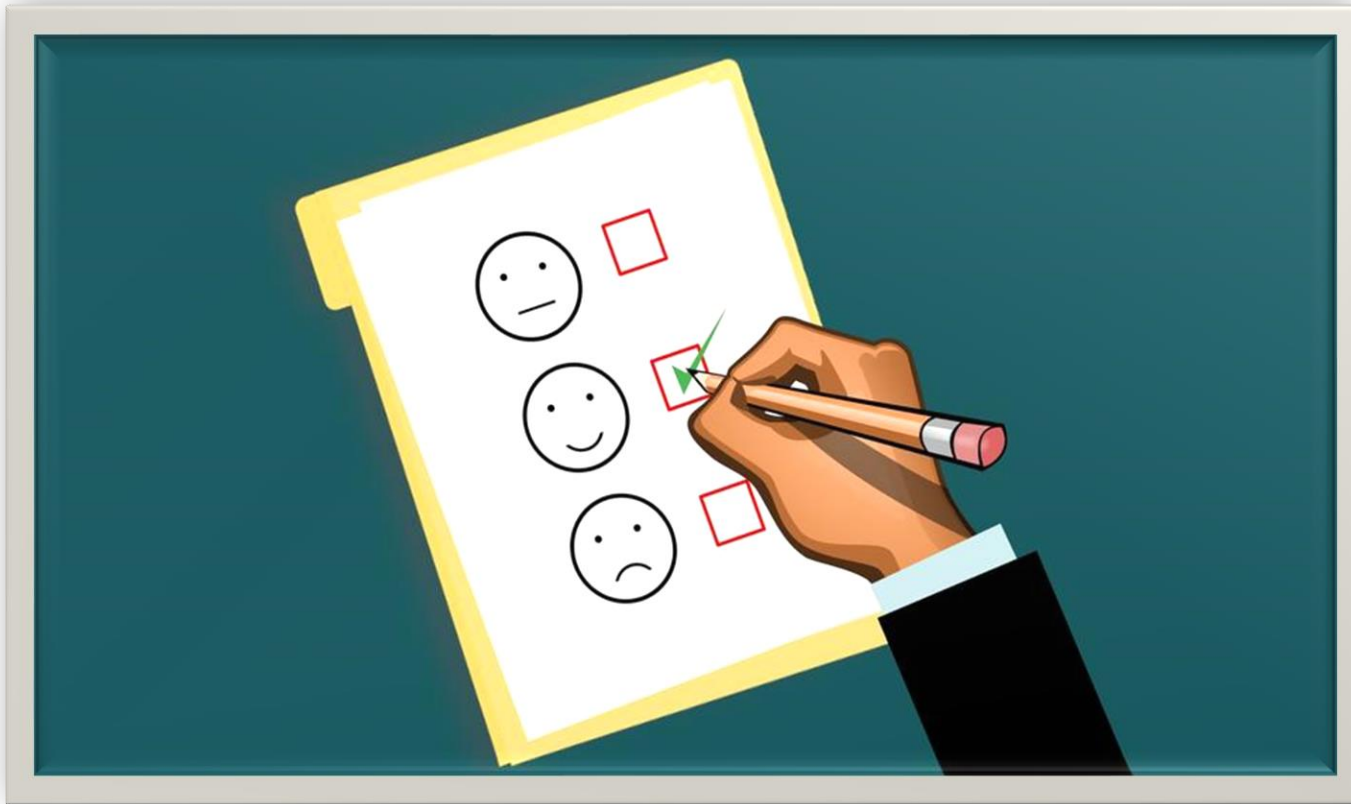


