

Predictive Analytics

2020. 11. 09

정 준 수 Ph.D.

과정 목표

★ 머신러닝 활용을 위한 모델구축 및 요소기술에 대한 기본지식 습득

- 분석 대상 데이터(텍스트, 영상 및 소리) 특성이해 및 분석 방법
- 머신러닝 학습모델 구축과 다양한 알고리즘 이해
- 머신러닝의 활용 및 향후과제



https://www.youtube.com/watch?v=0_gx3159OFI

자동차 자동변속기는 운전자의 손과 발을 편하게 만들어 엄마와 누나에게 운전의 길을 열어 주었다. 소프트웨어 로봇과 프로세스 오토메이션(RPA)은 모두에게 편리함을 제공하는 인공지능(AI) 세상으로 안내하는 길잡이다.

- 저자: 정 준 수 -

Agenda

What is predictive analytics

- How does it deliver value for marketing
 - Response modeling for acquisition
 - Churn modeling for retention
- How does it work
- Case studies

Predictive Analytics:



(data)

*Technology that learns from **experience** to
predict the future behavior of individuals
... in order to drive better decisions.*







Yesterday

male, CA, 10 purchases...



Today

response: YES



We reached out to the company for more current information, and while director of national media relations Ragan Dickens couldn't confirm Pop-Tarts are flying off shelves this season, he said ready-to-eat foods, as well as bottled water, batteries, fuel containers, and bread, are among the most frequently purchased items pre-hurricane.

[Market Research Tips](#)

How Target Used Data Analytics to Predict Pregnancies

December 6, 2016

|

[George Kuhn](#)

As crazy as that title may seem, the case study is absolutely true. Using big data and predictive analytics can tell a frighteningly accurate story. This Target story may be something you've seen or read over the years online or in [Charles Duhigg's book on The Power of Habit](#) but it's worth recapping. It was also featured in the [New York Times](#) and passed along to me by a colleague years ago. Either way, the story is pretty fascinating, so I am going to recap it here for you. For some readers new to the data analytics or market research field, this example displays the power of big data and how it is used to profile customers and predict purchase behavior.



리조트패키지



Dear, Mom

세상 하나뿐인 유일한 내 편

여자보다 엄마의 삶을 택한 그녀를 위해 준비한
세상 하나뿐인 특별한 여행!

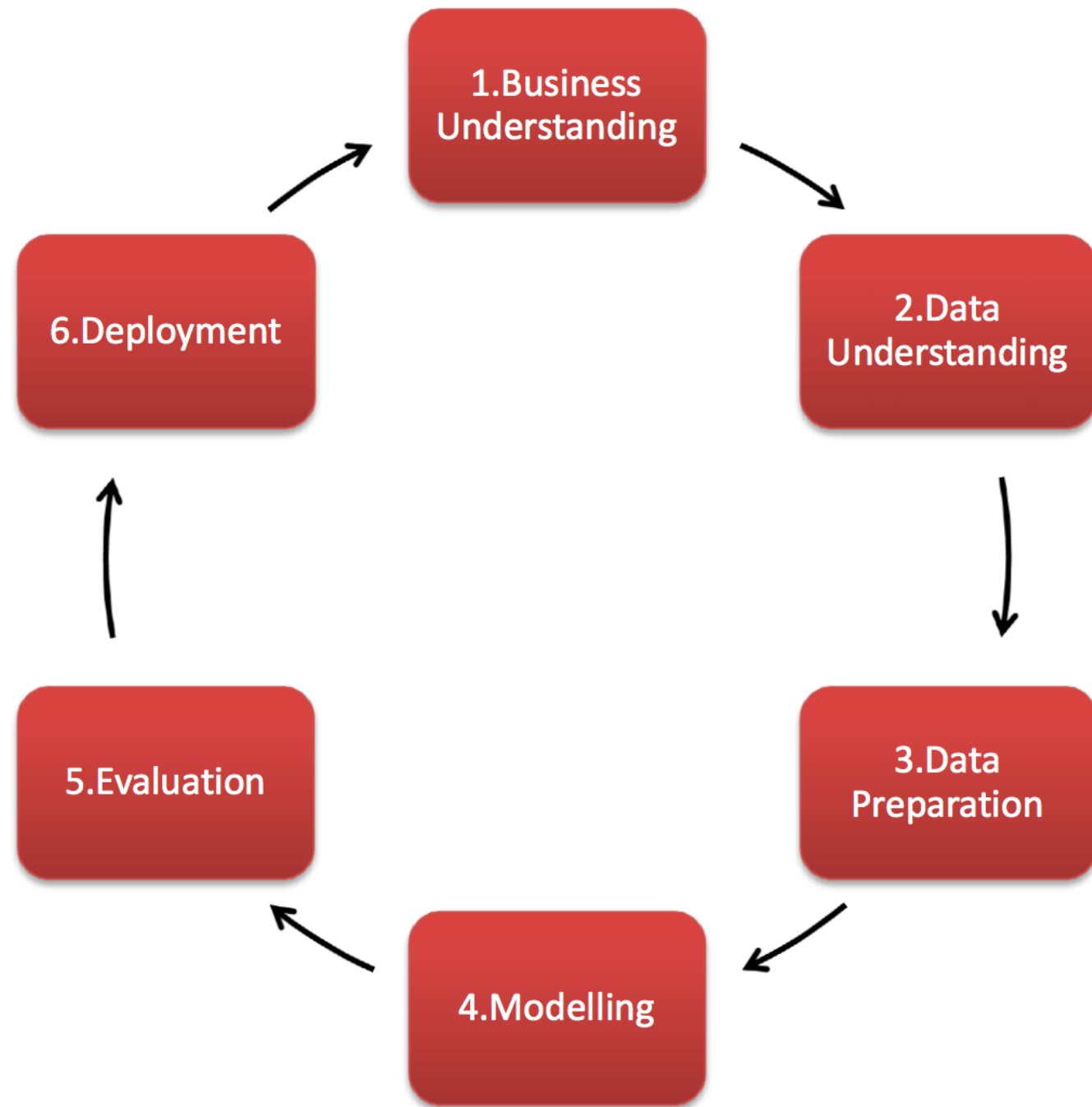
Predictive Analytics

1 무엇을 예측하는가?

