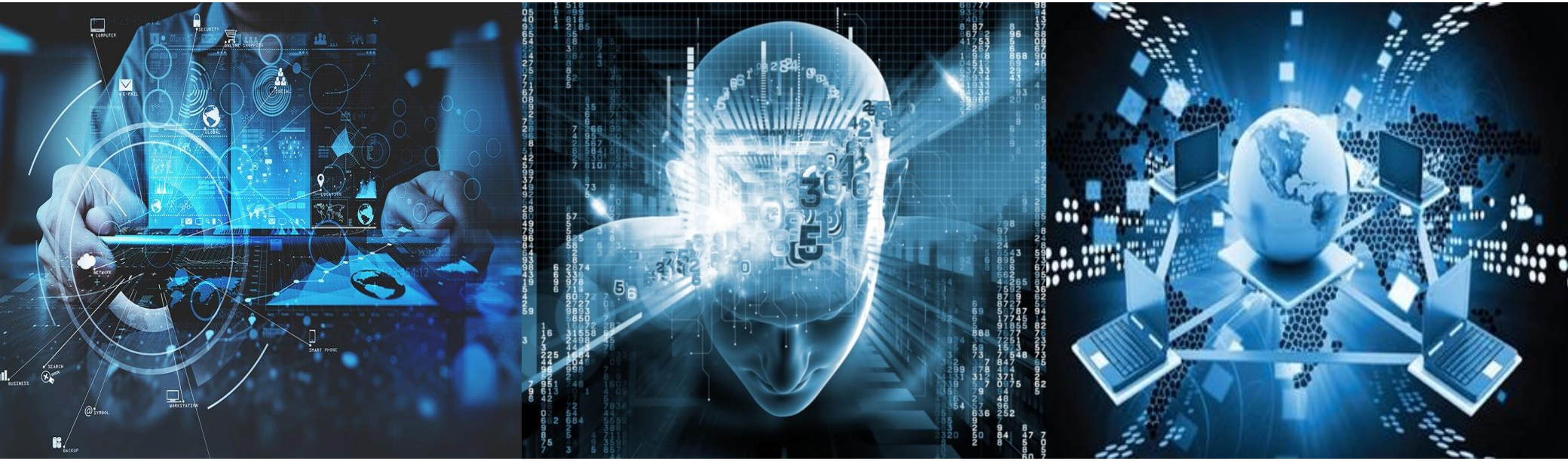
하나금융그룹

비파이 크래프트
Find Information to design valuable society



Machine Learning for BigData

Final Lecture

IT Competency Improvement Training
Kim Jin Soo

머신러닝(Machine Learning)

Using known data, develop a model to predict unknown data

알려진 데이터를 사용하여, 알려지지 않은 데이터를 예측하는 모델을 개발하는 기법

- Known Data :

  과거의 모든 빅데이터, 이전에 관측된 데이터,

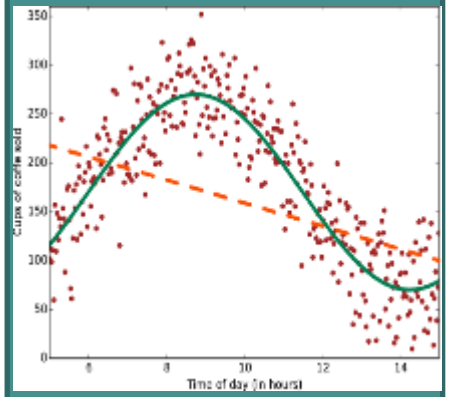
- Unknown Data :

