

Machine Learning

2022. 3. 11

정 준 수 Ph.D

Beginnings

Thresholded Logic Unit

1943

Perceptron

1957

Adaline

1960

XOR Problem

1969

1st Neural Winter

Multilayer Backprop

1982

CNNs

1986

LSTMs

1989

1997

2nd Neural Winter

SVMs

1995

Deep Nets

2006

Alex Net

2012

GPU Era

1940

1950

1960

1970

1980

1990

2000

2010



S. McCulloch - W. Pitts



R. Rosenblatt



B. Widrow - M. Hoff



M. Minsky - S. Papert



P. Werbos

D. Rumelhart - G. Hinton - R. Williams

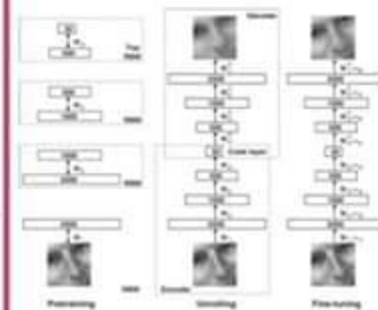
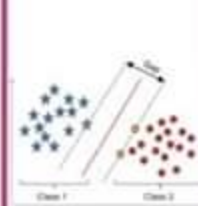
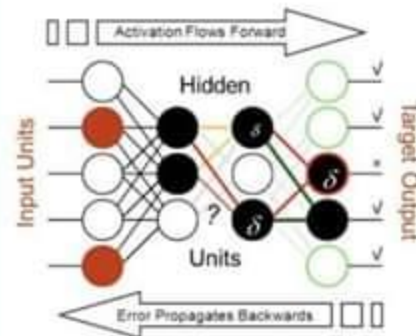
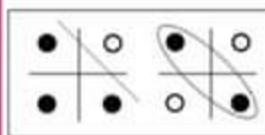
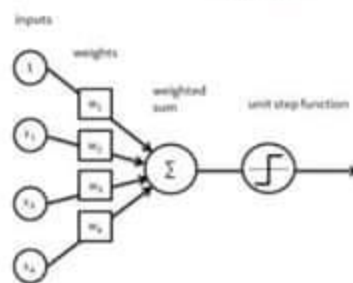
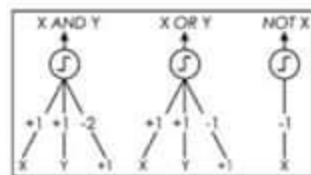
Y. Lecun - J. Schmidhuber



C. Cortes - V. Vapnik



R. Salakhutdinov - J. Hinton - A. Krizhevsky - I. Sutskever



Machine Learning에서 “예측”이란?

이전에 본적 없는 새로운 데이터에 대한 정확한 출력 예측

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.

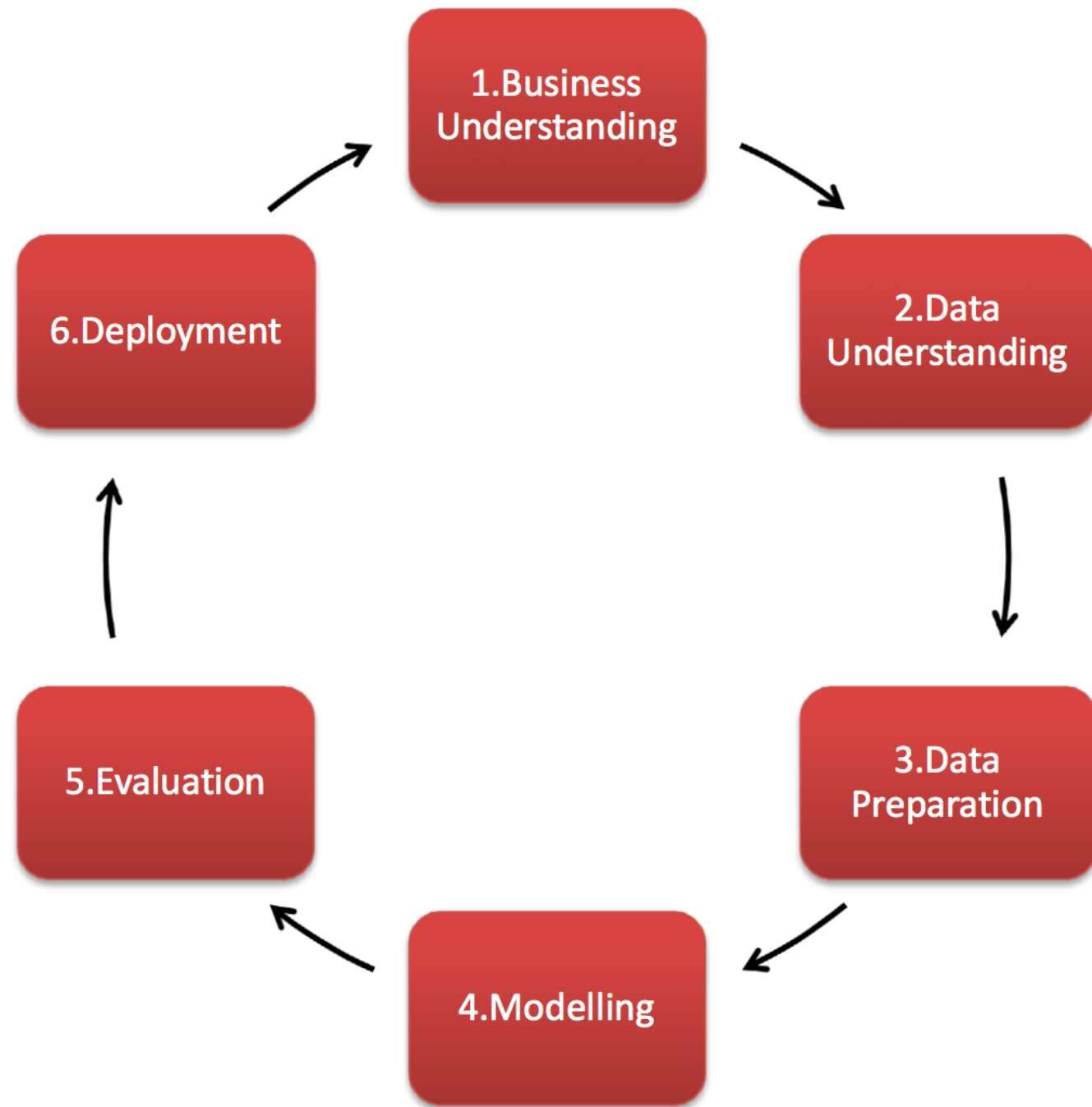
Predictive Analytics

1 무엇을 예측하는가?

