



Department of Computer Science & Information Technology,
University of Engineering & Technology Peshawar.

CS 318 - Data Science (Fall 2021)
DUE DATE: March 4th, 2022

Term Project
TOTAL POINTS : 10

Important:

- Submit the hand written solution of the assignment by the deadline.
- The total absolute marks of the term project are 10.
- In case two or more submissions are found similar, zero marks will be awarded.

Question 1

10 Points

Download Data from the following url;

<https://archive.ics.uci.edu/ml/datasets/Arrhythmia>

and perform the following tasks;

1. State the problem that is being given on the url.
2. Read the data (from arrhythmia.data) as dataframe and assign it column names based on the arrhythmia.names file
3. Identify which columns has missing values and solve the missing value. State your strategy as well.
4. The last column has values from 1 to 16, replace all values which are not 1 by 2.
5. Convert columns 2 (Sex: Sex (0 = male; 1 = female) into dummy variables. Name the columns as "Male" and "Female".
6. Find the statistical properties of the data
7. Apply data normalisation techniques on 2 columns of your choice
8. Draw a boxplot for column 4 (Weight)
9. Select 10 columns of your choice and predict the outcome using a logistic regression model.
10. Apply 10-fold cross validation and report accuracy of the model.

Note

For classification problem the model is judged based on accuracy, precision, recall and F-1 score.

Good Luck.