**EXAM 2**

**1. Load the Titanic dataset from CSV, and rename the columns if needed. Show the first 5 rows.**

**2. Use LabelEncoder to encode the 'Sex' column in the Titanic dataset. Show the updated column.**

**3. Fill missing values in the 'Age' column using the mean strategy with SimpleImputer.**

**4. Use MinMaxScaler to normalize the 'Fare' column and show the first 5 values.**

**5. Plot a histogram of the 'Age' column using matplotlib.pyplot. Label the axes.**

**6. Split the Titanic dataset into train and test sets using test\_size=0.3, with 'Survived' as the target. Show X\_train.head().**

**7. Extract the passenger title from the 'Name' column using .str.extract() and show the new 'Title' column.**

**8. Create a new column 'IsAlone' that equals 1 if SibSp and Parch are both 0, otherwise 0. Show the column.**

**9. From the Titanic dataset, create a bar chart using seaborn showing the average survival rate per 'Pclass'.**

**10. Load the Penguins dataset, drop the 'sex' column, and use StandardScaler on the numeric columns. Show the result.**

**✅ Section B – Theory Questions (30 Marks)**

**11. What does train\_test\_split() do in machine learning? Why do we use it?**  
train-to train the computer

Test –to test the computer

We use it to devide part test and part train

**12. List 4 types of machine learning and give one real-world example of each.**

**Supervised-**

**unsupervised - fan nadi elahly**

**reinforcement –self driving car**

**semi supervised- NLP**

**13. In the Titanic dataset, which three features are commonly used to help predict survival, and why might they be useful?**

**14. What is the purpose of SimpleImputer? When is it used?**

**To put mean in the place of missing data**

**When there is missing data**

**15. What are three common train-test split ratios? Write the training and testing percentages.**

70% 30%

80% 20%

50% 50%