CST/Math Data Mining	Fall 2021
Tu, Th, 2:00-3:15	HW 3 Programming

[6pts] Given the attached diabetes data set. Read the data set description – separate file.

1. Input the data
2. Separate the data into learning set (~2/3 of all records) and test data.
3. Built the classification trees using CART (rpart) and C4.5 (J48) trees.
4. Plot the rpart tree and C4.5 tree
5. Predict the classes of records in test data. What is the accuracy for each tree? Generate table of predicted classes vs known classes in test set
6. Are trees the same? If not which one is smaller? Why? Are accuracies the same?

Everything that you need to use as programming 'cut and paste' patterns for modification are posted in 'weekly lectures- >week 4' and 'weekly lectures- >week 5'.

Note that for class attribute imported from CSV file values 0,1 are interpreted as integers upon input. To apply any classification algorithm, they must be factors. So, you need to convert data types (i.e. integer to factor) similarly to the following example. Suppose aaa$type are integers in {1,2,3}:
aaa$type<-as.factor(aaa$type)