Module 5: Classification

Assignment

Module 5: Assignment

Analyze the information given in the Affairs dataset and create classifiers using it.

The dataset can be loaded into R using the library:

- library(AER)

affairs	gender	age	yearsmarried	children	religiousness	education	occupation	rating
0	male	37.0	10.000	no	3	18	7	4
0	female	27.0	4.000	no	4	14	6	4
0	female	32.0	15.000	yes	1	12	1	4
0	male	57.0	15.000	yes	5	18	6	5
0	male	22.0	0.750	no	2	17	6	3
0	female	32.0	1.500	no	2	17	5	5
0	female	22.0	0.750	no	2	12	1	3
0	male	57.0	15.000	yes	2	14	4	4
0	female	32.0	15.000	yes	4	16	1	2
0	male	22.0	1.500	no	4	14	4	5
0	male	37.0	15.000	yes	2	20	7	2

The description of the attributes in the dataset are as follows:

- affairs – numeric. How often engaged in extramarital sexual intercourse during the past year
- gender – Factor indicating gender.
- age – numeric variable coding age in years
- yearsmarried – numeric variable coding number of years married 0.125 = 3 months or less, 0.417 = 4–6 months, 0.75 = 6 months–1 year, 1.5 = 1–2 years, 4 = 3–5 years, 7 = 6–8 years, 10 = 9–11 years, 15 = 12 or more years.
- religiousness – numeric variable coding religiousness: 1 = anti, 2 = not at all, 3 = slightly, 4 = somewhat, 5 = very.
- Education - numeric variable coding level of education: 9 = grade school, 12 = high school graduate, 14 = some college, 16 = college graduate, 17 = some graduate work, 18 = master's degree, 20 = Ph.D., M.D., or other advanced degree.

Perform the following tasks on the dataset

Task 1: Using the affairs column of our data

→ Create a new column with nominal values "YES" and "NO".

→ Convert it into factor

Task 2: Create a classifier with our data using Decision tree algorithm

→ Plot the Decision tree

→ Calculate the accuracy using confusion matrix

Task 3: Create another classifier with our data using random forest algorithm

→ Calculate the accuracy using confusion matrix

→ Find out the importance of attributes using importance() function