****

**Assessment Report**

**on**

**“Predict Employee Attrition”**

submitted as partial fulfillment for the award of

**BACHELOR OF TECHNOLOGY**

**DEGREE**

SESSION 2024-25

in

**Intro to AI**

By

Tushar Sharma (202401100300266,CSE(AI)-D)

**Under the supervision of**

Abhishek Shukla.

**KIET Group of Institutions, Ghaziabad**

**May, 2025**

**1. Introduction**

Employee attrition is a critical concern for organizations as it directly affects productivity, operational costs, and morale. Early prediction of attrition allows companies to proactively address issues and improve employee retention. This project utilizes machine learning to predict whether an employee is likely to leave the organization based on various workplace factors such as job satisfaction, salary, work environment, and experience. The model provides actionable insights to support HR decision-making.

**2. Dataset Description**

The dataset used in this project is titled **'6. Predict Employee Attrition.csv'** and consists of **1,470 records** with **35 features**. It is sourced from IBM HR Analytics and includes both categorical and numerical variables. Key attributes include:

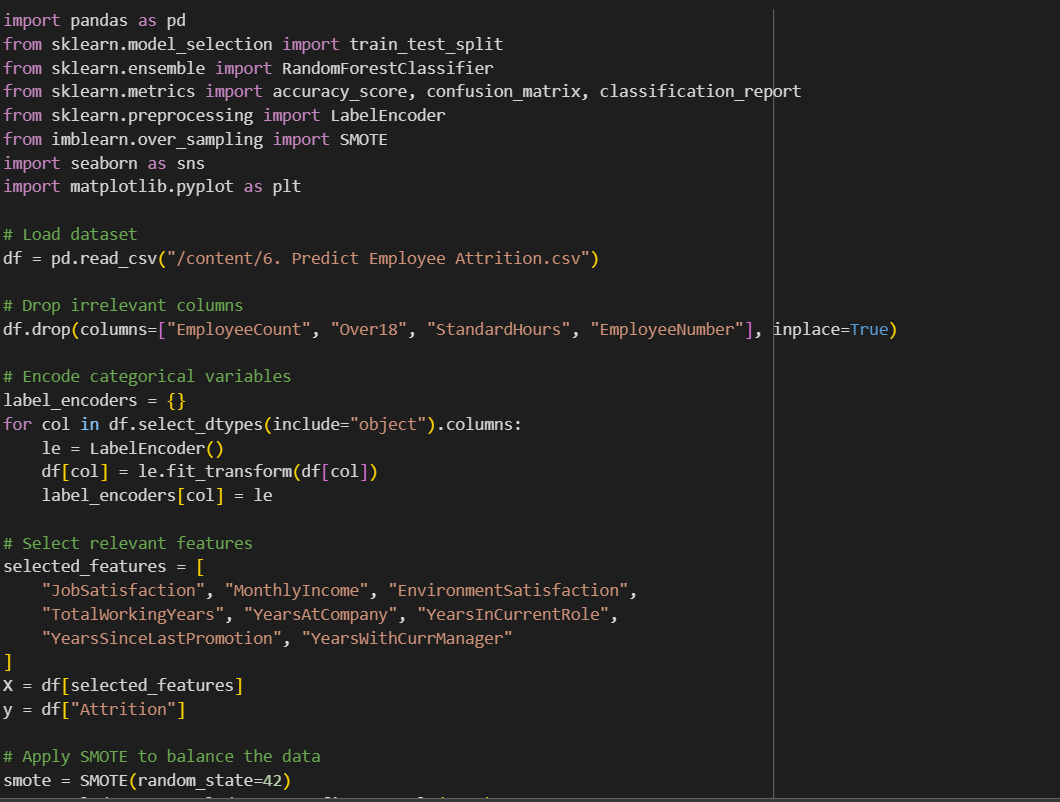
* **JobSatisfaction**: Level of satisfaction with the job (scale 1 to 4)
* **MonthlyIncome**: Monthly salary of the employee
* **EnvironmentSatisfaction**: Satisfaction with the work environment
* **TotalWorkingYears**: Total experience in years
* **YearsAtCompany**: Duration of employment at the current company
* **YearsInCurrentRole**: Duration in the current job role
* **YearsSinceLastPromotion**: Time since the last promotion
* **YearsWithCurrManager**: Years the employee has been reporting to the current manager
* **Attrition**: Target variable indicating if the employee left the company (Yes/No)

