

report

SMOTE 적용

SMOTE 적용 전후 class 분포는 다음과 같다.

[SMOTE 적용 전 Class 분포]

Class

0 7999

1 394

Name: count, dtype: int64

[SMOTE 적용 후 Class 분포]

Class

0 7999

1 7999

Name: count, dtype: int64

SMOTE를 적용하는 이유는 사기 거래 데이터가 정상 거래에 비해 너무 적어 모델이 정상 거래 위주로만 학습하여 클래스 불균형 문제가 발생하는 것을 방지하기 위함이다.

SMOTE 방법으로 새로운 데이터를 생성해 소수 클래스 수를 늘림으로서 클래스 불균형 문제를 해결할 수 있다.

최종 성능 평가

- 선정한 모델: Random Forest

--- [Classification Report] ---

	precision	recall	f1-score	support
0	0.99	1.00	1.00	2001
1	0.95	0.89	0.92	98
accuracy			0.99	2099
macro avg	0.97	0.94	0.96	2099
weighted avg	0.99	0.99	0.99	2099

Class 1 PR-AUC: 0.9537

Class 0 PR-AUC: 0.9980

- class 0
 - Recall: 1.00
 - F1-score: 1.00
 - PR-AUC: 0.9980
- class 1
 - Recall: 0.89 (>0.80)
 - F1-score: 0.92 (>0.88)
 - PR-AUC: 0.9537 (>0.90)
- 결과 분석

- class 0 (정상거래): 모든 지표가 1.00에 근접하는 성능을 보였다. 이는 모델이 정상거래를 오탐하지 않고 안정적으로 분류하고 있음을 의미한다.
- class 1 (사기거래): Random Forest모델과 SMOTE 오버샘플링을 적용한 결과 위와 같은 수치가 도출되었으며 이는 모든 목표 수치를 상회하는 우수한 성능을 보였다. 모델이 사기 거래를 변별해내는 능력이 탁월함을 입증한다.
- 별도의 하이퍼파라미터 튜닝이나 threshold 조정을 거치지 않은 기본 random forest 모델임에도 불구하고, SMOTE 기법을 통한 데이터 불균형 해소가 효과적으로 작용하였다.
- 추후 recall을 더 높여야 하는 상황이 온다면, threshold를 하향 조정하는 방식을 추가적으로 고려해 볼 수 있다.