**Project Summary: Employee Retention Analysis Using Python**

**Objective**

The goal of this project is to analyze employee satisfaction, performance, and workplace factors to predict employee retention. By leveraging machine learning techniques in Python, we aim to identify key indicators that influence employee attrition and develop predictive models to help organizations improve retention strategies.

**Dataset Overview**

The dataset includes the following attributes:

* **satisfaction\_level**: Employee job satisfaction level (0 to 1).
* **last\_evaluation**: Last performance evaluation score (0 to 1).
* **number\_project**: Number of projects the employee has worked on.
* **average\_montly\_hours**: Average monthly hours worked by the employee.
* **time\_spend\_company**: Number of years the employee has been with the company.
* **Work\_accident**: Whether the employee had a work-related accident (0 = No, 1 = Yes).
* **left**: Target variable indicating whether the employee left the company (0 = No, 1 = Yes).
* **promotion\_last\_5years**: Whether the employee was promoted in the last five years (0 = No, 1 = Yes).
* **Department**: The department in which the employee works.
* **salary**: Employee’s salary level (Low, Medium, High).

**Methodology**

1. **Data Preprocessing:**
   * Handling missing values, if any.
   * Encoding categorical variables (Department and salary).
   * Scaling numerical features for machine learning models.
2. **Exploratory Data Analysis (EDA):**
   * Analyzing correlations between features and employee attrition.
   * Visualizing distributions of satisfaction levels, working hours, and promotions.
   * Identifying trends in employee turnover across departments and salary levels.
3. **Handling Imbalance (if applicable):**
   * Using techniques like **SMOTE** to balance the dataset if the target variable is imbalanced.
4. **Feature Selection:**
   * Identifying the most relevant factors contributing to employee attrition using statistical tests or feature importance methods.
5. **Model Development:**
   * Training classification models such as **Random Forest, Gradient Boosting, Support Vector Machines (SVM), and Neural Networks**.
   * Evaluating model performance using metrics like **accuracy, precision, recall, and F1-score**.
6. **Interpretability & Insights:**
   * Identifying key drivers of employee turnover using **SHAP values** or feature importance analysis.
   * Providing actionable recommendations to HR departments based on model results.

**Expected Outcomes**

* A machine learning model that can predict employee attrition with high accuracy.
* Insights into which factors most strongly influence employee retention.
* Data-driven recommendations for improving employee satisfaction and reducing turnover.

**Tools & Technologies**

* **Python:** Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn
* **Machine Learning:** Random Forest, Gradient Boosting, SVM, Neural Networks
* **Data Processing:** SMOTE (if needed), Feature Scaling, One-Hot Encoding
* **Model Evaluation:** Precision, Recall, F1-Score, ROC Curve