**Case Study: Predicting Employee Attrition**

**Overview:** Employee attrition poses significant challenges for organizations, including financial costs, loss of expertise and disruption to team dynamics. On average, replacing an employee costs an organization 50-60% of their annual salary. HR departments are often under pressure to identify employees at risk of leaving and to implement targeted interventions to improve retention.

Imagine you are an HR consultant tasked with analyzing employee data for a Fortune 500 company. Your goal is to develop a data-driven solution to predict employee attrition and provide actionable recommendations to retain top talent.

Your task is to analyze the IBM HR Analytics dataset and develop a machine learning model to predict employee attrition. Beyond prediction, the goal is to uncover actionable insights from the data to help the HR team understand the key factors driving attrition and devise effective strategies to address them.

The success of your solution will be judged not only on the accuracy of the predictive model but also on your ability to explain the drivers of attrition and provide recommendations that can be implemented in a business setting.

**Objective:** The primary objective of this case study is to build a machine learning model that:

1. Accurately predicts whether an employee will leave the company (Attrition: Yes/No).
2. Provides actionable insights and explainability to help HR teams understand key factors influencing attrition.

**Dataset:** The dataset provided (IBM HR Analytics Employee Attrition & Performance) contains the following information:

**Key Variables:**

1. **Age**: The age of the employee.
2. **Attrition**: Indicates whether the employee left the company (Yes/No).
3. **BusinessTravel**: Frequency of business travel (Non-Travel, Travel\_Rarely, Travel\_Frequently).
4. **Department**: The department where the employee works (e.g., Sales, Research & Development).
5. **DistanceFromHome**: Distance from the employee's home to their workplace.
6. **EducationField**: Field of education (e.g., Life Sciences, Medical, Marketing).
7. **Gender**: Gender of the employee (Male/Female).
8. **JobRole**: The specific role of the employee (e.g., Sales Executive, Manager).
9. **JobSatisfaction**: Satisfaction level with the job (1 = Low, 4 = Very High).
10. **MonthlyIncome**: Monthly income of the employee.
11. **YearsAtCompany**: Number of years the employee has worked at the company.
12. **WorkLifeBalance**: Work-life balance satisfaction (1 = Bad, 4 = Best).

**Target Variable:**

* **Attrition**: Indicates whether an employee left the company (Yes/No).

For a full list of variables, please refer to the dataset file.

**Tasks:**

1. **Exploratory Data Analysis (EDA):**
   * Analyze and visualize the dataset to understand trends and relationships between variables.
   * Identify and handle missing values or anomalies, if any.
   * Summarize key insights from the data.
2. **Preprocessing:**
   * Encode categorical variables and scale numerical variables as required.
   * Handle imbalanced data if necessary.
3. **Model Development:**
   * Apply machine learning techniques to predict attrition.
   * Evaluate multiple models and select the best-performing one based on metrics such as accuracy, precision, recall and F1-score.
4. **Explainability and Insights:**
   * Use techniques such as SHAP (SHapley Additive exPlanations) or feature importance to identify key drivers of attrition.
   * Provide actionable recommendations based on your findings.
5. **Advanced Insights (Optional):**
   * Explore clustering to segment employees and understand unique patterns within specific groups.
   * Perform correlation analysis to identify strong relationships between variables.
6. **Presentation:**
   * Prepare a concise 5-slide PowerPoint presentation covering:
     1. Problem Statement and Objective.
     2. Key Insights from EDA.
     3. Model Development and Performance Metrics.
     4. Explainability Insights and Key Drivers.
     5. Recommendations and Conclusion.

**Deliverables:**

1. **Code:**
   * Submit a well-documented Python notebook or script with all steps clearly outlined.
2. **Presentation:**
   * A 3-4 slide PowerPoint presentation summarizing your approach, findings and recommendations.