**오버샘플링 및 표준화 순서**

\*\*SVM(Support Vector Machine)\*\*은 \*\*데이터의 스케일\*\*에 민감하기 때문에 \*\*표준화\*\*가 중요한 단계입니다. 또한, \*\*오버샘플링(SMOTE 등) 기법\*\*은 데이터 분포를 변경하는 작업이므로 \*\*표준화 이전에 진행하는 것이 일반적으로 권장\*\*됩니다.

### 추천 데이터 처리 순서

1. \*\*목표 및 설명 변수 분리\*\*: 먼저 전체 데이터에서 목표 변수와 설명 변수를 분리합니다.

2. \*\*Train/Test 데이터 분할\*\*: 설명 변수와 목표 변수를 분리한 후, 학습용 데이터와 테스트용 데이터를 분할합니다. 이렇게 해야 \*\*테스트 데이터에는 오버샘플링이 적용되지 않아\*\* 성능 평가에 신뢰성을 유지할 수 있습니다.

3. \*\*Train 데이터 오버샘플링\*\*: 학습 데이터(train)에서 오버샘플링을 수행합니다. 오버샘플링은 학습 데이터에서만 적용해야 하며, 테스트 데이터에는 절대 적용하지 않습니다.

4. \*\*표준화\*\*:

- \*\*Train 데이터\*\*에 표준화를 적용하고, 동일한 변환기를 사용해 \*\*Test 데이터\*\*를 변환합니다.

- 이 순서를 유지하면 오버샘플링된 학습 데이터와 원본 테스트 데이터 모두 같은 스케일로 변환됩니다.

### 요약된 순서

1. \*\*목표 및 설명변수 분리\*\* (`X`, `y` 분리)

2. \*\*Train/Test 데이터 분할\*\* (예: `train\_test\_split`)

3. \*\*Train 데이터 오버샘플링\*\* (예: SMOTE 등)

4. \*\*표준화 (Train 데이터에 학습 후 Test 데이터 변환)\*\*

이 순서를 따르는 것이 \*\*SVM 분석에서 오버샘플링과 표준화를 적용하는 올바른 방식\*\*입니다.