

Neil Wagner  
Dr. Brahma  
BSAN 6070  
26 March 2021

#### CA05A

Before creating the model, the first step was to remove the insignificant variables. These variables were identified by observing the p-value. If the value was over 0.05, then the variable is considered insignificant and was thus removed from the data analysis. The next step was to define X and y and build the model. The AUC score, which represents the probability of a classifier to distinguish between classes, returned at 0.66. This score is rather poor, as a value of 0.7 or above is preferable. The accuracy score, on the other hand, is about 0.61 which is also rather low.

The feature importance section identified which variables have the greatest influence on outcome. The larger the coefficient, the greater the importance of the feature. The results showed that waist size has the greatest influence, as the coefficient is by far the highest at 1.01. Hip measurement, with a coefficient value of 0.55, is the second most influential feature. These results can be extremely useful as they can identify which features and attributes are most associated with risk for cardiovascular disease. However, as mentioned above, the model performance scores are not as high as anticipated. Nonetheless, the results should not be ignored and the findings from the analysis can still be insightful and used to draw early conclusions.