2 무엇을 할 것인가?

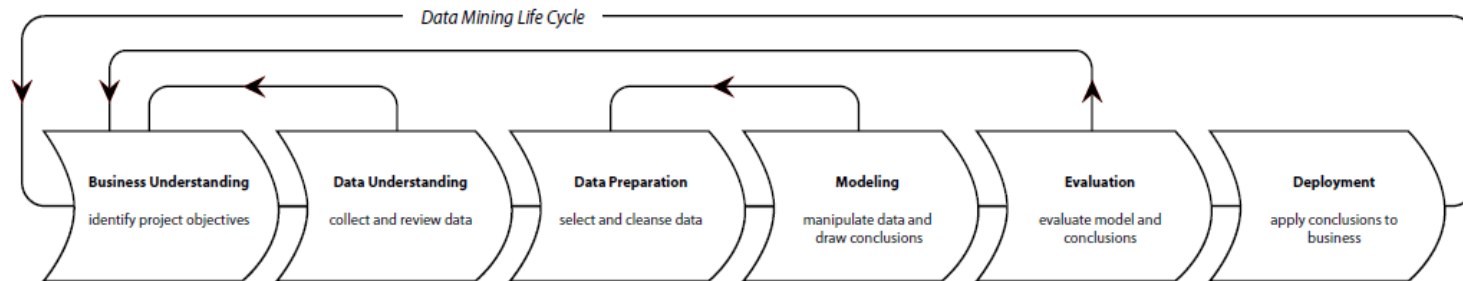
Machine Learning에서 “예측”이란?

이전에 본적 없는 새로운 데이터에 대한 정확한 출력 예측



CRISP-DM methodology

Phases



Determine Business Objectives

Background
Business Objectives
Business Success Criteria
(Log and Report Process)

Assess Situation

Inventory of Resources,
Requirements, Assumptions,
and Constraints
Risks and Contingencies
Terminology
Costs and Benefits
(Log and Report Process)

Determine Data Mining Goals

Data Mining Goals
Data Mining Success Criteria
(Log and Report Process)

Produce Project Plan

Project Plan
Initial Assessment of Tools and
Techniques
(Log and Report Process)

Collect Initial Data

Initial Data Collection Report
(Log and Report Process)

Describe Data

Data Description Report
(Log and Report Process)

Explore Data

Data Exploration Report
(Log and Report Process)

Verify Data Quality

Data Quality Report
(Log and Report Process)

Data Set

Data Set Description
(Log and Report Process)

Select Data

Rationale for Inclusion/
Exclusion
(Log and Report Process)

Clean Data

Data Cleaning Report
(Log and Report Process)

Construct Data

Derived Attributes
Generated Records
(Log and Report Process)

Integrate Data

Merged Data
(Log and Report Process)

Format Data

Reformatted Data
(Log and Report Process)

Select Modeling Technique

Modeling Technique
Modeling Assumptions
(Log and Report Process)

Generate Test Design

Test Design
(Log and Report Process)

Build Model Parameter Settings

Models
Model Description
(Log and Report Process)

Assess Model

Model Assessment
Revised Parameter
(Log and Report Process)

Evaluate Results

Align Assessment of Data
Mining Results with
Business Success Criteria
(Log and Report Process)

Approved Models

Review Process
Review of Process
(Log and Report Process)

Determine Next Steps

List of Possible Actions
Decision
(Log and Report Process)

Plan Deployment

Deployment Plan
(Log and Report Process)

Plan Monitoring and Maintenance

Monitoring and
Maintenance Plan
(Log and Report Process)

Produce Final Report

Final Report
Final Presentation
(Log and Report Process)

Review Project

Experience
Documentation
(Log and Report Process)

Generic Tasks

Specialized Tasks

(Process Instances)

