**Exam 1**

**Duration:** 60 minutes  
 **Total Marks:** 100

**Section A – Coding (70 Marks – 10 Questions)**

*Write your answers using Python syntax where applicable. You can use Pandas, Seaborn, Matplotlib, or sklearn as needed. (write the code answers in google colab and then send the file with exam)*

**1. (7 marks)**

**Load the Penguin dataset from penguins.csv, and rename the columns to:  
['species', 'island', 'bill\_length', 'bill\_depth', 'flipper\_length', 'body\_mass', 'sex'].  
Display the first 5 rows.**

**2. (7 marks)**

**Use LabelEncoder from sklearn.preprocessing to encode the 'species' column in the penguins dataset. Show the updated DataFrame.**

**3. (7 marks)**

**Apply OneHotEncoder from sklearn.preprocessing on the 'island' column and convert the result to a DataFrame.**

**4. (7 marks)**

**Normalize the numerical columns of the penguins dataset using MinMaxScaler. Show the resulting normalized DataFrame.**

**5. (7 marks)**

**Standardize the numerical columns using StandardScaler from sklearn.preprocessing. Show the output.**

**6. (7 marks)**

**Check for missing values in the penguins dataset. Then use SimpleImputer with strategy 'mean' to fill them. Show the DataFrame after imputation.**

**7. (7 marks)**

**Use KBinsDiscretizer with strategy 'quantile' and 5 bins on the flipper\_length column. Show the transformed output.**

**8. (7 marks)**

**Split the penguins dataset using train\_test\_split (70% train, 30% test), with species as the target. Show X\_train.head() and Y\_train.head().**

**9. (7 marks)**

**Create a scatter plot using seaborn.scatterplot() with x='flipper\_length', y='body\_mass', and hue='species' on the penguins dataset.**

**10. (7 marks)**

**From the Titanic dataset, extract Title from the Name column using .str.extract(), encode it using LabelEncoder, and show the updated DataFrame.**

**Section B – Theory (30 Marks – 5 Questions) (for the MCQ Highlight the correct answer with yellow color)**

**11. (6 Marks)**

**What is the difference between AI, ML, and DL?**  
**Answer:**

**12. (6 Marks)**

**What are the 4 types of machine learning? Give one real-world example for each.**  
**Answer:**

**13. (6 Marks)**

**True or False:**  
“Label encoding is better than one-hot encoding in all cases.”  
**Answer:**

* True
* False

**14. (6 Marks)**

**What are the steps in the machine learning lifecycle? List them.**  
**Answer:**

**15. (6 Marks)  
What are the three most commonly used train-test split ratios in machine learning? Write both the training and testing percentages for each.  
Answer:**