Python 기반 머신러닝/딥러닝 실습

(전북대학교/데이터 청년 캠퍼스 교육과정 – Machine Learning)

2021. 7. 22 ~ 7. 30

(총 28 시간, 4시간/일 x 7일)

정 준 수 Ph.D

과정 목표

복잡한 데이터 구조 패턴을 기계(컴퓨터)로 하여금 스스로 학습하게 하는 인공지능 기계학습 기술을 활용하여 현업의 빅데이터를 지도학습이나 자율학습형태로 분석하고 성능을 평가하여 그 결과를 실제 업무에 적용

1. 분석 방법론, 빅데이터 분석 사례, 파이썬을 이용한 빅데이터 분석 환경 구성
2. 지도학습 모델, 회귀분석, 시계열 분석, 예측 모델 개발, HMM, SVM
3. 지도학습 모델, 분류 분석, 인공신경망, 앙상블모델(XGBoost, LGBM)
4. 의사결정나무, SVM, 자율학습 모델, 데이터 군집 모델 개발, K-means Clustering
5. 빅데이터 모델 평가, 혼돈 메트릭스, AUC, RMSE, F-measure
6. 시각화 도구, 시각화 스토리 텔링, 분석 정보 시각화하기, 시각화 보정

Deep learning is not like pure mathematics. It is a heavily experimental field, so it's important to be a strong practitioner, not just a theoretician.

딥 러닝은 순수한 수학과는 다릅니다. 매우 실험적인 분야이기 때문에 이론가가 아닌 실무자가 되는 것이 중요합니다.

머신러닝 vs 딥러닝

머신러닝이란?

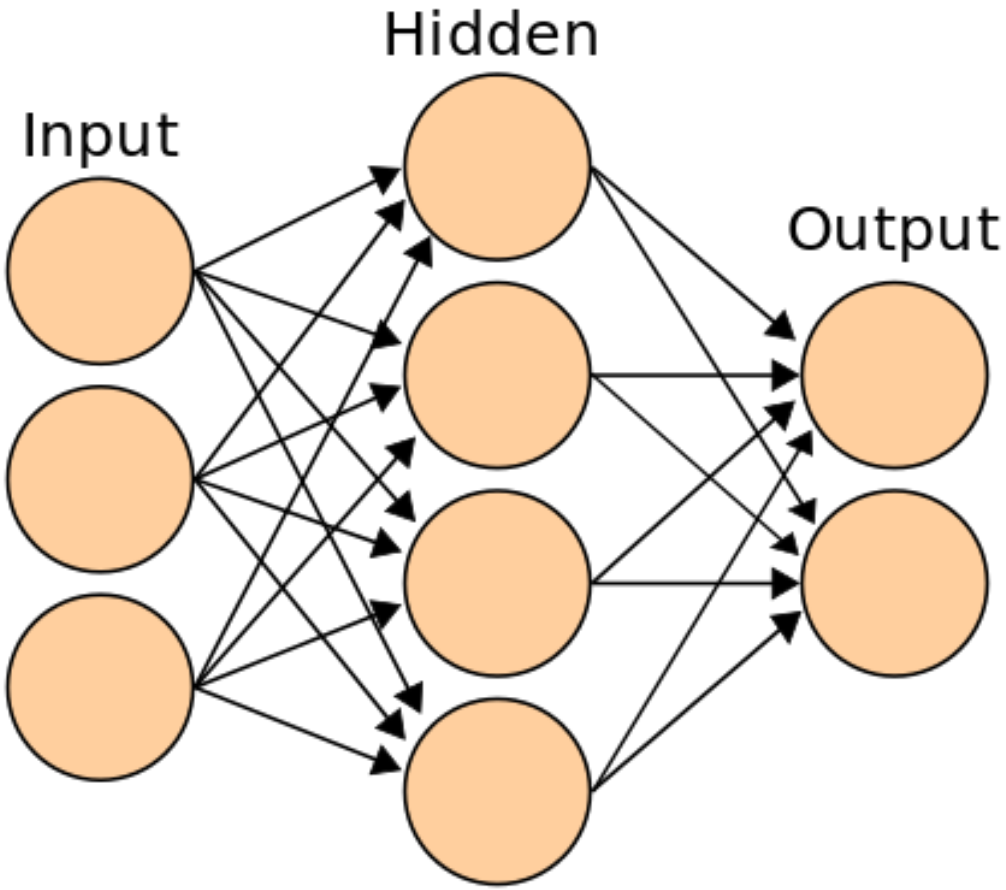
인간이 다양한 경험과 시행착오를 통해 지식을 배우는 것처럼, 컴퓨터에게 충분히 많은 데이터를 주고, 거기에서 일반적인 패턴을 찾아내게 하는 방법을 말합니다. (머신러닝의 대표적인 알고리즘이 딥러닝입니다.)

딥러닝이란?

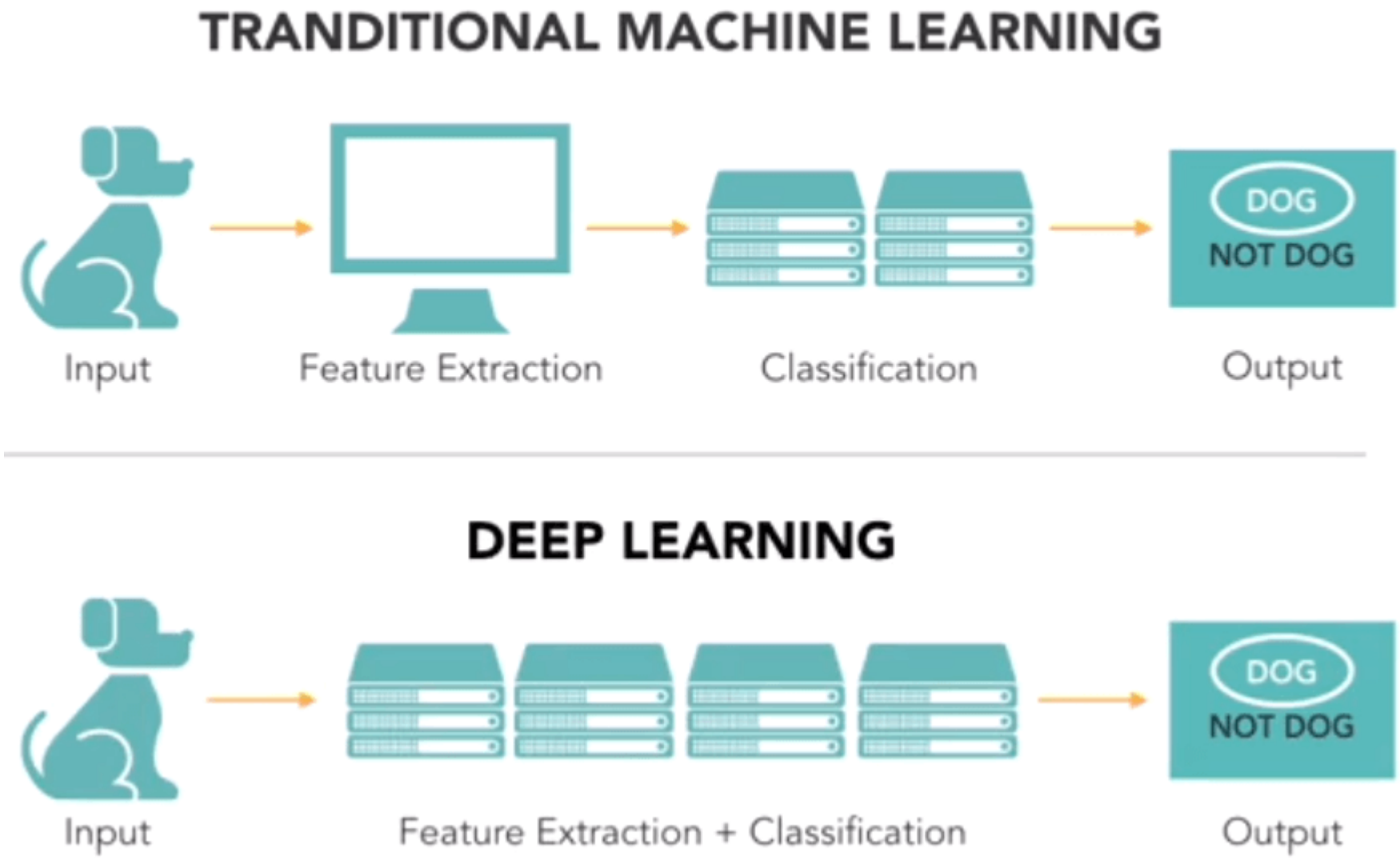
머신러닝의 대표적인 학습법이 다음장 그림처럼 여러 층을 거쳐 점점 추상화 단계로 접어드는 알고리즘 형태인 '딥 러닝(Deep Learning)'입니다.

패턴을 찾기 위해선, 수많은 데이터가 있어야 되겠지요. 패턴을 견고하게 만드는 일종의 학습용 훈련데이터 말입니다.

