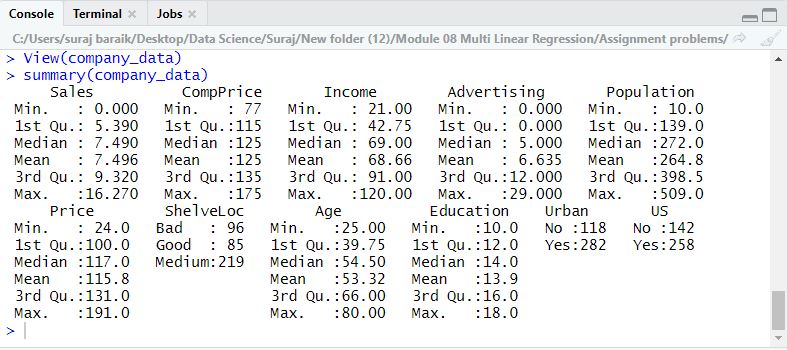
Decision Tree Documentation

Business Problem: To know about the segment or attributes contributing to high sale of a cloth manufacturing company

Objective: Built a decision tree with ‘Sale’ as target variable.

About the Data:

The file contain 11 columns and 400 rows entries of a company sales data.



Splitting data into training and testing. As the species are in order

Splitting the data based on species

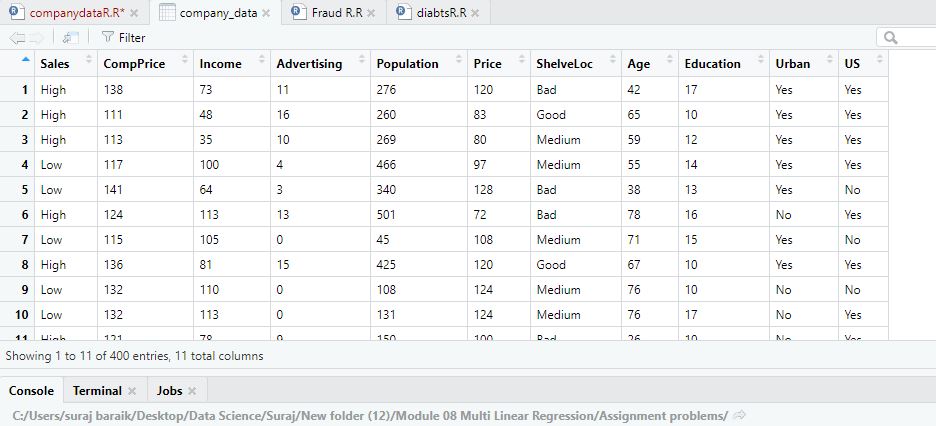
Change Sales to a qualitative variable by splitting it on the median.

we neeed to install C50 package to use ak

install.packages("C50")

use prp() to make cleaner plot with caret

library(rpart.plot)



#Split data into train / validation

set.seed(111)

split <- createDataPartition(y=company\_data$Sales, p=0.6, list=FALSE)

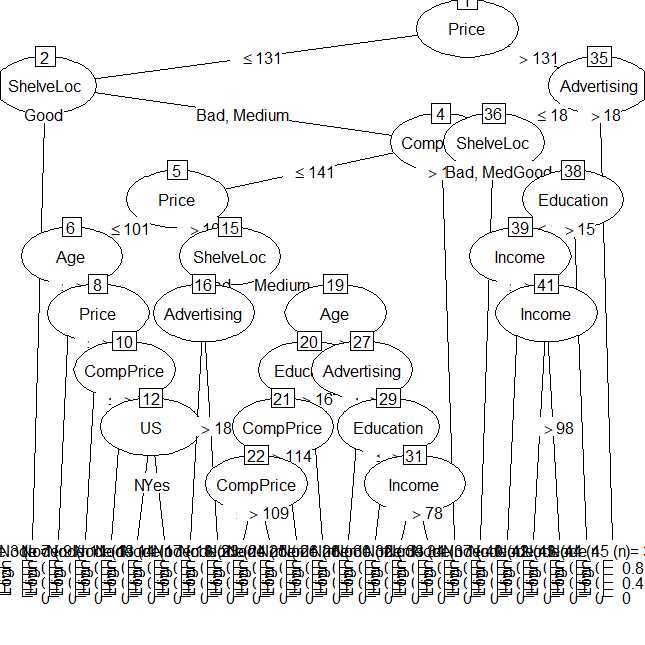
train <- company\_data[split,]

test <- company\_data[-split,]