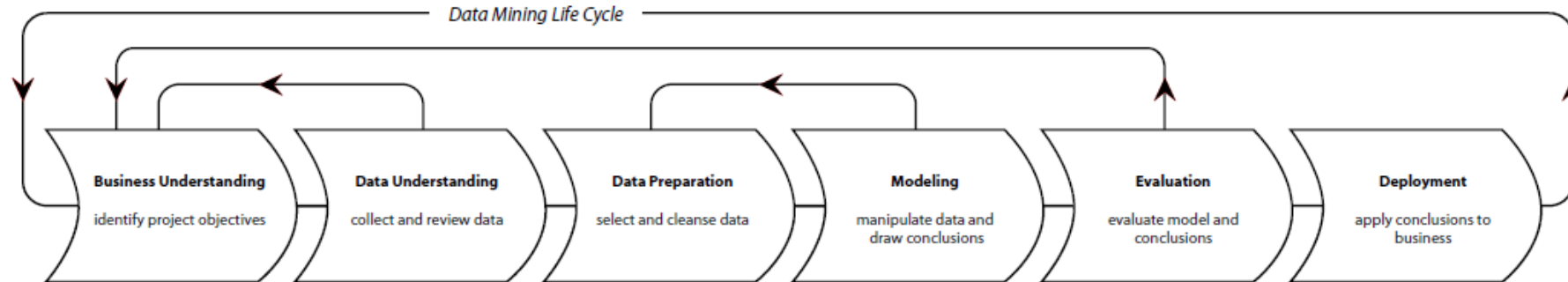
2 무엇을 할 것인가?

