

/\* elice \*/

# 문과생을 위한 머신러닝

4주차: 머신러닝 실무체험



David Oh 선생님

# 목차

1. 머신러닝 업무 익히기
2. 타겟 마케팅을 위한 머신러닝 업무

# 1. 머신러닝 업무 익히기

# 머신러닝 업무 리뷰

< 데이터 과학의 목표 >

Decision Making



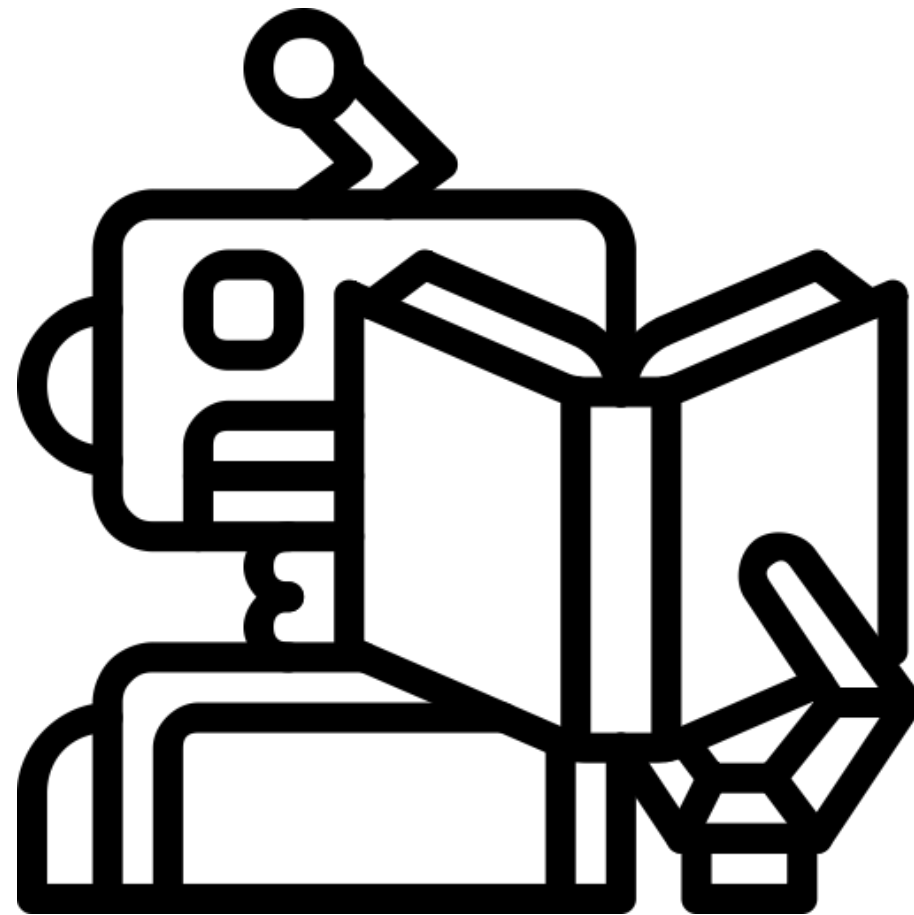
Monetization



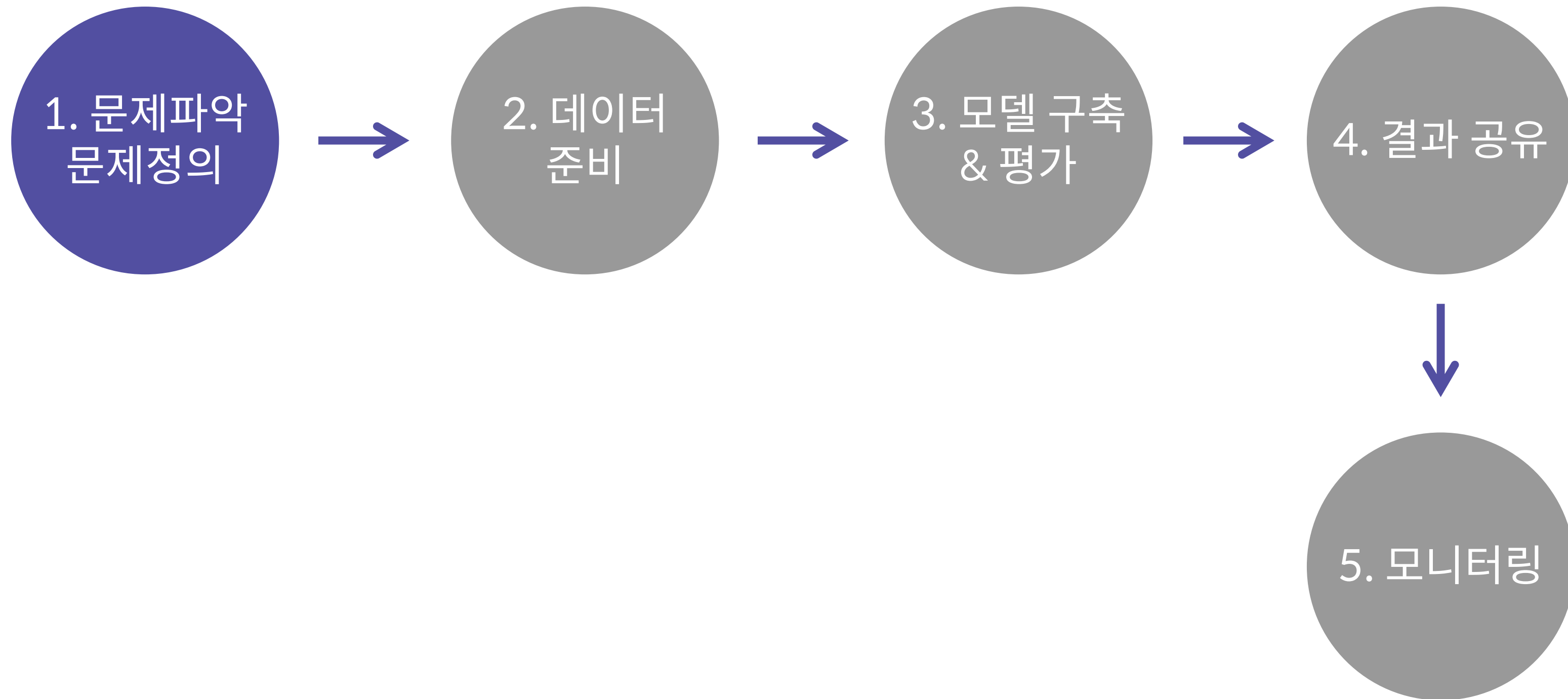
# 머신러닝 업무 리뷰

< 머신러닝의 목표 >

Prediction & Pattern Analysis