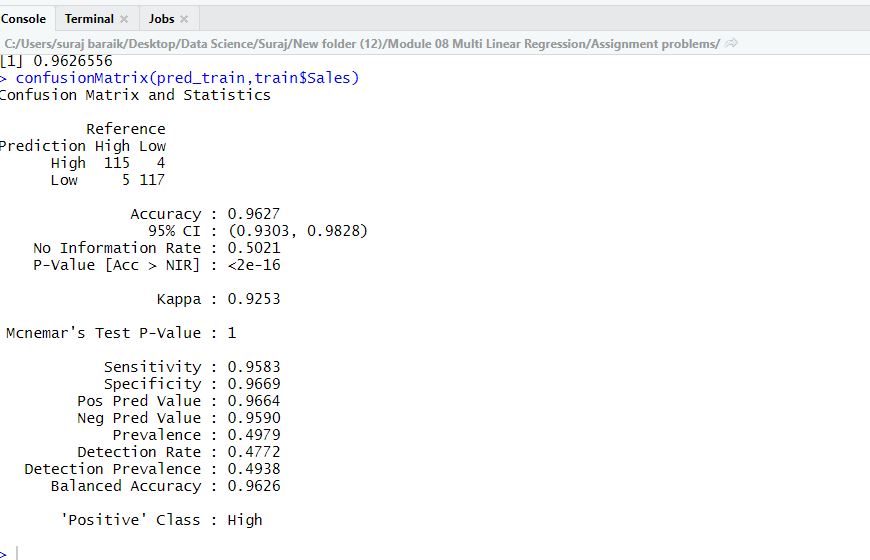
Building model on training data

sales\_train <- C5.0(Sales~.,data= train)

Tree Graph

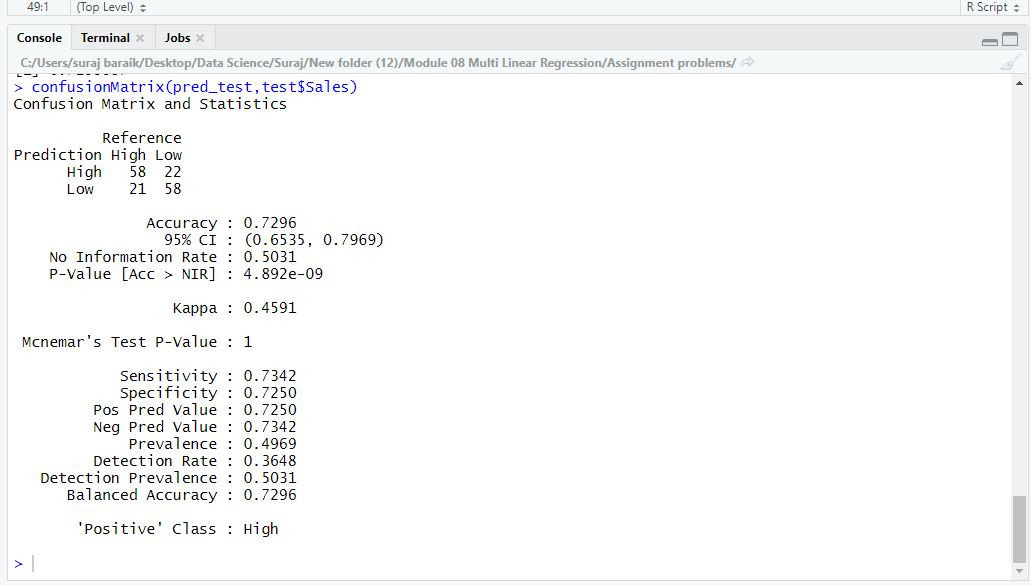


confusion matrix data



This train model is giving us 96% accuracy with error rate of 95% we will check with test model

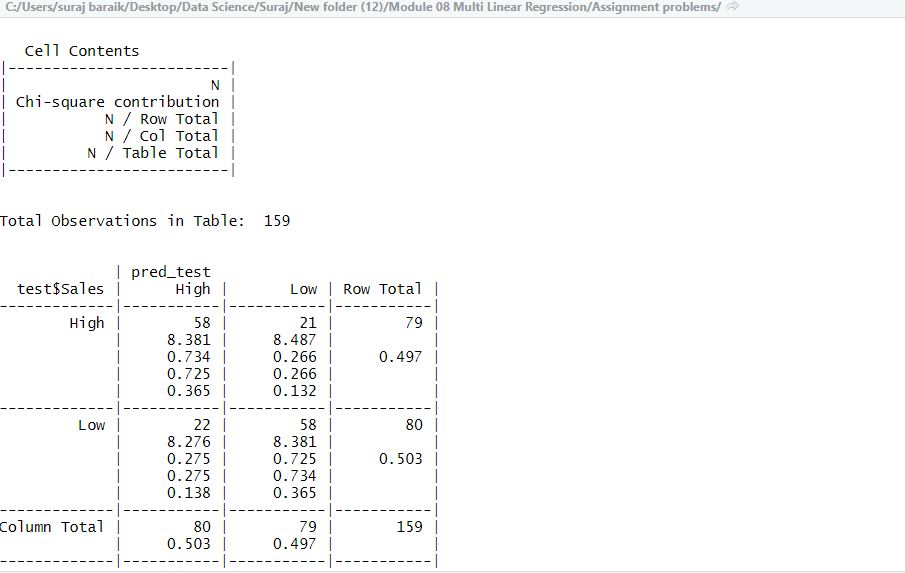
confusionMatrix(pred\_test,test$Sales)



This train model is giving us 72% accuracy with error rate of 73%, as model will have extremely low training error but a high testing error. This tells us that model is overfitting.

