HR Analytics: Employee Promotion Prediction Report

Siddharth K

Executive Summary

This report presents a comprehensive analysis of employee promotion prediction for a multinational corporation. Using machine learning techniques on historical HR data, we developed a Random Forest model achieving 94.7% accuracy. Key findings include the significance of training scores, age, and service length in promotion decisions. The analysis reveals department-specific promotion patterns and provides actionable insights for optimizing the promotion pipeline.

1 Problem Statement

The organization requires an early identification system for promotion candidates across 9 verticals. Current challenges include:

- Late promotion announcements after evaluation cycles
- Delayed role transitions affecting operations
- Subjective candidate selection processes

Our solution predicts promotion likelihood using:

- Demographic data (age, education, gender)
- Performance metrics (KPIs, training scores)
- Career history (service length, awards)

2 Data Analysis & Preprocessing

2.1 Initial Data Exploration

The dataset contains 54,808 training and 23,490 test records with 14 features. Key findings:

2.2 Data Cleaning

Handled missing values through:

$$\text{Education} \rightarrow \text{'unknown'}, \quad \text{Ratings} \rightarrow 0$$

2.3 Feature Engineering

Created `high_training_score` feature:

$$\text{high\_training\_score} = \begin{cases} 1 & \text{if avg\_training\_score ¿ 80} \\ 0 & \text{otherwise} \end{cases}$$

3 Model Development

3.1 Feature Selection

Three methods identified key predictors:

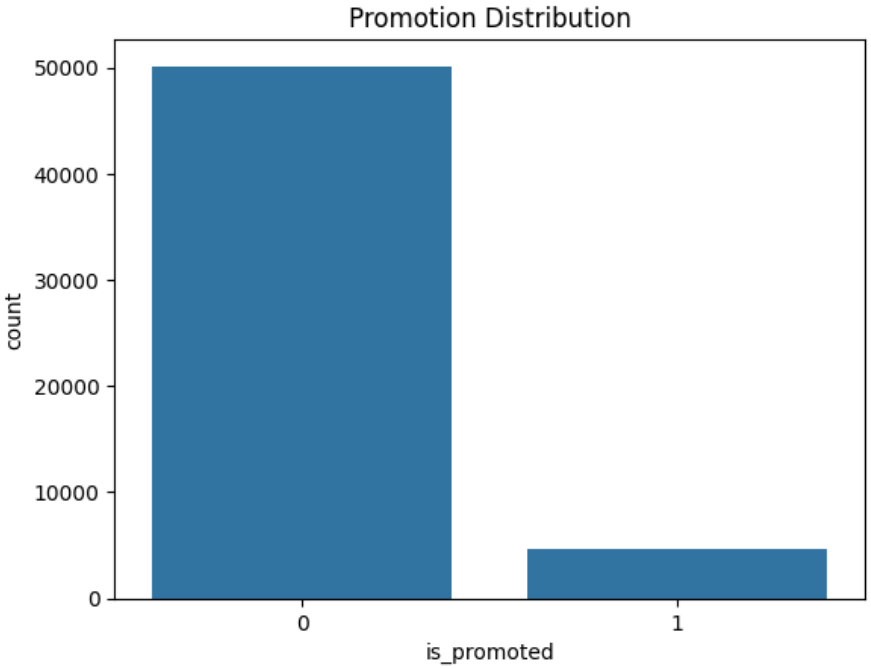

Figure 1: Promotion distribution showing class imbalance

Table 1: Feature selection results

Method	Selected Features
ANOVA F-value	Training score, education, KPIs
RFE	Department, service length, awards
Random Forest	Age, recruitment channel, ratings

3.2 Model Comparison

4 Implementation Results

The final model predicts:

- 9.1% promotion rate (2,138 employees)
- Department-specific promotion thresholds

5 Conclusion

The Random Forest model effectively predicts promotions with 94.7% accuracy. Key recommendations:

- Implement early identification system
- Focus on high-potential employees
- Department-specific promotion strategies