딥러닝 레이어

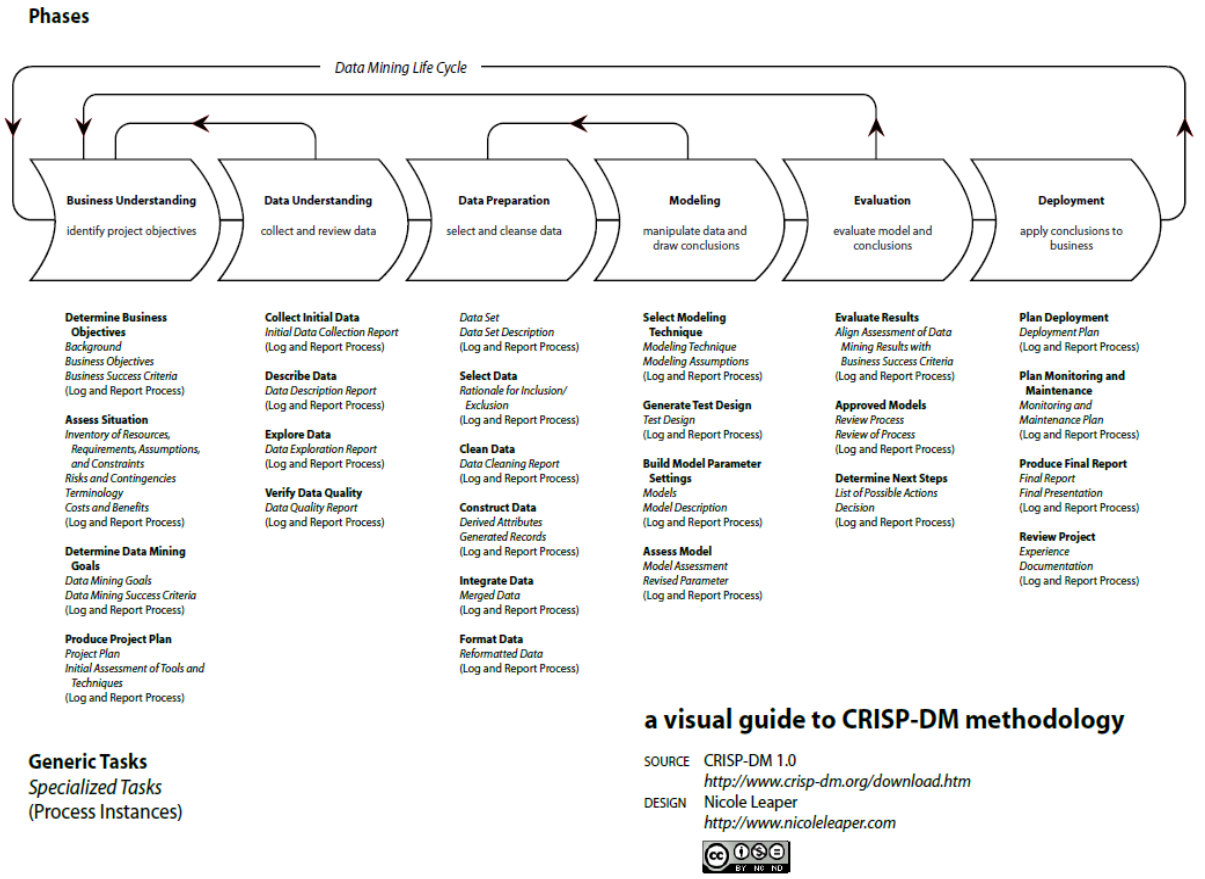


Machine Learning vs Deep Learning



CRISP-DM (Cross Industry Standard Process for Data Mining)

CRISP-DM(Cross Industry Standard Process for Data Mining)은 데이터 마이닝 전문가가 사용하는 일반적인 접근 방식을 설명한 가장 널리 사용되는 공개 표준 분석 모델입니다.



Machine Learning에서 “예측”이란?

이전에 본적 없는 새로운 데이터에 대한 정확한 출력 예측

Predictive Analytics

1 무엇을 예측하는가?

2 무엇을 할 것인가?

Machine Learning - Training

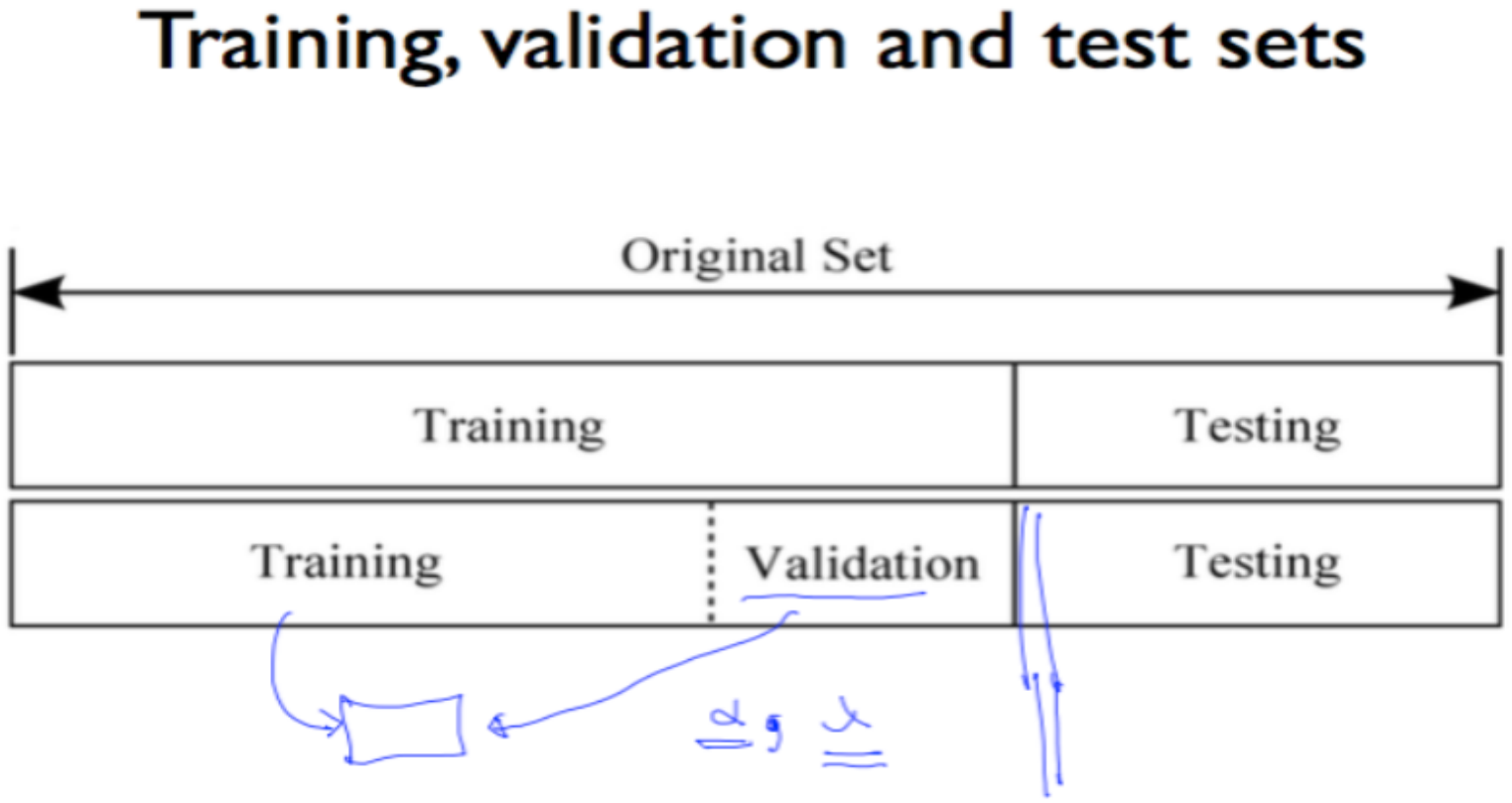
	Input			Output
Example 1	0	0	1	0
Example 2	1	1	1	1
Example 3	1	0	1	1
Example 4	0	1	1	0

New situation	1	0	0	?

```
from numpy import exp, array, random, dot
training_set_inputs = array([[0, 0, 1], [1, 1, 1], [1, 0, 1], [0, 1, 1]])
training_set_outputs = array([[0, 1, 1, 0]]).T
random.seed(1)
synaptic_weights = 2 * random.random((3, 1)) - 1
for iteration in range(10000):
    output = 1 / (1 + exp(-(dot(training_set_inputs, synaptic_weights))))
    synaptic_weights += dot(training_set_inputs.T, (training_set_outputs - output) * output * (1 - output))
print (1 / (1 + exp(-(dot(array([1, 0, 0]), synaptic_weights)))))
```

출처: https://medium.com/technology-invention-and-more/how-to-build-a-simple-neural-network-in-9-lines-of-python-code-cc8f23647ca1

머신러닝 학습 데이터의 구성



데이터 영역별 구성도다. 데이터가 충분하다면 validation set을 구성하지 않을 이유가 없다. 쉽게 보면 test를 한번 더 한다고 생각해도 좋을 것이다. 머신러닝의 교과서에 해당하는 mnist 예제는 train 55,000개, validation 5,000개, test 10,000개의 dataset으로 구성되어 있다.

예측 분석 응용: 이탈 모델링으로 고객 이탈 방지하기

1 무엇을 예측하는가?

어느 고객이 떠나갈 것인가.

2 무엇을 할 것인가?

떠날 위기에 있는 고객들을 타깃으로 한 고객 유지 마케팅을 수행한다.

예측 분석 응용: 부동산 담보대출 채권 가치 추산

1 무엇을 예측하는가?

부동산 담보대출 고객 중에서 누가 향후 90일 내에 조기 상환할 것인가.

2 무엇을 할 것인가?

부동산 담보대출 채권의 가치를 계산한 후 다른 은행에 팔아 넘길지 여부를 결정한다.

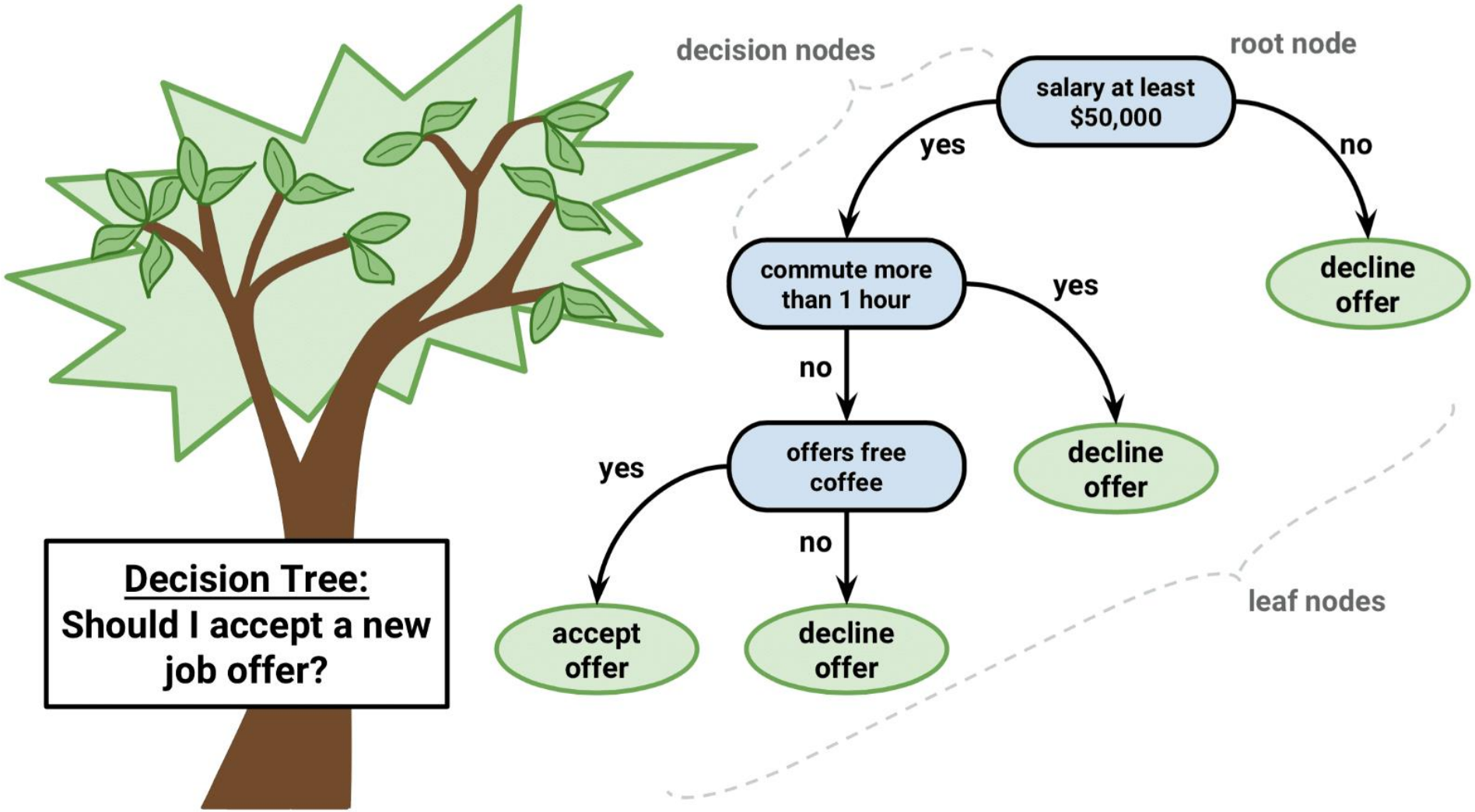
나무에서 돈이 자란다!

