

Practical Data Science (Explorative Data Analysis)

Write python scripts to solve the following problems.

Problem 1: Explore Car dataset

Download the car.data from datasets branch of algorithmica github repository. Here are the descriptions of the attributes of the car dataset:

buying: vhigh, high, med, low.
maint: vhigh, high, med, low.
doors: 2, 3, 4, 5more.
persons: 2, 4, more.
lug_boot: small, med, big.
safety: low, med, high.

The output class attribute can take one of the following values:

unacc, acc, good, vgood

Do the following tasks:

- Load the dataset into frame and convert all the attributes to factor type.
- Explore all the attributes individually using univariate numerics and graphics.
- What kind of preprocessing do you suggest after the univariate explorations.
- Explore all the bivariate relationships numerically and graphically.
- What features do you recommend for predicting class category and why?

Problem 2: Exploring Kidney data

Download the chronic_kidney_data.txt from datasets branch of algorithmica github repository. The description of the dataset can be found at following link:

http://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease

Do the following tasks:

- Load the dataset into frame and convert all the attributes to factor type.
- Explore all the attributes individually using univariate numerics and graphics.
- What kind of preprocessing do you suggest after the univariate exploration.
- What kind of missing value handling mechanism do u suggest and why?
- Explore all the bivariate relationships numerically and graphically.
- What features do you recommend for predicting the disease is chronic or not and why?

Problem 3: Exploring Adult DataSet

Download the adult_data.dat from datasets branch of algorithmica github repository. The description of the dataset can be found at following link:

<http://archive.ics.uci.edu/ml/datasets/Adult>

Do the following tasks:

Practical Data Science (Explorative Data Analysis)

- a. Load the dataset into frame and convert all the attributes to factor type.
- b. Explore all the attributes individually using univariate numerics and graphics.
- c. What kind of preprocessing do you suggest after the univariate exploration.
- d. What kind of missing value handling mechanism do u suggest and why?
- e. Explore all the bivariate relationships numerically and graphically.
- f. What features do you recommend for predicting the disease is chronic or not and why?