Employee Attrition Prediction Project Report

# 1. Objective

The objective of this project was to build a classification model to predict whether an employee would leave the company (attrition = Yes) based on various HR-related features, and to extract actionable insights that HR teams can use to reduce employee turnover.

# 2. Dataset Used

• Dataset: IBM HR Analytics Employee Attrition & Performance

• Source: Kaggle or IBM Sample Data

• Target Variable: Attrition (Yes/No)

# 3. Tools & Technologies

• Programming Language: Python

• Libraries: Pandas, NumPy, Seaborn, Matplotlib, Scikit-learn, SHAP

• IDE: Google Colab

# 4. Steps Followed

4.1 Data Cleaning & Preprocessing

Removed irrelevant features, encoded categorical variables using one-hot encoding, scaled numerical features, and handled class imbalance with stratified sampling.

4.2 Exploratory Data Analysis (EDA)

Identified patterns such as high attrition among employees working overtime, those with low income, or short tenure. Used visual tools like countplots, boxplots, and heatmaps.

4.3 Model Building

Trained Logistic Regression and Random Forest models. Evaluated them using accuracy, F1-score, precision, recall, and ROC-AUC.

4.4 Model Explainability

Used SHAP (SHapley Additive Explanations) to visualize global and local feature importances.

# 5. Personal Findings

Most important features contributing to attrition included OverTime, JobRole, EnvironmentSatisfaction, MonthlyIncome, and JobSatisfaction. SHAP helped visualize how each feature contributed to individual predictions.

# 6. Problems Faced & Counter-Actions

• Imbalanced classes: Used stratified sampling and considered SMOTE.  
• SHAP force\_plot error: Fixed by using .iloc[[0]] instead of .iloc[0].  
• Many categorical variables: Used one-hot encoding with drop\_first=True.  
• Visual clutter in EDA: Focused only on top correlated features.  
• SHAP plots not rendering: Used shap.initjs() and static image export.

# 7. Key Insights & Recommendations

• High attrition with OverTime = Yes: Implement better shift management.  
• Low income drives attrition: Offer raises or equity.  
• Short tenure employees attrite more: Enhance onboarding and mentoring.  
• Low satisfaction scores: Improve feedback cycles and work culture.

# 8. Outcome

Built an explainable, high-performing attrition prediction model. Identified HR strategies to reduce attrition. Demonstrated full machine learning pipeline from data cleaning to interpretability.

# 9. Future Work

• Implement SMOTE or ADASYN to balance data.  
• Test XGBoost, CatBoost models.  
• Deploy using Flask or Streamlit.  
• Add SHAP dashboard in the web app.