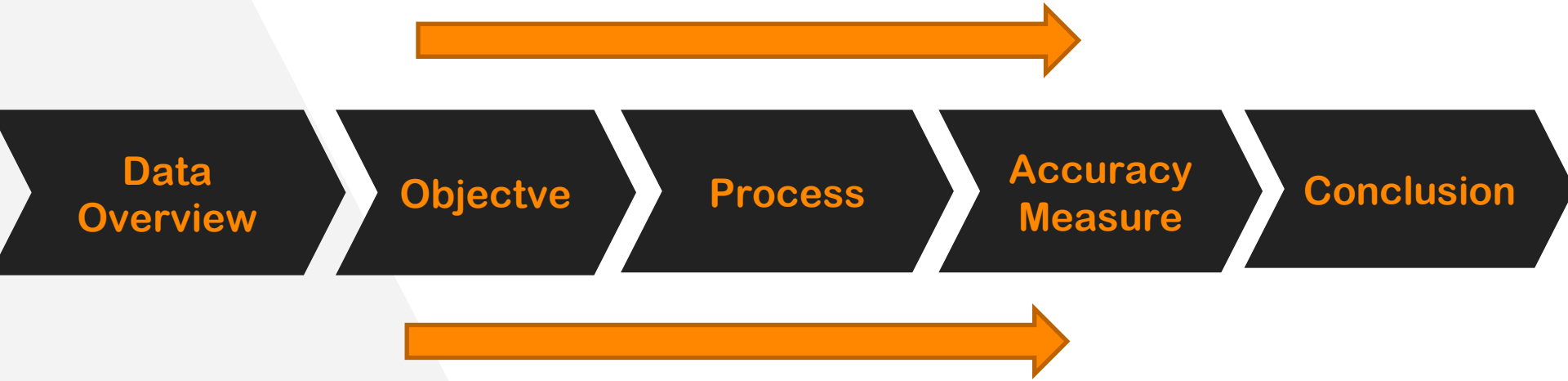
EMPLOYEE PROMOTION CLASSIFICATION PROJECT

By Prathmesh Chavan

TABLE OF CONTENT



Data Overview & Problem Statement

Problem Statement:

Every year, around 5% of the employees are promoted in the company. Based on the data of employee, find out the main variables that impact an employee's promotion, in turn contributing to more benefits to the Company.

Data Description:

Department : specific division where employee works in organization.
Education : Education qualification of the employees.
Previ year rating : previous year overall performance of an employee on a scale of 0 to 5.
KPIs_met >80% : Key performance Indicator (KPI) is higher than 80% or not.
Awards won : Award won or not.
Avg training score : Employee training score out of 100.
Is promoted : Employee Promotion (1 = Promoted and 0 = not Promoted).

“ Who Should Be Promoted? ”

OBJECTIVE

Our End Goal Is To Develop A Model Which Will Help HR Or Managers To Decide Which Employee Should Get Promotion Based On Various Factors.

It Will Save Organizations Time And Effort And Allow Managers To Focus On Other Majory Aspects In The Organization.

PROCESS

1 Data Preprocessing

Handling Missing Data, Type Casting , Feature Engineering, Handling Imbalanced Data , Feature Scaling.

2 Exploratory Data Analysis

Analysing Data, Hypothesis Testing, Feature Selection

3 Model Building

Comparing Multiple Models, Cross validation, Confusion Matrix,Selecting Best Algorithm Based On accuracy Score And Confusin Matrix

Data Preprocessing

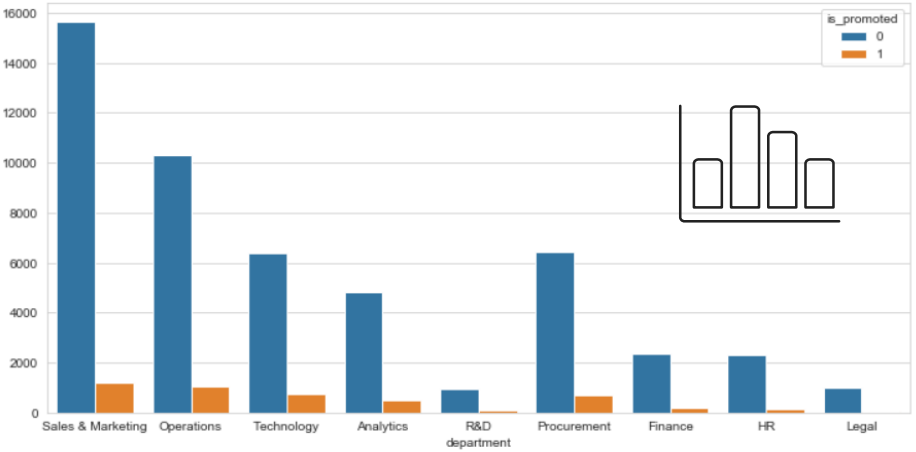
department	region	education	gender	recruitment_channel	no_of_trainings	age	previous_year
Sales & Marketing	region_7	Master's & above	f	sourcing	1	35	
Operations	region_22	Bachelor's	m	other	1	30	
Sales & Marketing	region_19	Bachelor's	m	sourcing	1	34	
Sales & Marketing	region_23	Bachelor's	m	other	2	39	

