모델 평가까지의 전체 과정을 포함한 파이프라인 문서화

1. 개요 - Kaggle의 타이타닉 데이터를 이용하여 생존자 예측 모델링을 위한 모델 파이프라인 구현 및 문서화를 진행합니다
    a. 목표 : 전처리된 타이타닉 데이터 기반으로 생존여부를 예측하는 모델을 구축하고 테스트 데이터를 실행해 봅니다
    b. 데이터셋
        i. https://www.kaggle.com/competitions/titanic
        ii. train.csv / test.csv
2. 알고리즘 선택 근거 : 로지스틱 회귀
    a. 모델 학습 후 각 피터가 생존(1) 또는 사망(0)의 확률에 얼마나 영향을 미치는지를 상관계수를 통해 해설 해볼 수 있습니다
    b. 단순히 0/1(이진)이 아닌 확률 값을 확인하기 위해서 입니다
    c. 복잡한 모델(앙상블 등)을 사용해보기 전에 기준선 모델이 될 수 있습니다
3. 평가지표
    a. 혼동행렬
    b. F1-Score
    c. ROC-AUC Score
4. EDA - 관련 내용은 별도 문서 참조
5. 모델 구현 및 평가는 별도 파일 참조
6. 결론
    a. 위에서 만들어진 모델은 테스트 데이터에서 약 80% Accuracy와 0.86 ROC-AUC 점수를 달성하였습니다.
    b. 상관계수를 분석을 통해 성별과 객실등급이 생존에 가장 큰 영향을 미치는 요인임을 확인 할 수 있습니다
    c. 추가적인 피처 엔지니어링을 통해 성능 개선을 해볼 수 있을것 같습니다
        i. 이름에서 성별을 추출 등
    d. 로지스틱 회귀 외의 비선형 모델을 사용해서 베이스라인 과 선능을 비교해볼 수 있을것 같습니다