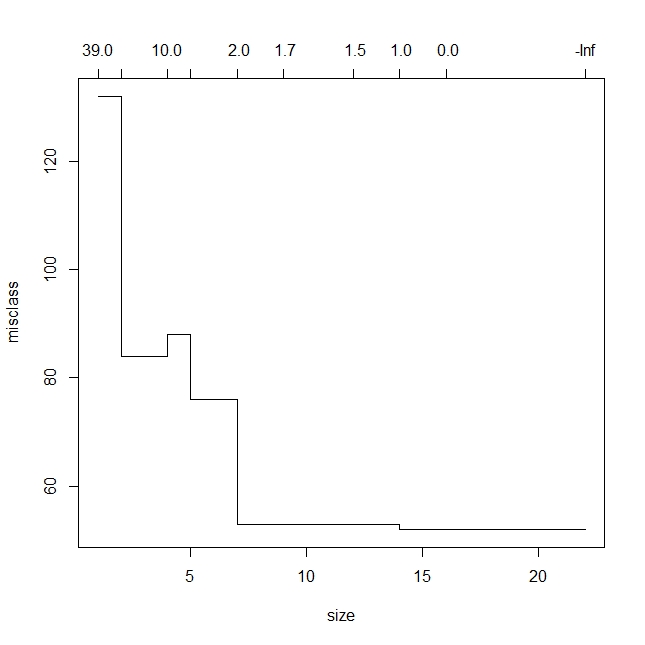
Cross table

CrossTable(test$Sales,pred\_test)

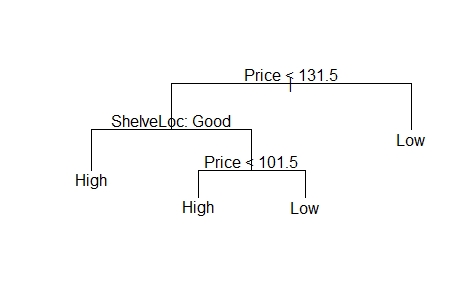


A test error rate of ~73.5% is pretty good! But we could potentially improve it with cross validation.

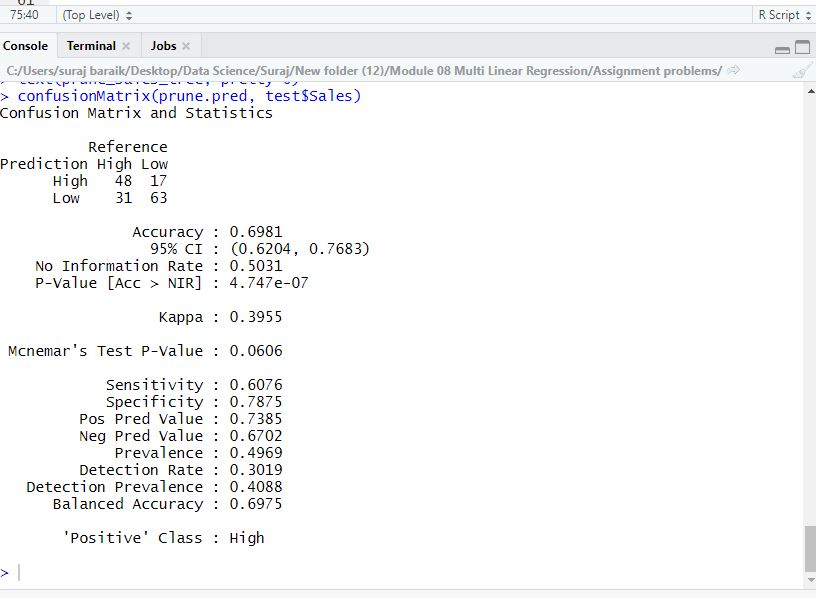
Cross Validating



Here we see that the the lowest / simplest misclassification error is for a 4 leaf model. We can now prune the tree to a 4 leaf model.



confusionMatrix(prune.pred, test$Sales)



This doesnt really improve our classification as accuracy dips down to 69%

We need to go with different splitting method or go with other variable for decision dependent atribute other than sales to improve accuracy

with 1st model we are getting 72% is test accuracy which is also good we will go with that.