

2024년 1학기 데이터마이닝 중간고사

학번 _____

이름 _____

■ `heart_train.csv`과 `heart_test.csv` 자료는 심장마비(heart attack)의 가능성과 연관된 자료이다.

Column
<ul style="list-style-type: none">age: Age of the personsex: Gender of the personcp: Chest Pain type<ul style="list-style-type: none">Value 1: typical anginaValue 2: atypical anginaValue 3: non-anginal painValue 4: asymptomatictrtbps: resting blood pressure (in mm Hg)chol: cholestoral in mg/dl fetched via BMI sensorfbs: (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)restecg: resting electrocardiographic results<ul style="list-style-type: none">Value 0: normalValue 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteriathalachh: maximum heart rate achievedexang: exercise induced angina (1 = yes; 0 = no)oldpeak: Previous peakslp: Slopecaa: number of major vessels (0-3)thall: Thal rateoutput: Target variable (0=less chance of heart attack, 1=more chance of heart attack)

1. `heart_train.csv` 자료로 주어진 모형(3개 이상)을 학습하고, `heart_test.csv` 자료로 평가한 지표를 계산하시오. [10점]

- 각 모형 학습 및 평가 코드 제출
- 변수 변환 및 데이터 처리 과정을 모두 포함한 코드 제출

Model	Accuracy	F1-score	AUC	Note / Parameters
로지스틱 회귀				
의사결정나무 - CART				
의사결정나무 - C4.5				
서포트벡터머신				
나이브 베이즈 분류				
신경망				
랜덤포레스트				

2. 위의 모형 평가 지표 중 AUC 기준으로 최종 모형을 선택하고, 모형의 결과를 해석하시오. [10점]

-끝-