  누락된 데이터, 보이지 않는 데이터, 존재하지 않는 미래데이터

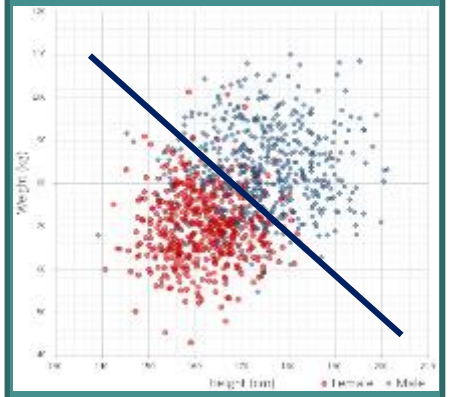
- Model : Know Data + Algorithms(ML algorithm)

머신러닝 알고리즘

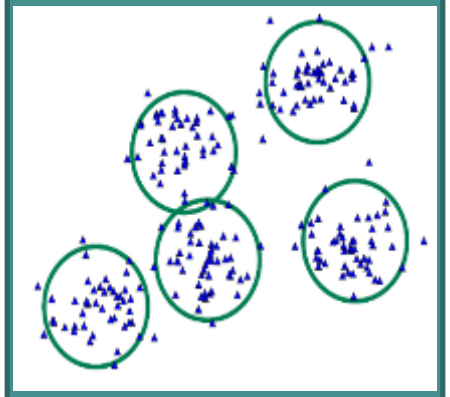
Regression



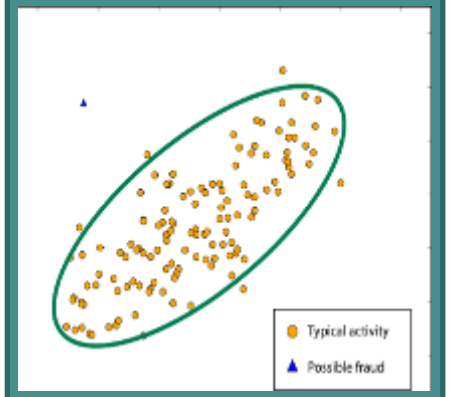
Classification



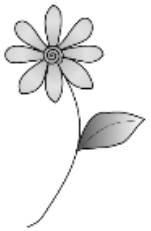
Clustering



Anomaly Detection



머신러닝 모델 : Regression

				
…	…	…	…	…
1990	50°F	30°F	68°F	95°F
2000	48°F	29°F	70°F	98°F
2010	49°F	27°F	67°F	96°F
2020	?	?	?	?

Using known data, develop a model to predict unknown data.