Modeling





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

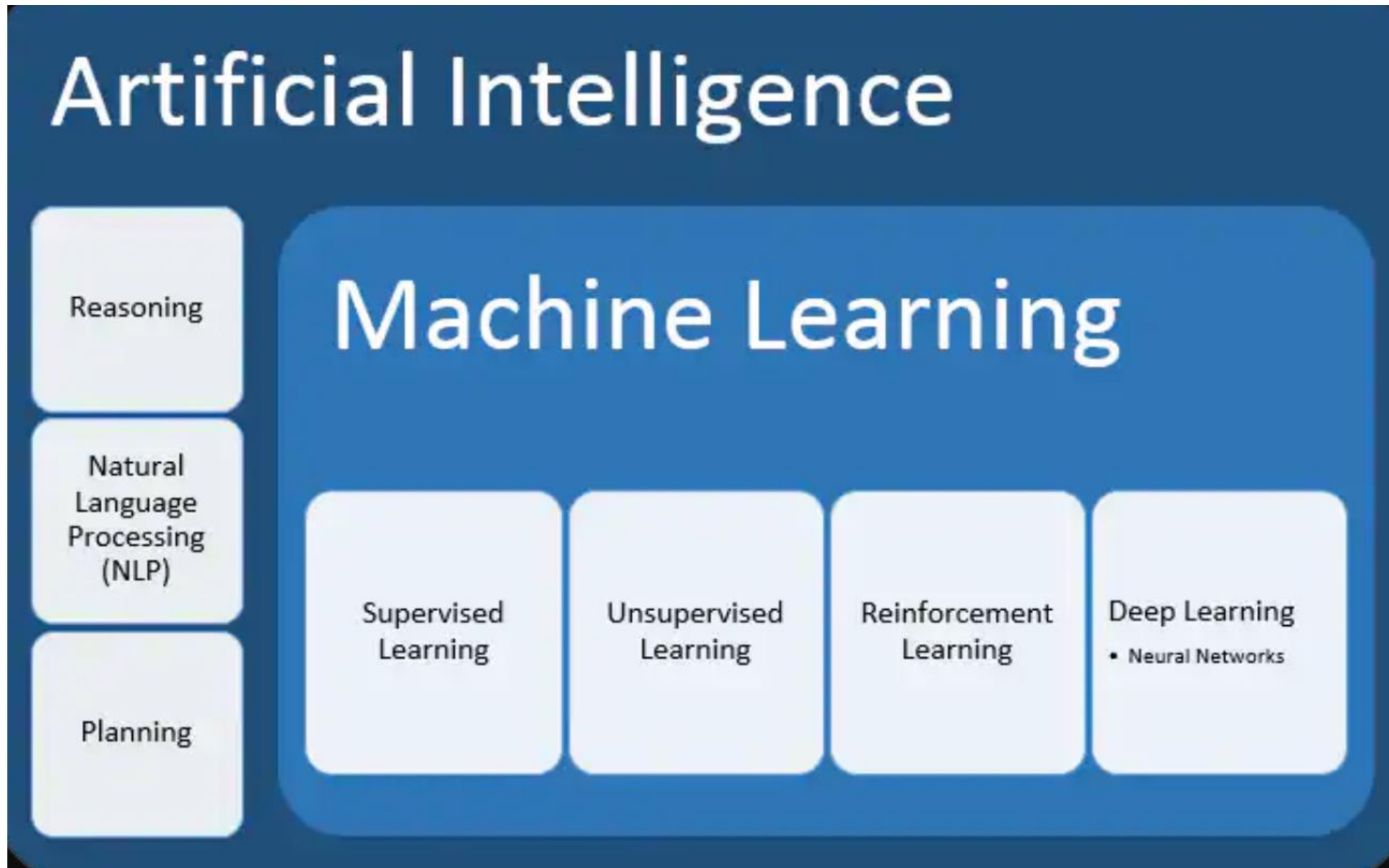
<http://www.crisp-dm.org/download.htm>

DESIGN Nicole Leaper

<http://www.nicoleleaper.com>

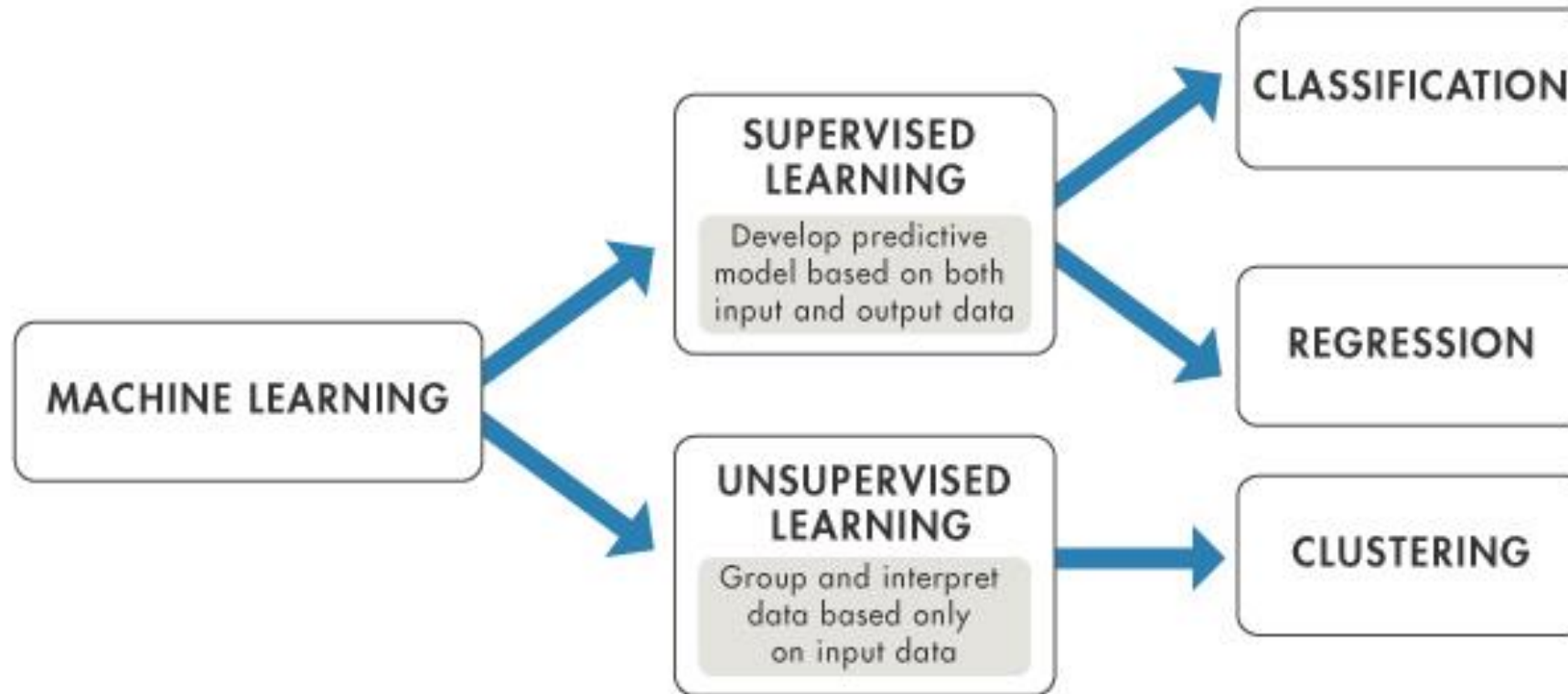


Machine Learning



Machine learning uses two types of techniques:

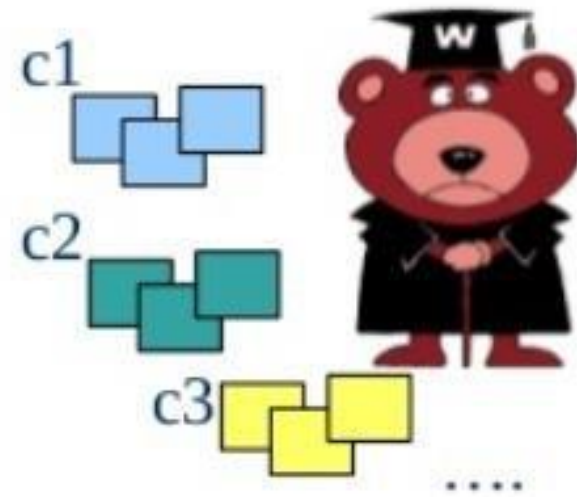
- Supervised learning, which trains a model on known input and output data so that it can predict future outputs,
- Unsupervised learning, which finds hidden patterns or intrinsic structures in input data.



Supervised Vs. Unsupervised

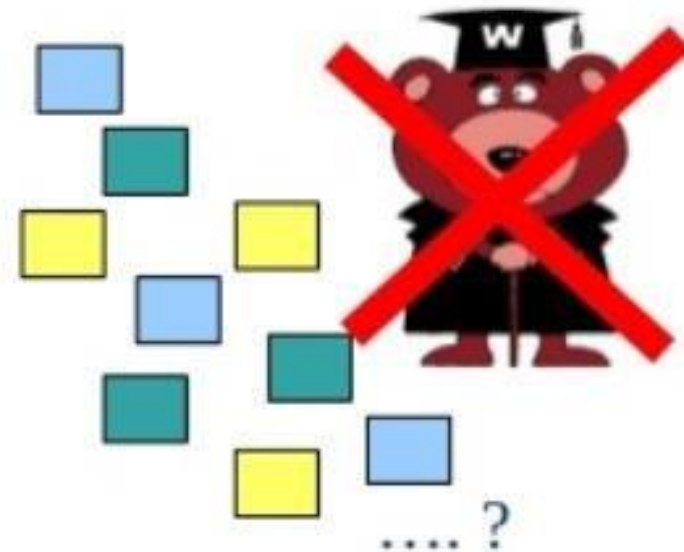
- **Supervised**

- **knowledge of output** - learning with the presence of an “expert” / teacher
 - data is **labelled** with a class or value
 - **Goal:** predict class or value label
 - e.g. Neural Network, Support Vector Machines, Decision Trees, Bayesian Classifiers

