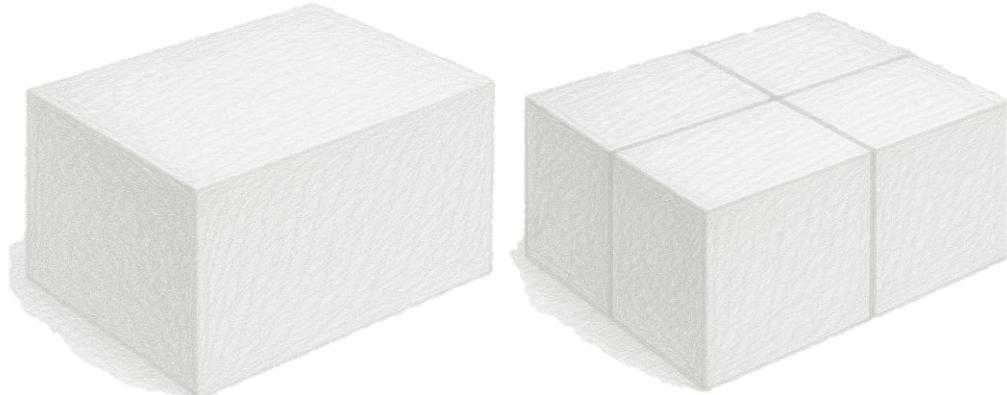


Titanic 승객 생존률 분류 모델: 전처리-모델링-검증

- 목적: 승객 특성으로 생존률을 분석 가능한 조합(선실x성별x승선항x나이대)을 분류하고, 생존률이 높은 그룹을 도출하는 지도학습 파이프라인 구축
- Background: 범주와 수치가 혼합된 타이타닉 데이터셋에 대해 결측 보정 및 전처리와 6개 분류기 비교(5-Fold 교차검증)로 성능을 평가
- 결과: SVM 최종 선정, 테스트 정확도 81.6%. Cherbourg에서 탑승한 1등석 19~35세의 여성이 생존률이 가장 높았던 것으로 확인



데이터 준비 및 전처리

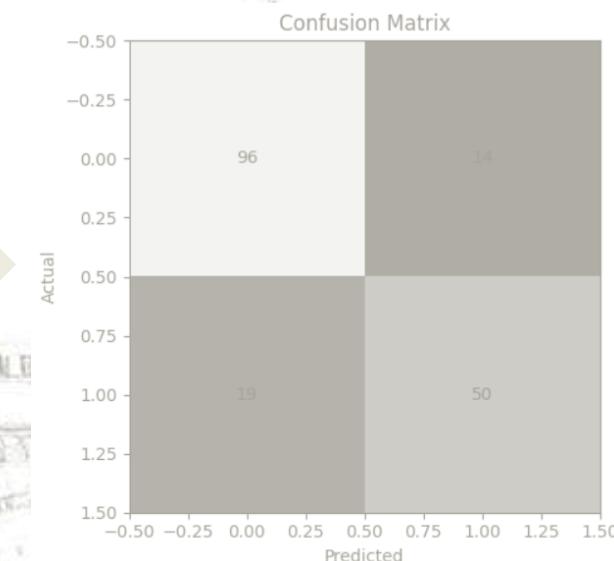
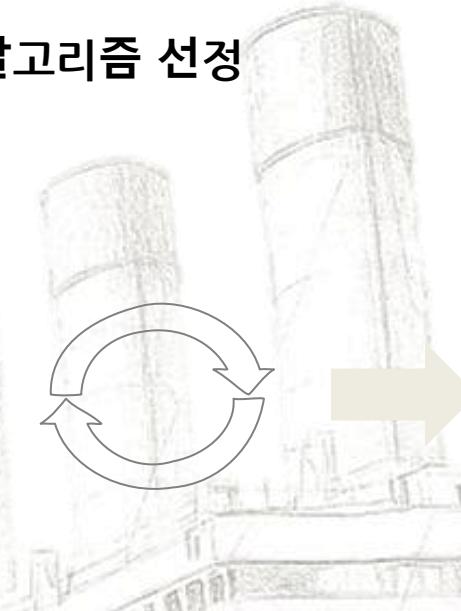
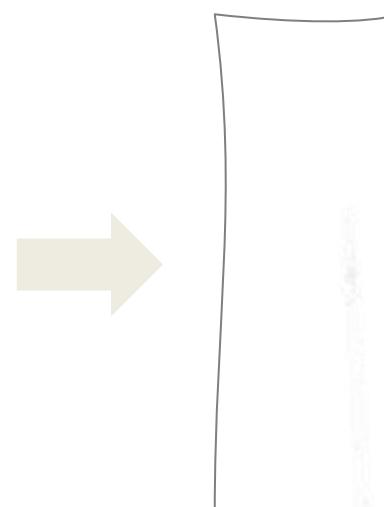
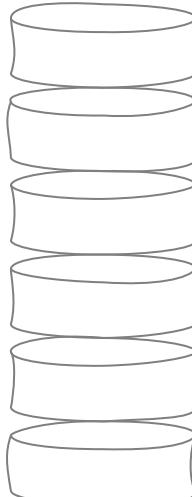


결측치 보정, 전처리 파이프라인 구성



모델 분석 및 평가

6종 알고리즘에 대해 비교 평가 후 최종 알고리즘 선정



- [데이터 출처] Kaggle이 제공하는 Titanic 머신러닝 Dataset 활용
- 고유값 컬럼 제거 및 결측 보정으로 학습 안정화

→ Pclass x Sex x Embarked x AgeBand의
해석 가능한 그룹을 정의

- 수치(표준화) 범주(원-핫) 전처리와 모델을 일괄 구성
- Naïve Bayes, SVM, K-NN, Decision Tree, RandomForest, Gradient Boosting 6종을 StratifiedFold 비교선정

→ SVM(RBF) 최종 선정 (교차검증 평균 0.8273)

- 최종 선택 모델로 홀드아웃 테스트에서 Accuracy, Precision, Recall, F1 산출
- 혼동행렬로 오류 패턴 진단

→ 테스트 정확도 81.6%