Before Data Processing

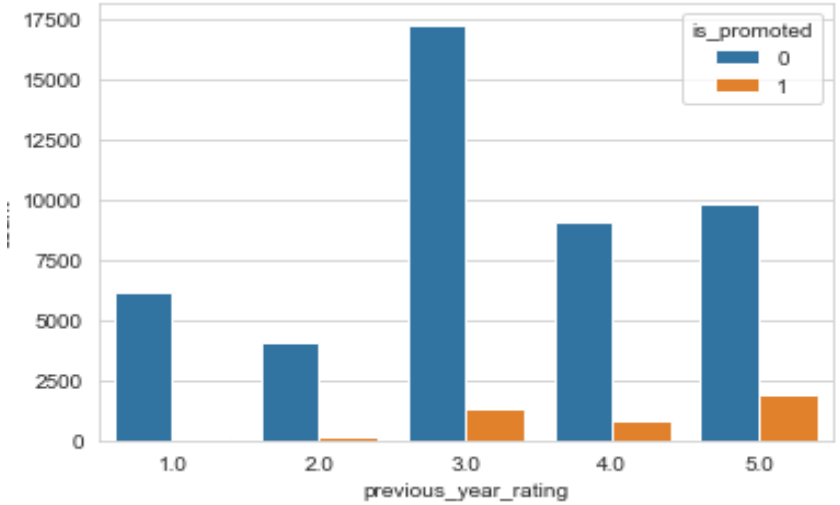
vg_training_score	region	awards_won?	KPIs_met >80%	previous_year_rating	department_Operations
-1.075931	-0.713357	-0.154018	1.356878	1.395766	-0.510993
-0.253282	0.773827	-0.154018	-0.736986	1.395766	1.956975
-1.001145	0.476390	-0.154018	-0.736986	-0.250651	-0.510993
-1.001145	0.872972	-0.154018	-0.736986	-1.897069	-0.510993
0.718939	1.170409	-0.154018	-0.736986	-0.250651	-0.510993

After Data Processing

Exploratory Data Analysis



Department Wise Promoted Employees, Orange Bar Represents Promoted Employee



Previous Rating Wise Promoted Employees, We Can See Employee Having Higher Rank Will Have Higher Chances Of Promotion

Model Building And Accuracy Measure

In Order To Find Best Model For Dataset I Have Tested Dataset With Multiple Algorithms.

RandomForestClassifier – 93.95%

XGBoost – 94.08%

SVM – 92.11%

KNN – 93.48%

LogisticRegression – 92.22%

XGBoost

```
xgb = XGBClassifier()
xgb.fit(x_train_raw , y_train_raw)
print(xgb.score(x_test_raw , y_test_raw))
scores = cross_val_score(xgb , x_raw , y_raw , cv = 5)
print(scores.mean())
```

```
0.940825883354619
0.9413042086936049
```

Confusion Matrix

```
y_pred = xgb.predict(x_test_raw)
cm = confusion_matrix(y_test_raw , y_pred)
cm
```

```
array([[14999,    36],
       [  937,   471]], dtype=int64)
```

Overall Xgboost gives best accuracy with 94% accuracy, also it's a generalized model with minimum type1 error and 1 type2 error.

Conclusion And **Business Findings**

Employees who got promotion had good ratings a year before hey got promoted.

Employees having Masters degree have good chance of promotion

In first 2 years employee will not get promotion

If Employees manages to meet KPI then there is high chance that the employee will get promotion.

THANKS!

Any questions?

You can find me at cprathmesh08@gmail.com

Github Repository

https://github.com/PrathmeshChavan/employee_promotion