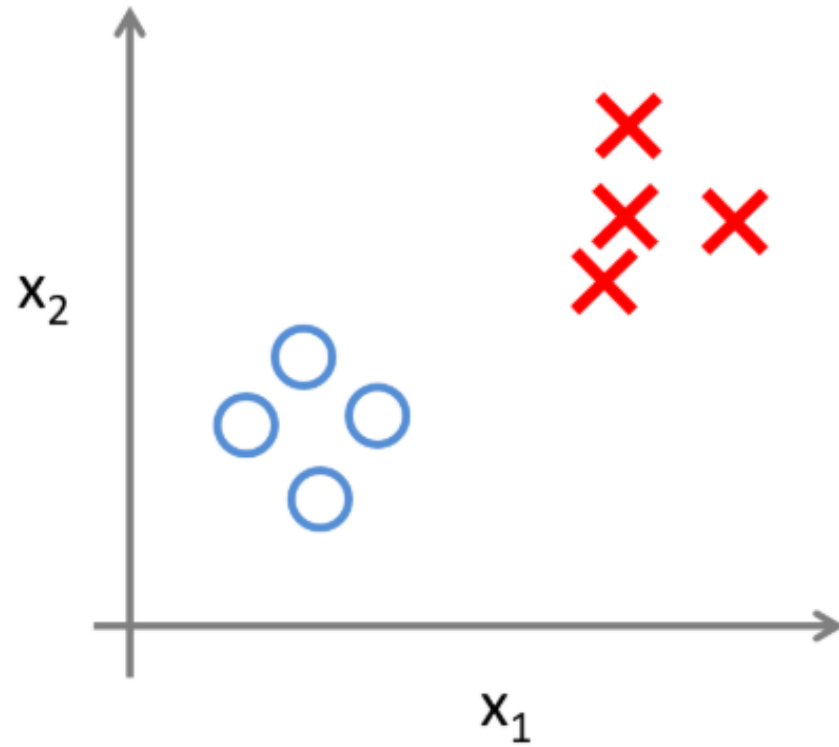


- **Unsupervised**

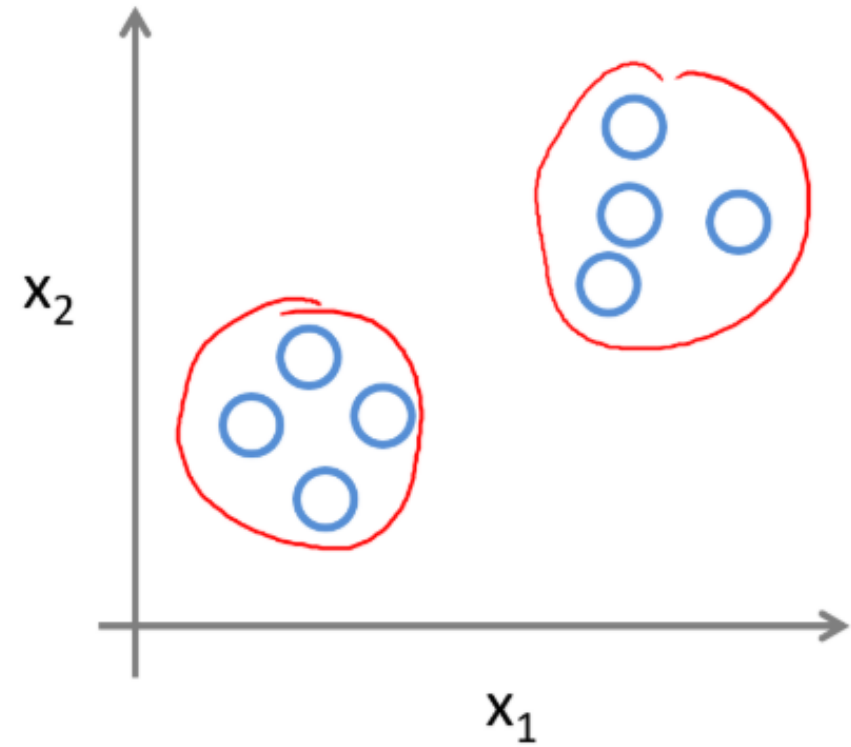
- **no knowledge of output** class or value
 - data is **unlabelled** or value un-known
 - **Goal:** determine data patterns/groupings
- Self-guided learning algorithm
 - (internal self-evaluation against some criteria)
 - e.g. k-means, genetic algorithms, clustering approaches ...