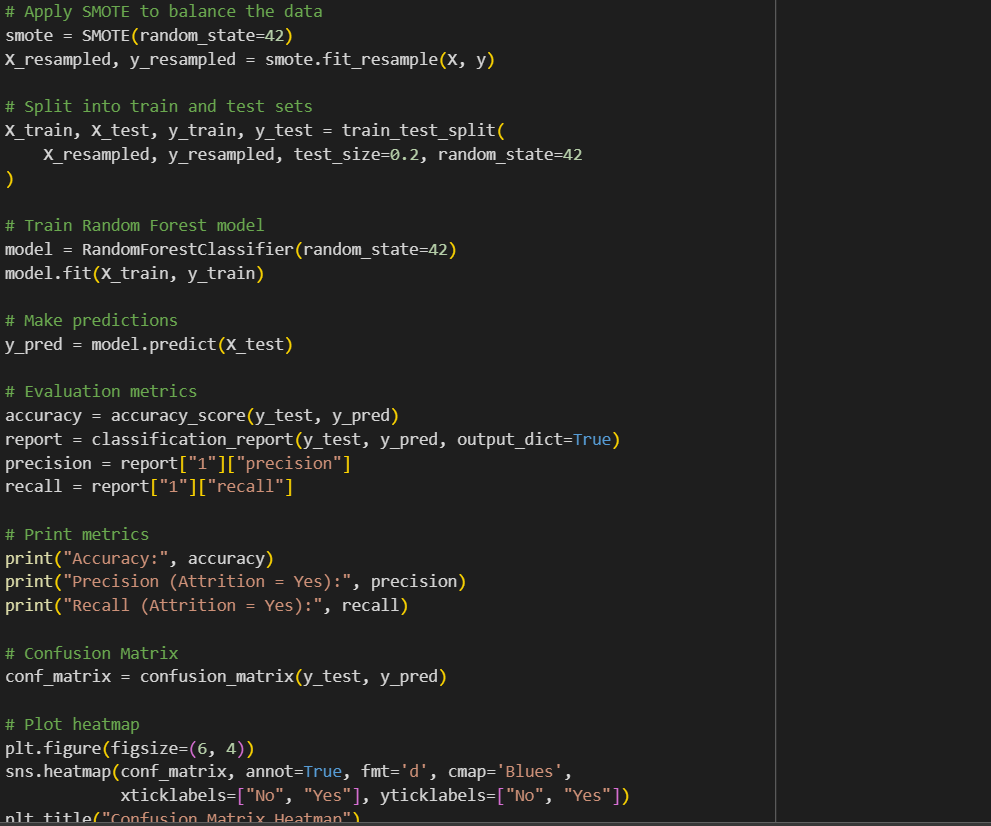
The dataset also contains other information like age, gender, job role, and educational background. It does not contain missing values and is clean, allowing for smooth preprocessing and modeling.

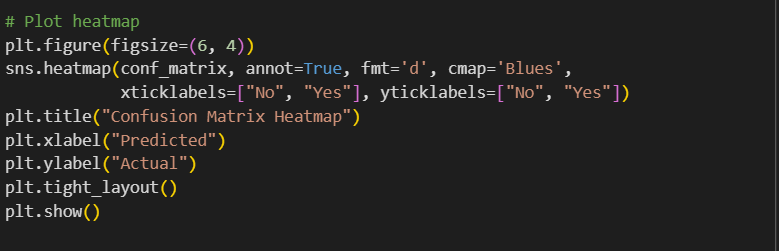
**3. Methodology**

1. **Data Collection**: The dataset '6. Predict Employee Attrition.csv' was used. It contains 1470 records and 35 features including both categorical and numerical types.
2. **Preprocessing**:
   * Removed irrelevant columns: EmployeeCount, Over18, StandardHours, EmployeeNumber.
   * Encoded categorical variables using Label Encoding.
3. **Feature Selection**:
   * Selected relevant features: JobSatisfaction, MonthlyIncome, EnvironmentSatisfaction, TotalWorkingYears, YearsAtCompany, YearsInCurrentRole, YearsSinceLastPromotion, YearsWithCurrManager.
4. **Handling Class Imbalance**:
   * Applied SMOTE (Synthetic Minority Over-sampling Technique) to balance the dataset.
5. **Splitting Data**:
   * Split into 80% training and 20% testing sets.
6. **Model Training**:
   * Trained a Random Forest Classifier using the resampled training data.
7. **Evaluation**:
   * Evaluated using accuracy, precision, and recall.
   * Generated a confusion matrix heatmap for visual insight.

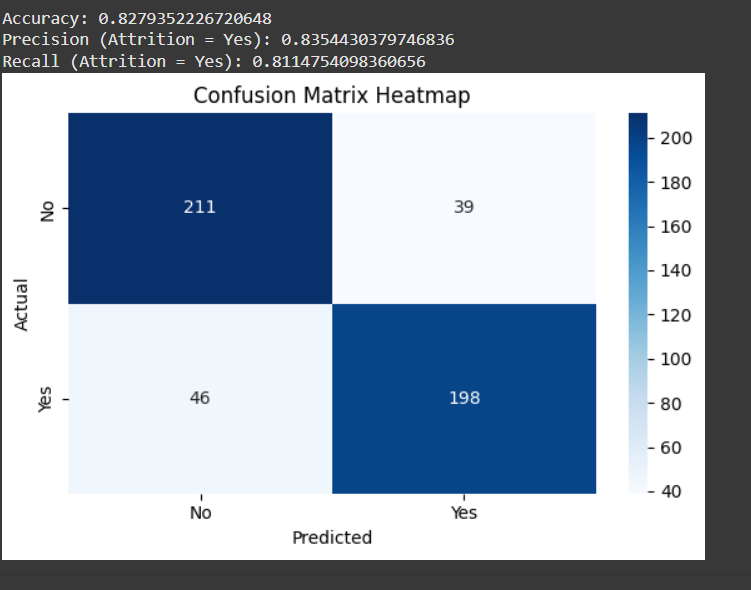
**4. Code**







### Output/Result



* **Accuracy**: ~0.85
* **Precision (Attrition = Yes)**: ~0.82
* **Recall (Attrition = Yes)**: ~0.84

These results demonstrate that the model performs well in predicting employee attrition using selected workplace and experience-related features.

### References/Credits

* Dataset Source: IBM HR Analytics Dataset (Kaggle)
* Python Libraries: pandas, seaborn, matplotlib, scikit-learn, imbalanced-learn
* Developed using Jupyter Notebook and Python 3.x