a visual guide to CRISP-DM methodology

SOURCE CRISP-DM 1.0

<http://www.crisp-dm.org/download.htm>

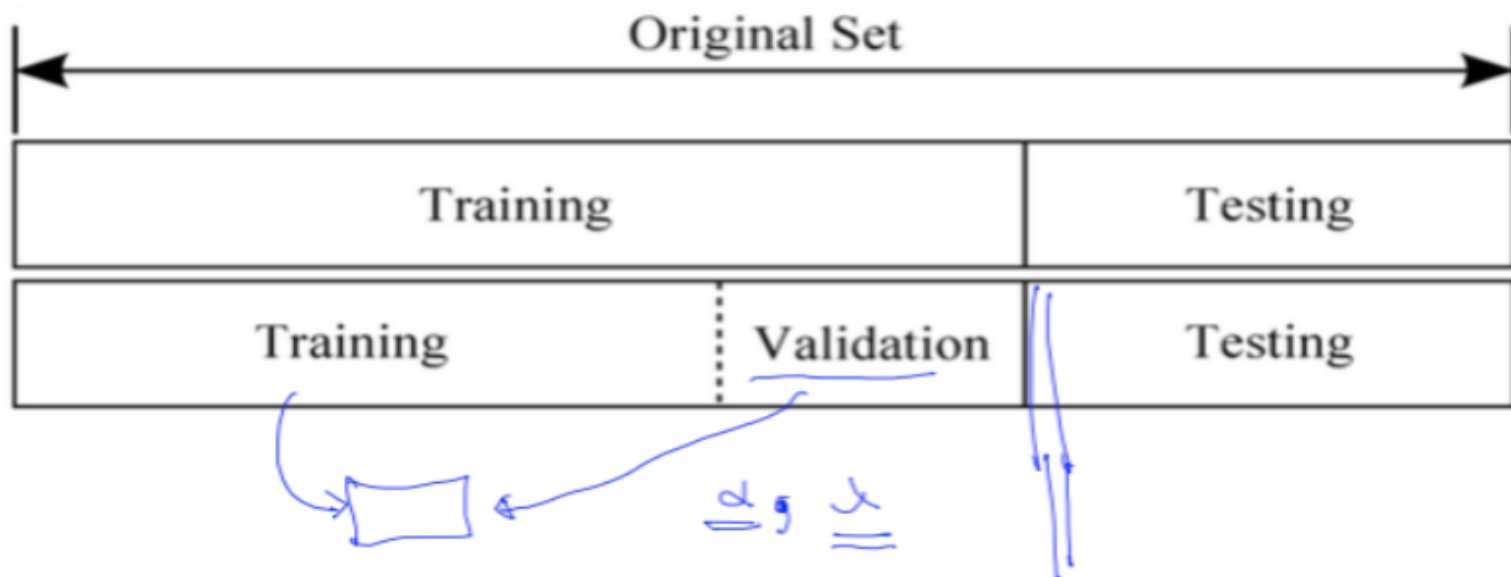
DESIGN Nicole Leaper

<http://www.nicoleleaper.com>



학습 데이터의 구성

Training, validation and test sets



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.

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>



<https://movie.naver.com/movie/bi/mi/mediaView.nhn?code=136842&mid=29267#tab>

Predictive Analytics:



(data)

*Technology that learns from **experience** to
predict the future behavior of individuals
... in order to drive better decisions.*



Predictive Analytics

1 무엇을 예측하는가?

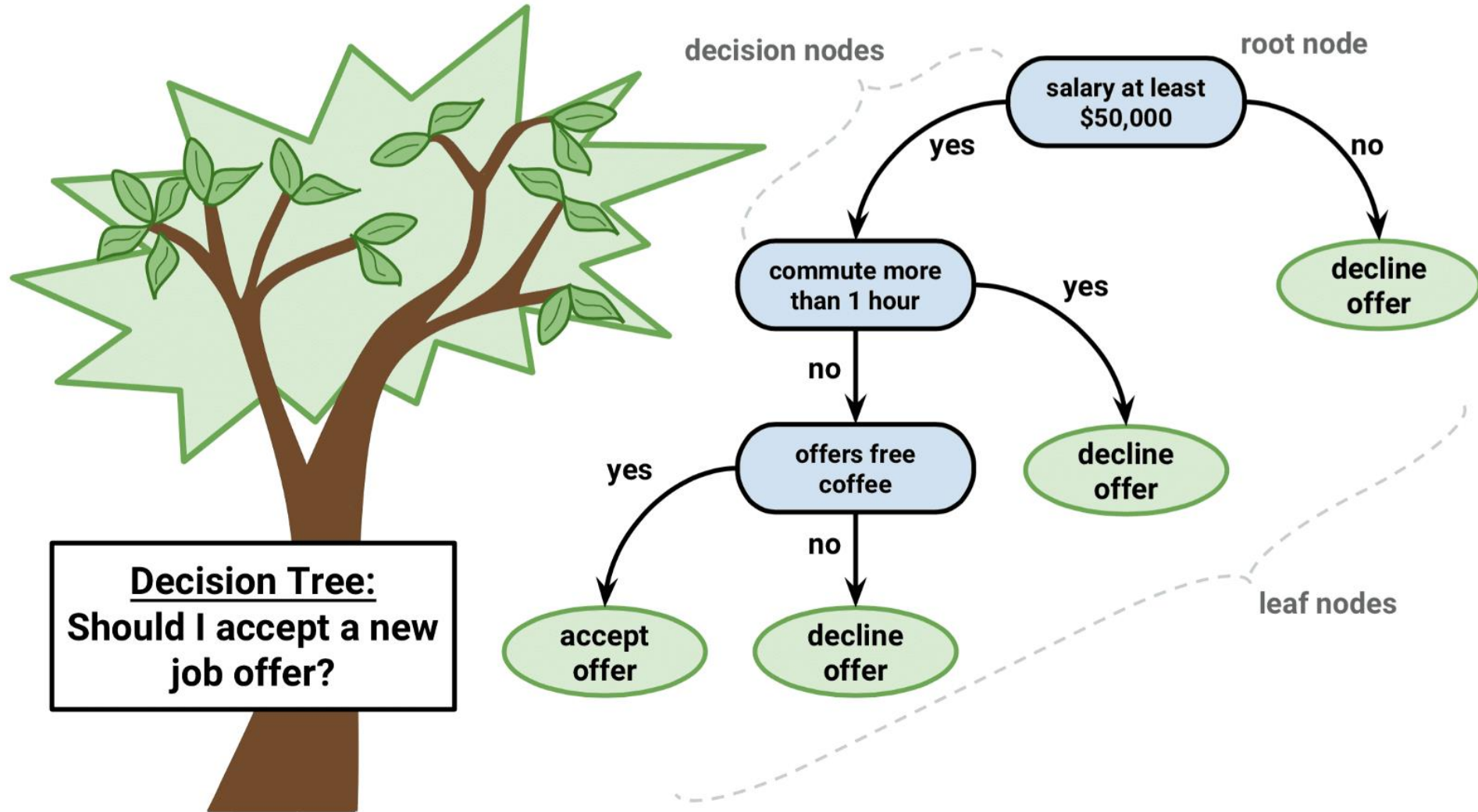
2 무엇을 할 것인가?

Machine Learning에서 “예측”이란?