# 머신러닝 업무 프로세스



# Define the Problem

머신러닝 프로젝트를 시작할 때  
해결해야 하는 **비즈니스 문제**를 명확하게 먼저 정의

## < 문제정의/문제파악을 위한 세부 프로세스 >



# Define the Problem

비즈니스 문제를 파악한 후에 이를 해결하기 위한  
데이터 과학과 머신러닝 문제로 전환



# Types of Machine Learning

## **Supervised Learning**

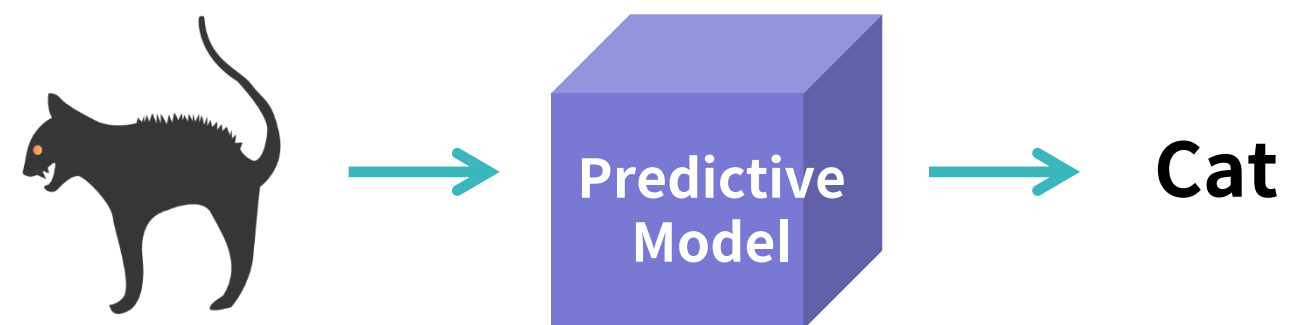
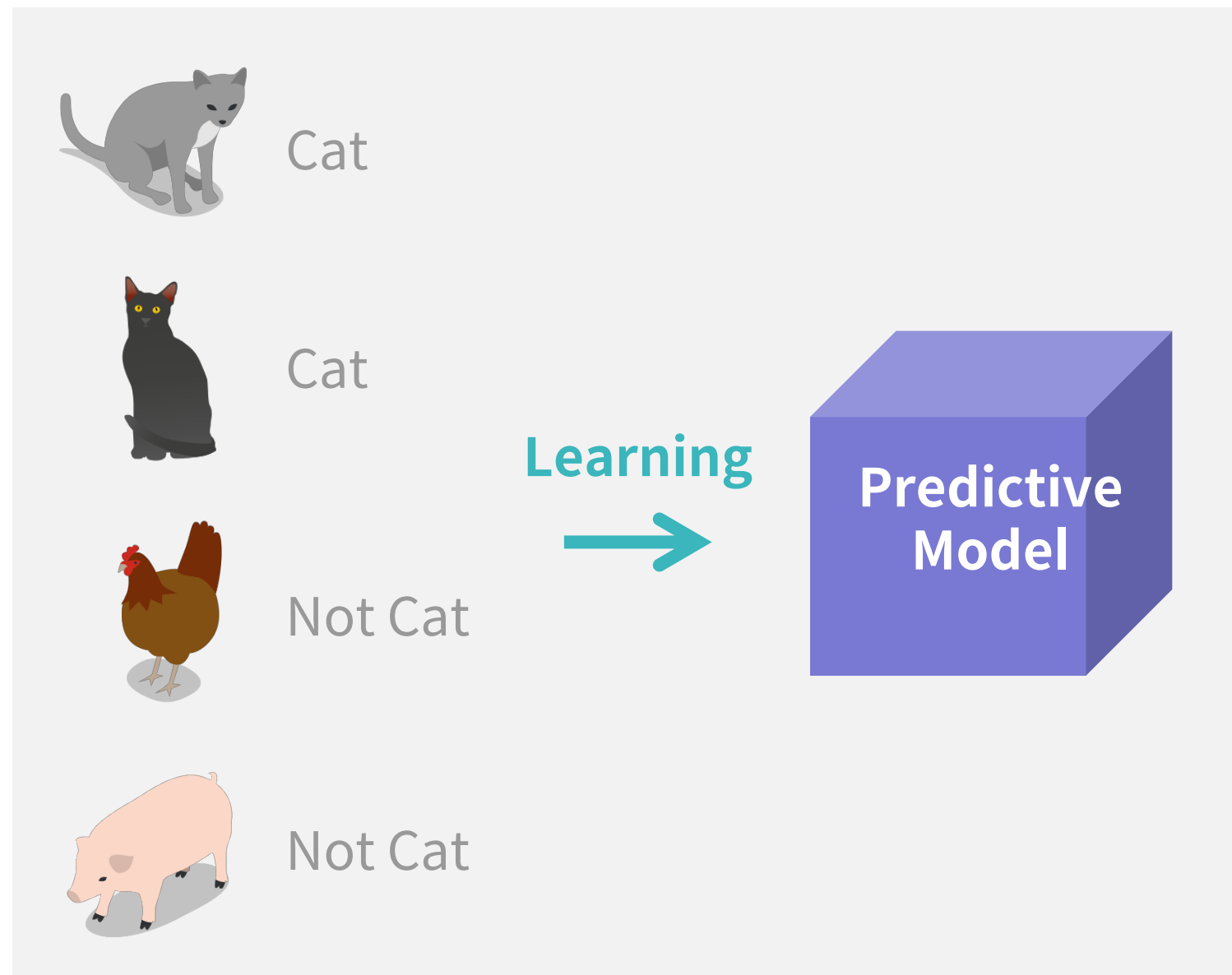
Develop Predictive Model  
based on Input & Output Data