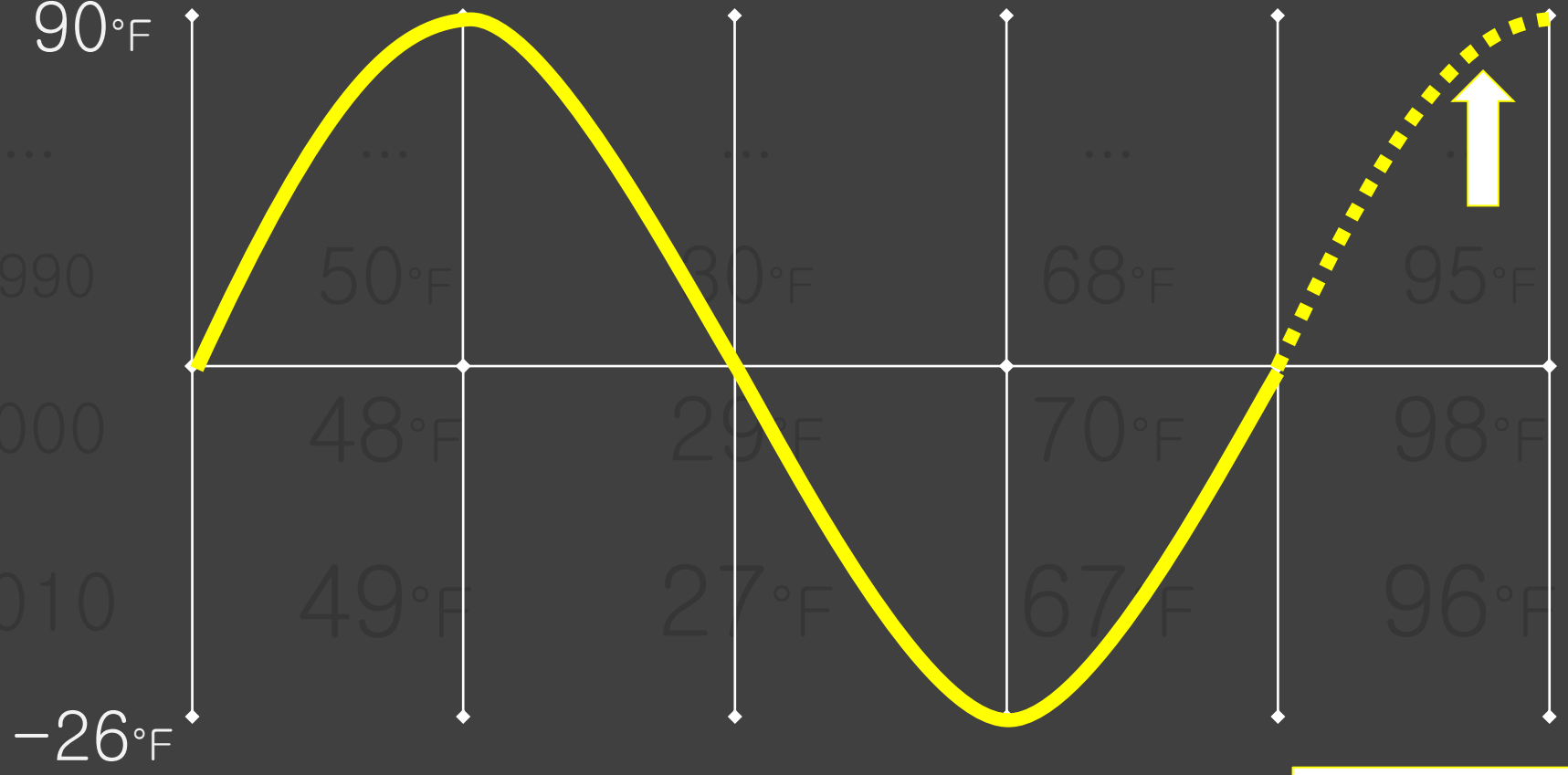
머신러닝 모델 : Regression




…	…	…	…	…
1990	50°F	30°F	68°F	95°F
2000	48°F	29°F	70°F	98°F
2010	49°F	27°F	67°F	96°F
2020	?	?	?	?

Using known data, develop a model to predict unknown data.