이전에 본적 없는 새로운 데이터에 대한 정확한 출력 예측



Decision Tree



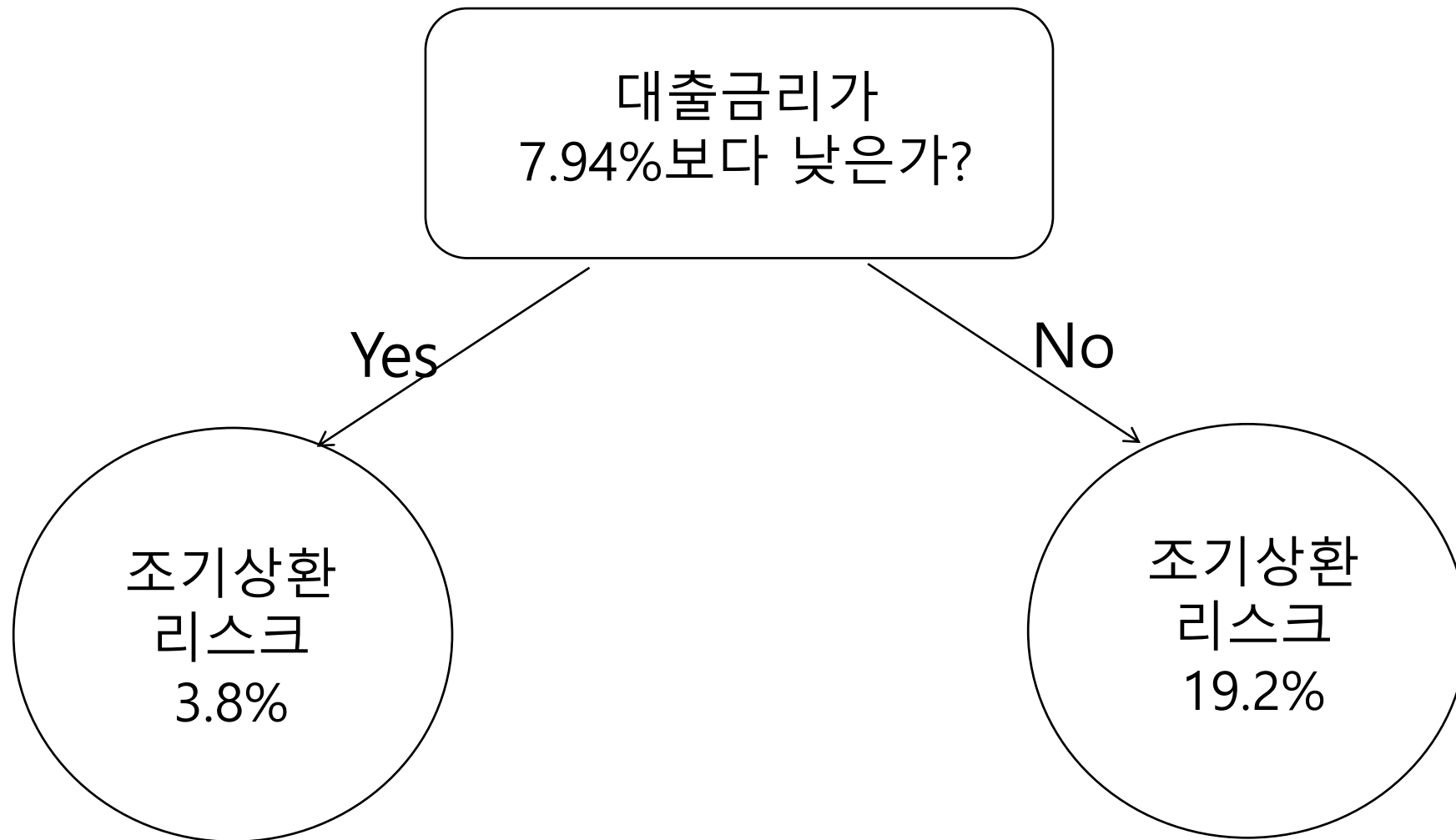
예측 분석 응용: 이탈 모델링으로 고객 이탈 방지하기

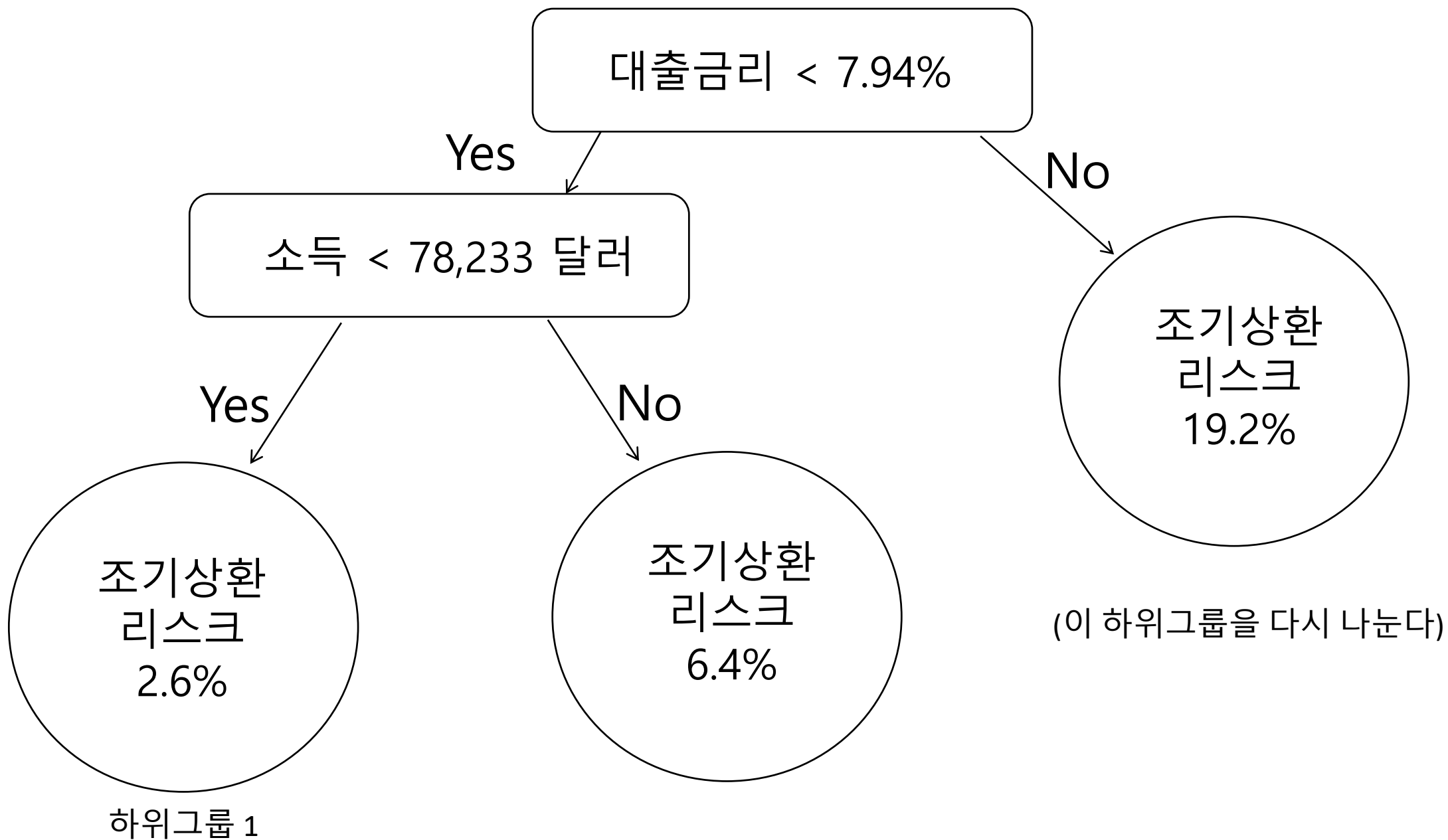
1 무엇을 예측하는가?