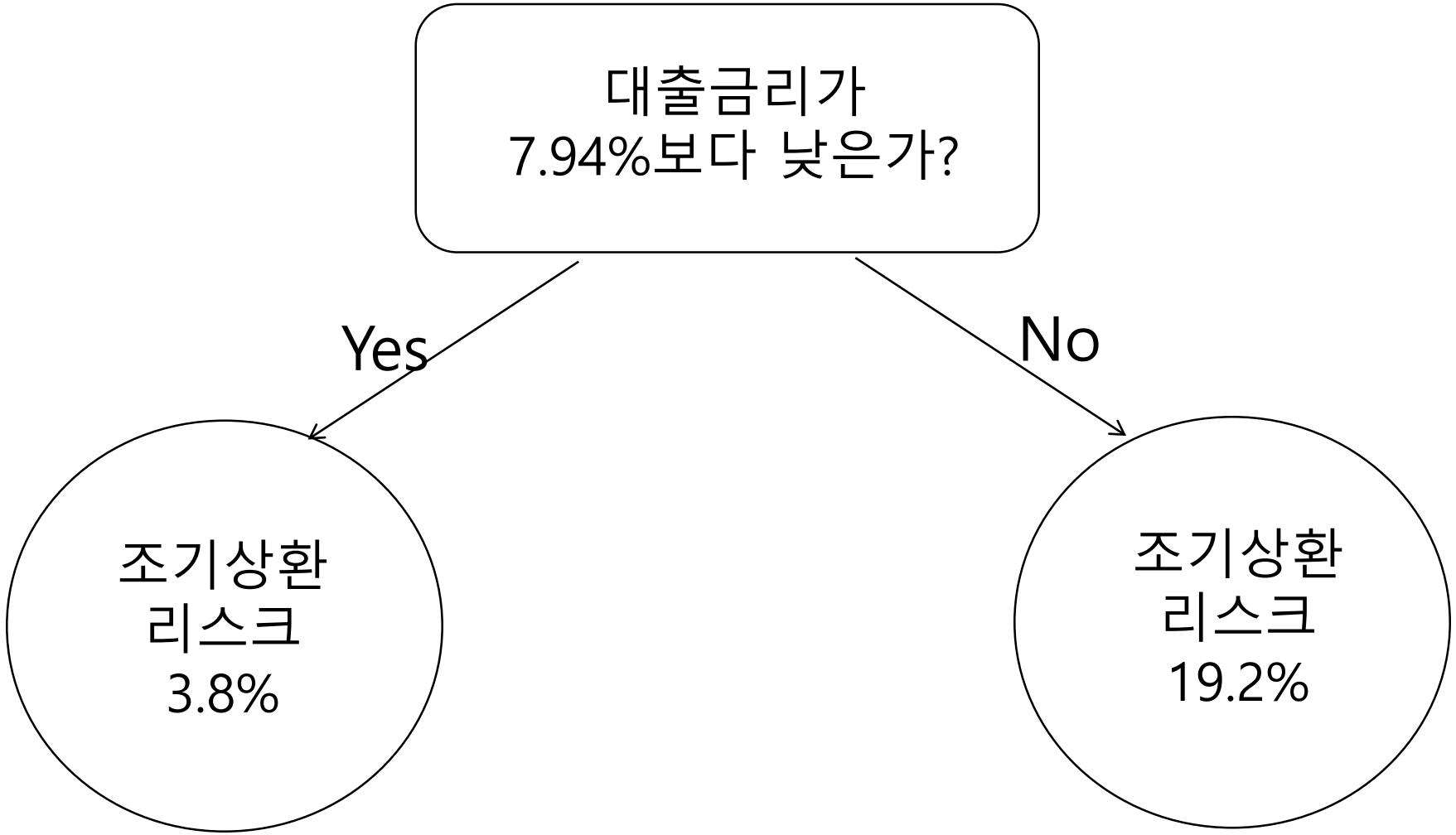
체이스 은행 예측모델은 부동산 담보대출 중에서 실제로

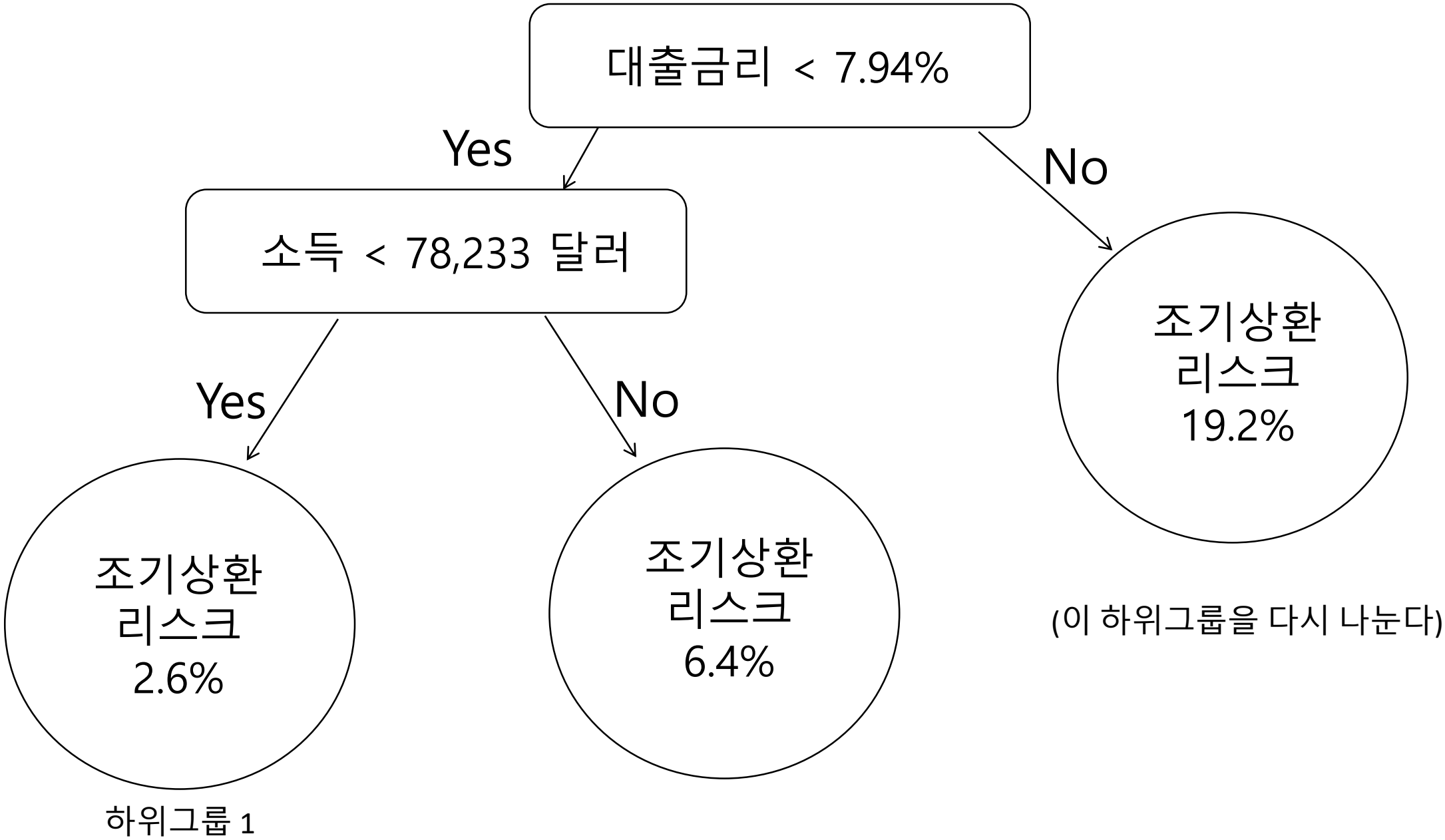
조기상환된 대출 건들 중 74%를 정확하게 인식해 내어서

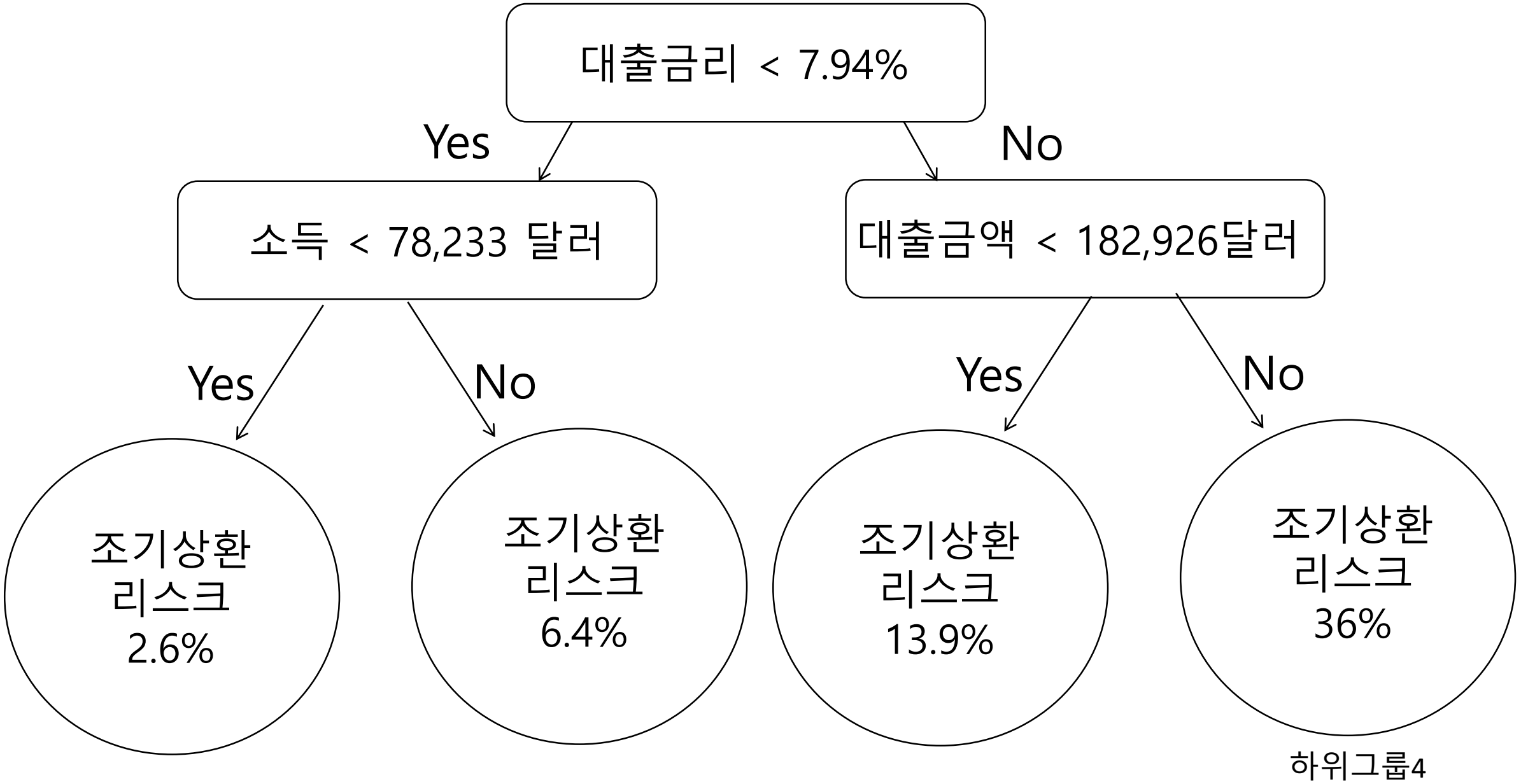
부동산 담보대출 포트폴리오를 성공적으로 관리함.

