**HR Employee Promotion Prediction Project**

**📋 Project Overview**

This project explores the key factors influencing employee promotions within a large organization. By analyzing various attributes such as department, region, education, age, and previous performance ratings, we aim to predict whether an employee will be promoted. Through machine learning techniques, we strive to uncover patterns that can optimize promotion processes and create more data-driven HR strategies.

**🎯 Objective**

The main objective of this project is to develop a predictive model that accurately classifies whether an employee will be promoted based on historical data. This can aid HR teams in understanding promotion trends and improving their decision-making processes.

**📂 Dataset Overview**

The dataset consists of 13 columns and 54,808 rows, with one dependent variable (whether the employee was promoted) and 12 independent variables that include:

* **Employee Information**: ID, Department, Region, Gender
* **HR Metrics**: No. of Trainings, Length of Service, Awards Won, Previous Year Rating
* **Education**: Highest level of education attained by the employee
* **Performance Scores**: Average Training Score, Previous Year Rating

**Key Features:**

1. **Previous Year Rating**: Ratings provided to employees in the prior year.
2. **Average Training Score**: The average score employees received in their training assessments.
3. **Length of Service**: Total years of service by the employee.
4. **No. of Trainings**: How many trainings the employee has attended.
5. **Awards Won**: Whether the employee has won any awards.
6. **Department & Region**: Organizational and geographical factors impacting promotion.

**⚙️ Data Preprocessing**

Before modeling, we handled missing values and outliers in the dataset:

* Missing values in **Education**, **Previous Year Rating**, and **Average Training Score** were imputed using **mode**, **median**, and **mean** respectively.
* Outliers in numerical features (e.g., Age, Length of Service) were replaced using interquartile ranges (IQR) to maintain data integrity.

**Feature Engineering:**

* **Log Transformation**: Applied to skewed features (e.g., Age, Length of Service, and Training Scores) to normalize their distribution.
* **One-Hot Encoding**: Applied to categorical features like department and region to make them suitable for machine learning models.

**🛠️ Modeling and Evaluation**

Several machine learning models were developed and evaluated:

1. **Logistic Regression**:
   * **Accuracy**: 93.60%
   * **Precision**: 93.08%
   * **Recall**: 63.24%
   * **AUC**: 0.7841
   * **F1 Score**: 0.6902
2. **Random Forest Classifier**:
   * **Accuracy**: 93.72%
   * **Precision**: 93.12%
   * **Recall**: 64.03%
   * **AUC**: 0.7772
   * **F1 Score**: 0.6999
3. **Decision Tree Classifier**:
   * **Accuracy**: 88.53%
   * **Precision**: 64.80%
   * **Recall**: 67.24%
   * **AUC**: 0.6724
   * **F1 Score**: 0.6588

The **Random Forest Classifier** performed the best overall, with the highest accuracy, precision, and F1 score.

**🔍 Key Insights and Findings**

* **Departmental Influence**: Employees in **Technology** and **Analytics** had the highest promotion rates, while **HR** and **Legal** departments saw the lowest.
* **Training and Promotion**: Higher training scores significantly correlated with a higher likelihood of promotion. However, attending more trainings beyond 2 did not necessarily increase promotion chances.
* **Regional and Gender Disparities**: Certain regions showed significantly higher promotion rates, and female employees had a slightly higher promotion rate compared to male employees.
* **Education Levels**: Employees with **Master’s & above** had the highest promotion rates, followed by those with **Below Secondary** education.

**🚀 Conclusion**

This project provides a powerful framework for predicting employee promotions based on historical data. By implementing machine learning models and careful data analysis, we can identify key factors that influence promotion outcomes, offering actionable insights to HR teams.