Employee Attrition Analysis Report:

---

1. Dataset Description & Preprocessing

Dataset Overview:

The dataset consists of employee records, including work experience, salary, job satisfaction, and attrition status. The key objective is to analyze factors influencing employee attrition and build predictive models to identify at-risk employees.

Preprocessing Steps:

- **Data Cleaning:**
  - Renamed column **'Departments '** to **'departments'** (fixed trailing space issue).
  - Checked for and removed duplicate records.
  - Verified no missing values in the dataset.
- **Feature Engineering:**
  - Separated independent variables (**X**) and target variable (**y = left**).
  - Encoded categorical variables such as **salary** and **departments**.
  - Standardized numerical variables using **StandardScaler**.

---

2. Models Implemented & Selection Rationale

1️⃣ Logistic Regression

✅ Selected as a baseline model due to its interpretability and efficiency in binary classification.
✅ Helps understand the impact of features on attrition probability.

2️⃣ Decision Tree Classifier

✅ Chosen for its ability to capture complex feature interactions. ✅ Provides decision rules useful for HR teams.

3️⃣ Random Forest Classifier

✅ Improves performance by aggregating multiple decision trees. ✅ Reduces overfitting and improves generalization.

4️⃣ XGBoost Classifier

✅ Boosting technique selected for high accuracy. ✅ Handles imbalanced data well and provides feature importance scores.

---

3. Key Insights & Visualizations

Attrition Trends

- **Employees with lower satisfaction levels are more likely to leave.**
- **High attrition rates in Sales and Technical departments.**
- **Employees with low salaries have higher attrition risk.**

Salary & Attrition

- Employees with **low salaries** have the highest attrition rate.
- Competitive pay and performance-based bonuses may help retain employees.

Workload Impact

- Employees working **excessive hours** or handling too many projects tend to leave.
- Work-life balance improvements can help reduce turnover.

SHAP Feature Importance Analysis

- **Top factors influencing attrition:**
  - Satisfaction Level
  - Number of Projects
  - Salary Level
  - Promotion in Last 5 Years
  - Workload (Avg Monthly Hours)

---

4. Challenges Faced & Solutions

1️⃣ Data Cleaning Issues

- **Challenge:** Extra spaces in column names led to errors.
- **Solution:** Renamed columns properly before preprocessing.

2️⃣ Imbalanced Target Variable

- **Challenge:** More employees stayed than left, creating imbalance.
- **Solution:** Used **stratified sampling** and **balanced weighting techniques**.

3️⃣ Model Performance Variations

- **Challenge:** Logistic Regression had lower accuracy compared to tree-based models.
- **Solution:** Used ensemble models like **Random Forest and XGBoost** for improved prediction accuracy.

Conclusion & Recommendations

- **Invest in employee satisfaction programs.**
- **Improve salary structures and bonuses.**
- **Provide career growth opportunities.**
- **Enhance workload management to prevent burnout.**
- **Monitor high-risk employees using predictive models.**

This analysis provides valuable insights into attrition drivers, helping HR teams take proactive steps to reduce turnover.