



# Machine Learning Project Report

**Name :** MD ZAKRIYA AHMED

**Project Title:** Predicting Employee Promotions in an MNC

## 1. Problem Statement

A multinational company (MNC) wants to predict which employees (manager and below) are likely to be promoted based on historical performance, demographics, training outcomes, and KPIs. The goal is to use machine learning to expedite the internal promotion cycle by identifying high-potential employees at an early stage.

## 2. Business Objective

Predict is\_promoted (1 if the employee is to be promoted, 0 otherwise).

Improve internal HR operations by reducing delay in promotion announcements.

## 3. Dataset Description

Column	Description
employee_id	Unique ID for employee
department	Department of employee
region	Region of employment
education	Education Level
gender	Gender of Employee
recruitment_channel	Channel of recruitment
no_of_trainings	Number of trainings completed in last year
age	Age of the employee
previous_year_rating	Employee's performance rating last year

Column	Description
length_of_service	Number of years with the organization
KPIs_met >80%	Whether the employee met >80% of KPIs (1/0)
awards_won?	Whether employee won awards last year (1/0)
avg_training_score	Average score in current training evaluations
is_promoted	Target Variable – Whether employee was promoted (1) or not (0)

## 4. Data Preprocessing

### Missing Values:

Imputed education with mode.

Imputed previous\_year\_rating with median (assumed from code).

### Categorical Encoding:

Converted categorical features like gender, region, department, and recruitment\_channel to numeric using Label Encoding.

No feature scaling was required for tree-based models.

## 5. Exploratory Data Analysis (EDA)

### Used histograms and countplots to analyze:

Promotion rates by department, education, awards, and KPI performance.

### Observations:

Employees who met >80% KPIs or won awards were more likely to be promoted.

Some departments had higher promotion rates than others.

**Imbalanced dataset:** majority class is not promoted.

## 6. Feature Engineering

Created no new features but ensured relevant ones were cleaned and encoded.

Converted all inputs to numeric formats suitable for modeling.

## 7. Model Building

Used multiple classification models and compared their performance:

Model	Highlights
Logistic Regression	Baseline model
Decision Tree Classifier	Captures non-linear patterns
Random Forest	Performed better due to ensembling

**Best Performing Model:** Random Forest (from notebook)

## 9. Final Predictions

Applied the selected model to the test dataset.

**Saved predictions in the format:**

```
csv
Copy
Edit
employee_id,is_promoted
8724,0
74430,0
...
```

**Final file:** promotion\_predictions\_2025-05-18.csv

## 10. Conclusion & Future Work

### Conclusion:

Built a classification model to predict employee promotions.

Handled missing data, encoding, and model selection effectively.

Best model deployed on the test set to generate is\_promoted.

### ✓ Deliverables

✓ Code in IEP TASK 2.ipynb

✓ **Final Predictions:** promotion\_predictions\_2025-05-18.csv

✓ This Report