

IE 8990 Distributed Machine Learning Techniques
Assignment (5+6)
Due: 12/1/2021 11:59 PM

Use the **Haberman.csv** file to answer the following questions:

Question 1: Check if there is any missing value in the dataset. Perform some basic exploratory data analysis.

Question 2: Using “Survival Status for patients died within 5 years” as your target variable, implement the following classification algorithms: (1) Logistics Regression (2) Decision Tree (3) Random Forest and (4) Support Vector Machines (5) Adaboost (6) Naïve Bayes Classifier.

Question 3: Make a comparison of the prediction performance of the different algorithms implemented on **Question 2**.

Question 4: There is a problem with the dataset. Try detecting the problem and correct your prediction.