**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.