Decision Tree

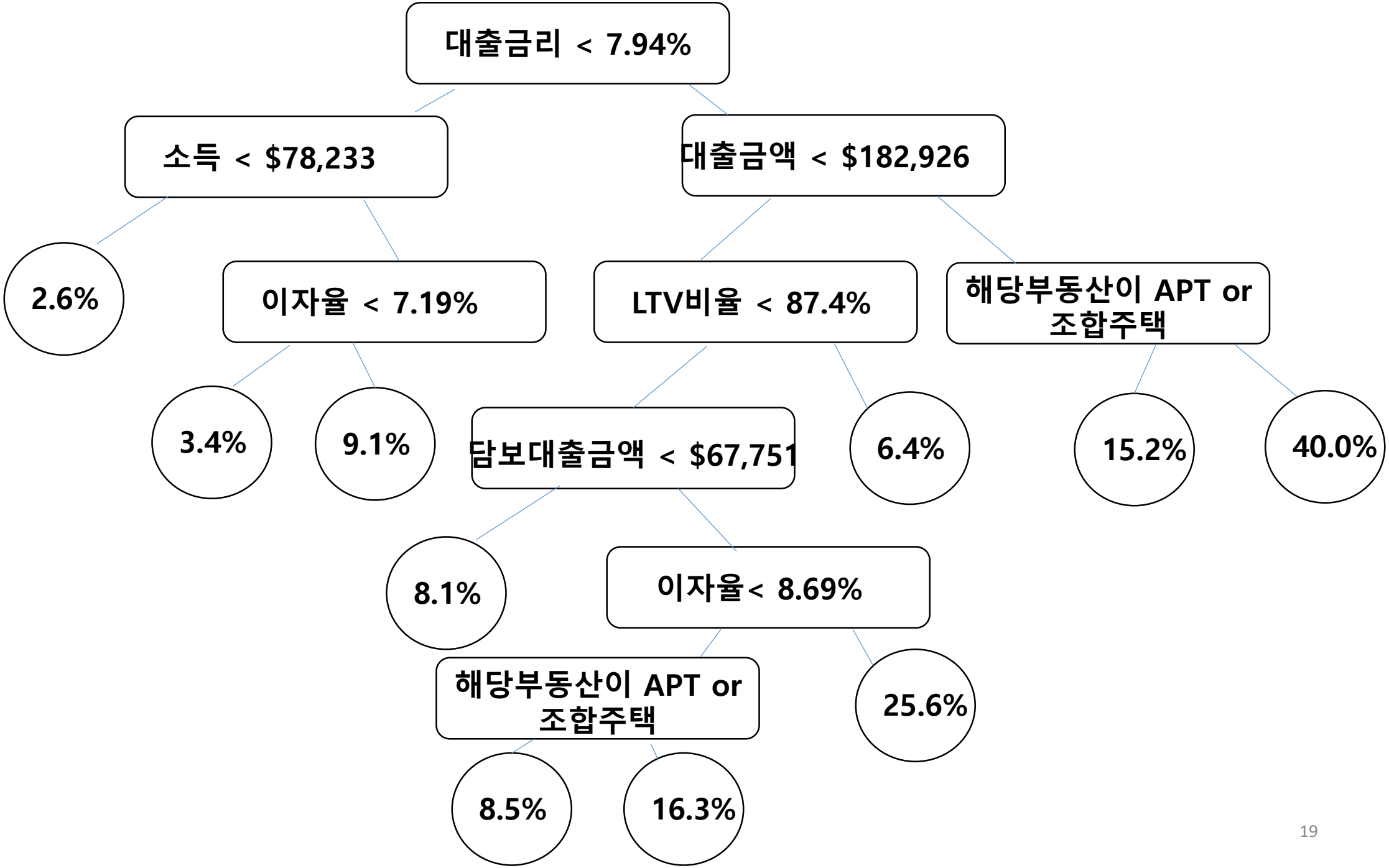



대출금리가
7.94%보다 낮은가?
Yes
No
조기상환
리스크
3.8%
조기상환
리스크
19.2%


대출금리 < 7.94%
Yes
No
소득 < 78,233 달러
Yes
No
조기상환
리스크
19.2%
조기상환
리스크
2.6%
조기상환
리스크
6.4%
(이 하위그룹을 다시 나눈다)
하위그룹 1


대출금리 < 7.94%
Yes
No
소득 < 78,233 달러
대출금액 < 182,926달러
Yes
No
Yes
No
조기상환 리스크 2.6%
조기상환 리스크 6.4%
조기상환 리스크 13.9%
조기상환 리스크 36%

하위그룹4


대출금리 < 7.94%
소득 < $78,233
대출금액 < $182,926
2.6%
이자율 < 7.19%
LTV비율 < 87.4%
해당부동산이 APT or 조합주택
3.4%
9.1%
담보대출금액 < $67,751
6.4%
15.2%
40.0%
8.1%
이자율< 8.69%
해당부동산이 APT or 조합주택
25.6%
8.5%
16.3%

만약(IF):

부동산 담보대출 금액이 67,751 달러와 같거나 그보다 더 많고 182,926 달러보다 작다.

그리고(AND):

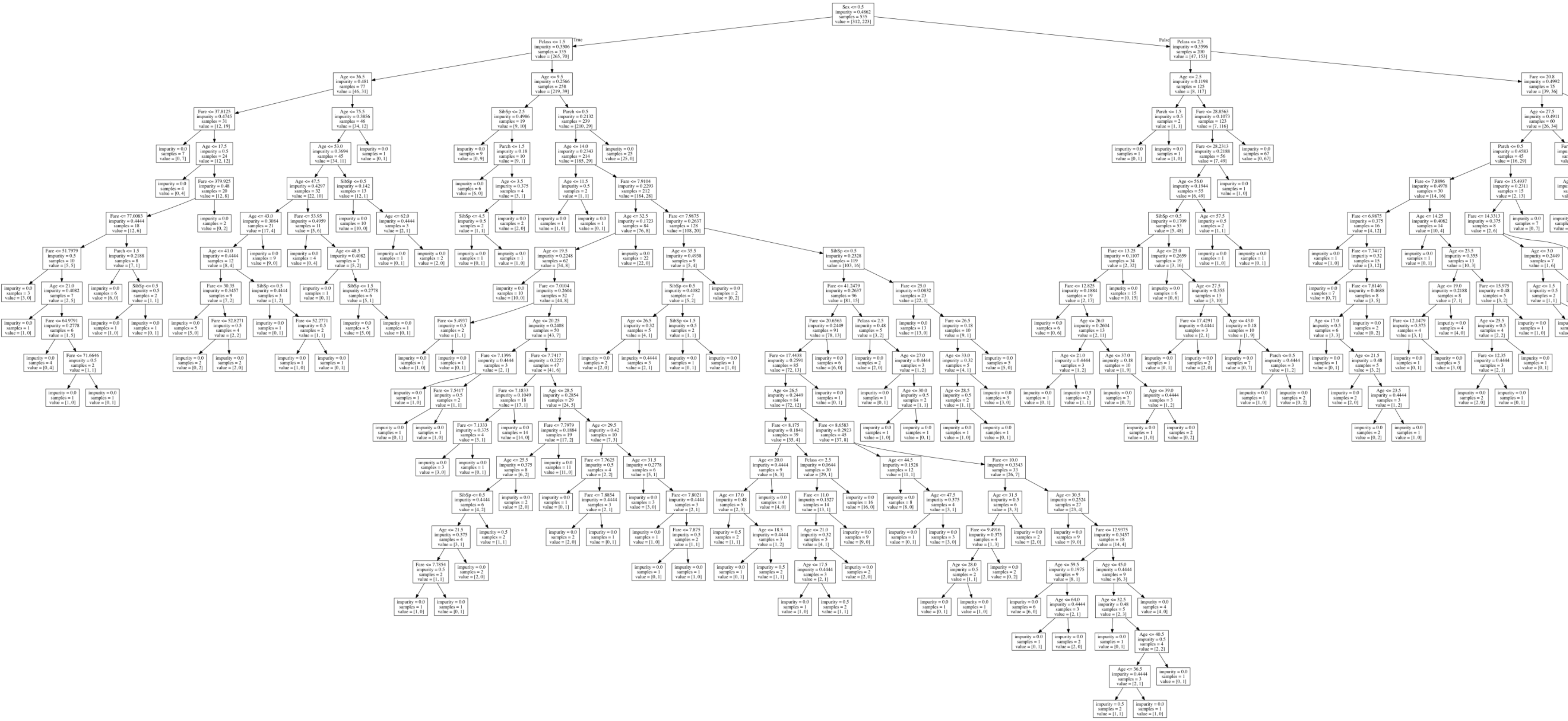
이자율이 8.69%와 같거나 그보다 더 높다.

그리고(AND):

부동산 자산가치 대비 대출금액의 비율이 87.4% 보다 작다.

