IBM HR Analytics Employee Attrition Analysis Report

1. Introduction

Employee **attrition** refers to the loss of employees from an organization, either voluntarily (resignation) or involuntarily (termination, retirement). High attrition rates can lead to increased recruitment and training costs, loss of institutional knowledge, and reduced morale. Therefore, understanding and predicting attrition is crucial for effective human resource management and organizational stability.

The **target variable** in this analysis is **Attrition**, a binary indicator showing whether an employee has left the company (Yes) or stayed (No).

2. Data Overview and Preprocessing

- The dataset consists of **1470 employee records** with 35 features including demographics (Age, Gender), job-related attributes (Department, JobRole, JobSatisfaction), compensation (MonthlyIncome, PercentSalaryHike), and performance indicators.
- Categorical variables such as Department, Gender, JobRole, and OverTime were encoded using label encoding or one-hot encoding to prepare for modeling.
- Analysis of the target variable showed an **imbalance**: approximately 16% of employees left (Attrition = Yes), while 84% stayed.
- To address this imbalance, **SMOTE (Synthetic Minority Over-sampling Technique)** was optionally applied during model training to generate synthetic samples of the minority class.

3. Exploratory Data Analysis (EDA)

- **Department-wise attrition rates** revealed that the **Sales** and **Research & Development** departments had higher attrition compared to Human Resources.
- **Job satisfaction** showed a strong correlation with attrition: employees with lower satisfaction scores were more likely to leave.
- **Age distribution** indicated younger employees had a higher tendency to attrite.
- Other factors such as **OverTime** (employees working extra hours) and **DistanceFromHome** (commute distance) also correlated with higher attrition.

Visualizations such as bar charts and box plots were used to illustrate these relationships clearly.

4. Model Building

- A **Random Forest Classifier** was trained using a stratified train-test split to preserve the class distribution.
- An **XGBoost Classifier** was also trained for comparison, showing slightly improved accuracy.
- Both models were tuned to optimize performance metrics.

5. Model Evaluation

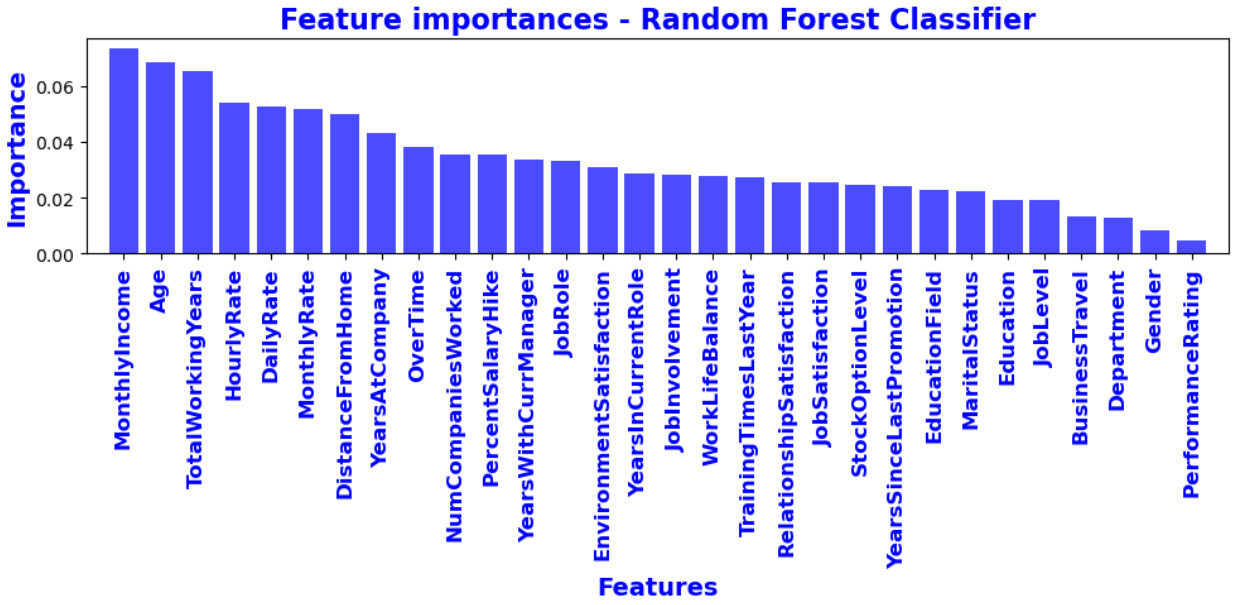
- The best performing model achieved:
    - **Accuracy:** ~85%
    - **Precision, Recall, and F1-Score:** Balanced metrics indicating good classification of both attrition and non-attrition cases.
- The **confusion matrix** showed a low false negative rate, critical for identifying employees at risk of leaving.
- The **ROC curve** demonstrated strong discrimination ability with an AUC score above 0.85.

6. Feature Importance and Explainability

The top 5 features influencing attrition prediction were:

Feature	Impact on Attrition
OverTime	Employees working overtime were more likely to leave.
JobSatisfaction	Lower job satisfaction increased attrition risk.
MonthlyIncome	Lower income correlated with higher attrition.
YearsAtCompany	Shorter tenure employees were more likely to leave.
DistanceFromHome	Longer commute distances increased attrition risk.

Bar plots from the notebook visualize these feature importances, highlighting their relative influence on the model's decisions.



7. Business Insights and Recommendations

- **Manage Workload:** Employees working overtime are at higher risk of leaving. Implement workload balancing and monitor overtime hours.
- **Enhance Job Satisfaction:** Regularly survey employees to identify dissatisfaction drivers and improve workplace environment.
- **Competitive Compensation:** Review salary structures to ensure competitive pay, especially for employees with lower monthly income.
- **Support New Employees:** Strengthen onboarding and mentoring programs to increase retention among employees with fewer years at the company.
- **Flexible Work Arrangements:** Offer options such as remote work or flexible hours to mitigate the impact of long commutes.

Focus retention efforts primarily on the Sales and Research & Development departments, which exhibit higher attrition rates.

8. Conclusion

This analysis successfully identified key predictors of employee attrition using machine learning models and provided actionable insights for HR management. By addressing the highlighted factors, IBM can reduce turnover, retain talent, and improve organizational performance.