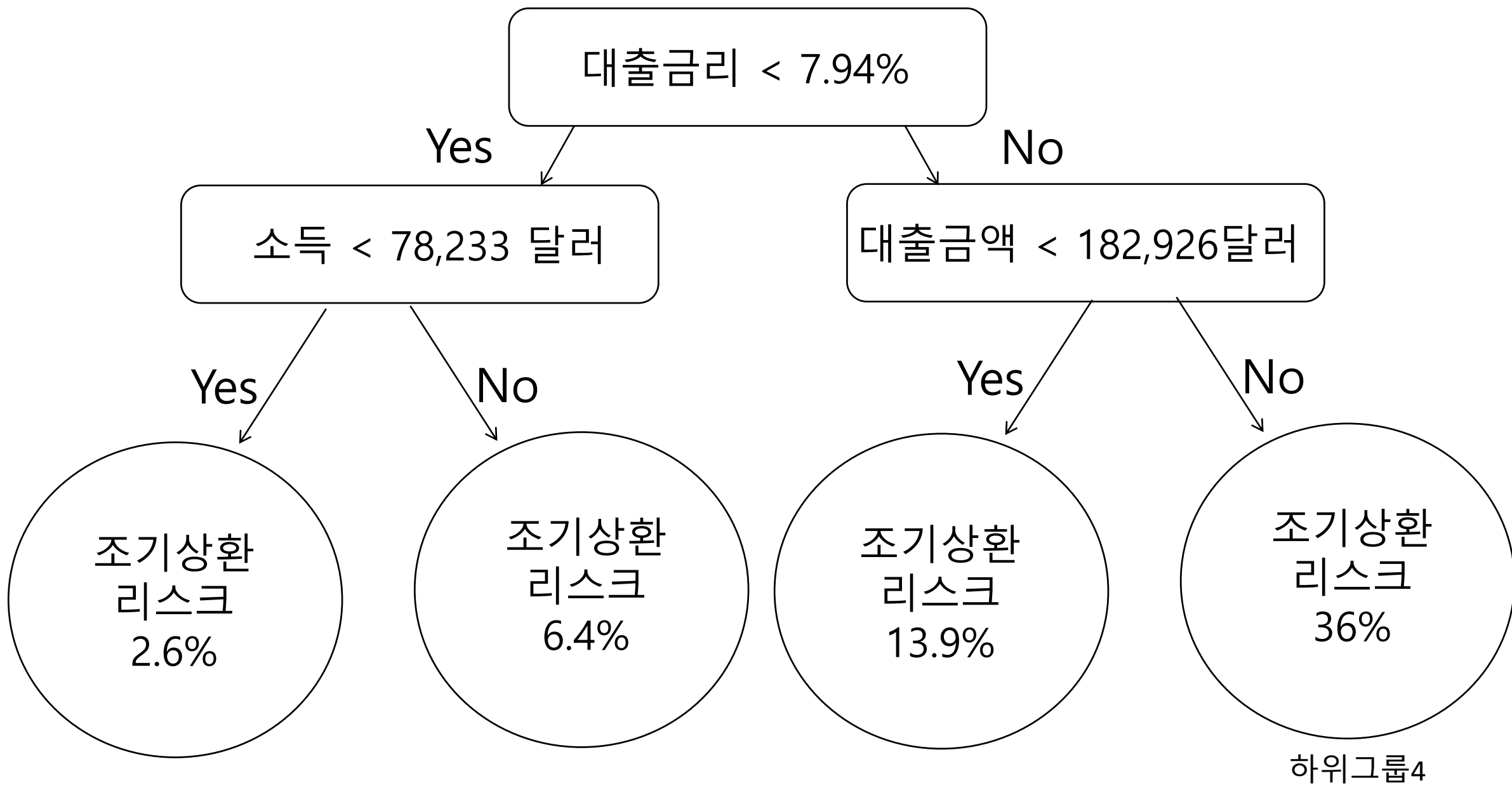
어느 고객이 떠나갈 것인가.

