Introduction:

The goal of this project was to predict the probability of having a heart attack using 14 variables (age, sex, chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar, resting ECG results, max heart rate, exercise induced angina, old peak heart rate, slope of the peak, number of major vessels, thal, and target heart rate) given in the heart.csv data-set. The classification models chosen to analyze this data-set were: random forest, bootstrap/bagging, support vector machines (SVM), and k-nearest neighbor (KNN). After exploring these various classification methods, we can analyze and interpret the results of each method to determine which might be the “best” classifier. Additionally, the logistic regression methods, lasso and elastic net, were taken into consideration to explore which variables are most important. (add sentence on the importance of this)

Analysis of random forest: