Assignment 4: Part 2

Use the Diabetes Readmissions data (on Canvas)

Tree Models

Use the same training and test samples and the same response variable for the Diabetes Readmission data set that you used for Logistic Regression Assignment.

1. Build various Classification Tree models by varying the number of nodes in the tree, for the Training data set for predicting the "Readmitted" variable, using the other variables.
2. Generate the confusion matrix of predictions in train and test data (actual readmitted variable versus the predicted readmitted variable labels).
3. Choose the winning tree model as the model with (a) the highest train accuracy AND the lowest fall in accuracy from train to test. In short – best performance in train and equally good performance in test.
4. Interpret the best tree.
    a. How many interactions do you see?
    b. Can you interpret the tree? Do you like it? Comment

5. Summarize your results. Which has performed better for the Diabetes Readmissions data – tree or logistic regression?

SAVE your predictions of tree models – both for training and test data sets. **You will need them later.**

Points: 5. 1 per question.