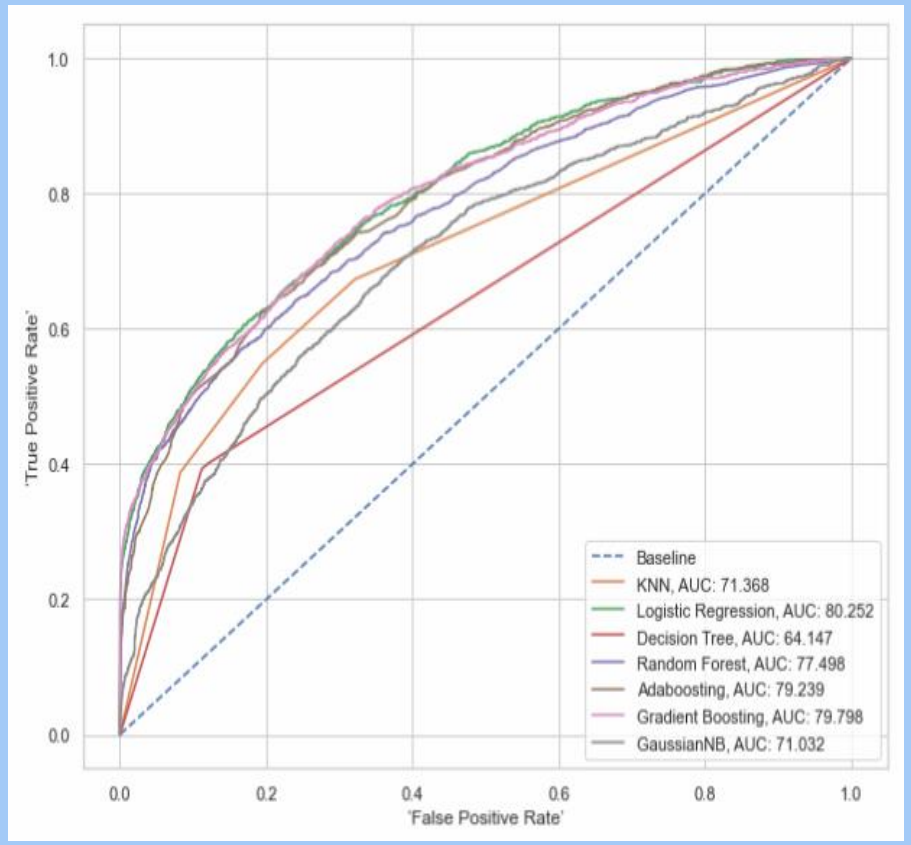


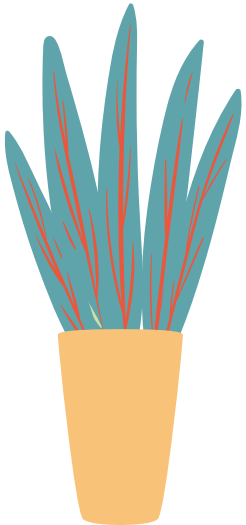


Classification Algorithms

- KNN
- KNN GridSearchCV
- Logistic Regression
- Decision Tree
- Random Forest
- Adaboosting
- Gradiant Boosting
- GaussianNB
- Stacking(LR,KNN,RF AND ET)

Analysis & Results





Best Model Logistic Regression

conclusion

- **Improvement:**

  Increase Dataset

- **Recommendations for business owner**
- **Tools**
- **libraries** (Pandas, Numpy, os, pickle, sklearn, imblearn.over_sampling, seaborn and matplotlib)

THANKS!

Any Questions?

Banan Alhethlool **Randa Mohammed**