MUSHROOMS PROBLEMS

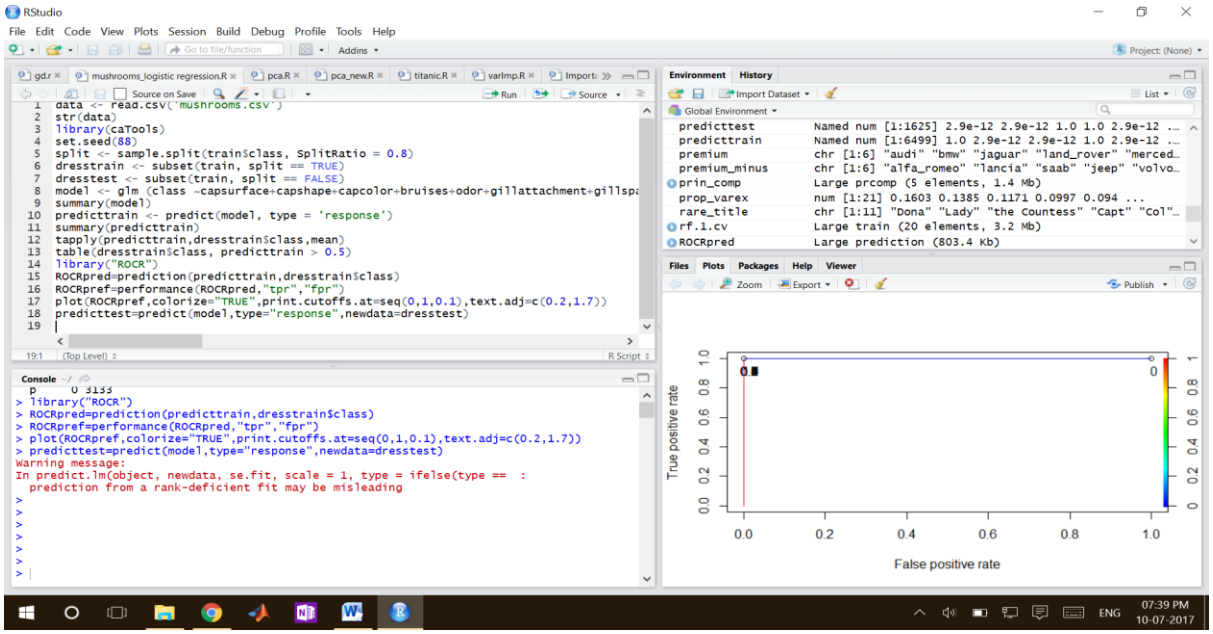
In the given problem, we have to classify the mushrooms between edible and poisonous. The dataset include 23 attributes (variables) of mushroom.

Twenty three attributes are -

capsurface,capshape,capcolor,bruises,odor,gillattachment,gillspacing,gillsize,gillcolor,stalkshape,stalkroot,stalksurfaceabovering,stalksurfacebelowring,stalkcolorabovering,stalkcolorbelowring,veilcolor,ringnumber,ringtype,sporeprintcolor,population,habitat.

LOGISTIC REGRESSION

Here is the code of logistic regression:



From the above image, we can see that the datasets are properly classified.

Accuracy - 100%.