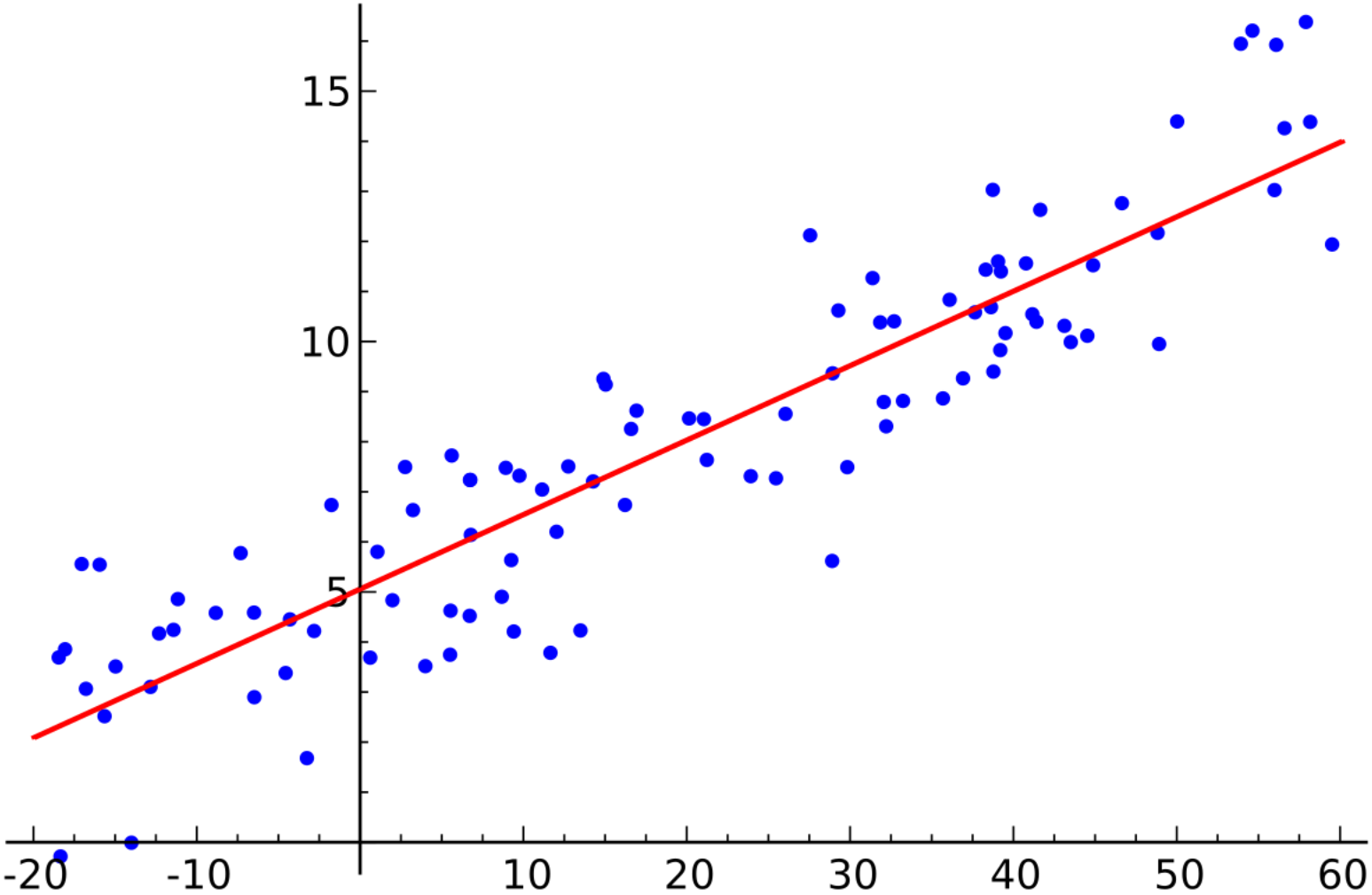
Supervised Learning



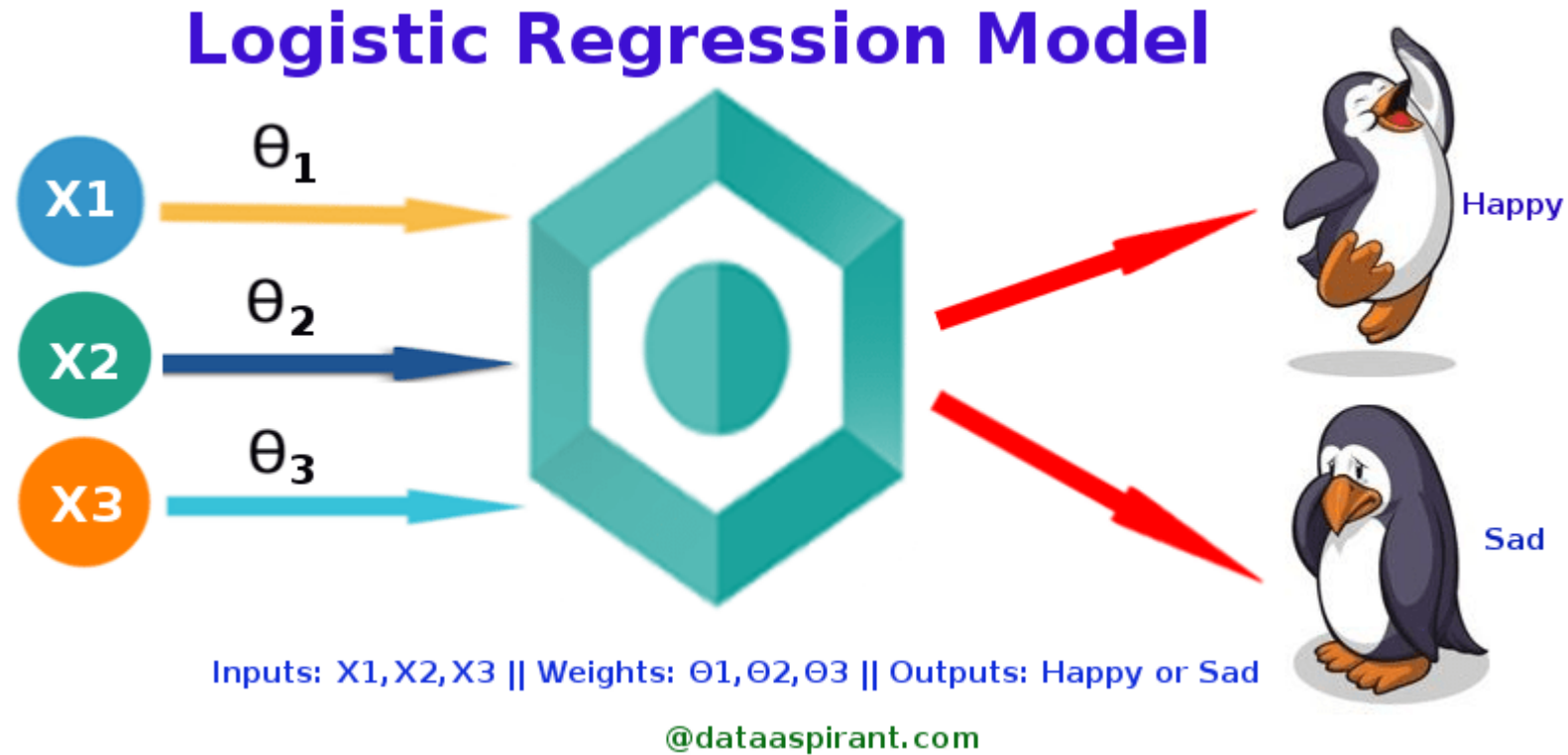
Unsupervised Learning



Linear Regression

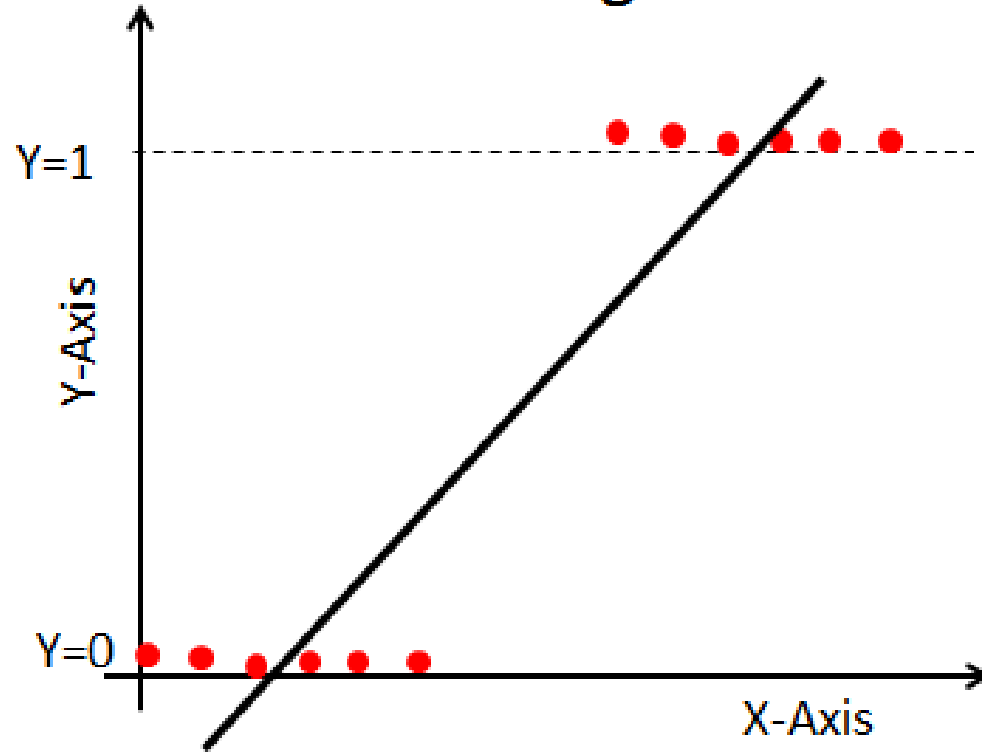


Logistic Regression

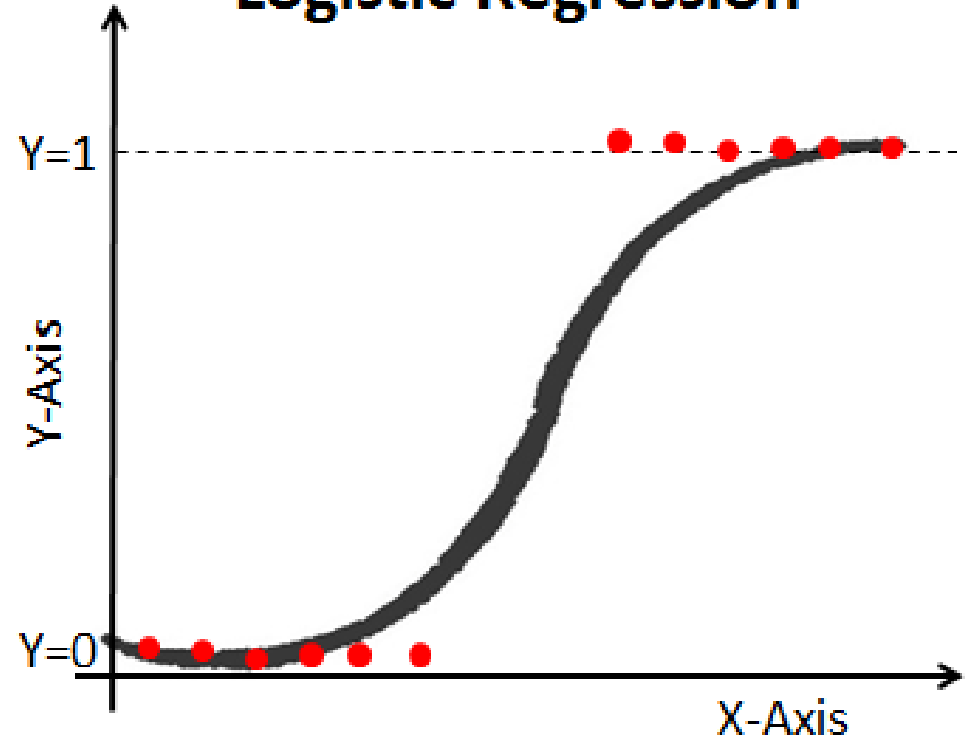


For example,
To predict whether an email is spam (1) or (0)
Whether the tumor is malignant (1) or not (0)