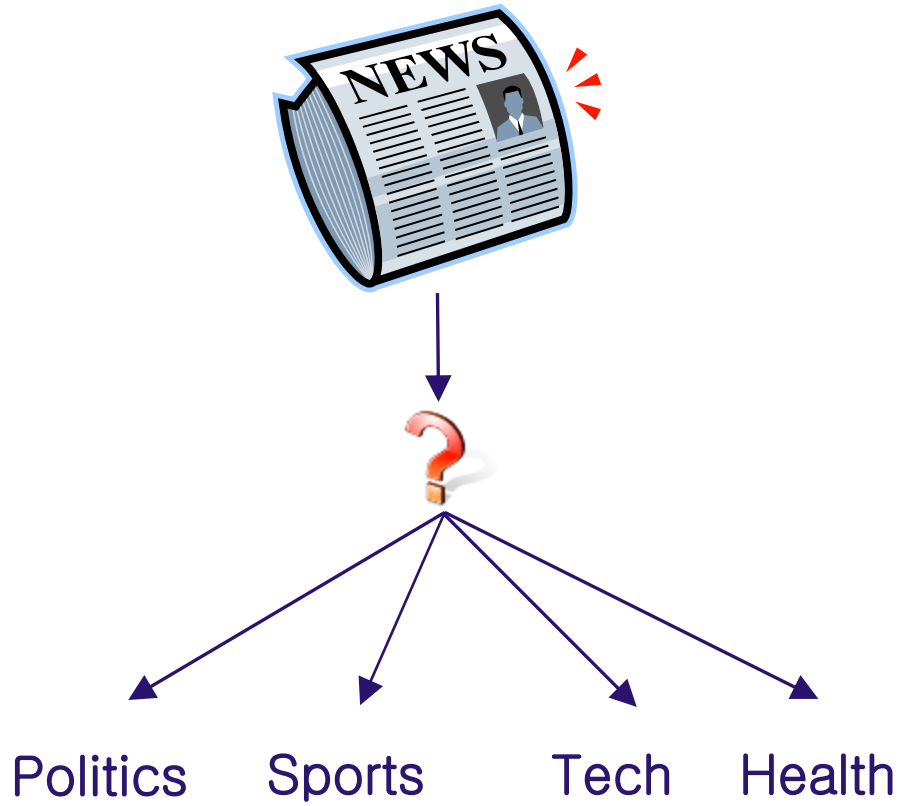
머신러닝 모델 : Regression



Using known data, develop a model to predict unknown data.

머신러닝 모델 : Classification

Classify a news article as (politics, sports, technology, health, ...)



머신러닝 모델 : Classification

Documents consist of unstructured text. Machine learning typically assumes a more structured format of examples

Process the raw data

머신러닝 모델 : Classification

Known data (Training data)