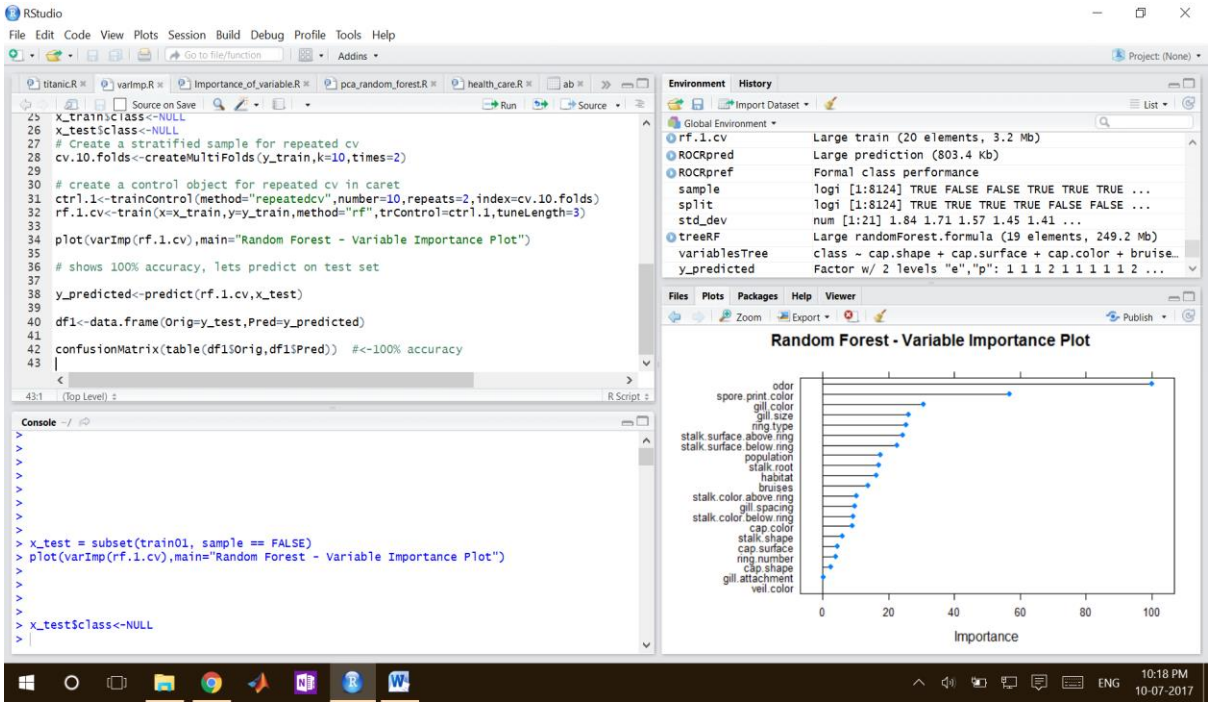
RANDOM FOREST

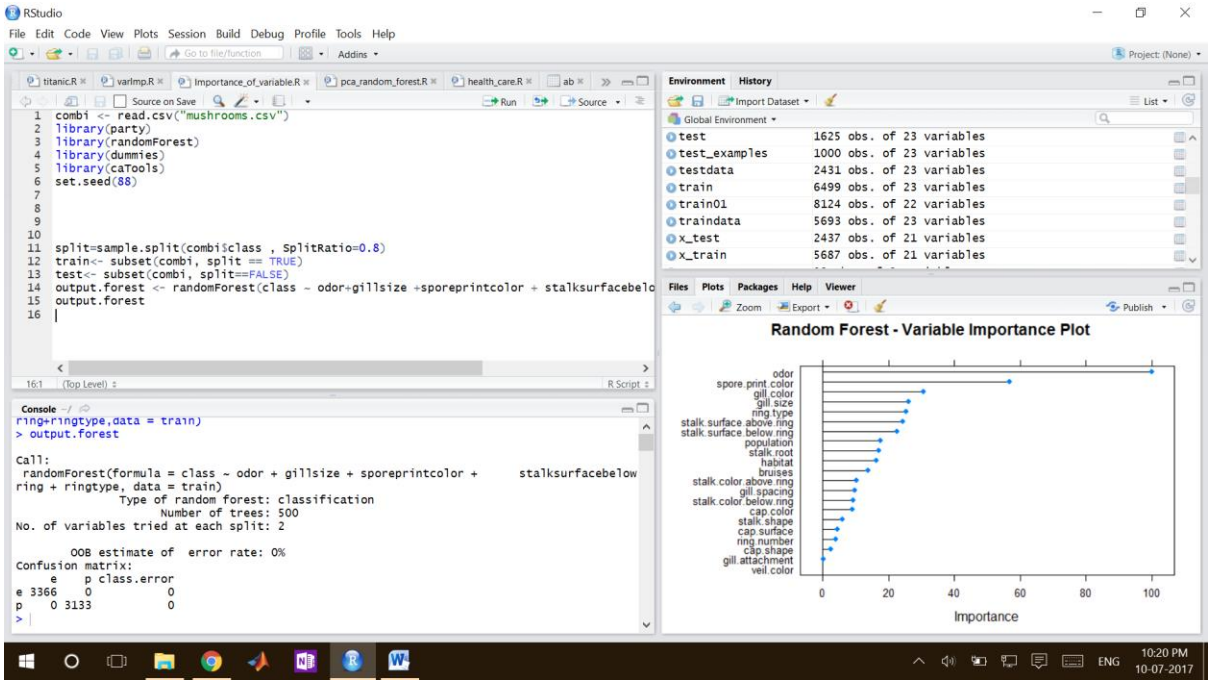
I have also used Random Forest Model to classify the mushrooms with the accuracy of 100%. Then, I calculated the importance of variables using random forest. Apart from that I also have reduced the variables from 23 to 5 which also gives 100% accuracy.

Five variables are: odor, gillsize ,sporeprintcolor , stalksurfacebelowring,ringtype.

Random forest code:

MUSHROOMS PROBLEMS





PRINCIPAL COMPONENT ANALYSIS (PCA)

I have also done the principal component analysis to reduce the variables. From the image, we can infer that there are eight principal components which give 68% accuracy.

MUSHROOMS PROBLEMS