## **Unsupervised Learning**

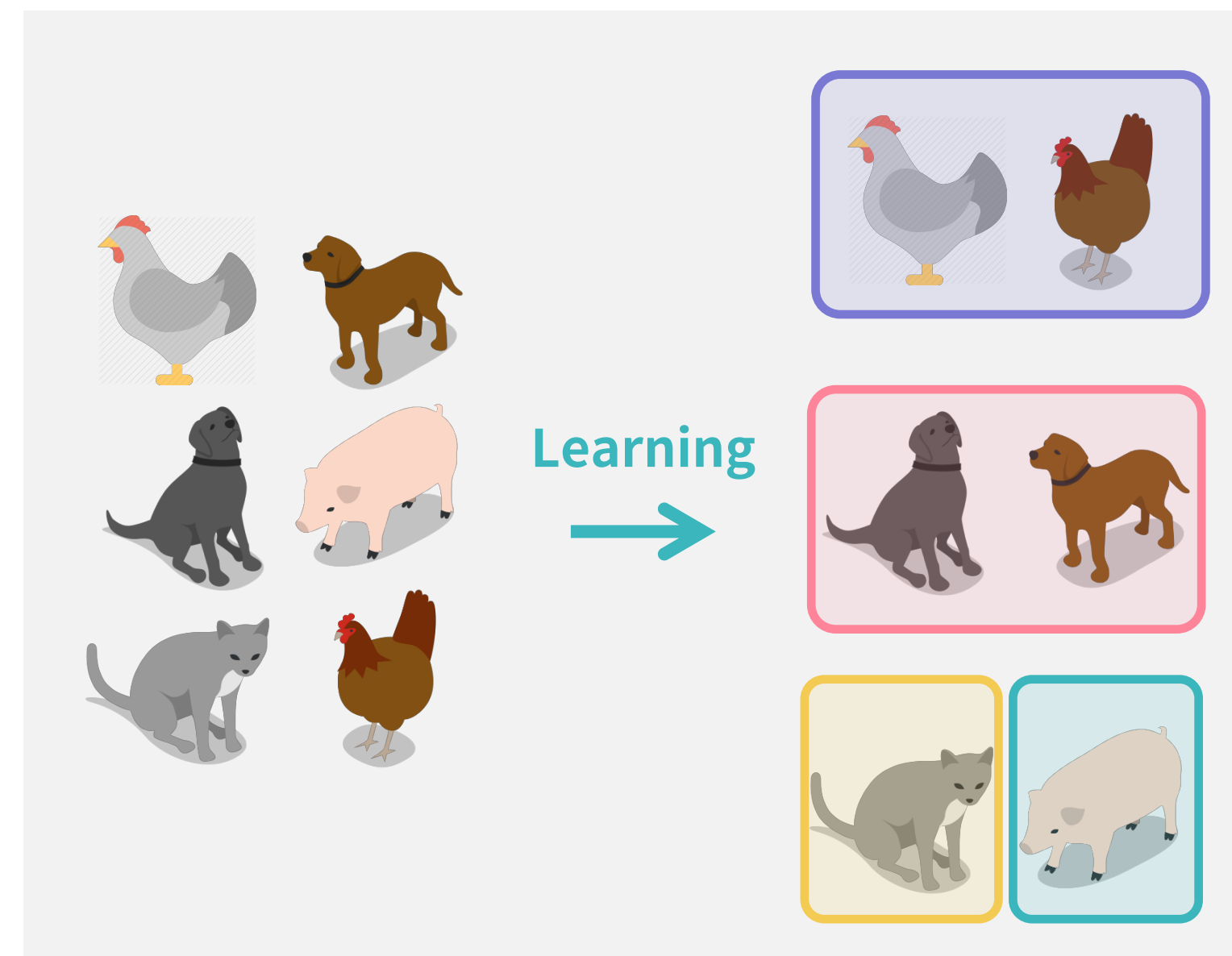
Group and Interpret Data  
based on only Input Data

