

IBM HR Analytics: Employee Attrition & Performance

Data Science Internship Project – Unified Mentor



Introduction

Objective of the Project:

To analyze employee attrition patterns and predict which employees are likely to leave the company.

Significance:

Employee attrition directly impacts productivity and cost. HR teams need insights to retain talent and reduce turnover.

Context:

IBM has provided a real-world HR dataset that includes various features influencing employee satisfaction and retention.

Approach:

Use machine learning techniques to discover patterns, build predictive models, and provide data-driven insights.



Internship Overview

Organization: Unified Mentor

Role: Data Science Intern

Duration: 1st June'2025 - 31st July'2025

Responsibilities:

- Worked on real-world datasets
- Built ML models to extract insights
- Delivered predictive analytics projects

Dataset Description

Dataset: IBM HR Analytics Employee Attrition & Performance

Records: 1,470 employees

Features:

- **Demographics:** Age, Gender, MaritalStatus
- **Job-related:** JobRole, Department, MonthlyIncome
- **Behavioral:** JobSatisfaction, OverTime, WorkLifeBalance
- **Target:** Attrition (Yes/No)

Data Cleaning

- Checked for missing values – none found
- Removed any duplicate rows (if present)
- Verified data types for all columns
- Cleaned inconsistent or irrelevant entries
(e.g., columns like EmployeeCount, Over18 were dropped due to uniform values)

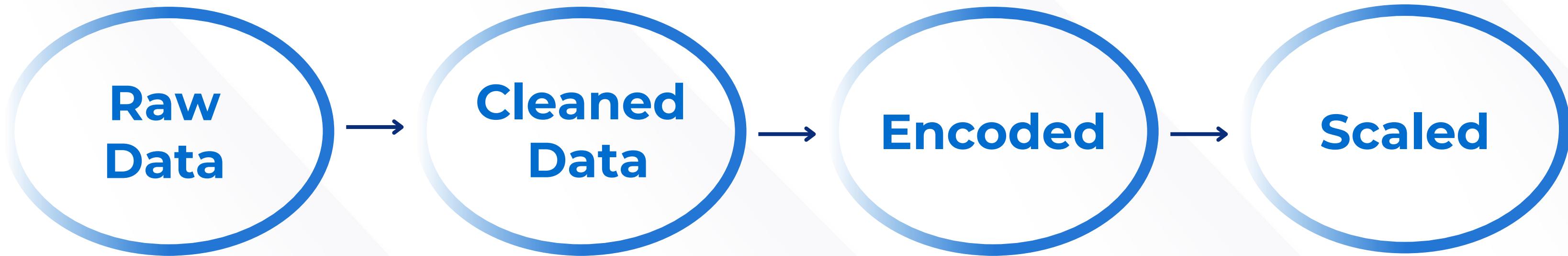
Data Preprocessing

Encoding:

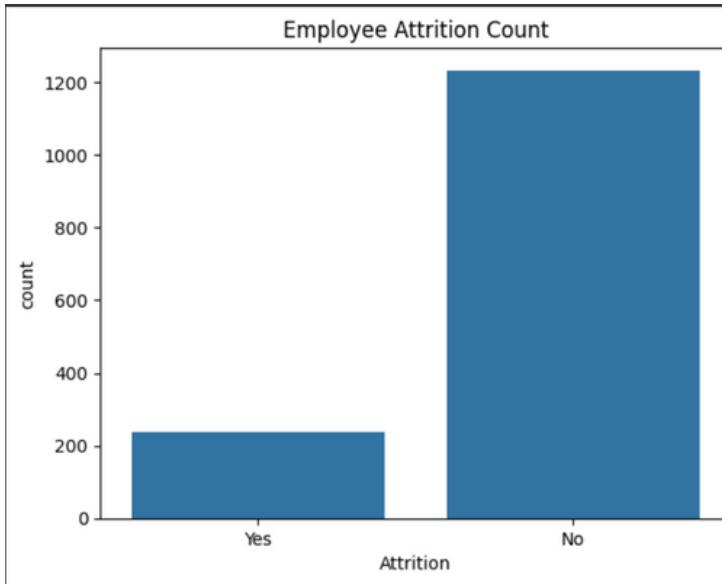
- Used Label Encoding for binary features
- Used OneHotEncoding for categorical features with multiple categories