Linear Regression



Logistic Regression



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)))))
```

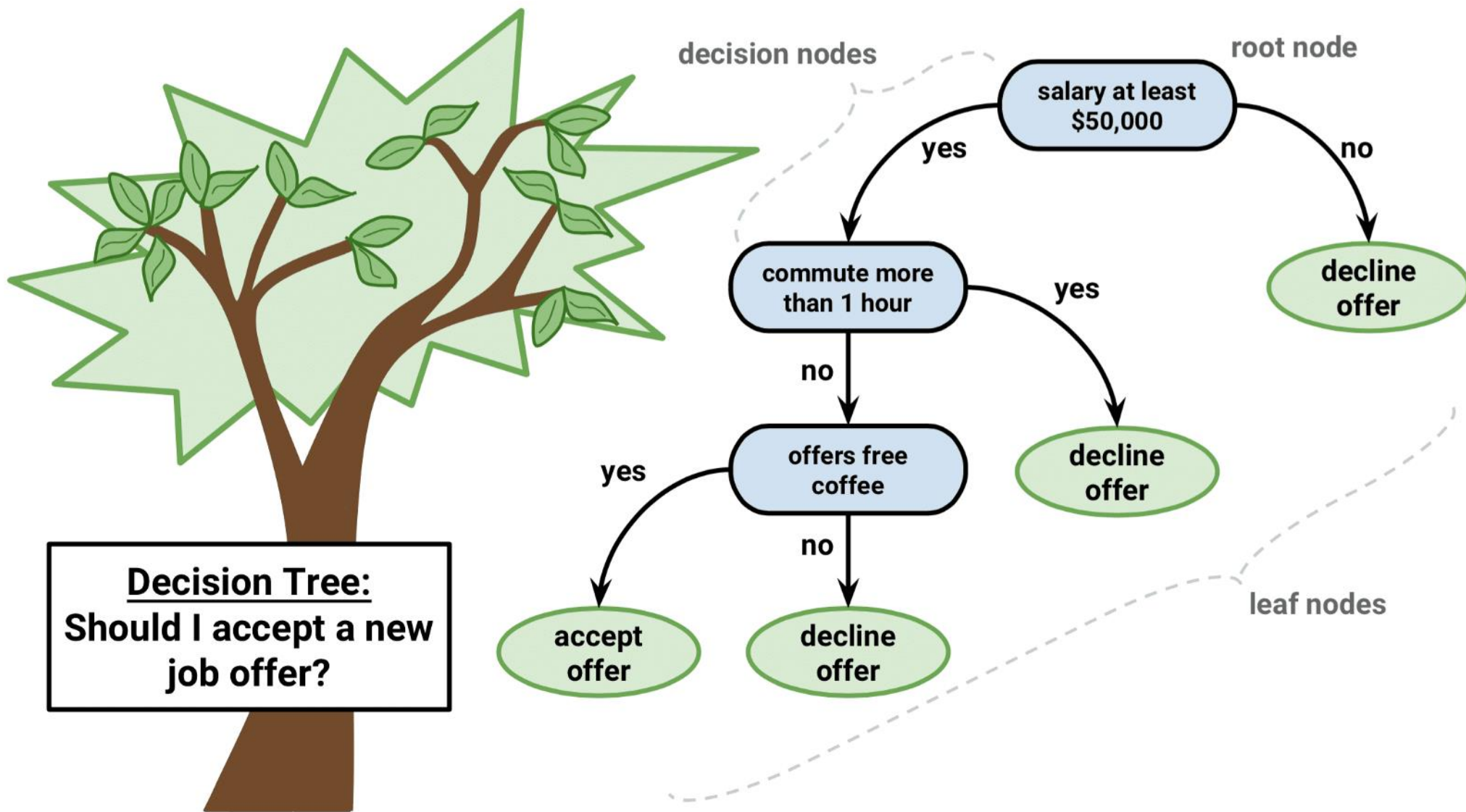
<https://medium.com/technology-invention-and-more/how-to-build-a-simple-neural-network-in-9-lines-of-python-code-cc8f23647ca1>

■ Confusion Matrix

		True condition	
Total population		Condition positive	Condition negative