# Types of Machine Learning

## Supervised Learning



## Unsupervised Learning



# Supervised Learning

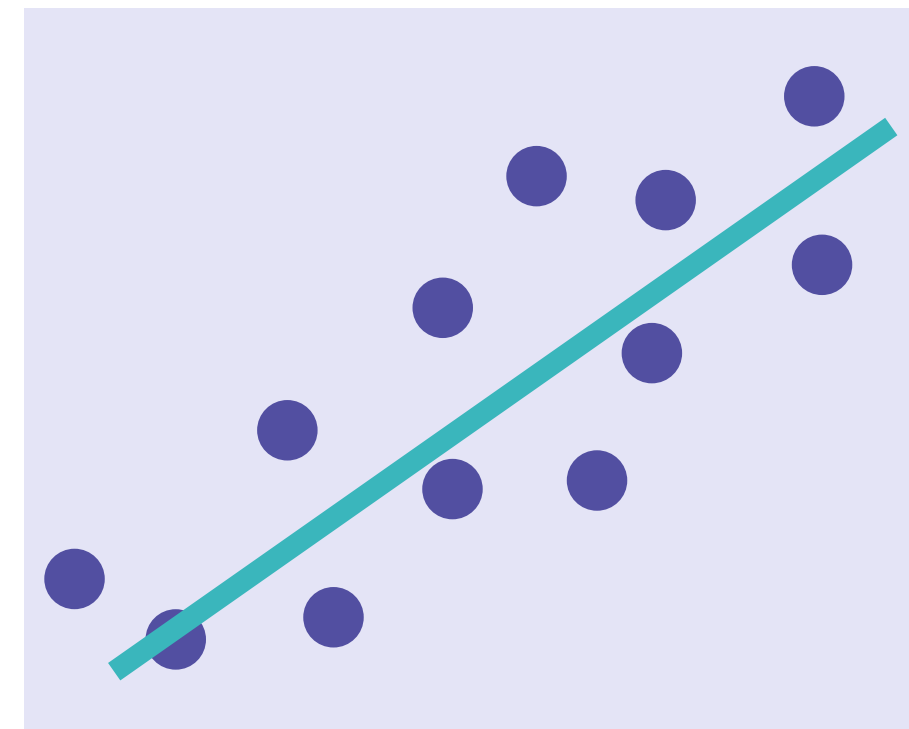
## Classification

분류  
범주를 예측



## Regression

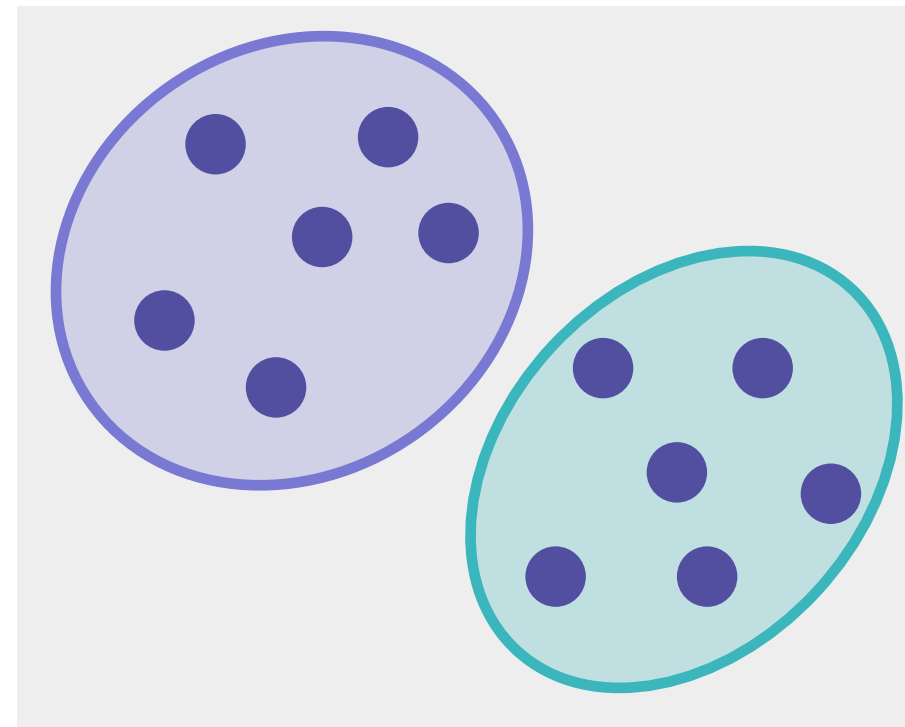
회귀  
숫자를 예측



# Unsupervised Learning

## Clustering

유사한 그룹끼리 군집화



# 현실의 문제를 머신러닝 문제로

Business Problem	Target/Output	ML Problem
고객이 서비스를 이탈할 것인가	범주 : 이탈여부	Classification
내년도 서비스 예상 매출액은 얼마인가	숫자: 매출액	Regression
사용자 정보와 구매이력 기반 고객 세분화	-	Clustering

