

Artificial Intelligence

Homework Assignment 6.

1. pima-indians-diabetes.csv파일을 사용하여 다음에 해당하는 코드를 작성하시오.
 - (a) CSV 파일(pima-indians-diabetes.csv)을 DataFrame 형태로 적재하여 읽고, shape와 첫 20 행을 출력하시오. (1점)
 - (b) Column 1 ~ 5까지의 0값을 NaN로 치환하고 첫 20행과 NaN으로 치환된 데이터 셀의 개수를 출력하시오 (1점)
 - (c) Column 0(임신 횟수)을 0회, 1 ~ 3회, 3회 이상의 3개 구간으로 나누고 one-hot encoding을 수행한 후, 중간 20행을 sparse matrix, numpy array, DataFrame 형태로 각각 출력하시오. (3점)
 - (d) 결측치를 포함하는 행을 모두 삭제한 후, 삭제 전/후 행과 열의 개수를 출력하시오. (1점)
 - (e) 결측치를 각 column별 평균으로 대체하고, 치환이 발생한 행을 선별적으로 출력하고 치환이 발생한 행의 개수를 출력하시오. (1점)
 - (f) CSV 파일(pima-indians-diabetes.csv)을 train dataset 100건과 test dataset 100건을 선택하여 각각 별도의 csv 파일의 형태로 나누어 저장하되, 샘플을 선택하는 방식은 SRSWOR(simple random sample without replacement)을 취하고 positive와 negative 비율은 각각 5:5가 되도록 구성하시오. (3점)

Submitting your assignment :

- Due date: Zip your file and upload it at <https://lms.mju.ac.kr/> by 24:00 Monday April 25th, 2022.
- Your homework cover page must be of the form provided by the courseware.
- You must zip the homework with the cover page including your jupyter notebook assignment file(*.ipynb), two train and test dataset files generated from problem (f), and screen capture image of the execution results.
- Both of your file names must be of the form "hw6_StudentId_StudentName.ipynb", i.e., hw6_60063539_한지수.ipynb
- You must protect your homework from others. Any form of academic dishonesty will not be tolerated. If you get caught, you will receive -10 points for this homework!
- This assignment is 10 points total and the late penalty is 2 points per day.