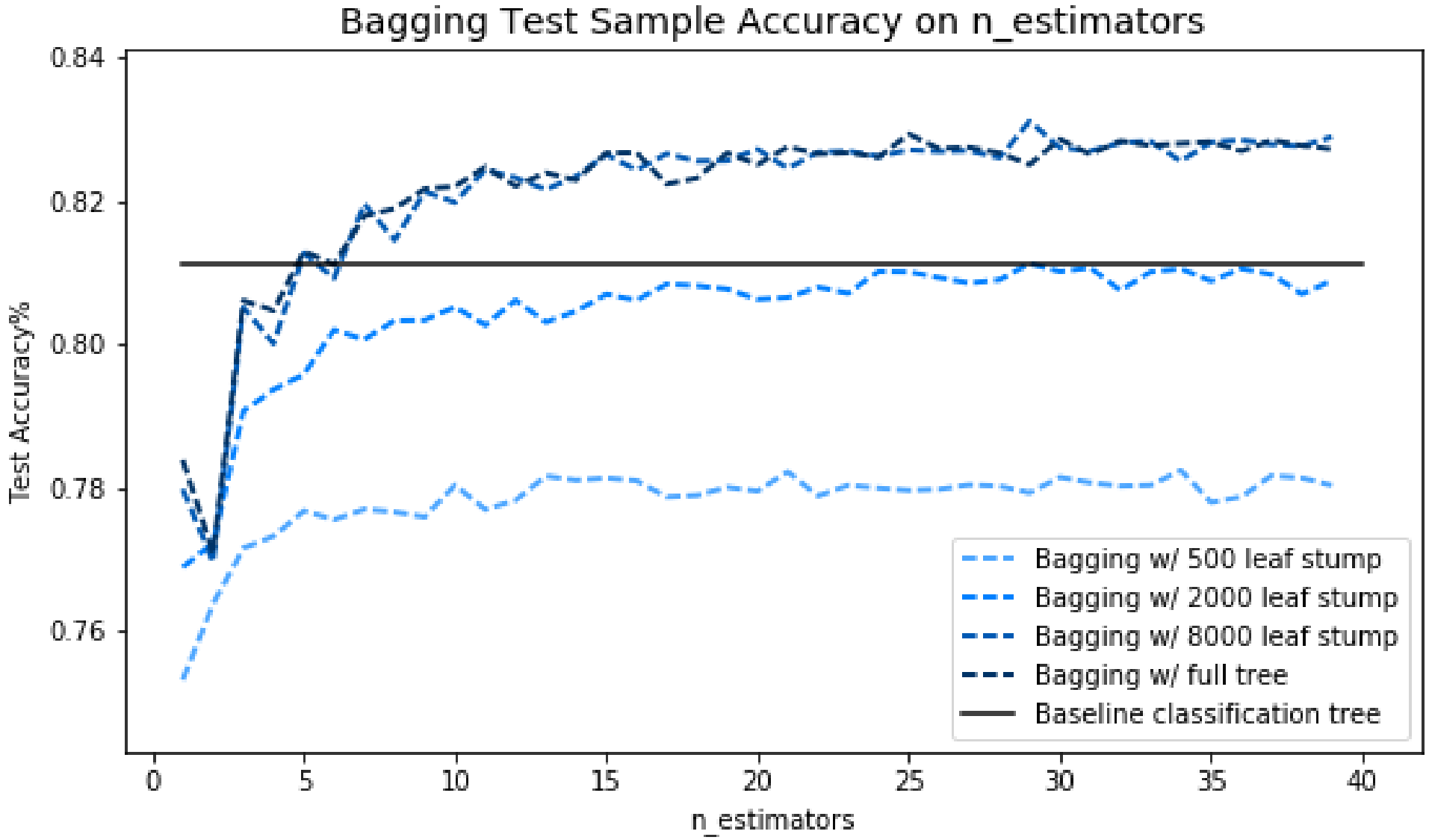
그러면(THEN):

조기상환 확률은 25.6% 이다.