# 기타 머신러닝 문제

**Recommender  
System**

**Anomaly  
Detection**

**Network  
Analysis**

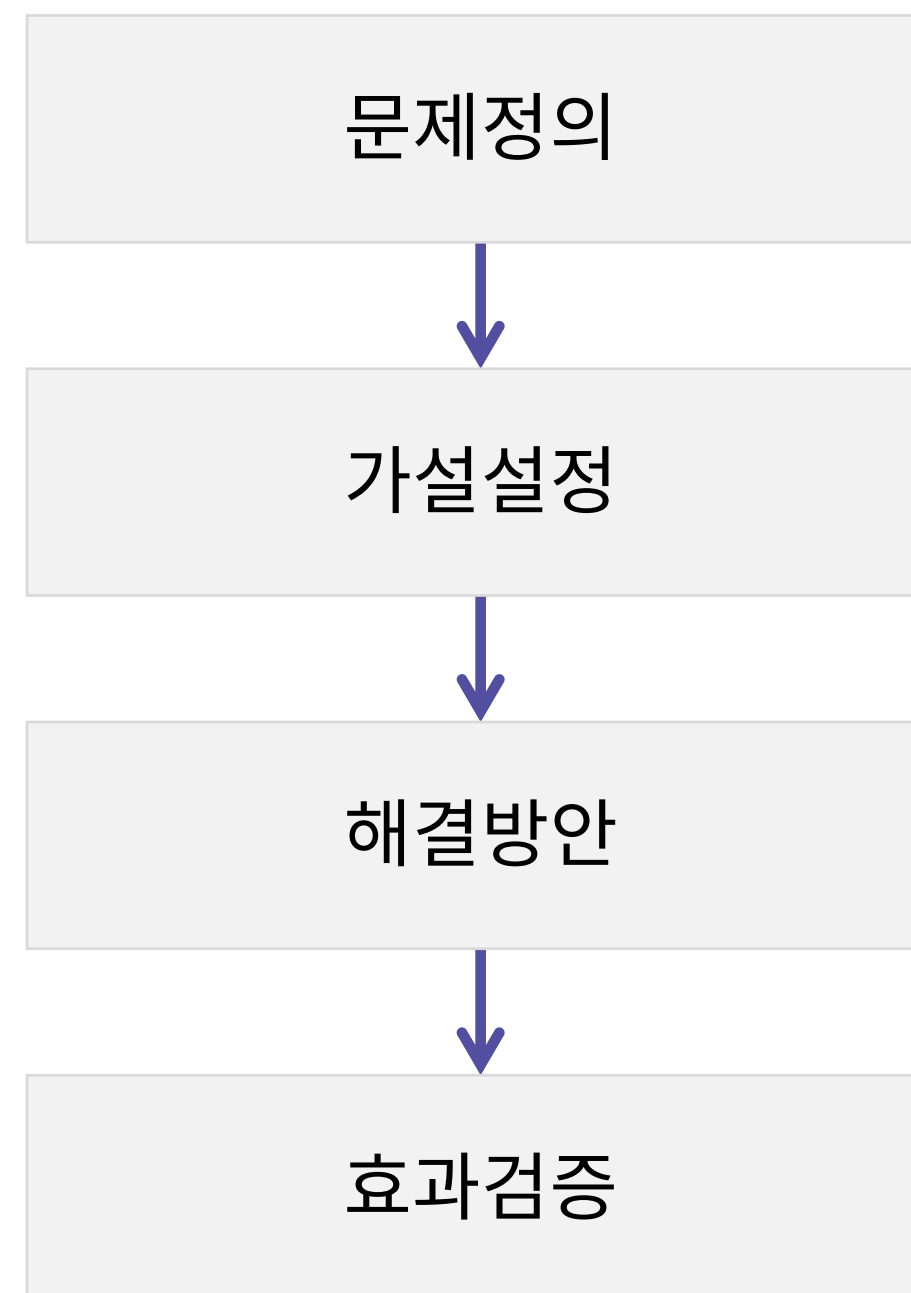
**Dimensionality  
Reduction**

**Profiling**

**Time series  
Forecasting**

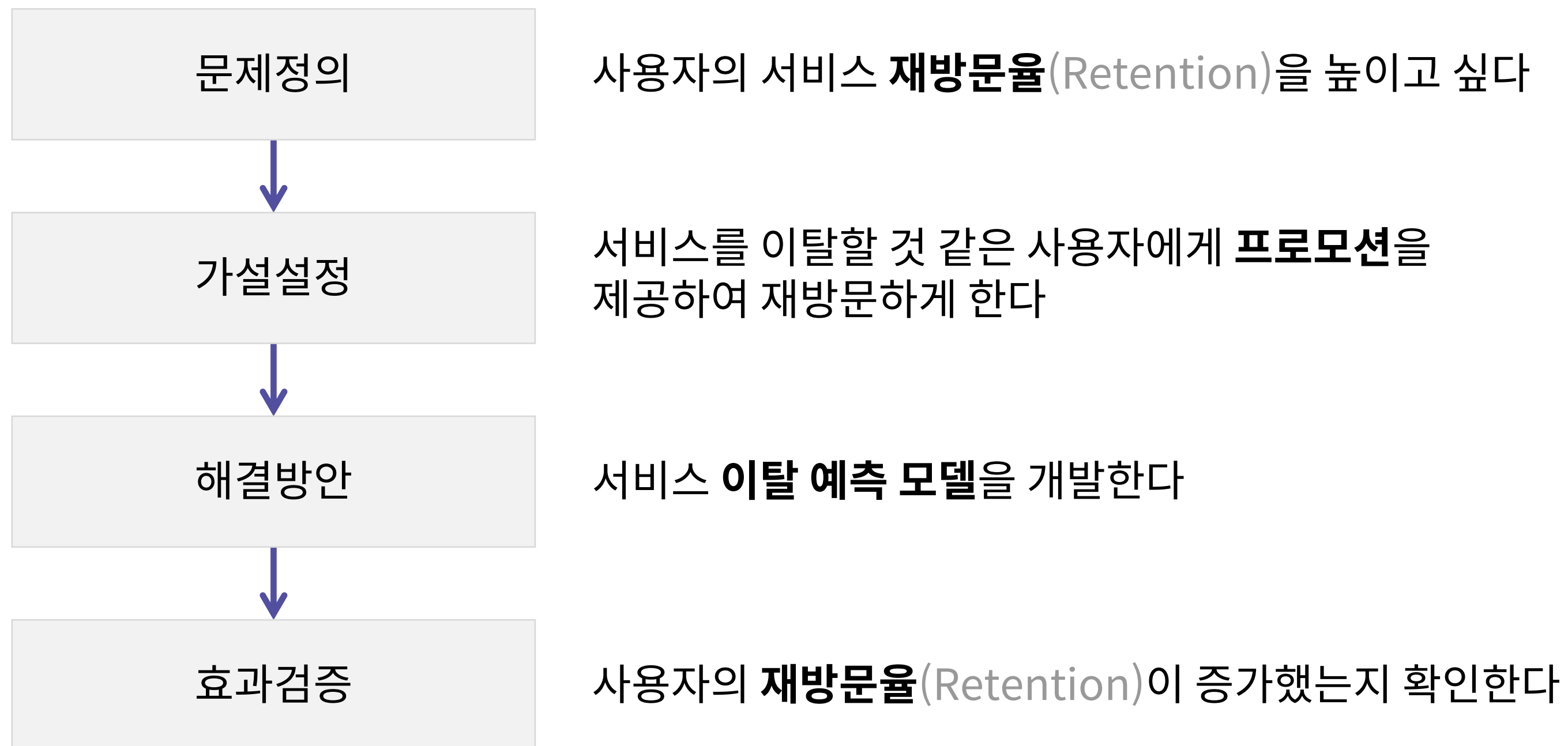
# 효과검증 설계 예시

머신러닝 도입에 따른 효과 검증 프레임워크



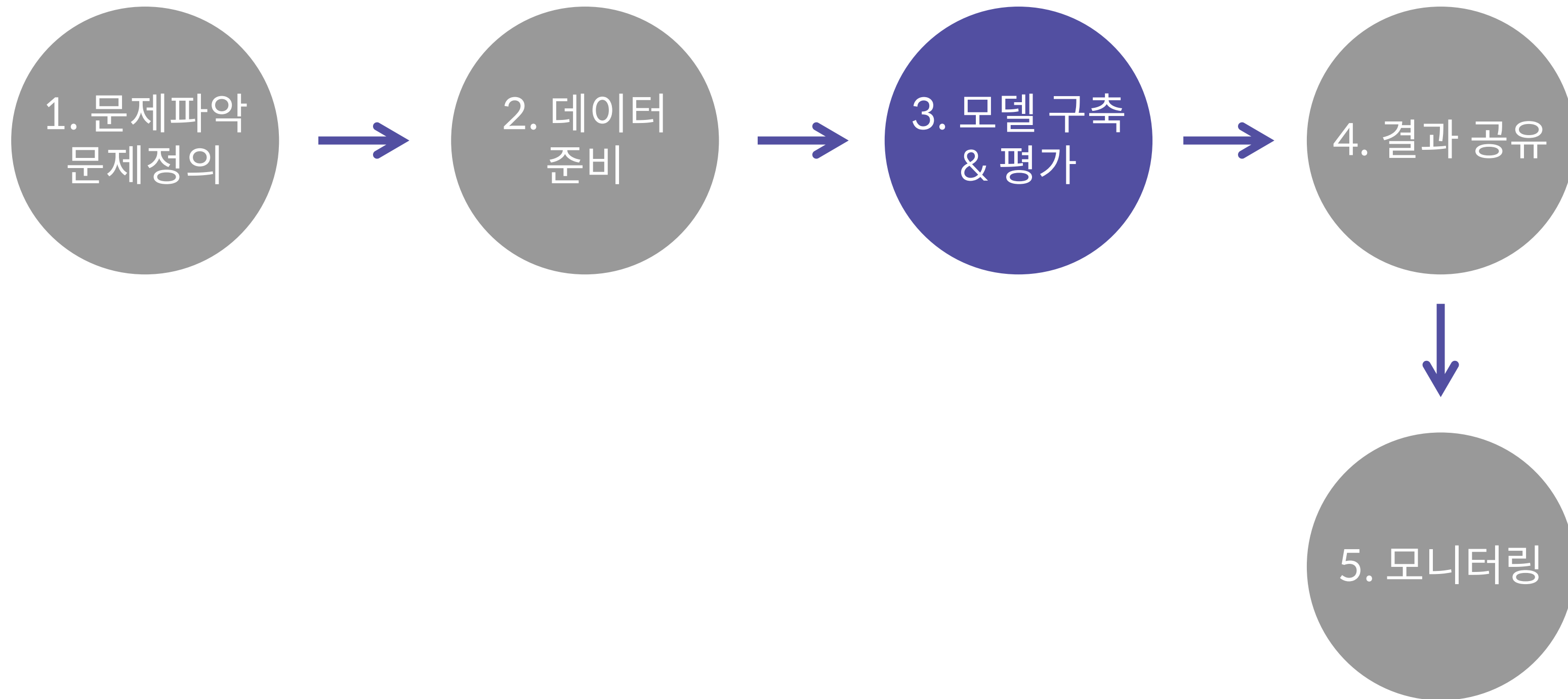
# 효과검증 설계 예시

## 머신러닝 도입에 따른 효과 검증 프레임워크





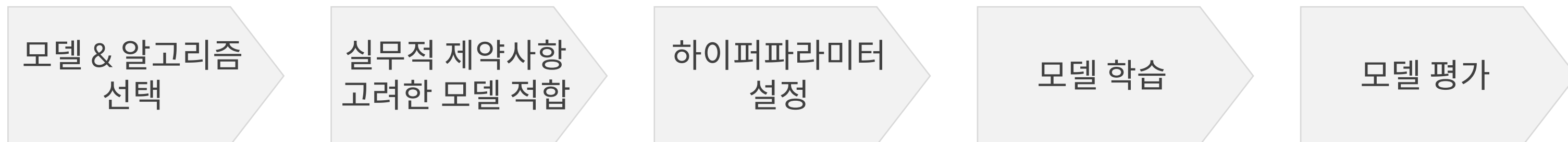
# 머신러닝 업무 프로세스



# Build Model & Evaluation

머신러닝 문제로 전환하고 데이터 준비를 마친 이후에는  
적절한 머신러닝 모델 & 알고리즘을 선택하여 **모델을 구축하고 평가**

