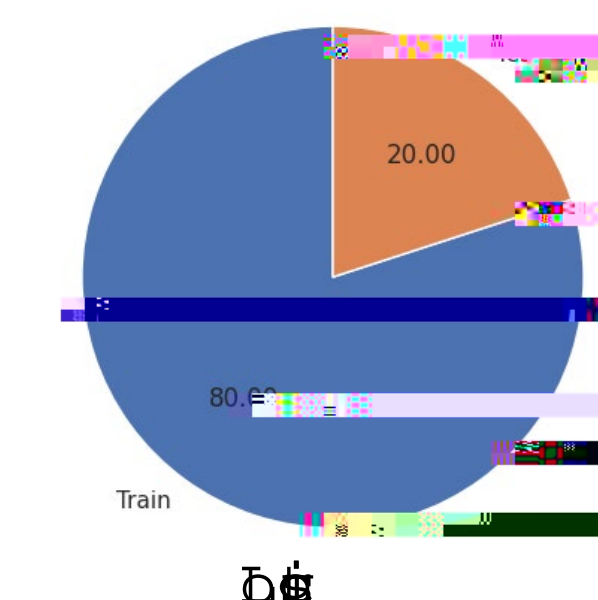


Predicting the Probability of Heart Disease... A Comprehensive Guide to Model Building

Greg Balkcom (Master's), Dr. MinJae Woo (Instructor, Python Class)

- Heart disease is the leading cause of death in the United States.



METHODS

- Data set from www.kaggle.com
- 918 records, tested the performance of each model on test section and put results in a confusion matrix
- 7) Evaluated model performance
 - ROC: Area under the "Receiver Operating Curve"
 - Accuracy: $(\text{True Pos} + \text{True Neg}) / \text{Total}$
 - Precision: $\text{True Pos} / (\text{True Pos} + \text{False Pos})$
 - Recall: $\text{True Pos} / (\text{True Pos} + \text{False Neg})$
 - F1 Score: $2 * ((\text{Prec} * \text{Rec}) / (\text{Prec} + \text{Rec}))$
- 8) Identified important factors in predicting the probability of heart disease