6/27 강의노트

p-value : 모델 생성에서 가장 중요. 0.05 이상이면 유의미한 모델이나 변수가 아님

스크린샷, 텍스트, 도표, 그래프이(가) 표시된 사진

자동 생성된 설명

계수 시각화 : 0에 가까울수록 유의미하지 않은 변수

패턴, 스크린샷, 사각형, 다채로움이(가) 표시된 사진

자동 생성된 설명

텍스트, 영수증, 폰트, 화이트이(가) 표시된 사진

자동 생성된 설명

히트맵 : 절댓값이 높을수록 머신러닝에 부정적인 영향이 큼 (피처간의 연관성임, 각 피처는 독립성을 가져야 올바른 분석이 가능하므로 계수와는 반대로 0에 가까워야 독립적이며, 머신러닝에 유효함)

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

AIC / BIC : 정보소실량, 상대적으로 낮을수록 좋음

텍스트, 종이, 흑백, 번호이(가) 표시된 사진

자동 생성된 설명

p-value가 낮은 변수를 제거해 나가면서 모델 성능을 측정하고, 다음으로는 VIF가 높은 변수를 제거

그 과정에서 모델 성능이 낮아지는 순간 해당 작업을 종료

독립변수가 모두 제거되면 종속변수를 예측하는데 유의미한 데이터가 아니다 라고 해석 가능

의사결정나무

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

로지스틱 회귀 모델의 경우 기울기 등으로만으로는 모델에 관한 설명을 하기 어려움