

모델 선택 실전 QA

Q1. 고객 이탈 예측을 하려고 합니다. 어떤 모델이 적절할까요?

A1.

1. 목적: 이탈할 가능성이 있는 고객을 사전에 분류
2. 추천 모델: 로지스틱 회귀, 랜덤 포레스트, XGBoost
3. 이유:
 - 로지스틱 회귀는 해석이 쉬워 주요 영향 요인 설명 가능
 - 트리 기반 모델은 변수 간 복잡한 상호작용을 반영하고 예측 정확도도 높음
 - 고객 수가 많고, 피쳐도 다양할 경우 앙상블 모델이 유리

Q2. 웹 서비스의 버튼 위치 A/B 테스트를 했습니다. 효과를 측정하려면 어떤 방법을 써야 하나요?

A2.

1. 목적: 두 그룹의 클릭율 차이가 유의미한지 검정
2. 추천 방법: 독립 이항 비율의 Z-검정 또는 카이제곱 검정
3. 이유:
 - 클릭은 이항 데이터이므로 비율 기반 검정이 적절
 - 샘플 수가 크면 정규 근사 기반 Z-검정, 작으면 피셔의 정확 검정 고려

Q3. 고객 세분화를 하려면 어떤 모델이 좋을까요?

A3.

1. 목적: 유사한 고객끼리 군집으로 나누어 마케팅 타겟 설정
2. 추천 모델: K-Means, GMM, DBSCAN
3. 이유:
 - K-Means는 빠르고 직관적이거나, 군집 수 사전 정의 필요
 - GMM은 군집 형태가 원형이 아닐 때 적합
 - DBSCAN은 밀도 기반으로 잡음 데이터에도 강함

Q4. 캠페인 반응률이 낮은 고객군을 예측하려면 어떤 모델을 써야 하나요?

A4.

1. 목적: 캠페인 효율을 높이기 위해 반응 가능성이 낮은 대상 선별
2. 추천 모델: XGBoost (클래스 불균형 처리 우수), 로지스틱 회귀 + SMOTE
3. 이유:
 - 불균형한 이진 분류 문제에서 XGBoost는 클래스 가중치 조정 기능 내장
 - 성능과 해석의 균형을 위해 로지스틱 회귀 + SMOTE도 효과적

Q5. 판매량과 관련된 요인을 파악하고 싶습니다. 어떤 모델이 좋을까요?

A5.

1. 목적: 판매량과 독립 변수 간 관계 해석 및 예측
2. 추천 모델: 다중 선형 회귀, Lasso
3. 이유:
 - 선형 회귀는 변수 영향력 해석 가능
 - Lasso는 불필요한 변수를 제거하면서 예측 성능 향상

Q6. 날씨나 요일에 따라 매장 방문객 수가 변합니다. 예측 모델은 무엇을 써야 하나요?

A6.

1. 목적: 요인 기반으로 방문자 수 예측
2. 추천 모델: 선형 회귀, XGBoost, 랜덤 포레스트 회귀
3. 이유:
 - 변수 수가 많고 비선형 관계가 존재할 가능성 → 트리 기반 회귀 사용
 - 변수 간 상호작용 자동 학습 가능

Q7. 실시간 이상 거래를 탐지하려면 어떤 모델이 적합할까요?

A7.

1. 목적: 정상 패턴에서 벗어난 이례 거래 실시간 탐지
2. 추천 모델: Isolation Forest, One-Class SVM, Autoencoder
3. 이유:
 - 비지도 학습으로 이상 패턴을 학습
 - 실시간 대응이 필요할 경우 예측 속도가 빠른 모델 우선 적용

Q8. 신입사원의 이직 가능성을 예측하고 싶습니다. 어떤 모델이 좋을까요?

A8.

1. 목적: 조직 안정성을 위한 이직 위험군 선별
2. 추천 모델: 로지스틱 회귀, 랜덤 포레스트
3. 이유:
 - 이진 분류 문제
 - 로지스틱 회귀는 HR 실무자가 변수 영향력을 쉽게 이해할 수 있음
 - 랜덤 포레스트는 다양한 변수 조건 하에서도 강건한 성능

Q9. 월별 수요량을 예측하고 싶습니다. 어떤 모델을 쓰는 게 좋을까요?

A9.

1. 목적: 제품 재고 계획 수립
2. 추천 모델: ARIMA, Prophet, LightGBM + 시간 변수
3. 이유:
 - 계절성/추세가 뚜렷하면 ARIMA 또는 Prophet
 - 외부 변수까지 포함할 경우 회귀 기반 모델 활용 가능

Q10. 직원의 성과를 평가하는 모델이 필요합니다. 어떤 모델을 선택할 수 있나요?

A10.

1. 목적: 성과 예측 및 보상 체계 개선
2. 추천 모델: 선형 회귀, 결정트리 회귀
3. 이유:
 - 정량 지표 기반으로 성과 점수 예측
 - 선형 회귀는 각 요소의 기여도 해석 가능
 - 트리 기반 회귀는 조건 분기 기준으로 해석 가능