

Dataset available: http://10.30.151.151/lab/SJT516/

This data set ***Diabetes*** have 8 baseline variables body mass index, average blood pressure, and six blood serum measurements were obtained for each of n = 442 diabetes patients, as well as the response of interest, a quantitative measure of disease progression one year after baseline.

bmi body mass index

bp average blood pressure

s1 tc, total serum cholesterol

s2 ldl, low-density lipoproteins

s3 hdl, high-density lipoproteins

s4 tch, total cholesterol / HDL

s5 ltg, possibly log of serum triglycerides level

s6 glu, blood sugar level

-Handle the missing value using any statistical method.

-For bp and hdl, Apply data scaling methods.

- Build a Decision tree by implementing the CART algorithm and test accuracy, precision, recall and ROC curve for at least 3 different parameters.