Sex <= 0.5 impurity = 0.4862 samples = 535 value = [312, 223]
True
False
Pclass <= 1.5 impurity = 0.3306 samples = 335 value = [265, 70]
Pclass <= 2.5 impurity = 0.3596 samples = 200 value = [47, 153]
Age <= 36.5 impurity = 0.481 samples = 77 value = [46, 31]
Age <= 9.5 impurity = 0.2566 samples = 258 value = [219, 39]
Age <= 2.5 impurity = 0.1198 samples = 125 value = [8, 117]
Fare <= 20.8 impurity = 0.4992 samples = 75 value = [39, 36]
Fare <= 37.8125 impurity = 0.4745 samples = 31 value = [12, 19]
Age <= 75.5 impurity = 0.3856 samples = 46 value = [34, 12]
SibSp <= 2.5 impurity = 0.4986 samples = 19 value = [9, 10]
Parch <= 0.5 impurity = 0.2132 samples = 239 value = [210, 29]
Parch <= 1.5 impurity = 0.5 samples = 2 value = [1, 1]
Fare <= 28.8563 impurity = 0.1073 samples = 123 value = [7, 116]
Age <= 27.5 impurity = 0.4911 samples = 60 value = [26, 34]
Fare <= 31.3312 impurity = 0.2311 samples = 15 value = [13, 2]
impurity = 0.0 samples = 7 value = [0, 7]
Age <= 17.5 impurity = 0.5 samples = 24 value = [12, 12]
Age <= 53.0 impurity = 0.3694 samples = 45 value = [34, 11]
impurity = 0.0 samples = 1 value = [0, 1]
impurity = 0.0 samples = 9 value = [0, 9]
Parch <= 1.5 impurity = 0.18 samples = 10 value = [9, 1]
Age <= 14.0 impurity = 0.2343 samples = 214 value = [185, 29]
impurity = 0.0 samples = 25 value = [25, 0]
impurity = 0.0 samples = 1 value = [0, 1]
impurity = 0.0 samples = 1 value = [1, 0]
Fare <= 28.2313 impurity = 0.2188 samples = 56 value = [7, 49]
impurity = 0.0 samples = 67 value = [0, 67]
Parch <= 0.5 impurity = 0.4583 samples = 45 value = [16, 29]
Fare <= 14.85 impurity = 0.4444 samples = 15 value = [10, 5]
impurity = 0.0 samples = 9 value = [9, 0]
Fare <= 32.8812 impurity = 0.4444 samples = 6 value = [4, 2]
impurity = 0.0 samples = 4 value = [0, 4]
Fare <= 379.925 impurity = 0.48 samples = 20 value = [12, 8]
Age <= 47.5 impurity = 0.4297 samples = 32 value = [22, 10]
SibSp <= 0.5 impurity = 0.142 samples = 13 value = [12, 1]
impurity = 0.0 samples = 6 value = [6, 0]
Age <= 3.5 impurity = 0.375 samples = 4 value = [3, 1]
Age <= 11.5 impurity = 0.5 samples = 2 value = [1, 1]
Fare <= 7.9104 impurity = 0.2293 samples = 212 value = [184, 28]
Age <= 56.0 impurity = 0.1944 samples = 55 value = [6, 49]
impurity = 0.0 samples = 1 value = [1, 0]
Fare <= 7.8896 impurity = 0.4978 samples = 30 value = [14, 16]
Fare <= 15.4937 impurity = 0.2311 samples = 15 value = [2, 13]
Age <= 54.0 impurity = 0.1975 samples = 9 value = [8, 1]
Age <= 38.5 impurity = 0.4444 samples = 6 value = [2, 4]
impurity = 0.0 samples = 2 value = [0, 2]
impurity = 0.0 samples = 4 value = [4, 0]
Fare <= 77.0083 impurity = 0.4444 samples = 18 value = [12, 6]
impurity = 0.0 samples = 2 value = [0, 2]
Age <= 43.0 impurity = 0.3084 samples = 21 value = [17, 4]
Fare <= 53.95 impurity = 0.4959 samples = 11 value = [5, 6]
impurity = 0.0 samples = 10 value = [10, 0]
Age <= 62.0 impurity = 0.4444 samples = 3 value = [2, 1]
SibSp <= 4.5 impurity = 0.5 samples = 2 value = [1, 1]
impurity = 0.0 samples = 2 value = [2, 0]
impurity = 0.0 samples = 1 value = [1, 0]
impurity = 0.0 samples = 1 value = [0, 1]
Age <= 32.5 impurity = 0.1723 samples = 84 value = [76, 8]
Fare <= 7.9875 impurity = 0.2637 samples = 128 value = [108, 20]
SibSp <= 0.5 impurity = 0.1709 samples = 53 value = [5, 48]
Age <= 57.5 impurity = 0.5 samples = 2 value = [1, 1]
Fare <= 6.9875 impurity = 0.375 samples = 16 value = [4, 12]
Age <= 14.25 impurity = 0.4082 samples = 14 value = [10, 4]
Fare <= 14.3313 impurity = 0.375 samples = 8 value = [2, 6]
impurity = 0.0 samples = 7 value = [0, 7]
impurity = 0.0 samples = 8 value = [8, 0]
impurity = 0.0 samples = 1 value = [0, 1]
Fare <= 16.45 impurity = 0.32 samples = 5 value = [1, 4]
impurity = 0.0 samples = 1 value = [1, 0]
Fare <= 51.7979 impurity = 0.5 samples = 10 value = [5, 5]
Parch <= 1.5 impurity = 0.2188 samples = 8 value = [7, 1]
Age <= 41.0 impurity = 0.4444 samples = 12 value = [8, 4]
impurity = 0.0 samples = 9 value = [9, 0]
impurity = 0.0 samples = 4 value = [0, 4]
Age <= 48.5 impurity = 0.4082 samples = 7 value = [5, 2]
impurity = 0.0 samples = 1 value = [0, 1]
impurity = 0.0 samples = 2 value = [2, 0]
impurity = 0.0 samples = 1 value = [0, 1]
impurity = 0.0 samples = 1 value = [1, 0]
Age <= 19.5 impurity = 0.2248 samples = 62 value = [54, 8]
impurity = 0.0 samples = 22 value = [22, 0]
Age <= 35.5 impurity = 0.4938 samples = 9 value = [5, 4]
SibSp <= 0.5 impurity = 0.2328 samples = 119 value = [103, 16]
Fare <= 13.25 impurity = 0.1107 samples = 34 value = [2, 32]
Age <= 25.0 impurity = 0.2659 samples = 19 value = [3, 16]
impurity = 0.0 samples = 1 value = [1, 0]
impurity = 0.0 samples = 1 value = [0, 1]
impurity = 0.0 samples = 1 value = [1, 0]
Fare <= 7.7417 impurity = 0.32 samples = 15 value = [3, 12]
impurity = 0.0 samples = 1 value = [0, 1]
Age <= 23.5 impurity = 0.355 samples = 13 value = [10, 3]
Age <= 3.0 impurity = 0.2449 samples = 7 value = [1, 6]
impurity = 0.0 samples = 1 value = [1, 0]
Age <= 30.5 impurity = 0.5 samples = 2 value = [1, 1]
impurity = 0.0 samples = 3 value = [0, 3]
impurity = 0.0 samples = 3 value = [3, 0]
Age <= 21.0 impurity = 0.4082 samples = 7 value = [2, 5]
impurity = 0.0 samples = 6 value = [6, 0]
SibSp <= 0.5 impurity = 0.5 samples = 2 value = [1, 1]
Fare <= 30.35 impurity = 0.3457 samples = 9 value = [7, 2]
SibSp <= 0.5 impurity = 0.4444 samples = 3 value = [1, 2]
impurity = 0.0 samples = 1 value = [0, 1]
SibSp <= 1.5 impurity = 0.2778 samples = 6 value = [5, 1]
impurity = 0.0 samples = 10 value = [10, 0]
Fare <= 7.0104 impurity = 0.2604 samples = 52 value = [44, 8]
SibSp <= 0.5 impurity = 0.4082 samples = 7 value = [5, 2]
impurity = 0.0 samples = 2 value = [0, 2]
Fare <= 41.2479 impurity = 0.2637 samples = 96 value = [81, 15]
Fare <= 25.0 impurity = 0.0832 samples = 23 value = [22, 1]
Fare <= 12.825 impurity = 0.1884 samples = 19 value = [2, 17]
impurity = 0.0 samples = 15 value = [0, 15]
impurity = 0.0 samples = 6 value = [0, 6]
Age <= 27.5 impurity = 0.355 samples = 13 value = [3, 10]
impurity = 0.0 samples = 7 value = [0, 7]
Fare <= 7.8146 impurity = 0.4688 samples = 8 value = [3, 5]
Age <= 19.0 impurity = 0.2188 samples = 8 value = [7, 1]
Fare <= 15.975 impurity = 0.48 samples = 5 value = [3, 2]
Age <= 1.5 impurity = 0.5 samples = 2 value = [1, 1]
impurity = 0.0 samples = 5 value = [0, 5]
impurity = 0.0 samples = 1 value = [0, 1]
impurity = 0.0 samples = 1 value = [1, 0]
impurity = 0.0 samples = 1 value = [1, 0]
Fare <= 64.9791 impurity = 0.2778 samples = 6 value = [1, 5]
impurity = 0.0 samples = 1 value = [1, 0]
impurity = 0.0 samples = 1 value = [0, 1]
impurity = 0.0 samples = 5 value = [5, 0]
Fare <= 52.8271 impurity = 0.5 samples = 4 value = [2, 2]
impurity = 0.0 samples = 1 value = [0, 1]
Fare <= 52.2771 impurity = 0.5 samples = 2 value = [1, 1]
impurity = 0.0 samples = 5 value = [5, 0]
impurity = 0.0 samples = 1 value = [0, 1]
Fare <= 5.4937 impurity = 0.5 samples = 2 value = [1, 1]
Age <= 20.25 impurity = 0.2408 samples = 50 value = [43, 7]
Age <= 26.5 impurity = 0.32 samples = 5 value = [4, 1]
SibSp <= 1.5 impurity = 0.5 samples = 2 value = [1, 1]
Fare <= 20.6563 impurity = 0.2449 samples = 91 value = [78, 13]
Pclass <= 2.5 impurity = 0.48 samples = 5 value = [3, 2]
impurity = 0.0 samples = 13 value = [13, 0]
Fare <= 26.5 impurity = 0.18 samples = 10 value = [9, 1]
impurity = 0.0 samples = 6 value = [0, 6]
Age <= 26.0 impurity = 0.2604 samples = 13 value = [2, 11]
Fare <= 17.4291 impurity = 0.4444 samples = 3 value = [2, 1]
Age <= 43.0 impurity = 0.18 samples = 10 value = [1, 9]
Age <= 17.0 impurity = 0.5 samples = 6 value = [3, 3]
impurity = 0.0 samples = 2 value = [0, 2]
Fare <= 12.1479 impurity = 0.375 samples = 4 value = [3, 1]
impurity = 0.0 samples = 4 value = [4, 0]
Age <= 25.5 impurity = 0.5 samples = 4 value = [2, 2]
impurity = 0.0 samples = 1 value = [1, 0]
impurity = 0.0 samples = 1 value = [0, 1]
impurity = 0.0 samples = 1 value = [1, 0]
impurity = 0.0 samples = 4 value = [0, 4]
Fare <= 71.6646 impurity = 0.5 samples = 2 value = [1, 1]
impurity = 0.0 samples = 2 value = [0, 2]
impurity = 0.0 samples = 2 value = [2, 0]
impurity = 0.0 samples = 1 value = [1, 0]
impurity = 0.0 samples = 1 value = [0, 1]
impurity = 0.0 samples = 1 value = [1, 0]
impurity = 0.0 samples = 1 value = [0, 1]
Fare <= 7.1396 impurity = 0.4444 samples = 3 value = [2, 1]
Fare <= 7.7417 impurity = 0.2227 samples = 47 value = [41, 6]
impurity = 0.0 samples = 2 value = [2, 0]
impurity = 0.4444 samples = 3 value = [2, 1]
impurity = 0.0 samples = 1 value = [0, 1]
impurity = 0.0 samples = 1 value = [1, 0]
Fare <= 17.4438 impurity = 0.2591 samples = 85 value = [72, 13]
impurity = 0.0 samples = 6 value = [6, 0]
impurity = 0.0 samples = 2 value = [2, 0]
Age <= 27.0 impurity = 0.4444 samples = 3 value = [1, 2]
Age <= 33.0 impurity = 0.32 samples = 5 value = [4, 1]
impurity = 0.0 samples = 5 value = [5, 0]
Age <= 21.0 impurity = 0.4444 samples = 3 value = [1, 2]
Age <= 37.0 impurity = 0.18 samples = 10 value = [1, 9]
impurity = 0.0 samples = 1 value = [0, 1]
impurity = 0.0 samples = 2 value = [2, 0]
impurity = 0.0 samples = 7 value = [0, 7]
Parch <= 0.5 impurity = 0.4444 samples = 3 value = [1, 2]
impurity = 0.0 samples = 1 value = [0, 1]
Age <= 21.5 impurity = 0.48 samples = 5 value = [3, 2]
impurity = 0.0 samples = 1 value = [0, 1]
impurity = 0.0 samples = 3 value = [3, 0]
Fare <= 12.35 impurity = 0.4444 samples = 3 value = [2, 1]
impurity = 0.0 samples = 1 value = [0, 1]
impurity = 0.0 samples = 1 value = [1, 0]
impurity = 0.0 samples = 1 value = [0, 1]
impurity = 0.0 samples = 1 value = [1, 0]
Fare <= 7.5417 impurity = 0.5 samples = 2 value = [1, 1]
Fare <= 7.1833 impurity = 0.1049 samples = 18 value = [17, 1]
Age <= 28.5 impurity = 0.2854 samples = 29 value = [24, 5]
Age <= 26.5 impurity = 0.2449 samples = 84 value = [72, 12]
impurity = 0.0 samples = 1 value = [0, 1]
impurity = 0.0 samples = 1 value = [0, 1]
Age <= 30.0 impurity = 0.5 samples = 2 value = [1, 1]
Age <= 28.5 impurity = 0.5 samples = 2 value = [1, 1]
impurity = 0.0 samples = 3 value = [3, 0]
impurity = 0.0 samples = 1 value = [0, 1]
impurity = 0.5 samples = 2 value = [1, 1]
impurity = 0.0 samples = 7 value = [0, 7]
Age <= 39.0 impurity = 0.4444 samples = 3 value = [1, 2]
impurity = 0.0 samples = 1 value = [1, 0]
impurity = 0.0 samples = 2 value = [0, 2]
impurity = 0.0 samples = 2 value = [2, 0]
Age <= 23.5 impurity = 0.4444 samples = 3 value = [1, 2]
impurity = 0.0 samples = 2 value = [2, 0]
impurity = 0.0 samples = 1 value = [0, 1]
impurity = 0.0 samples = 1 value = [0, 1]
impurity = 0.0 samples = 1 value = [1, 0]
Fare <= 7.1333 impurity = 0.375 samples = 4 value = [3, 1]
impurity = 0.0 samples = 14 value = [14, 0]
Fare <= 7.7979 impurity = 0.1884 samples = 19 value = [17, 2]
Age <= 29.5 impurity = 0.42 samples = 10 value = [7, 3]
Fare <= 8.175 impurity = 0.1841 samples = 39 value = [35, 4]
Fare <= 8.6583 impurity = 0.2923 samples = 45 value = [37, 8]
impurity = 0.0 samples = 1 value = [1, 0]
impurity = 0.0 samples = 1 value = [0, 1]
impurity = 0.0 samples = 1 value = [1, 0]
impurity = 0.0 samples = 1 value = [0, 1]
impurity = 0.0 samples = 1 value = [1, 0]
impurity = 0.0 samples = 2 value = [0, 2]
impurity = 0.0 samples = 2 value = [0, 2]
impurity = 0.0 samples = 1 value = [1, 0]
impurity = 0.0 samples = 3 value = [3, 0]
impurity = 0.0 samples = 1 value = [0, 1]
Age <= 25.5 impurity = 0.375 samples = 8 value = [6, 2]
impurity = 0.0 samples = 11 value = [11, 0]
Fare <= 7.7625 impurity = 0.5 samples = 4 value = [2, 2]
Age <= 31.5 impurity = 0.2778 samples = 6 value = [5, 1]
Age <= 20.0 impurity = 0.4444 samples = 9 value = [6, 3]
Pclass <= 2.5 impurity = 0.0644 samples = 30 value = [29, 1]
Age <= 44.5 impurity = 0.1528 samples = 12 value = [11, 1]
Fare <= 10.0 impurity = 0.3343 samples = 33 value = [26, 7]
SibSp <= 0.5 impurity = 0.4444 samples = 6 value = [4, 2]
impurity = 0.0 samples = 2 value = [2, 0]
impurity = 0.0 samples = 1 value = [0, 1]
Fare <= 7.8854 impurity = 0.4444 samples = 3 value = [2, 1]
impurity = 0.0 samples = 3 value = [3, 0]
Fare <= 7.8021 impurity = 0.4444 samples = 3 value = [2, 1]
Age <= 17.0 impurity = 0.48 samples = 5 value = [2, 3]
impurity = 0.0 samples = 4 value = [4, 0]
Fare <= 11.0 impurity = 0.1327 samples = 14 value = [13, 1]
impurity = 0.0 samples = 16 value = [16, 0]
impurity = 0.0 samples = 8 value = [8, 0]
Age <= 47.5 impurity = 0.375 samples = 4 value = [3, 1]
Age <= 31.5 impurity = 0.5 samples = 6 value = [3, 3]
Age <= 30.5 impurity = 0.2524 samples = 27 value = [23, 4]
Age <= 21.5 impurity = 0.375 samples = 4 value = [3, 1]
impurity = 0.5 samples = 2 value = [1, 1]
impurity = 0.0 samples = 2 value = [2, 0]
impurity = 0.0 samples = 1 value = [0, 1]
impurity = 0.0 samples = 1 value = [1, 0]
Fare <= 7.875 impurity = 0.5 samples = 2 value = [1, 1]
impurity = 0.5 samples = 2 value = [1, 1]
Age <= 18.5 impurity = 0.4444 samples = 3 value = [1, 2]
Age <= 21.0 impurity = 0.32 samples = 5 value = [4, 1]
impurity = 0.0 samples = 9 value = [9, 0]
impurity = 0.0 samples = 1 value = [0, 1]
impurity = 0.0 samples = 3 value = [3, 0]
Fare <= 9.4916 impurity = 0.375 samples = 4 value = [1, 3]
impurity = 0.0 samples = 2 value = [2, 0]
impurity = 0.0 samples = 9 value = [9, 0]
Fare <= 12.9375 impurity = 0.3457 samples = 18 value = [14, 4]
Fare <= 7.7854 impurity = 0.5 samples = 2 value = [1, 1]
impurity = 0.0 samples = 2 value = [2, 0]
impurity = 0.0 samples = 1 value = [0, 1]
impurity = 0.0 samples = 1 value = [1, 0]
impurity = 0.0 samples = 1 value = [0, 1]
impurity = 0.5 samples = 2 value = [1, 1]
Age <= 17.5 impurity = 0.4444 samples = 3 value = [2, 1]
impurity = 0.0 samples = 2 value = [2, 0]
Age <= 28.0 impurity = 0.5 samples = 2 value = [1, 1]
impurity = 0.0 samples = 2 value = [0, 2]
Age <= 59.5 impurity = 0.1975 samples = 9 value = [8, 1]
Age <= 45.0 impurity = 0.4444 samples = 9 value = [6, 3]
impurity = 0.0 samples = 1 value = [1, 0]
impurity = 0.0 samples = 1 value = [0, 1]
impurity = 0.0 samples = 1 value = [1, 0]
impurity = 0.5 samples = 2 value = [1, 1]
impurity = 0.0 samples = 1 value = [0, 1]
impurity = 0.0 samples = 1 value = [1, 0]
impurity = 0.0 samples = 6 value = [6, 0]
Age <= 64.0 impurity = 0.4444 samples = 3 value = [2, 1]
Age <= 32.5 impurity = 0.48 samples = 5 value = [2, 3]
impurity = 0.0 samples = 4 value = [4, 0]
impurity = 0.0 samples = 1 value = [0, 1]
impurity = 0.0 samples = 2 value = [2, 0]
impurity = 0.0 samples = 1 value = [0, 1]
Age <= 40.5 impurity = 0.5 samples = 4 value = [2, 2]
Age <= 36.5 impurity = 0.4444 samples = 3 value = [2, 1]
impurity = 0.0 samples = 1 value = [0, 1]
impurity = 0.5 samples = 2 value = [1, 1]
impurity = 0.0 samples = 1 value = [1, 0]