Scaling:

Applied StandardScaler to normalize numerical features like MonthlyIncome, Age

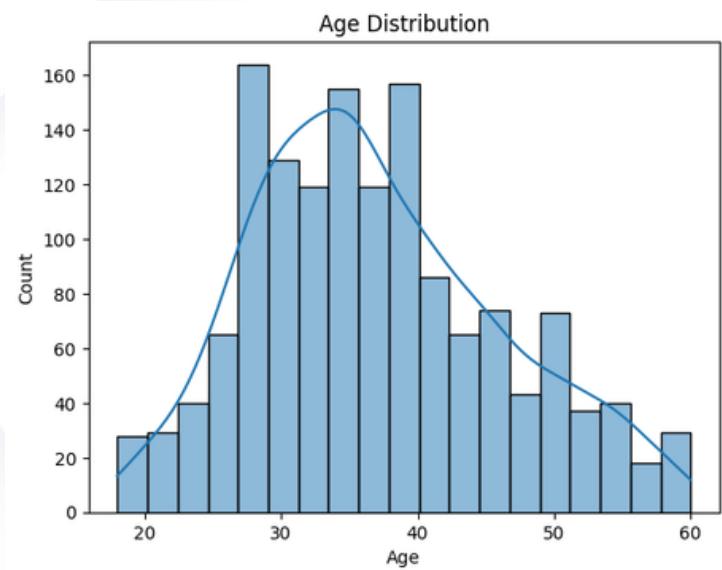


Exploratory Data Analysis (EDA)



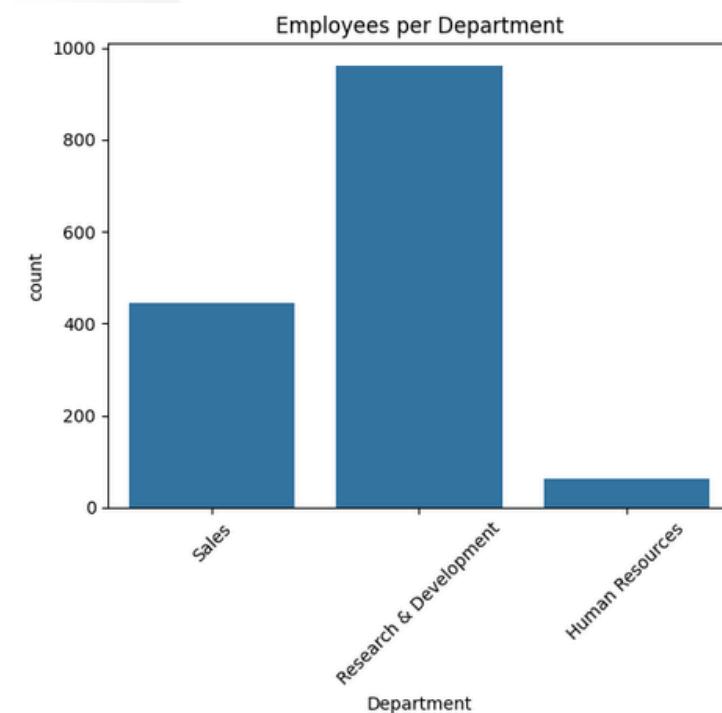
Attrition Distribution:

- Majority stayed; minority left → Imbalanced Dataset



Age Distribution:

- Younger employees had higher attrition rates



Employees per Department:

- R&D has the most employees; HR the fewest

ML pipeline :