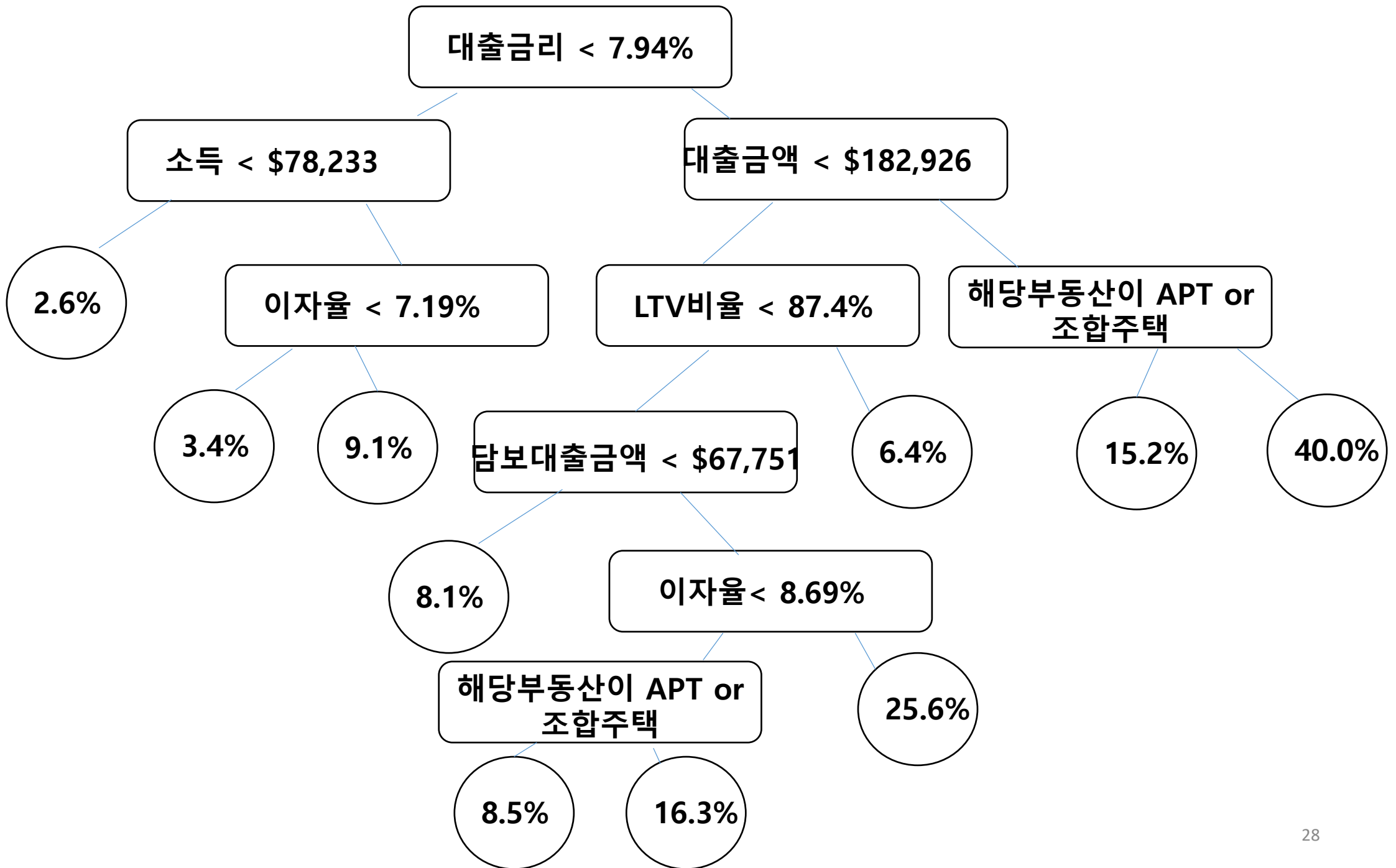
2 무엇을 할 것인가?

떠날 위기에 있는 고객들을 타겟으로 한 고객 유지 마케팅을 수행한다.









만약(IF):

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

그리고(AND):

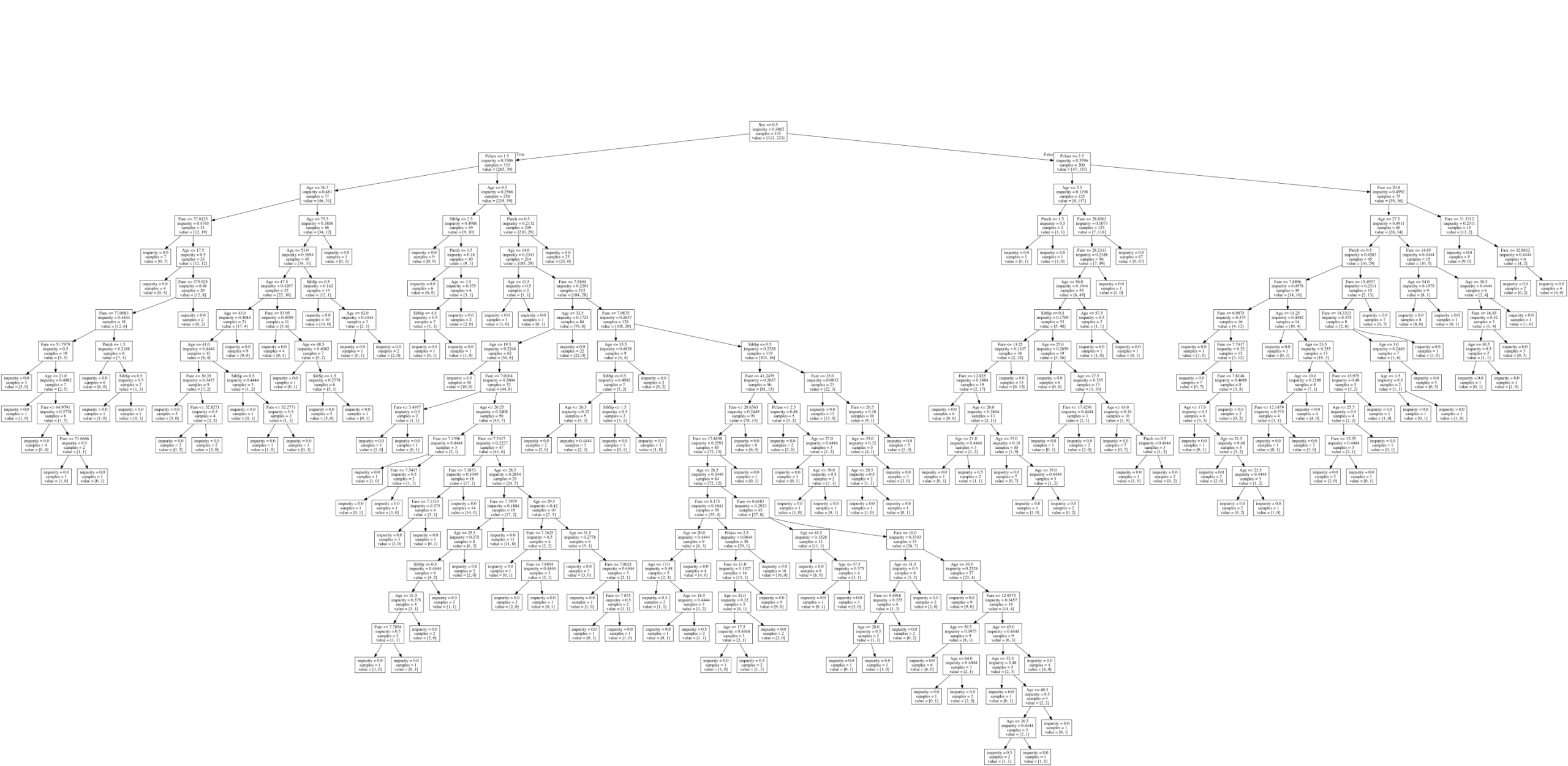
이자율이 8.69%와 같거나 그보다 더 높다.

