신용카드 고객 세그먼트 분류

평가기준 : 데이터 분석 결과, 정확도 / 머신러닝 결과 X

**[1단계] 프로젝트 목표 정의 및 역할 분담**

* **목표 요약**: 고객의 Segment 컬럼을 예측하는 AI 기반 알고리즘을 만들되, 핵심은 EDA 및 분석 인사이트 도출.
* **역할 분담** (4명 기준 예시):  
  1. **회원/신용정보 전담** → 인적 정보 기반 세그먼트 분석
  2. **승인매출/청구입금 전담** → 소비패턴, 카드 이용행태 분석
  3. **잔액/성과/마케팅 전담** → 수익성, 마케팅 반응성 분석
  4. **채널/전체 통합/모델링 전담** → 채널 접근 경로 및 전체 통합/최종 모델링

✅ Tip: 각자 분석한 내용을 “세그먼트 A는 이런 소비패턴이 특징이다”처럼 정리하면 좋음

**[2단계] 데이터 구조 파악 및 병합**

* train/test 모두 ID 기준으로 8개 테이블 존재(1.회원정보 2.신용정보 3.승인매출정보 4.청구입금정보 5.잔액정보 6.채널정보 7.마케팅정보 8.성과정보)
* 공통키(ID) 기준으로 병합

#### **[3단계] EDA (탐색적 데이터 분석)**

**여기서 제일 시간 많이 쓰는 게 핵심**

* **Segment별 분포 분석** → 전체 비율, 클래스 불균형 여부
* **회원/신용정보 기반 분석**
  + 성별/연령/Life Stage 등 별로 Segment 차이?
  + 신용등급/카드한도 등과 Segment의 상관성?
* **소비/입금 패턴 분석**
  + 월별 승인금액 총합, 업종별 소비 유형
  + Segment별 소비습관 차이 (고액결제? 정기결제? 어떤 업종에서 많이 쓰나?)
* **잔액/성과/마케팅 반응 분석**
  + 고객 충성도, 리텐션 지표, 마케팅 반응률

📊 시각화 추천: boxplot, violinplot, countplot, heatmap, lineplot

#### **[4단계] 주요 인사이트 요약**

* Segment 별 특징을 문장으로 정리  
  + A는 30대 이상, 소비활동 활발, 마케팅 응답 높음
  + B는 젊고 소비금액 적음, 이탈률 높음

👉 이걸 발표/보고서에 **핵심 분석 인사이트**로 강조!

#### **[5단계] 모델 개발**

* **XGBoost/CatBoost 등으로 분류**
  + 주요 피처 중요도 시각화 (feature\_importances\_)
  + 예측 정확도보다는 “중요 변수”에 더 집중

💡 예: “Segment 예측에 가장 영향을 미친 변수는 월평균 승인금액, Life\_Stage, 성별 등이다”

#### **[6단계] 최종 보고서 / 발표 정리**

* 배경 → 분석방법 → 인사이트 요약 → 모델 결과 요약 → 제안사항
* 특히, **"Segment별 마케팅 전략/CRM 전략 제안"** 같은 실질적 인사이트 포함하면 최고!

**\* 모델 개발 과정**

Macro F1 특성상 A와 B 클래스 예측 정확도가 전체 성적에 큰 영향을 줌

-> 그래서 AB 클래스만 따로 학습하는 모델을 만듦

✅ [1] 메인 모델 파이프라인 (Segment C/D/E 분류 목적)

1. 회원정보 ~ 성과정보 통합

2. 상수/중복 컬럼 제거 (train: 746, test: 745 컬럼)

3. Segment A, B 제거 (train = train[~train['Segment'].isin(['A', 'B'])])

4. 라벨 인코딩 및 데이터 분리, 범주형 지정

5. Optuna 하이퍼파라미터 튜닝

6. 최적 파라미터 저장 (.json)

7. 최종 모델 학습 (CatBoostClassifier.fit())

8. 예측 결과: C/D/E 예측 및 제출

→ 🎯 Segment C, D, E 분류 모델 (공식 평가 제출용)

✅ [2] VIP 모델 파이프라인 (Segment A,B 예측 목적)

1. 동일하게 통합된 train, test에서 Segment A,B만 필터링

2. A Segment에서 고정된 패턴 컬럼 탐색 (nunique() == 1)

3. 해당 패턴과 6개월간 일치하는 ID만 추출

4. 그 ID들로 구성된 train\_vips\_A.csv, test\_vips\_A.csv 저장

5. 이 필터링된 데이터에서 다시 A~E 전체 Segment 예측 모델 학습

6. 예측 결과에서 Segment == 'A'인 고객 ID 추출

→ 🎯 VIP (A Segment) 후보 식별용 보조 모델