Documents Labels

Tech

Health

Politics

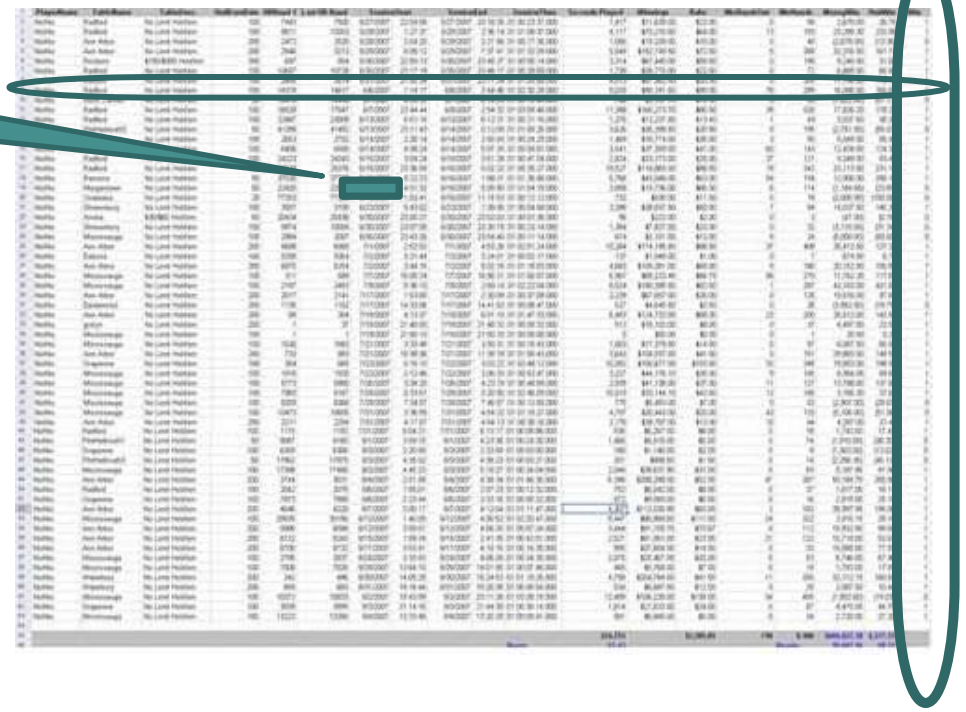
Politics

Sports

Process each data instance to represent it as a feature vector

Documents Labels

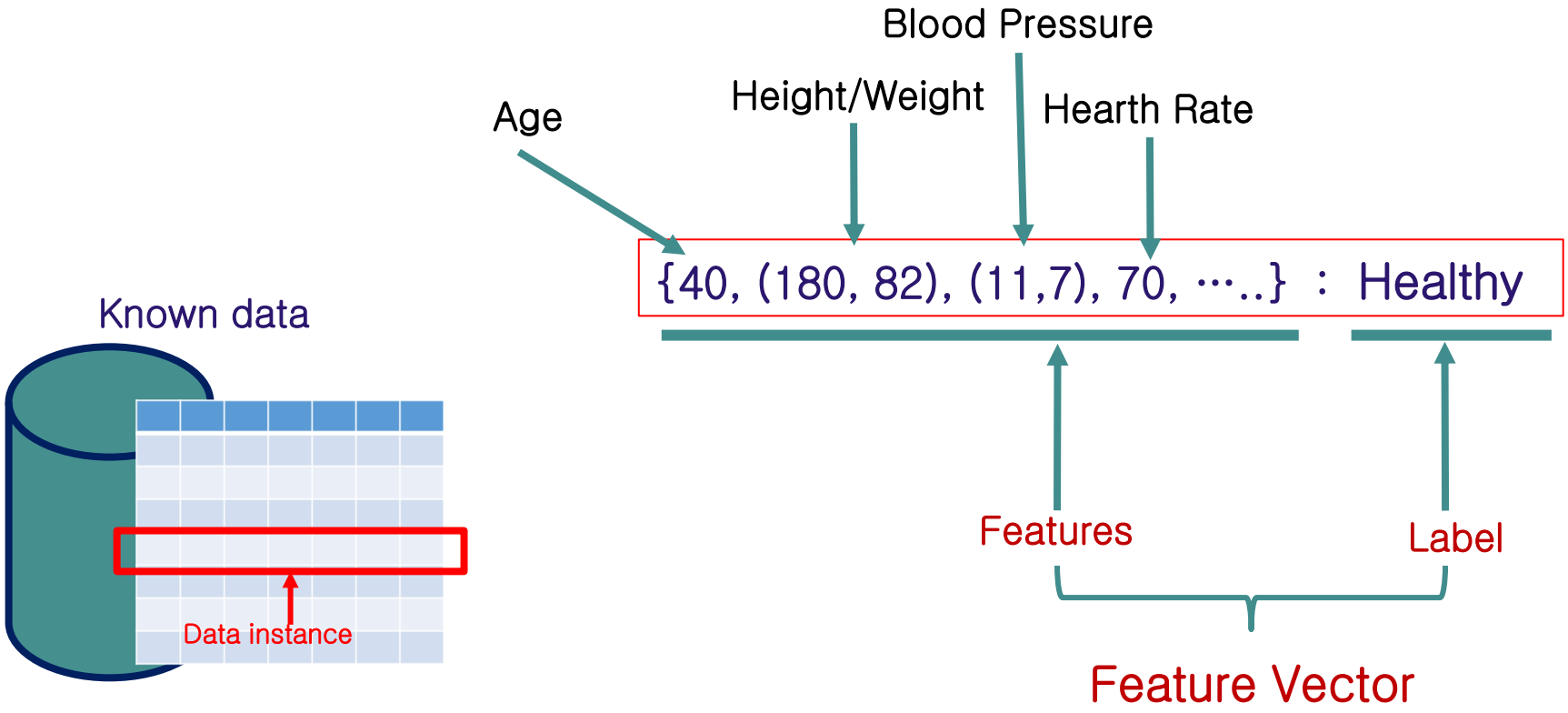
Feature



머신러닝 모델 : Classification

Feature vector

i.e.



