

## Data Analytics

### Assignment 1

#### Practical Question: Machine Learning Using Decision Tree on Employment Dataset

##### Objective:

You are provided with an **Employment Dataset** containing information about candidates who applied for jobs. Your task is to build a **Decision Tree Classification Model** to **predict whether a candidate should be employed or not** based on various features.

<https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset>

Form groups with a minimum of 4 and a maximum of 6 members to complete the task.

---

---

##### Dataset Description

Each row in the dataset represents a job applicant. The dataset includes the following features:

- age  
The age of the employee in years.
- education\_level  
The highest education level attained by the employee (e.g., High School, Bachelor's, Master's, PhD).
- years\_of\_experience  
Total number of years the employee has worked professionally.
- technical\_test\_score  
Score obtained by the employee in a technical assessment (out of 100).
- interview\_score  
Score obtained by the employee during the interview process (out of 10).
- previous\_employment  
Whether the employee had previous employment experience (Yes/No).
- suitable\_for\_employment (*Target*)  
Indicates if the candidate is suitable for employment (Yes/No).

## Tasks to Perform:

### 1. Data Loading and Exploration

- Load the dataset using Python libraries (e.g., pandas).
- Display the first few rows of the dataset.
- Perform basic EDA (Exploratory Data Analysis): Check for null values, data types, and distribution of features.

### 2. Data Preprocessing

- Convert categorical variables into numeric format (e.g., one-hot encoding or label encoding).
- Split the dataset into **training** and **testing** sets (e.g., 80% train, 20% test).

### 3. Model Building

- Train a **Decision Tree Classifier** using the training data to predict `suitable_for_employment`.

### 4. Model Visualization

- Visualize the decision tree using appropriate tools like `plot_tree()` or `graphviz`.

### 5. Model Testing and Prediction

- Predict the labels for the test dataset.
- Test the model using at least 3 hypothetical candidate profiles and interpret the predictions.

### 6. Model Evaluation

- Evaluate the model using:
  - **Accuracy Score**
  - **Confusion Matrix**
  - **Classification Report** (Precision, Recall, F1-Score)

## Bonus Task (Optional):

- Perform **feature importance analysis** to determine which features contribute most to the employment decision.

### **Required Libraries:**

pandas, numpy, sklearn, matplotlib, seaborn

### **Expected Output:**

- Clear and well-commented Python code
- Visualized decision tree
- Model performance metrics
- Interpretation of predictions