

Green University of Bangladesh
Department of Computer Science & Engineering
CLP - 1, Time: 40 minutes

Course Code: CSE 424

Course Title: Data Mining Lab

Total: 15 Marks

1. Load the dataset named “titanic.csv” into a DataFrame and print the first 8 rows of the DataFrame. **2 Marks**

2. Drop the following attributes from the DataFrame, print the first 5 rows of the modified DataFrame and also print the information about the modified DataFrame. 1) Name 2) Ticket 3) PassengerId 4) Fare 5) Cabin 6) Embarked. **3 Marks**

3. Create a new modified DataFrame and fill up the missing values of the ‘Age’ attribute column, using the mode value of the ‘Age’ attribute column. Print the information about the modified DataFrame. **5 Marks**

4. Create a new column named ‘Categorical_Age’ in the DataFrame. Use the values from the ‘Age’ column to fill out the values of the ‘Categorical_Age’ column. Use the partitioning information given below:

Categorizing Age into 4 Categories

Baby/Toddler: (0,4], 0 is excluded & 4 is included

Child: (4,18], 4 is excluded & 18 is included

Adult: (18,64], 18 is excluded & 64 is included

Elderly: (64,100], 64 is excluded & 100 is included

Print the information about the modified DataFrame.

5 Marks