## < 모델 구축 & 평가를 위한 세부 프로세스 >

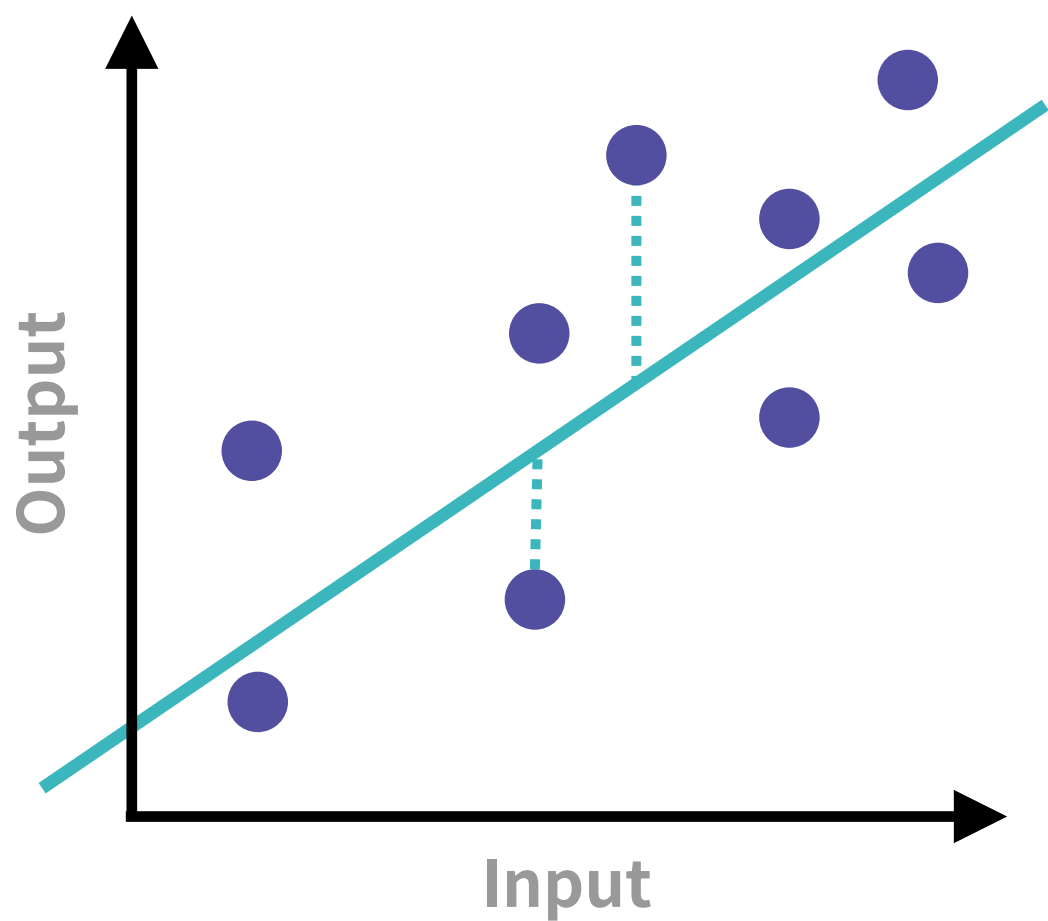


# 모델 & 알고리즘 선택

ML Model	Algorithm	Result
Classification	Logistic Regression Decision Tree Support Vector Machine	범주 예측
Regression	Linear Regression Ridge Regression Lasso Regression	숫자 예측
Clustering	K-means DBscan	군집

# 모델 평가

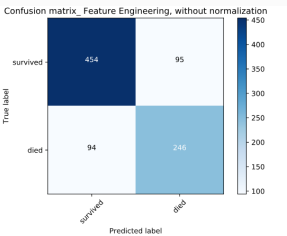
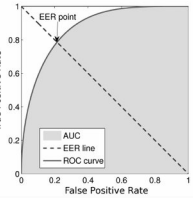
Regression은 실제 값( $y$ )과 예측한 값( $\hat{y}$ )의 차이  
오차(Loss/Cost/Error)를 통해 모델의 성능 평가



Acroynm	Full Name	Description