그리고(AND):

부동산 자산가치 대비 대출금액의 비율이 87.4% 보다 작다.

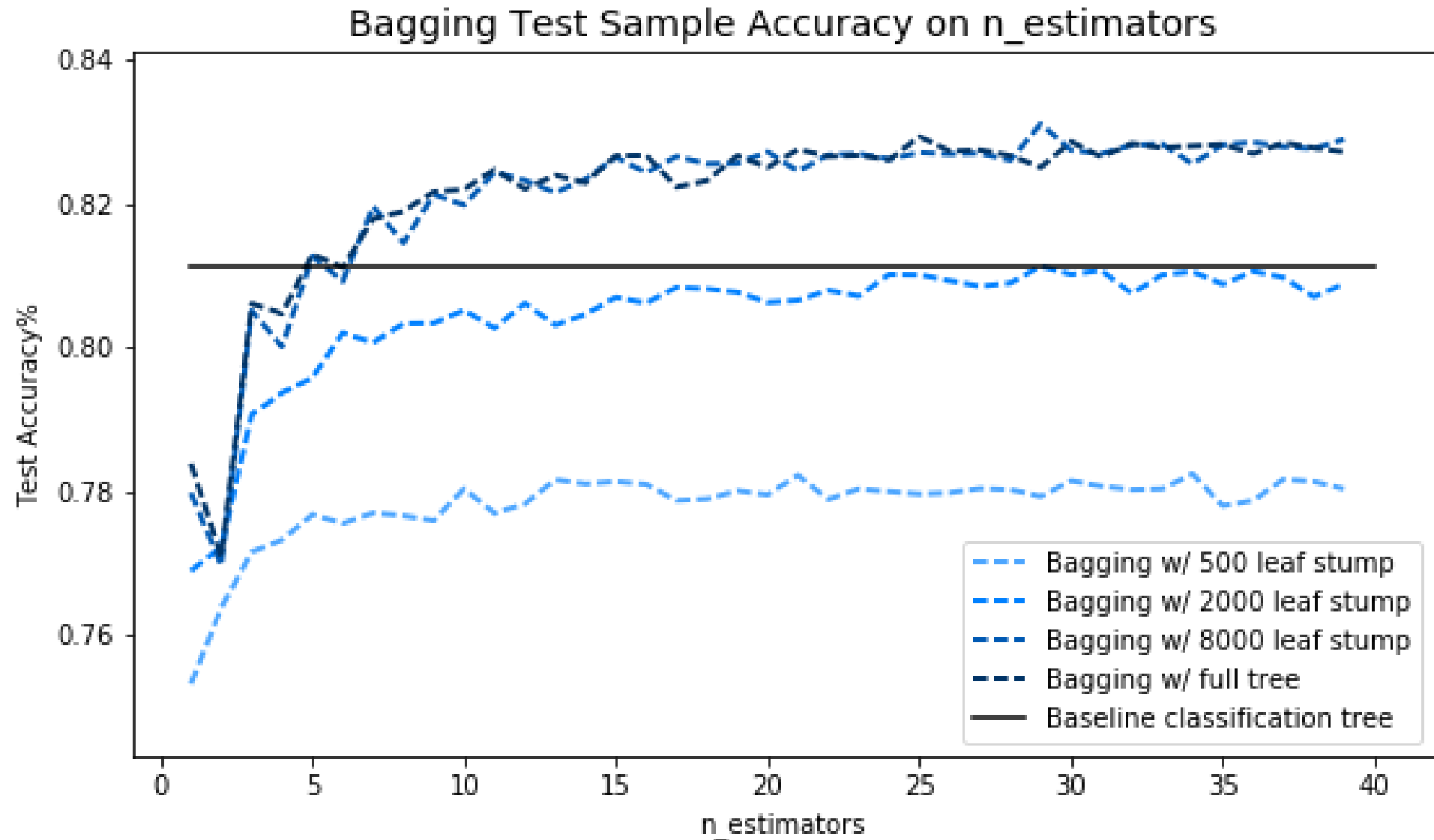
그러면(THEN):

조기상환 확률은 25.6% 이다.



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



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.