Predicted Condition	Predicted condition positive	True positive	False positive
	Predicted condition negative	False negative	True negative

	Predicted Non-Default	Predicted Default
Actual Non-Default	True Negative	False Positive
Actual Default	False Negative	True Positive

Decision Tree



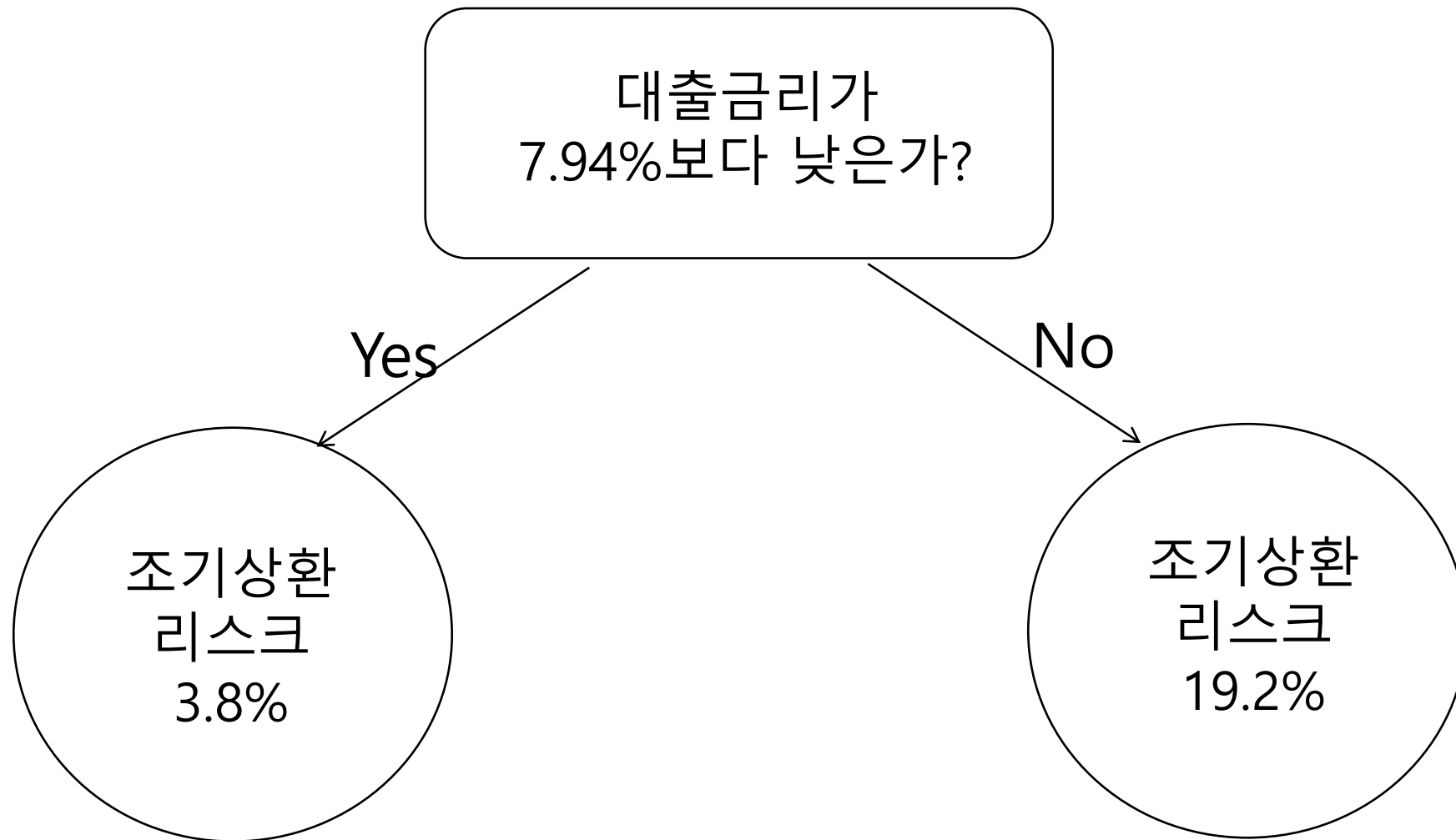
예측 분석 응용: 이탈 모델링으로 고객 이탈 방지하기

