

## Iris 데이터셋을 활용한 품종별 Petal Length 평균 차이 검정

### 1. 데이터 불러오기 및 구조 확인

``sns.load_dataset('iris')``로 데이터 불러옴

총 150 개 샘플, 5 개 변수 (꽃받침/꽃잎 길이·너비 + 품종)

결측치 없음

### 2. 기술통계량 분석

Species	Count	Mean	Std Dev	Min	25%	50%	75%	Max
setosa	50	1.462	0.174	1.0	1.4	1.50	1.575	1.9
versicolor	50	4.260	0.470	3.0	4.0	4.35	4.60	5.1
virginica	50	5.552	0.552	4.5	5.1	5.55	5.88	6.9

### 3. 시각화(Boxplot)

Virginica 가 가장 길고, Setosa 가 가장 짧음

### 4. 정규성 검정

Species	W	p	정규성
setosa	0.9550	0.0548	0
versicolor	0.9660	0.1585	0
virginica	0.9622	0.1098	0

### 5. 등분산성 검정

$p = 0.0000 \rightarrow$  등분산성 기각 (과제 지침상 가정하고 진행)

### 6~7. ANOVA

F	p
1180.1612	2.8568e-91

### 8. Tukey HSD

Group1	Group2	MeanDiff	p	차이
setosa	versicolor	2.798	0.000	0

setosa	virginica	4.090	0.000	0
versicolor	virginica	1.292	0.000	0

## 9. 결론

**\*\*Setosa\*\***: 가장 짧음 ( $\approx 1.46\text{cm}$ )

**\*\*Virginica\*\***: 가장 김 ( $\approx 5.55\text{cm}$ )

Petal Length 는 품종 분류에 유의미

# 신용카드 사기 거래 탐지

## 1. 데이터 로드 및 탐색

신용카드 거래 데이터셋 (creditcard.csv)을 불러와 구조를 확인한 결과, 총 284,807 개의 거래 중 정상 거래는 284,315 건(99.83%), 사기 거래는 492 건(0.17%)으로 심각한 클래스 불균형이 존재함을 확인하였다.

## 2. 샘플링

사기 거래(Class=1)는 전체 유지하고, 정상 거래(Class=0)는 10,000 건만 무작위로 샘플링하여 분석용 데이터셋을 구성하였다. 샘플링 후 클래스 비율은 Class 0: 95.31%, Class 1: 4.69%로 다소 완화되었다.

## 3. 데이터 전처리

Amount 변수에 대해 StandardScaler 로 표준화를 수행하여 Amount\_Scaled 변수로 대체하였다. 그 후 X와 y로 Feature와 Target을 분리하였다.

## 4. 학습/테스트 데이터 분할

Stratified 방식으로 학습셋과 테스트셋을 8:2로 분할하였으며, 클래스 비율을 유지하기 위해 stratify=y 옵션을 적용하였다.

## 5. SMOTE 적용

학습 데이터에 대해 SMOTE(Synthetic Minority Over-sampling Technique)를 적용하여 소수 클래스(Class=1)를 오버샘플링하였다. 이는 모델이 클래스 불균형으로 인해 소수 클래스를 무시하지 않도록 하기 위함이다.

## 6. 모델 학습 및 평가

RandomForestClassifier를 사용하여 모델을 학습하였다. 이 모델은 다수의 결정 트리를 앙상블하여 예측하는 방식으로, 클래스 불균형에 robust하고 높은 정확도와 재현율을 기대할 수 있다. class\_weight='balanced' 옵션을 통해 불균형 문제를 보정하였다.

모델 성능:

- Precision (Class 1): 0.9647
- Recall (Class 1): 0.8367
- F1-score (Class 1): 0.8962

- PR-AUC: 0.9203

## 7. 최종 성능 평가

모델의 성능은 목표 기준( $\text{Recall} \geq 0.80$ ,  $F1 \geq 0.88$ ,  $\text{PR-AUC} \geq 0.90$ )을 모두 충족하였다. 따라서 해당 모델은 실전 상황에서도 사기 거래를 효과적으로 탐지할 수 있을 것으로 기대된다.