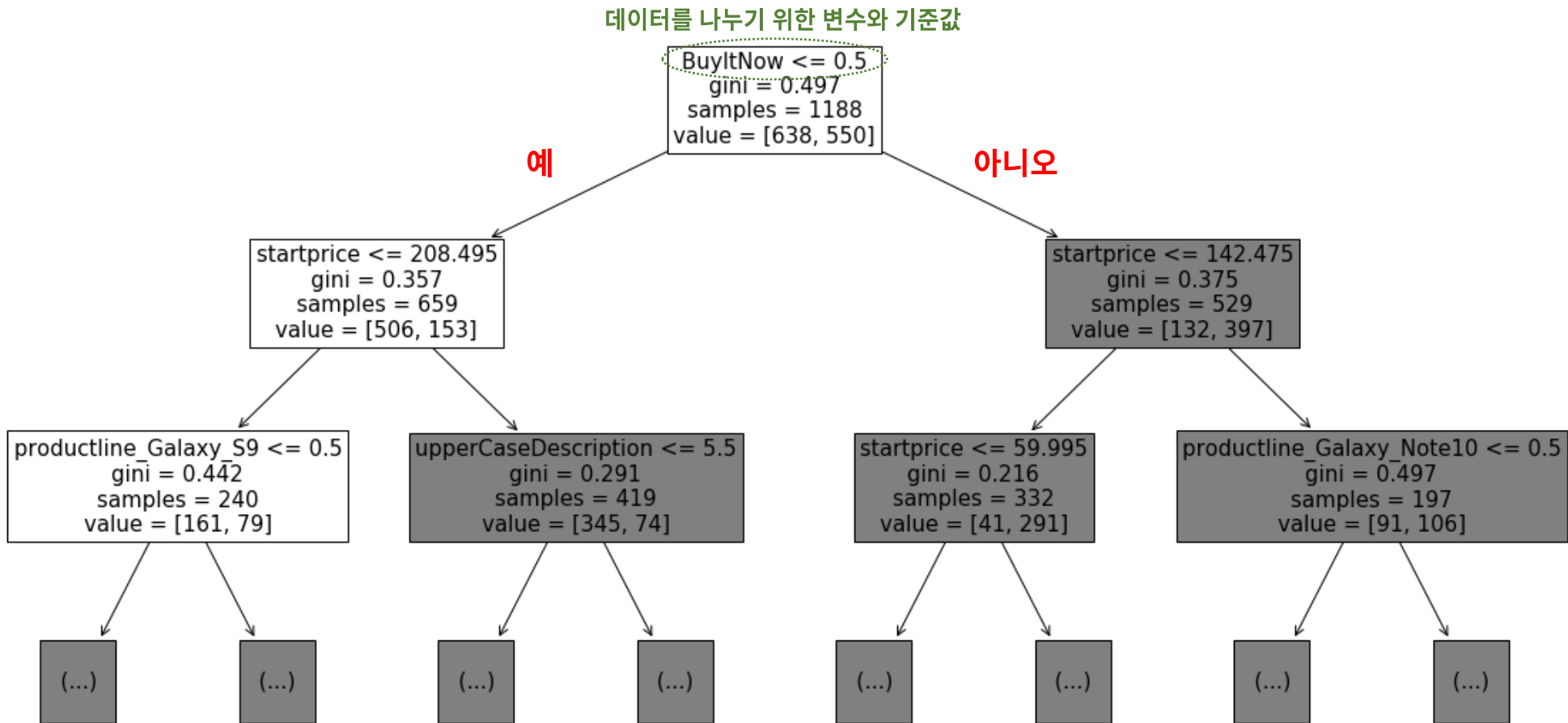
## Chapter 05. 구매 요인 분석 (Decision Tree)