Decision Tree: 20% 기준 다양한 향상도 성과 기록

Decision Tree(Leafs)	20% 기준 향상도
4개의 그룹	2.5
10개의 그룹	3.0
39개의 그룹	3.0


Bagging Test Sample Accuracy on n_estimators
Test Accuracy%
0.84
0.82
0.80
0.78
0.76
0
5
10
15
20
25
30
35
40
n_estimators
Bagging w/ 500 leaf stump
Bagging w/ 2000 leaf stump
Bagging w/ 8000 leaf stump
Bagging w/ full tree
Baseline classification tree


CJ E&M
20th Mnet
보·이·는·대·로·믿·지·마·라!
대/반/전/ 음악추리쇼
너의 목소리가 보여 2
Mnet tvN 공동 방송
10월 22일 목요일 | 밤 9시 40분

6명 중에는 음치도 있고 노래를 잘하는 실력자(정상)도 있다.

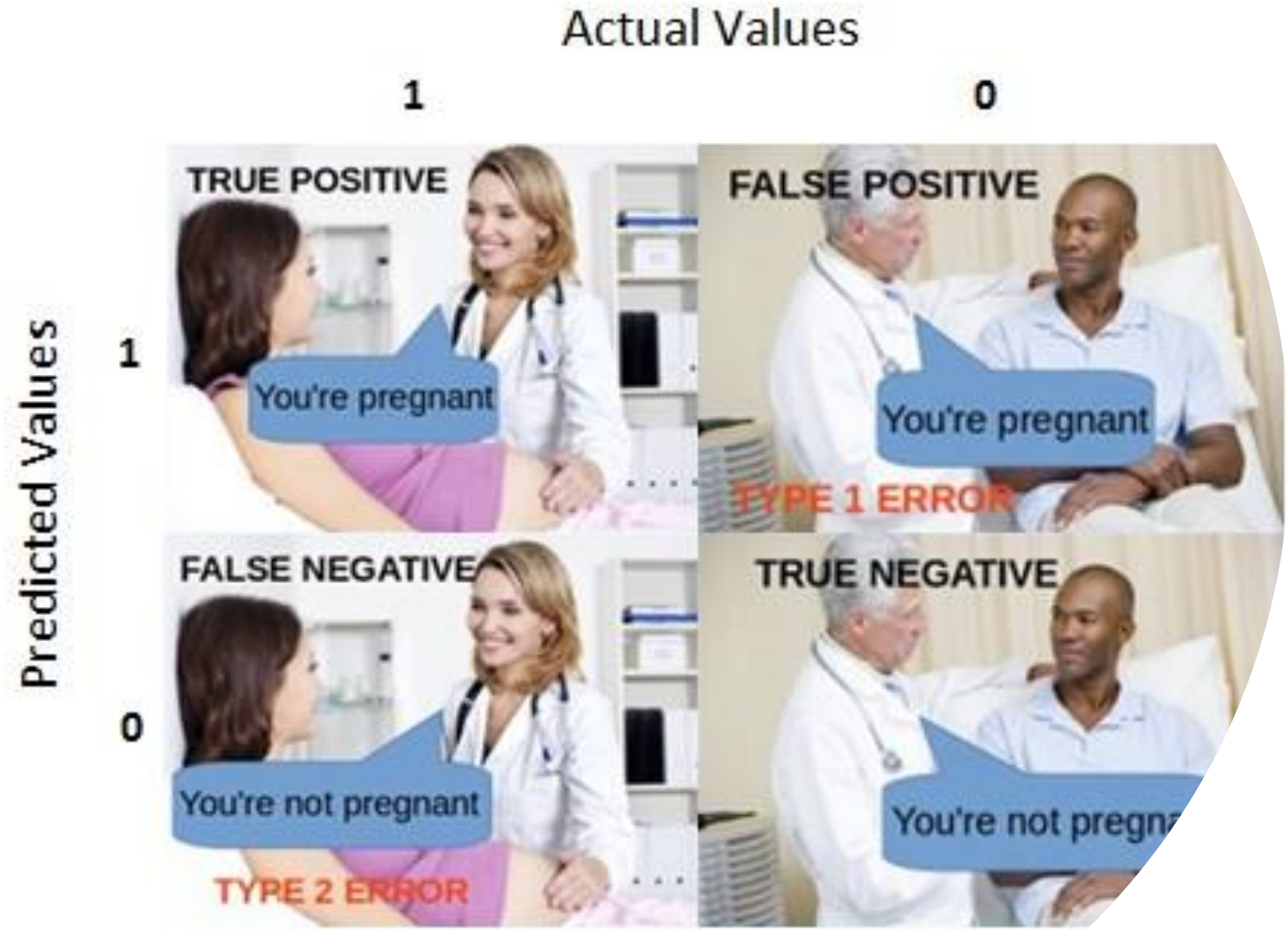
번호 [1, 2, 3, 4, 5, 6]

정답 : [**음치,음치,음치,음치,정상,정상**] 가 있다.

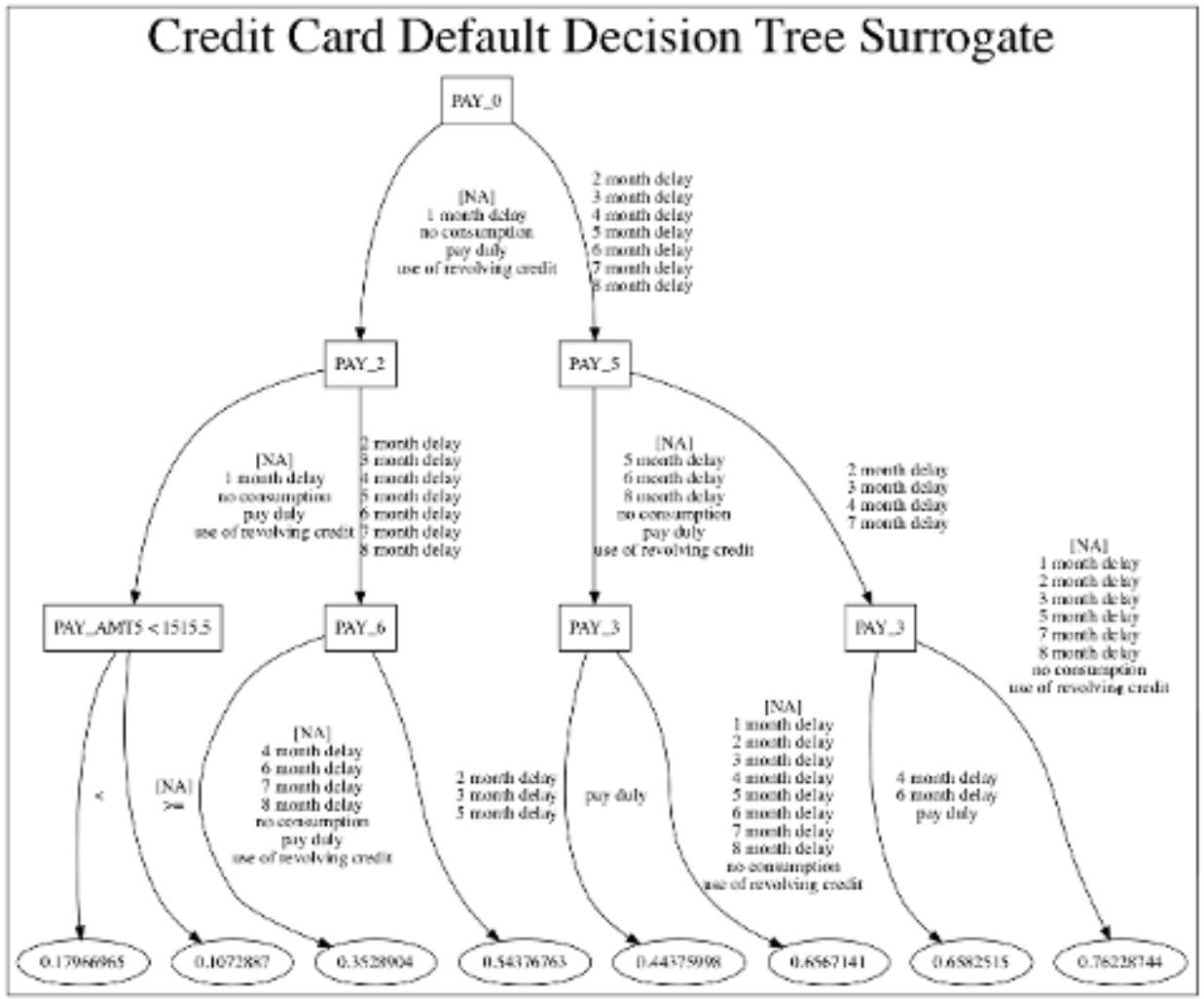
누가 음치인지 겉모습만 보고 맞춰야 한다.

