

HOMework ASSIGNMENT II

- Work individually.
- Submit a code and a short explanation along the code. The names of the files should be YOURNAME.py (or YOURNAME.ipynb) and YOURNAME.docx (e.g. OrsanOzener.py, OrsanOzener.docx)
- Any type of plagiarism will not be tolerated and will lead to disciplinary actions.
- **Due Date: 10th of January 2020, 13:00. Please submit your report through LMS. Only ONE submission per person.**

1 Introduction

In this assignment, you are supposed to work on a binary classification problem in which we need to predict the chronic kidney disease (ckd or notckd). The data is on the following csv.file: “chronic_kidney_disease_full.csv”.

The performance criteria will be the “Accuracy Score”. In this homework, you are supposed (i) perform the cross-validation scheme to select a single classification method, (ii) boost the performance of the method that you have selected by performing a hyper-tuning step. You are not allowed use any models other than the ones discussed in the class.

Data Set Information:

age - age
bp - blood pressure
sg - specific gravity
al - albumin
su - sugar
rbc - red blood cells
pc - pus cell
pcc - pus cell clumps
ba - bacteria
bgr - blood glucose random
bu - blood urea
sc - serum creatinine
sod - sodium
pot - potassium

hemo - hemoglobin
pcv - packed cell volume
wc - white blood cell count
rc - red blood cell count
htn - hypertension
dm - diabetes mellitus
cad - coronary artery disease
appet - appetite
pe - pedal edema
ane - anemia
class - class

Attribute Information:

We use 24 + class = 25 (11 numeric ,14 nominal)

- 1.Age(numerical) age in years
- 2.Blood Pressure(numerical) bp in mm/Hg
- 3.Specific Gravity(nominal) sg - (1.005,1.010,1.015,1.020,1.025)
- 4.Albumin(nominal) al - (0,1,2,3,4,5)
- 5.Sugar(nominal) su - (0,1,2,3,4,5)
- 6.Red Blood Cells(nominal) rbc - (normal,abnormal)
- 7.Pus Cell (nominal) pc - (normal,abnormal)
- 8.Pus Cell clumps(nominal) pcc - (present,notpresent)
- 9.Bacteria(nominal) ba - (present,notpresent)
- 10.Blood Glucose Random(numerical) bgr in mgs/dl
- 11.Blood Urea(numerical) bu in mgs/dl
- 12.Serum Creatinine(numerical) sc in mgs/dl
- 13.Sodium(numerical) sod in mEq/L
- 14.Potassium(numerical) pot in mEq/L
- 15.Hemoglobin(numerical) hemo in gms
- 16.Packed Cell Volume(numerical)
- 17.White Blood Cell Count(numerical) wc in cells/cumm
- 18.Red Blood Cell Count(numerical) rc in millions/cmm
- 19.Hypertension(nominal) htn - (yes,no)
- 20.Diabetes Mellitus(nominal) dm - (yes,no)
- 21.Coronary Artery Disease(nominal) cad - (yes,no)
- 22.Appetite(nominal) appet - (good,poor)
- 23.Pedal Edema(nominal) pe - (yes,no)
- 24.Anemia(nominal) ane - (yes,no)
- 25.Class (nominal) class - (ckd,notckd)

If you have any questions, please send an email to: orsan.ozener@ozyegin.edu.tr