innovaccer

Here are two problems that you need to solve. Please submit any codes that you write, along with a ppt (max 5 slides) explaining the algorithm (word limit 100) used along with the results obtained (define any metric that can best explain the accuracy of your model).

Dataset 1:

This dataset contains TripAdvisor scores along with features describing the reviewers and hotels located in The Strip, Las Vegas.

Question 1: Build a classification/regression model to predict scores for any new reviewer.

Question 2: Identify features that are most relevant in the model.

Dataset 2:

This breast cancer database was obtained from Dr. Wolberg's office at the University of Wisconsin Hospitals, Madison. Each record here contains values for different morphological and pathological features of a tumor dissected from any given patient. The class column indicates whether the patient has been characterized as the benign tumor or a malignant tumor.

Question 1: Build a classifier to identify patients with benign or malignant tumor based on the tumor characteristics

Question 2: As an oncologist, you would want to reduce your false positives as well as false negatives.

a. Identify the number of false positive and false negatives
b. Improve your classification model to reduce patients who are being predicted as having benign tumor but actually have malignant tumor.