J. Jedediah Smith

BIFX-551

Assignment 4

5/10/2022

1. Prepare the provided Diabetes dataset, implement ONE of the classification techniques and evaluate the performance using the proper metrics for the binary classification of the variable

“Outcome.”

Data was prepared by setting the outcome variable to a factor, then ensuring it was properly divided into training and test data. After verifying that the outcome proportions were similar in both the training and test data, a model was created using the training data. It was classified via decision tree. Then a prediction was generated using the test data. Finally, the model performance was assessed using a cross table of the prediction results. As shown below, the model had approximately 77% chance of making the right call. It seems to be biased towards 0, the non-diabetic cases, due to the larger number of those in the data set.

Text

Description automatically generated

Table

Description automatically generated