

[기말고사 연습문제]

(문제) 전자상거래 배송 데이터 제품 배송 시간에 맞춰 배송되었는지 예측모형을 만들고자 한다. 학습용 데이터 (X_train, Y_train)을 이용하여 배송 예측 모형을 만든 후, 이를 평가용 데이터(X_test)에 적용하여 얻은 예측값을 다음과 같은 형식의 CSV파일로 생성하시오.

(제출한 모델의 성능은 roc_auc 평가지표에 따라 채점)

(파일 : XX_train.csv, YY_train.csv, 기말고사코드.py)

(방법) LogisticRegression을 사용하며, X_train과 Y_train을 80%:20%로 분할하여 모델을 만든 다음 테스트 데이터로 검증 값을 구하여 지표를 계산한다.

(데이터 설명)

• ID	고객의 ID 번호
• 창고 블록	회사는 A,B,C,D,E와 같은 블록으로 나누어진 큰 창고
• 배송 방식	회사는 선박, 항공 및 육로와 같은 다양한 방법으로 제품을 배송
• 고객센터 통화	배송조회 문의전화 건수
• 고객 평가	1은 가장 낮음(Worst), 5는 가장 높음(Best)
• 제품 비용	제품 비용(미국 달러)
• 이전 구매	이전 구매 횟수
• 제품 중요도	회사는 낮음, 중간, 높음과 같은 다양한 매개 변수로 제품을 분류
• 성별	남성과 여성
• 제공되는 할인	해당 특정 제품에 대해 제공되는 할인
• 무게	(gms): 그램 단위의 무게
• 정시 도달(Target)	1은 제품이 정시에 도달하지 않았음을 나타내고 0은 정시에 도달