# 분석의 목적

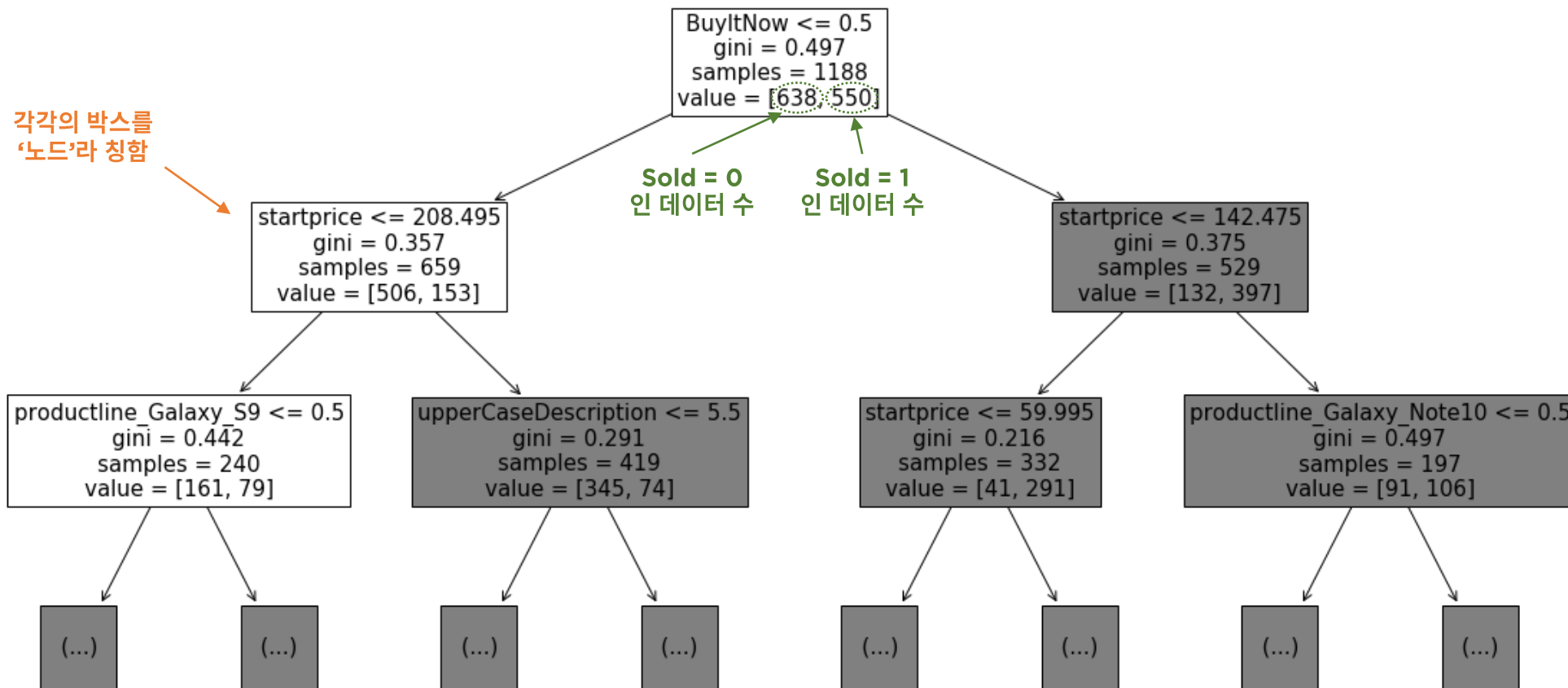
# 다시전 트리 모델을 통하여 온라인 경매 아이템 판매여부를 예측하고 각 변수의 영향도를 확인



- 특정 변수와 특정 기준점을 잡아서, '예', '아니오'에 따라 데이터를 분류
- 가장 중요하다고 여겨지는 변수와 기준점부터 적용됨 (알고리즘이 알아서 찾아줌)
- 중요하다는 기준은 Gini index로 결정



- **Gini index**: 각 노드가 얼마나 순수한가의 평가 지표. 0 ~ 0.5의 값을 가지며, 모든 데이터가 한쪽 클래스에만 속하면 0, 데이터가 정확히 반반일 경우 0.5의 값을 보여줌. 즉, 낮은 값일수록 노드 안의 데이터가 한쪽 클래스에 많이 쏠려있음을 의미
- 그러므로 이 알고리즘은 다음 노드의 **Gini 값이 낮아지는 방향으로 데이터를 분류**



# Decision Tree

Non Parametric

Feature Power  $X$

# Logistic Regression

Parametric

Feature Power  $0$