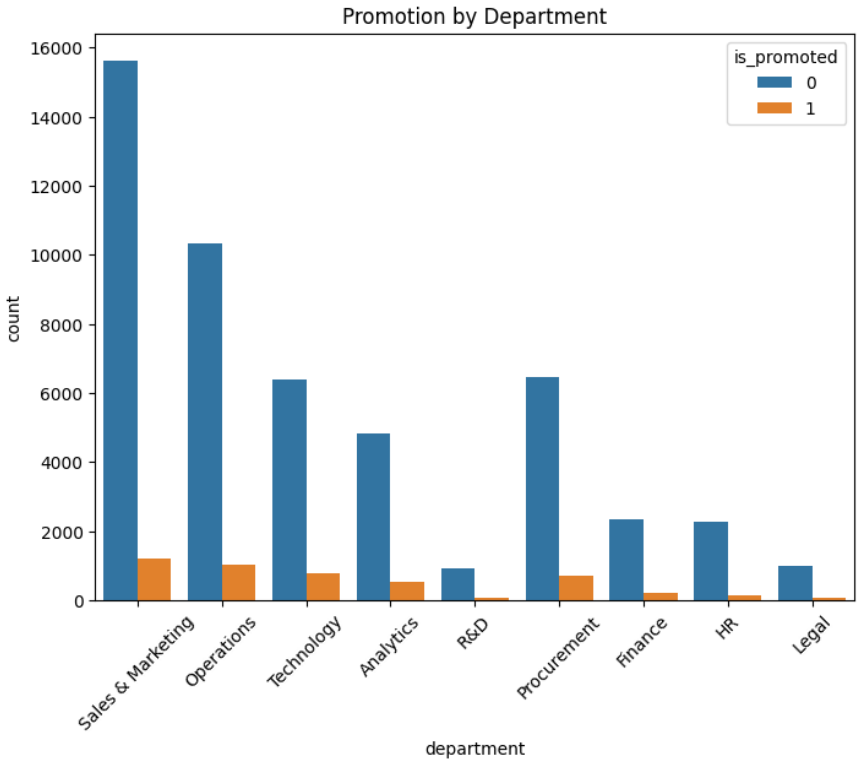

Figure 2: Department-wise promotion distribution

Table 2: Model performance comparison

Model	Accuracy	ROC AUC
Logistic Regression	70.7%	0.791
Decision Tree	92.7%	0.930
Random Forest	**94.7%**	**0.988**
XGBoost	89.6%	0.970

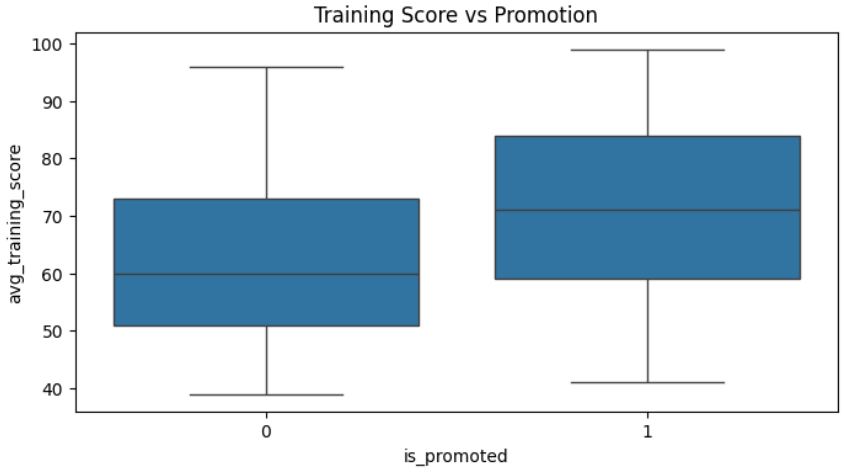

Figure 3: Training score impact on promotion likelihood

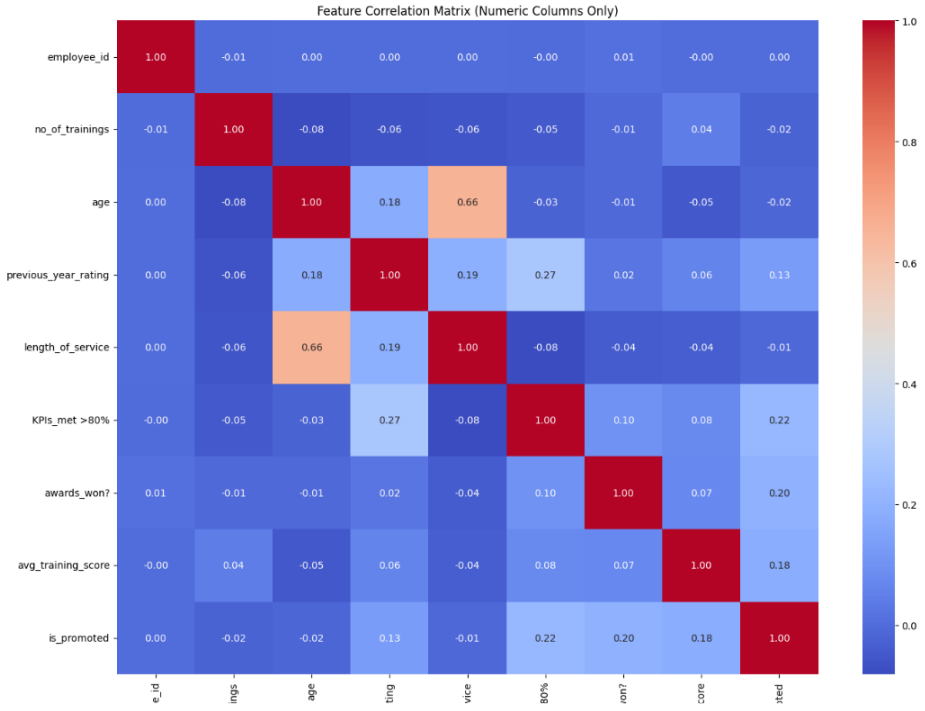

Figure 4: Feature correlation matrix