Preprocessing → Model → Evaluation

Model Building and Evaluation

Logistic Regression

Accuracy:
0.5986394557823129
ROC AUC Score:
0.5888534757515721

Random Forest

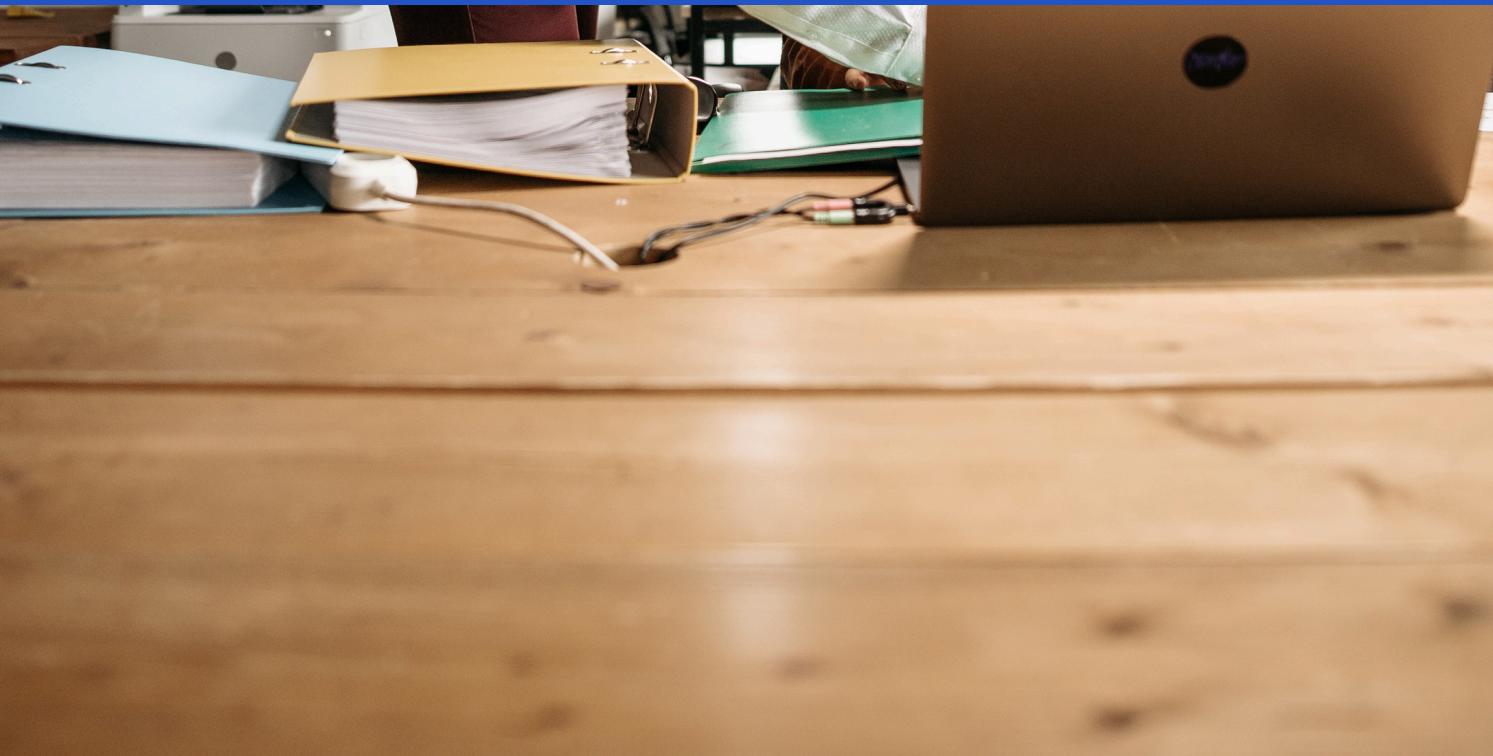
Accuracy:
0.7993197278911565
ROC AUC Score:
0.5876905848910329