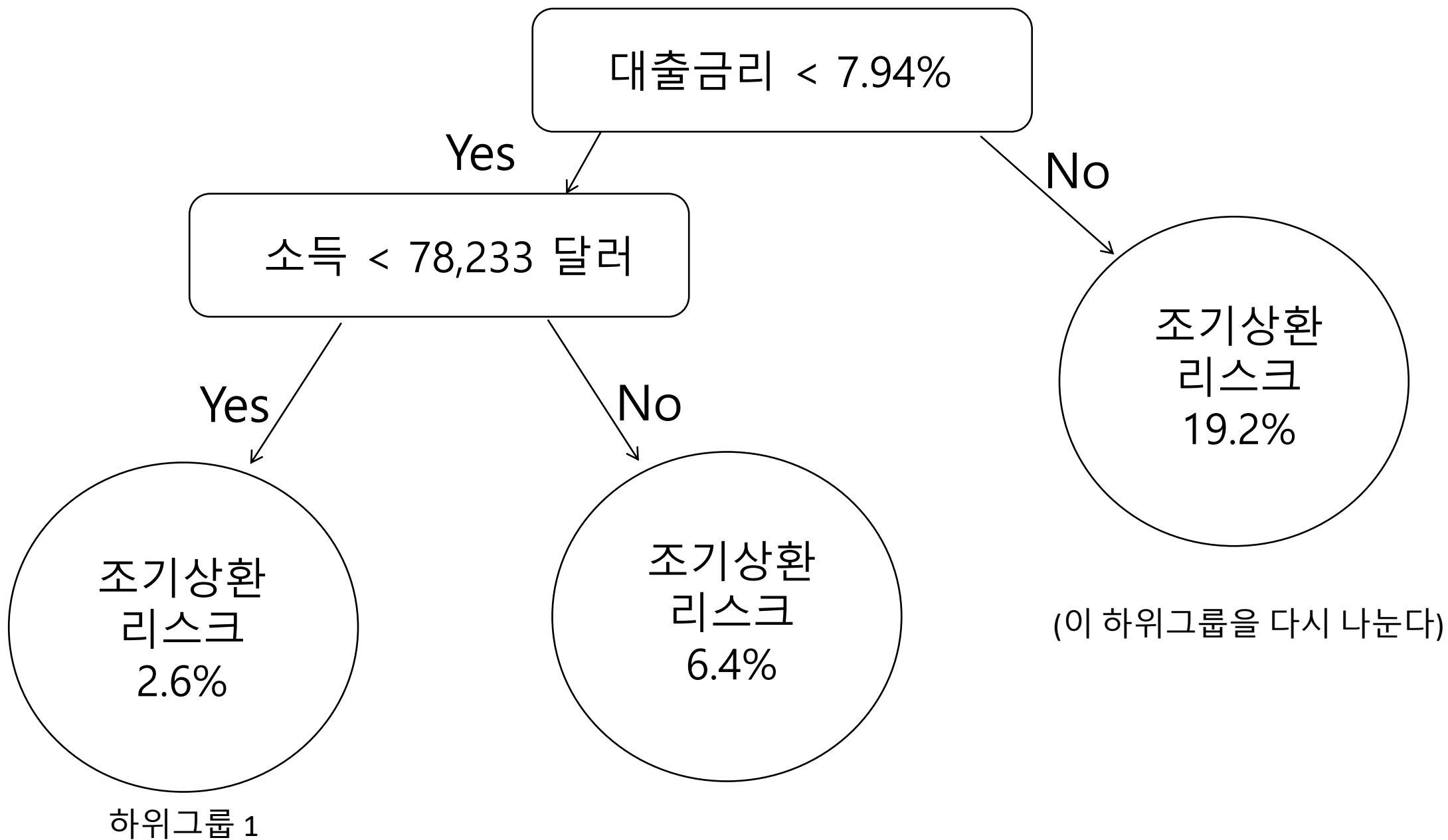
1 무엇을 예측하는가?

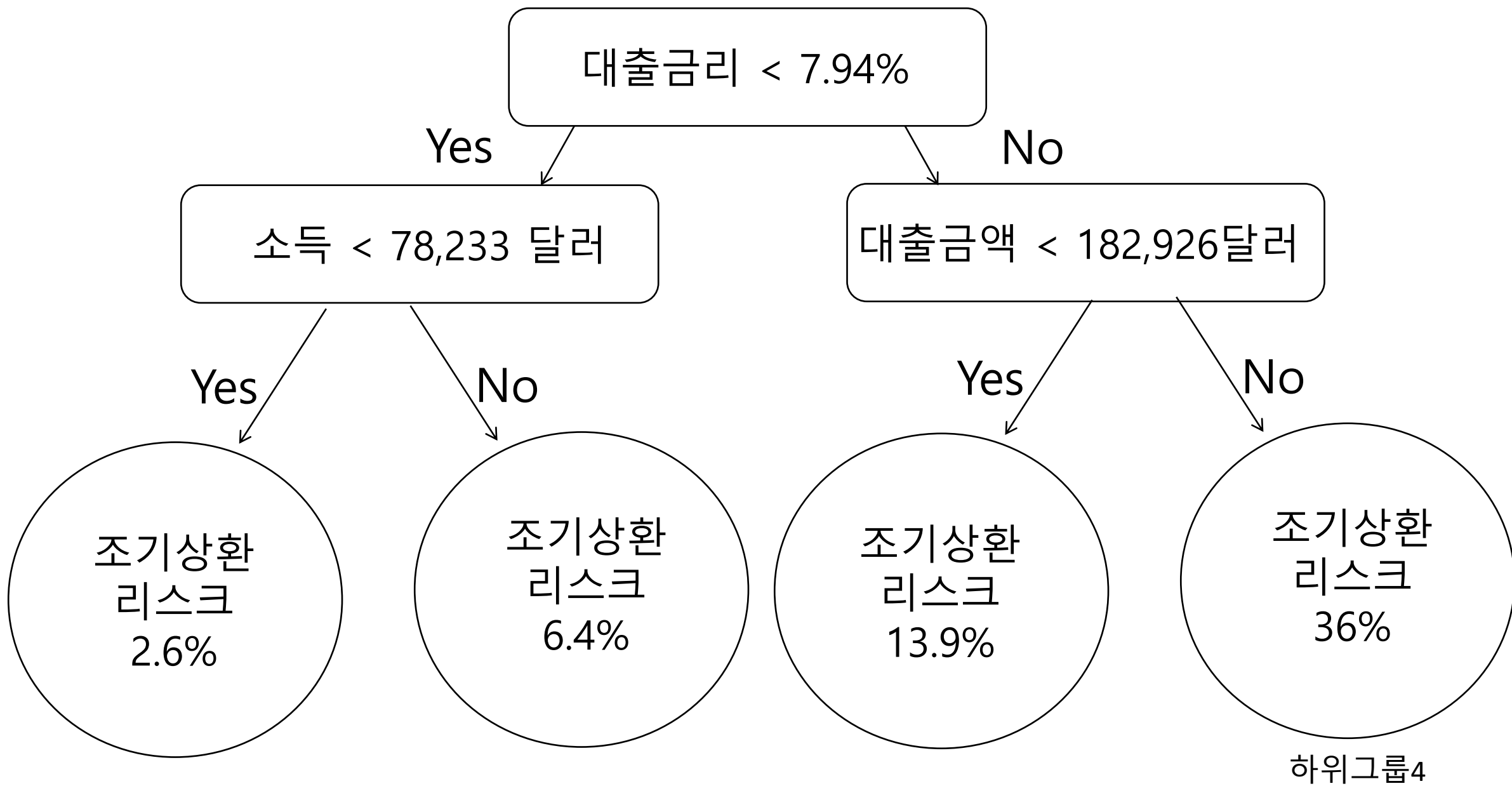
어느 고객이 떠나갈 것인가.

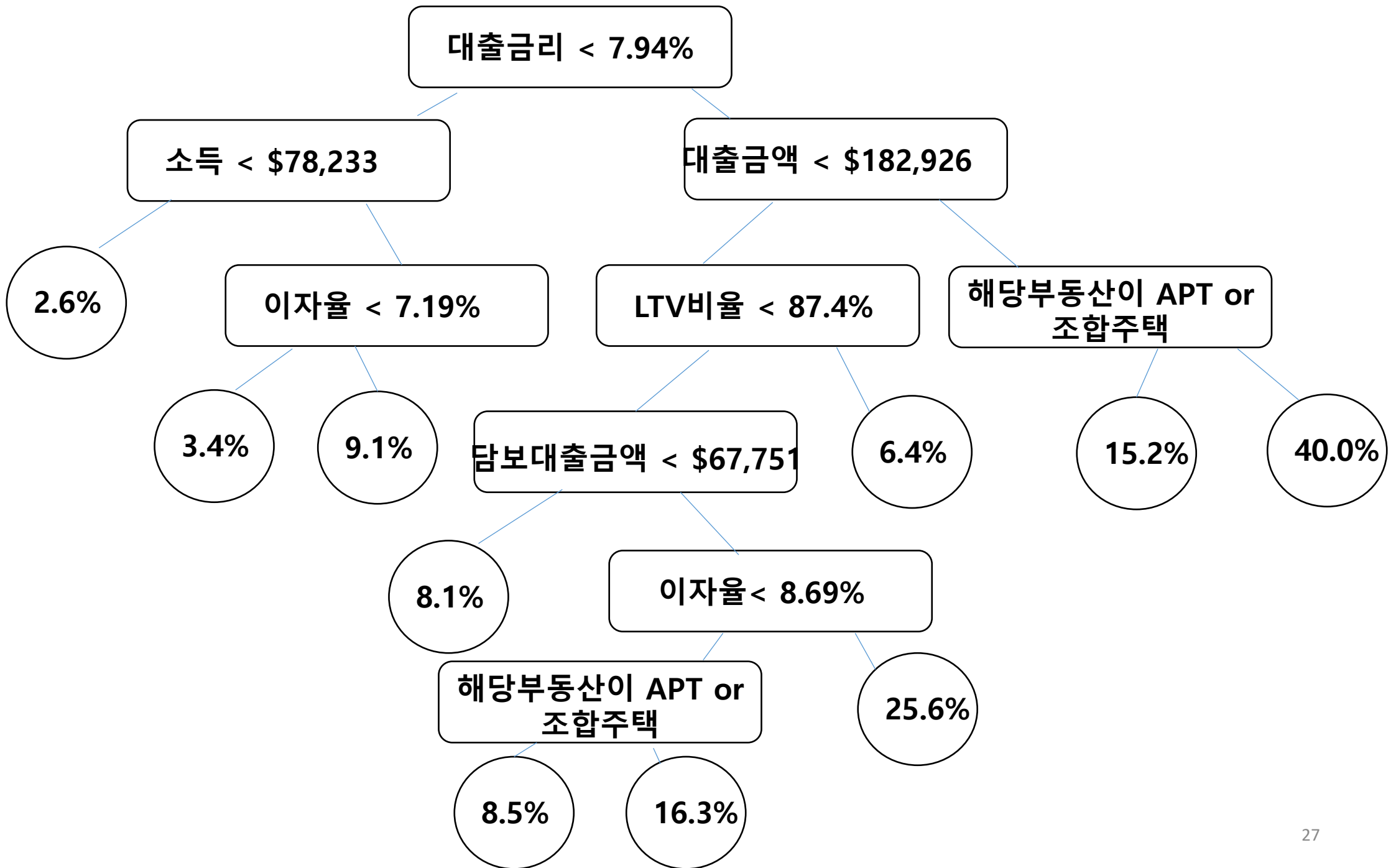
2 무엇을 할 것인가?

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.

정 준 수 / Ph.D (jsjeong@hansung.ac.kr)

- 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: Computer Vision, 머신러닝(ML), RPA
- <https://github.com/JSJeong-me/>

