**Task Guideline**

**About the project**

- The analysis is to predict employee churn using the HR dataset.

**Columns in the dataset**

**Satisfaction level:** Employee satisfaction rating

**Last evaluation:** Performance rating

**Number of projects:** Number of projects involved in

**Average monthly hours:** Average working hours

**Time spend company:** Years in the company

**Work accident:** Whether involved in a work accident

**Left:** Whether the employee left the company or is still in the company

**Promotion last 5 years:** Whether promoted in the last 5 years

**Department:** Department of the employee

**Salary:** Salary level (low, medium, high)

# **Steps Followed in the Analysis**

## **1. Loading the Dataset:**

- Imported necessary libraries such as `pandas`, `numpy`, and various `sklearn` modules.

- Loaded the dataset `HR\_Dataset.csv`.

## **2. Data Exploration and Cleaning:**

- Displayed a sample of the dataset and checked the column names.

- Renamed the column "Departments " to "departments".

- Checked the shape, information, and overall statistics of the dataset.

- Identified and handled null values.

- Checked for and removed duplicates from the dataset.

## **3. Data Visualization:**

- Plotted the distribution of the target variable "left".

## **4. Feature Engineering**:

- Created a feature matrix `X` and the response variable `y`.

## **5. Data Preprocessing:**

- Implemented pipelines for scaling numerical features and encoding categorical features using `StandardScaler`, `OneHotEncoder`, and `OrdinalEncoder`.

- Combined these preprocessing steps using `ColumnTransformer`.

## **6. Model Training:**

- Split the dataset into training and testing sets using `train\_test\_split`.

- Trained multiple models including `LogisticRegression`, `DecisionTreeClassifier`, `RandomForestClassifier`, and `XGBClassifier`.

## **7. Model Evaluation:**

- Evaluated the models using accuracy, precision, and recall scores.

**Findings and Recommendations**

# **1. Introduction:**

- The analysis is to predict employee churn using the HR dataset.

# **2. Data Overview:**

- The dataset contained information about employees, including their departments, salaries, and whether they left the company.

**The dataset contains 14999 rows and 9 Columns**

# 

# **4. Exploratory Data Analysis:**

- Initial exploration revealed that a significant proportion of employees had left the company.

- I visualised the distribution of the target variable to understand the churn rate 1991 employees had already left the company.

.

# **5. Feature Engineering:**

- Numerical features were scaled using StandardScaler.

, - categorical features were encoded to

-Encoding ensures compatibility with machine learning algorithms.

# 

# **6. Model Training and Evaluation:**

- I split the data into training and testing sets to evaluate model performance.

- Several models were trained: Logistic Regression, Decision Tree, Random Forest, and XGBoost.

- Models were evaluated using accuracy, precision, and recall metrics to determine their effectiveness.

# 

# **7. Findings:**

- Logistic Regression: Provided a baseline with moderate accuracy but loIr precision and recall.

- Decision Tree: ShoId high variance, indicating potential overfitting.

- Random Forest: Balanced model with better accuracy, precision, and recall compared to Logistic Regression and Decision Tree.

- XGBoost: Outperformed other models with the highest accuracy, precision, and recall, making it the best model for predicting churn.

# **8. Recommendations:**

- Model Deployment: Based on our evaluation, I recommend deploying the XGBoost model for predicting employee churn.

- Further Analysis: Conduct deeper analysis on feature importance to identify key factors contributing to churn. This can inform HR strategies to improve employee retention.

- Regular Monitoring: Implement a system to regularly monitor model performance and update it with new data to maintain accuracy.

By implementing these recommendations, I can better predict employee churn and develop targeted strategies to retain valuable employees, ultimately reducing turnover costs and improving organizational stability.