Peter Gurka

26 March 2021

CA05A Part 3 Explanation and Conclusion

BSAN 6070

Explanation

The performance overall can be defined through a series of facts. One is the accuracy score, which in this case is .755. The confusion matrix demonstrates there are 163 + 327 correct predictions and 61 + 98 incorrect predictions, which is not perfect, but obviously closer to correct than incorrect. In the classification, we understand through the support metric that the dataset is not perfectly balanced, rather the data is leaning towards 1 rather than 0 due to a larger value. Precision, recall, and f1 score were all very close, around 75 to 76%. This demonstrates that 75% of patients have some sort of CVD risk. Lastly, the ROC curve stays somewhat far away from the dotted line, demonstrating the model is doing a good job of identifying true positives and true negatives.

Conclusion

Overall, this model is not perfect, but the overall motion is in a positive direct. The performance factors demonstrate that this model does a fairly good job in determining the likelihood of Yes vs No for CVD risk. The largest positive factor was waist at 1.11, whereas the largest negative factor in the logistic regression is tea15. Therefore, it may be important that medical staff work with patients to maintain a certain waist metric in order to allow this influence to be a preventative tool against CVD.