**어렵듯하게라도 볼 수 있는 것이 완전히
깜깜한 암흑 상태보다는 훨씬 우수한 결과를 낳는다!**

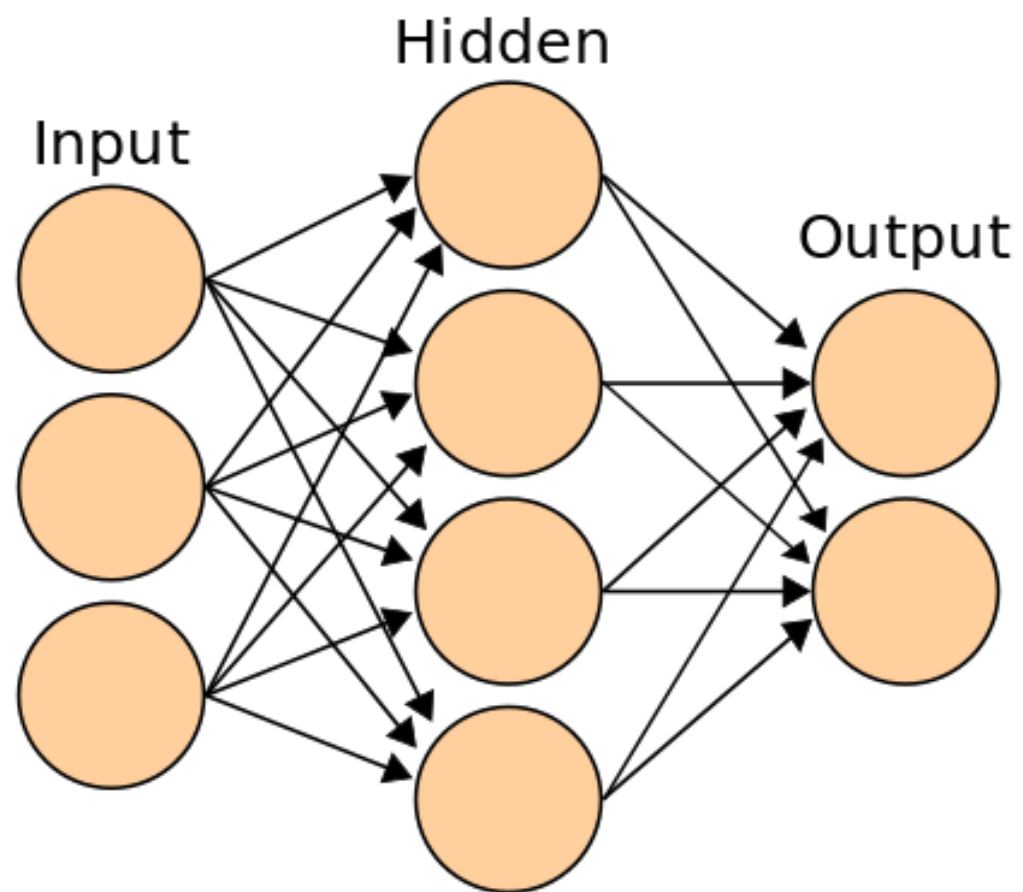
머신러닝이란?

인간이 다양한 경험과 시행착오를 통해 지식을 배우는 것처럼, [컴퓨터](#)에게 충분히 많은 데이터를 주고, 거기에서 일반적인 패턴을 찾아내게 하는 방법을 말합니다. (머신러닝의 대표적인 [알고리즘](#)이 딥러닝입니다.)

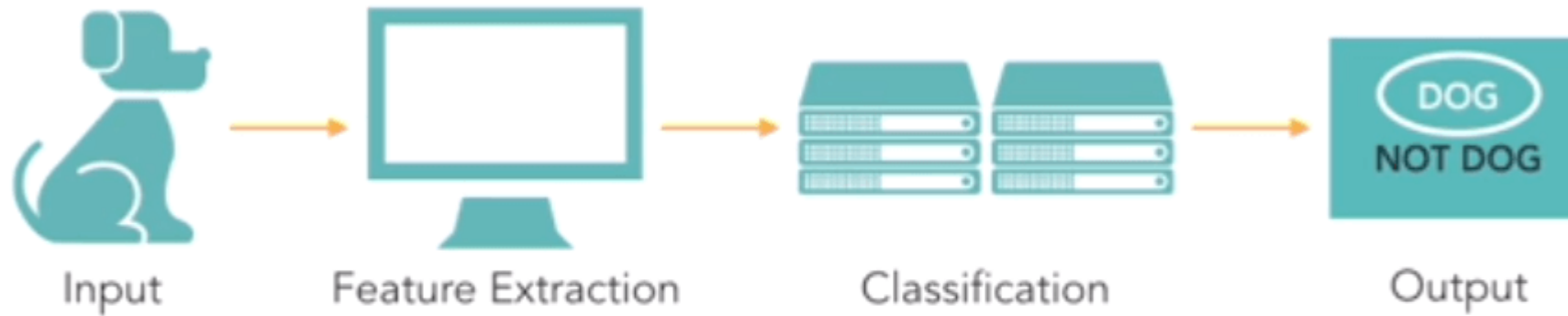
딥러닝이란?

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.



TRADITIONAL MACHINE LEARNING



DEEP LEARNING



보·이·는·대·로·**밀**·지·마·라!

대/반/전/ 음악추리쇼



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

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 pregnant

0

강사 소개

정 준 수 / Ph.D. (heinem@naver.com)

- 前) 삼성전자 연구원
- 前) 삼성의료원 (삼성생명과학연구소)
- 前) 삼성SDS (정보기술연구소)
- 現) (사)한국인공지능협회, AI, 머신러닝 전문강의, AI 교재출판 전문위원
- 現) 한국소프트웨어산업협회, AI, 머신러닝 SW전문강의
- 現) 한성대학교 교수(겸)
- 전문분야: 시각 모델링, 머신러닝(ML), RPA
- <https://github.com/jsjeong-me>

