**Mid Term Exam Batch 7,8**

**Machine Learning**

**Task 1:**

**Step 1: Dataset Loading and Understanding**

1. Load the dataset into a Pandas DataFrame.
   * Verify the dataset's structure using .info() and .head().
   * Identify the target variable (e.g., depression levels).
   * Check for missing values.

**Step 2: Data Preprocessing**

1. Handle missing or null values in the dataset:
   * Replace them with mean/median/mode or drop the rows/columns as appropriate.
2. Encode categorical features (if any) using techniques like encoding.
3. Normalize or standardize the numerical features.

**Step 3: Exploratory Data Analysis (EDA)**

1. Perform basic EDA to understand the data:
   * Plot the distribution of the target variable.
   * Create visualizations (e.g., histograms, boxplots, pair plots) to analyze feature relationships.
   * Boxplots to compare depression levels across categorical variables (e.g., gender, study habits, etc.).
   * Scatter plots to analyze the relationship between numerical features (e.g., hours of sleep vs. depression levels).
   * Bar plots to compare the depression levels across different categories (e.g., academic performance levels).
   * Calculate correlations between features.
2. Identify key patterns or trends in the dataset that may help in prediction.

**Step 4: Model Building and Training**

1. Split the dataset into training (70%) and testing (30%) subsets.
2. Train a classification model on the training set.

**Step 5: Model Evaluation**

1. Evaluate the model on the testing set using:
   * Accuracy
   * Confusion Matrix
   * Precision, Recall, and F1-score
2. If performance is not satisfactory, apply techniques like hyperparameter tuning or feature engineering.

**Step 6: Generate Html file using Profiling Libarary**

**Task 2:**

* Discuss the following (Any 3)
* Overfitting
* Under fitting
* Bias
* Variance

**Task 3:**

#### ****Step 1: Dataset Loading and Inspection****

1. Load the dataset (Placement\_Data\_Full\_Class.csv) into a Pandas DataFrame.
   * Display the first few rows using .head() to understand the structure.
   * Use .info() and .describe() to summarize the dataset.
   * Identify the target variable (salary) and feature variables.

#### ****Step 2: Data Preprocessing****

1. **Handle Missing Data:**
   * Identify and handle missing or null values in the **salary** column and other features.
   * Explain your approach (e.g., replacing with median or removing rows).
2. **Feature Engineering:**
   * Encode categorical variables (e.g., gender, specialization, etc.) using one-hot encoding or label encoding.
3. **Data Cleaning:**
   * Handle outliers in numerical columns (e.g., salary) using methods like capping or removal.
   * Normalize or standardize numerical features (if required).

#### ****Step 3: Exploratory Data Analysis (EDA)****

1. Perform EDA to understand patterns in the dataset:
   * Visualize the distribution of the **salary** column.
   * Analyze relationships between features and salary using scatter plots, boxplots, and correlation heatmaps.
   * Highlight key insights from the dataset.

#### ****Step 4: Model Building****

1. **Data Splitting:**
   * Split the dataset into training (80%) and testing (20%) subsets.
2. **Model Training:**
   * Train model.

#### ****Step 5: Model Evaluation****

1. Evaluate the models on the testing set using:
   * Mean Absolute Error (MAE)
   * Mean Squared Error (MSE)
   * Root Mean Squared Error (RMSE)
   * R-squared (R²)