MAE	Mean Absolute Error	$\frac{1}{n} \sum  y - \hat{y} $
MSE	Mean Square Error	$\frac{1}{n} \sum (y - \hat{y})^2$
RMSE	Root Mean Square Error	$\sqrt{\frac{\sum (y - \hat{y})^2}{n}}$
MAPE	Mean Percentage Error	$\frac{100\%}{n} \sum \left  \frac{y - \hat{y}}{y} \right $

# 모델 평가

Classification은 **실제 범주**(*class*)와 **예측한 범주**(*class*)의 **정확도**(Accuracy)를 통해 모델의 성능 평가

Name	Description	Etc.
Accuracy	옳게 분류한 정확도	$\frac{\text{correct prediction}}{\text{total data points}} \times 100\%$
Confusion Matrix	분류 결과를 2x2의 표로 정리한 혼동행렬	
F-measure	precision과 recall의 조화평균 *precision: 예측한 범주에서 실제 True 범주 비율 *recall: 실제 범주에서 옳게 True라고 예측한 범주 비율	$F = \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$
AUC	TPR과 FPR을 각각 x축과 y축으로 했을 때의 생성되는 ROC curve 아래의 면적 *TPR: True Positive Rate 옳게 예측한 비율 *FPR: True라고 잘못 예측한 비율	

## 2. 타겟 마케팅을 위한 머신러닝 업무

# [실습] 누구에게 프로모션을 제공해야 할까?



/\* elice \*/

문의 및 연락처

[academy.elice.io](https://academy.elice.io)

[contact@elice.io](mailto:contact@elice.io)

[facebook.com/elice.io](https://facebook.com/elice.io)

[medium.com/elice](https://medium.com/elice)