**📊 HR Attrition Prediction Project Report**

**1. Introduction**

Employee attrition is a critical challenge for organizations, directly impacting productivity, recruitment costs, and overall business performance. In this project, we analyze the **IBM HR Analytics Attrition Dataset** to identify patterns, explore key factors influencing attrition, and build predictive models. The ultimate goal is to gain actionable insights that help HR teams reduce turnover.

**2. Data Collection**

* Dataset Source: **Kaggle – IBM HR Analytics Attrition Dataset**
* File: WA\_Fn-UseC\_-HR-Employee-Attrition.csv
* Shape: **1470 rows × 35 columns**

**3. Data Preprocessing**

**Actions Taken:**

1. **Dropped Useless Columns**
   * Removed: EmployeeCount, Over18, StandardHours, EmployeeNumber
   * Reason: These features have no variance or predictive power.
2. **Target Variable Transformation**
   * Converted Attrition into binary (Yes=1, No=0).
3. **One-Hot Encoding**
   * Converted categorical features into numeric form (dropping first category to avoid multicollinearity).
4. **Imbalanced Data Handling**
   * Problem: Attrition is imbalanced (No: ~84%, Yes: ~16%).
   * Solution: Applied **SMOTE (Synthetic Minority Over-sampling Technique)** on training data.

✅ Balanced dataset achieved: equal representation of Attrition = Yes and No.

**4. Exploratory Data Analysis (EDA)**

* **Attrition Rate**:
  + No: 83.9%
  + Yes: 16.1%
* **Key Categorical Insights**:
  + More attrition among employees working overtime.
  + Sales and Research & Development have higher turnover compared to HR.
  + Single employees showed higher attrition compared to married.
* **Key Numerical Insights**:
  + Younger employees had higher attrition.
  + Lower job satisfaction and lower monthly income correlated with leaving.

**5. Modeling & Evaluation**

**Models Used:**

1. **Logistic Regression**
   * Provides interpretable coefficients.
   * Top Features: OverTime\_Yes, JobSatisfaction, MonthlyIncome, MaritalStatus\_Single.
2. **Decision Tree (max\_depth=4)**
   * Clear decision rules for HR teams.
   * Example Rule: Employees with low job satisfaction **and** overtime are more likely to leave.
3. **Random Forest (200 trees)**
   * Best predictive performance.
   * Accuracy: **82%**
   * Recall for Attrition=Yes: **26%** (still challenging due to imbalance, but improved with SMOTE).

**Top 10 Features (Random Forest):**

* + Stock Option Level
  + Monthly Income
  + Job Satisfaction
  + Marital Status (Married)
  + Years with Current Manager
  + Age
  + Environment Satisfaction
  + Monthly Rate
  + Job Involvement
  + Department (R&D)

**6. Insights & Business Implications**

* **Financial Incentives Matter**: Employees with lower income and stock option levels are more likely to leave.
* **Work-Life Balance**: Overtime strongly correlates with attrition. HR should monitor workloads.
* **Satisfaction Factors**: Job and environment satisfaction rank high in importance. Improving these can lower attrition.
* **Demographics**: Younger and single employees are more likely to quit—suggesting career growth and engagement strategies are needed.

**7. Conclusion**

This analysis demonstrates that **employee attrition can be partially predicted** using machine learning. Random Forest performed best, identifying financial rewards, satisfaction, and workload balance as critical drivers of turnover.