감으로 예측을 한다.

예측 : [**음치, 음치, 정상, 정상, 정상, 정상**]

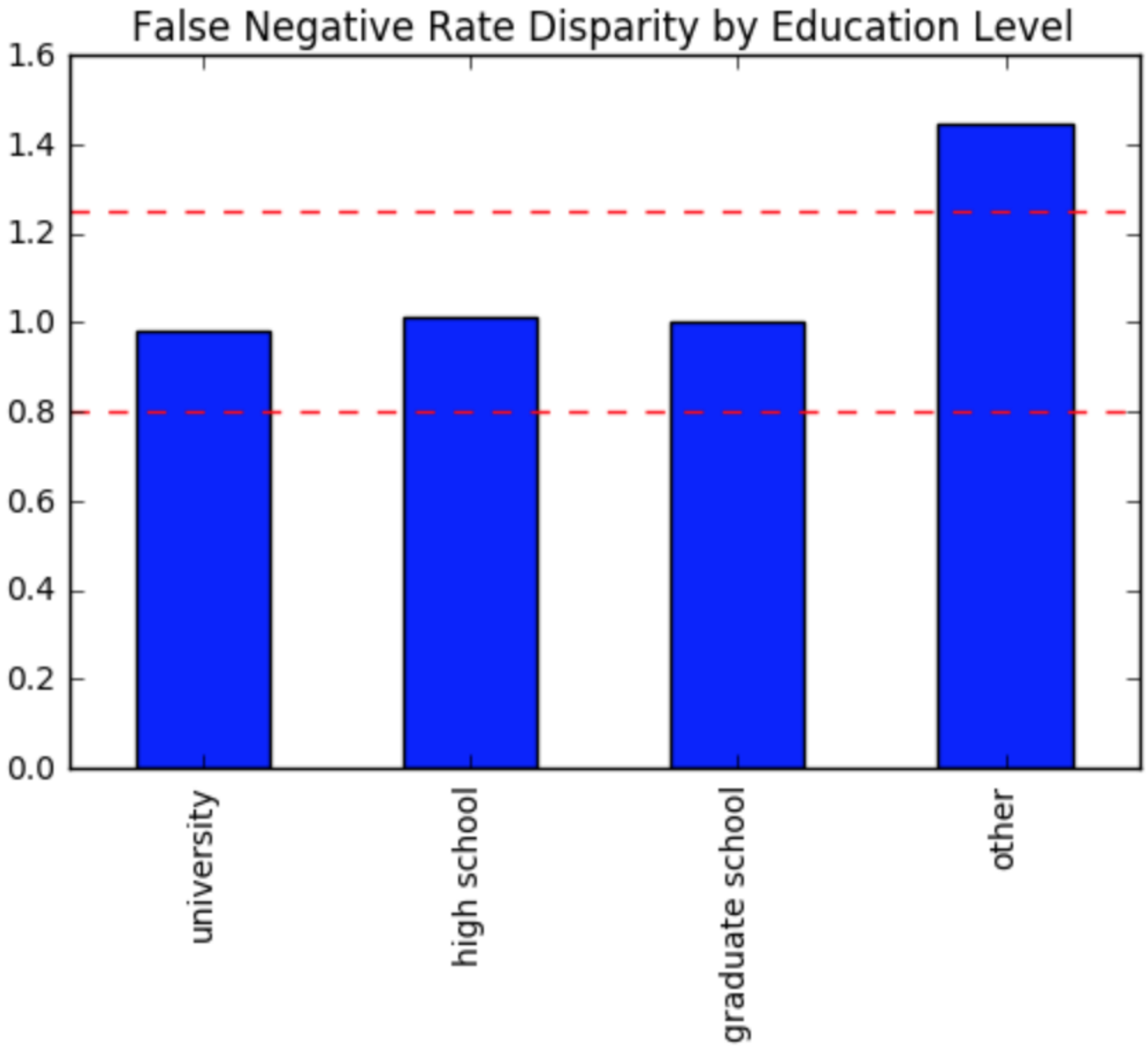

Actual Values
1
0
Predicted Values
1
0
TRUE POSITIVE
You're pregnant
FALSE POSITIVE
You're pregnant
TYPE 1 ERROR
FALSE NEGATIVE
You're not pregnant
TYPE 2 ERROR
TRUE NEGATIVE
You're not pregna

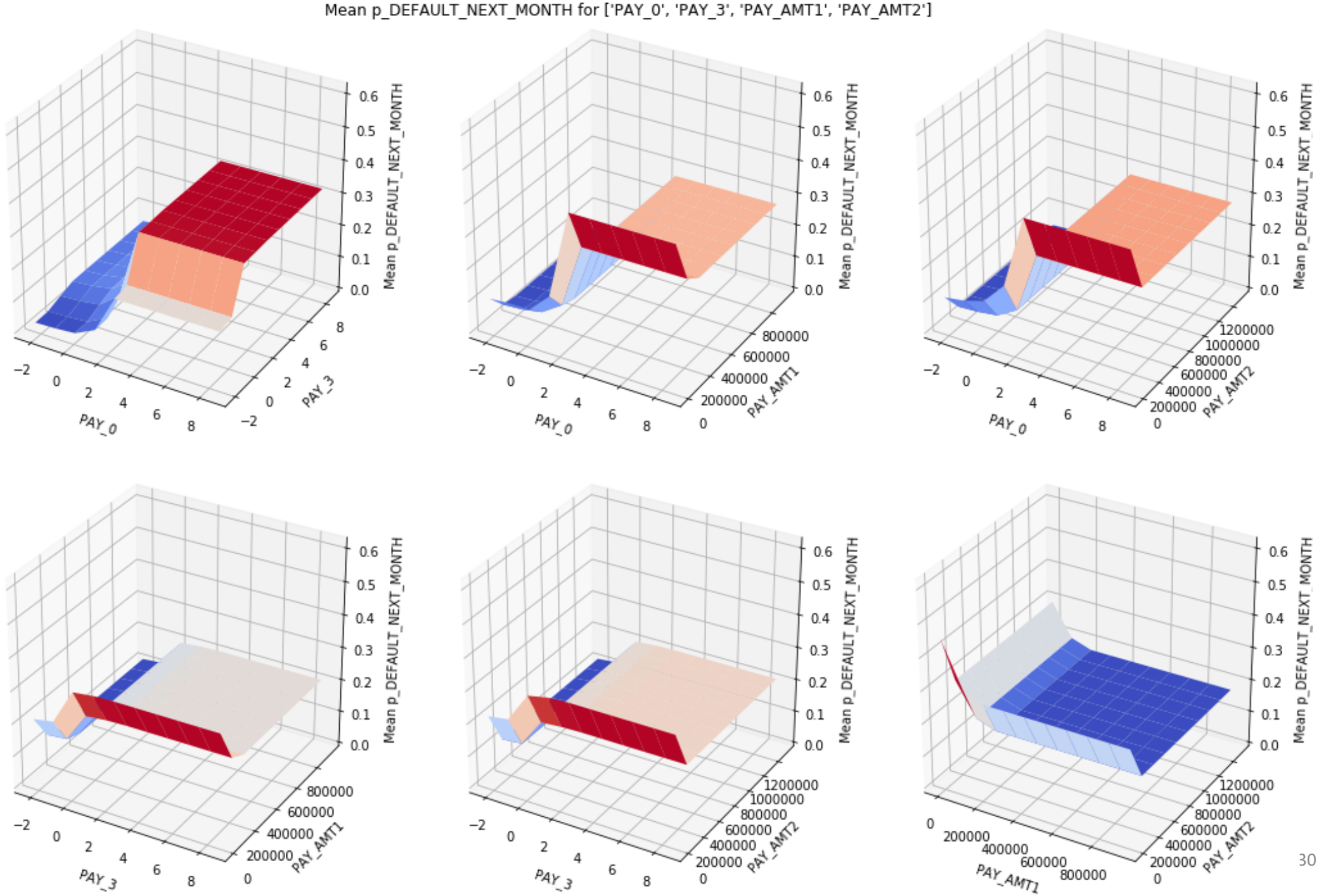
머신러닝 결과 분석(Decision Tree)



Error Metrics for PAY_0

PAY_0	Prevalence	Accuracy	True Positive Rate	Precision	Specificity	Negative Predicted Value	False Positive Rate	False Discovery Rate	False Negative Rate	False Omissions Rate
-2	0.049	0.857	0.3	0.119	0.885	0.961	0.115	0.881	0.7	0.039
-1	0.117	0.805	0.383	0.267	0.861	0.913	0.139	0.733	0.617	0.087
0	0.05	0.864	0.345	0.143	0.891	0.963	0.109	0.857	0.655	0.037
1	0.822	0.457	0.368	0.93	0.871	0.229	0.129	0.07	0.632	0.771
2	1	0.709	0.709	1	0.5	0	0.5	0	0.291	1
3	1	0.748	0.748	1	0.5	0	0.5	0	0.252	1
4	1	0.571	0.571	1	0.5	0	0.5	0	0.429	1
5	1	0.444	0.444	1	0.5	0	0.5	0	0.556	1
6	1	0.25	0.25	1	0.5	0	0.5	0	0.75	1
7	1	0.5	0.5	1	0.5	0	0.5	0	0.5	1
8	1	0.75	0.75	1	0.5	0	0.5	0	0.25	1


False Negative Rate Disparity by Education Level
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
university
high school
graduate school
other


Mean p_DEFAULT_NEXT_MONTH for ['PAY_0', 'PAY_3', 'PAY_AMT1', 'PAY_AMT2']
Mean p_DEFAULT_NEXT_MONTH
PAY_0
PAY_3
PAY_AMT1
PAY_AMT2

강사 소개

정 준 수 / Ph.D (heinem@naver.com)

- 前) 삼성전자 연구원
- 前) 삼성의료원 (삼성생명과학연구소)
- 前) 삼성SDS (정보기술연구소)
- 現) (사)한국인공지능협회, AI, 머신러닝 강의
- 現) 한국소프트웨어산업협회, AI, 머신러닝 강의
- 現) 서울디지털재단, AI 자문위원
- 現) 한성대학교 교수(겸)
- 전문분야: 시각 모델링, 머신러닝(ML), RPA
- https://github.com/JSJeong-me/