XGBoost

Accuracy:
0.8129251700680272
ROC AUC Score:
0.6646998018778533



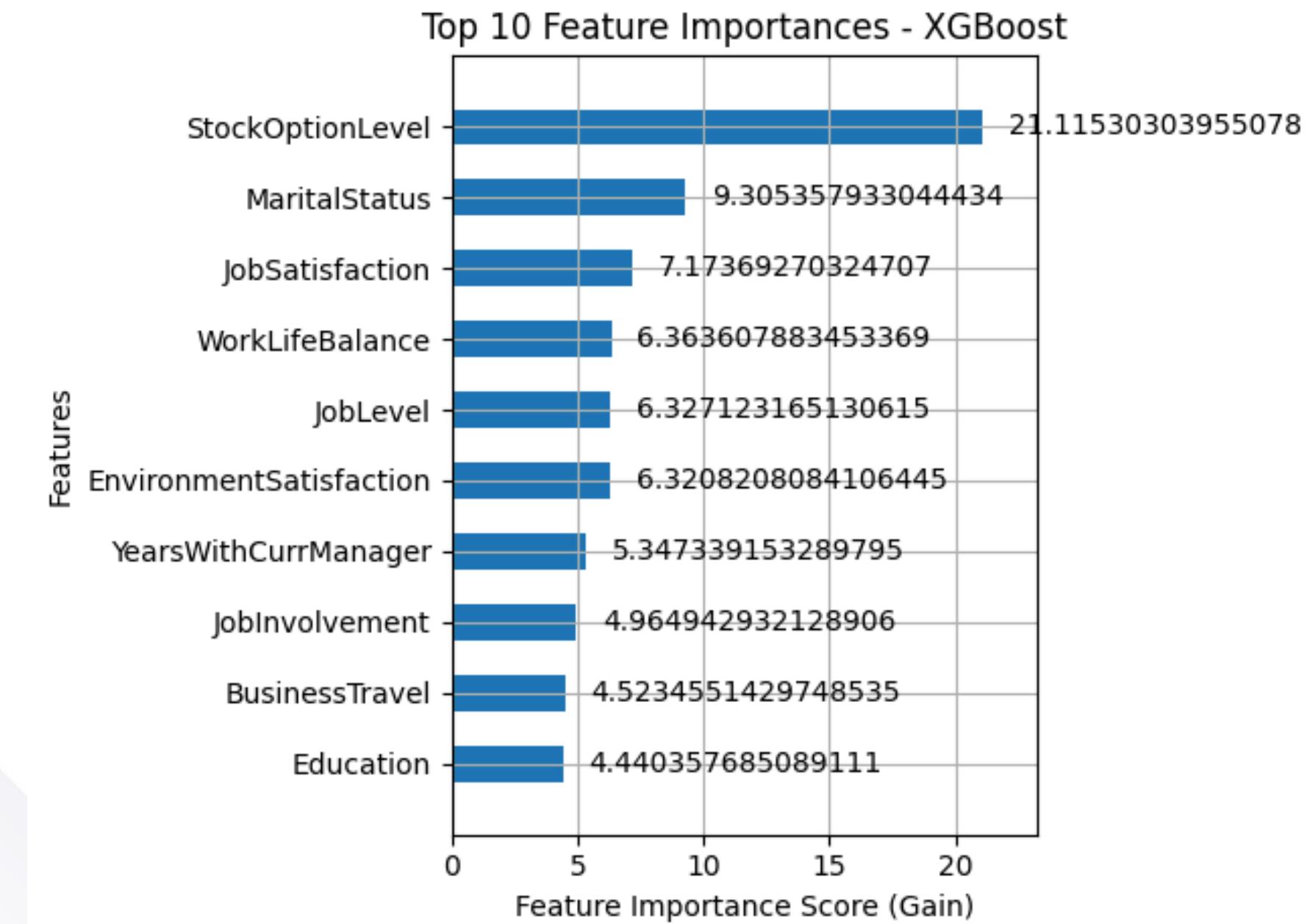
Best Model



Feature Importance



Employees with no stock options, low satisfaction, or poor work-life balance are more prone to leave.





Insights & Business Interpretation

Key Findings:

- Lack of stock options increases attrition
- Low job satisfaction & poor work-life balance lead to exits
- Single employees are more likely to leave
- Lower job levels show higher attrition

Suggestions:

- Offer stock-based incentives to boost retention.
- Prioritize work-life balance programs and employee wellness.
- Encourage career growth opportunities to improve job satisfaction.

Thank You.

Name: Srshti Jain
Email: srshtijain17@gmail.com
Project Title: IBM HR Analytics:
Employee Attrition & Performance

“Open to questions and feedback”