< 머신러닝 20220413 >

# 중간고사 : 다음주 금요일

1. Linear Regression :

데이터 : petrol\_consumption

- 드라이버/드라이빙/가솔린 등에 관한 속성을 가진 데이터

- petrol tax

- average income

- paved highways : 고속도로 주행 ( 아마 중요)

- 운전면허를 가진 인구비율

- DV : Consumption of petrol : the consumption of petrol

문제 :

1. 기술통계 및 요약을 사용하여 데이터를 설명하라

2. 데이터셋 시각화

A - 상관관계 그래프

B - 히스토그램 -seaborn

3. 모델설계 / 결정계수 찾기

4. 모델평가 : MAE, MSE,RMSE

5. 모델에 부정확성이 있다면, 이유는 무엇인가? : 주석으로 작성

2. Logistic Regression :

데이터 : 고객정보 (binary/ continuous/ categorical )

-나이

-직업

-결혼여부

-교육

-자가 ... etc

문제 :

1. 데이터탐색

A. value\_counts 함수를 사용하는 y값 보여주기

B. countplot

2. 데이터 전처리

3. 시각화

4. 모델 설게 및 결정계수 찾기

5.

- 5개의 변수에 관해

2. Logistic Regression