머신러닝 모델 : Classification

Developing a Model

Training data

Documents Labels

Feature Vectors

Tech

Health

Politics

Politics

Sports

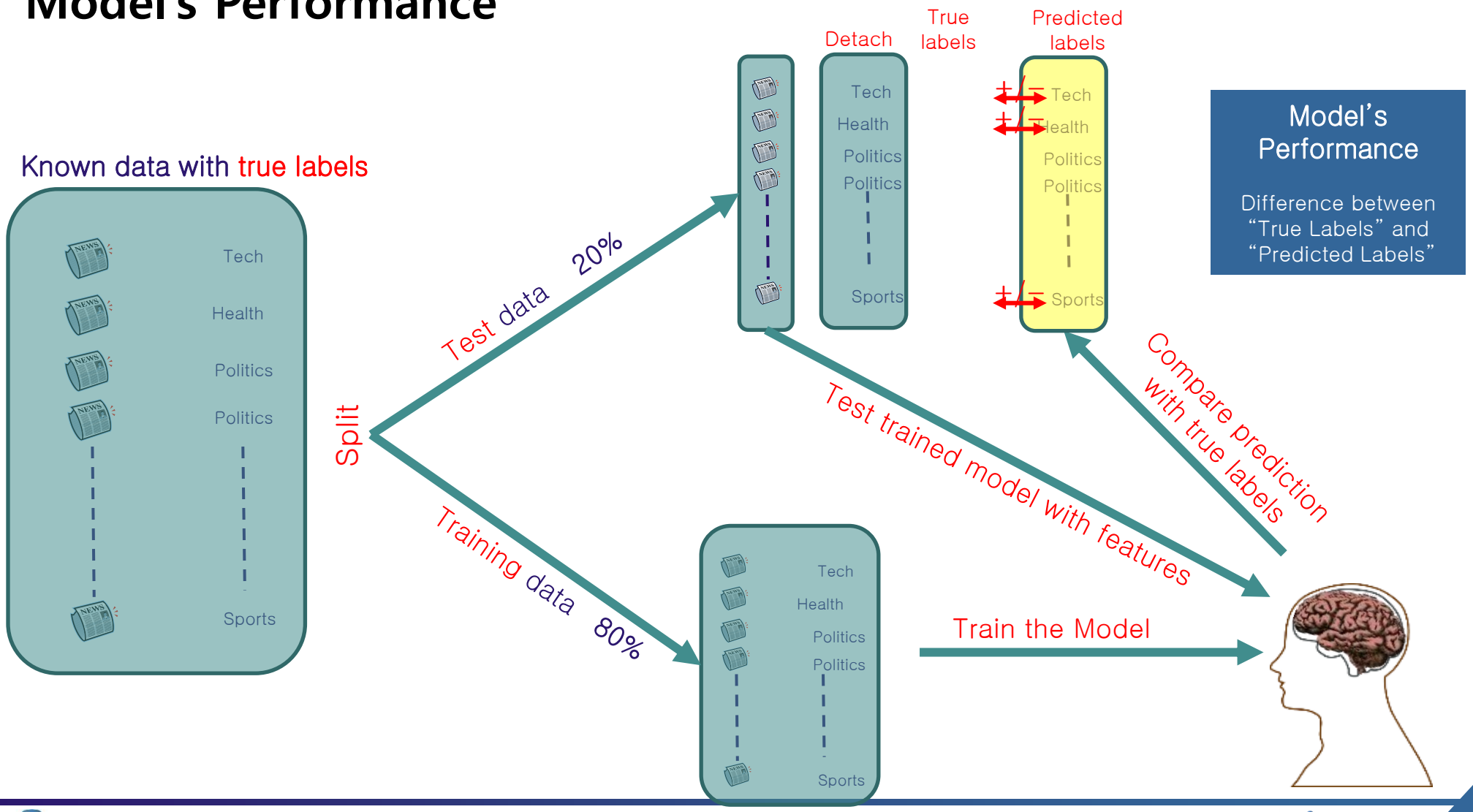
Train
the Model

Base
Model

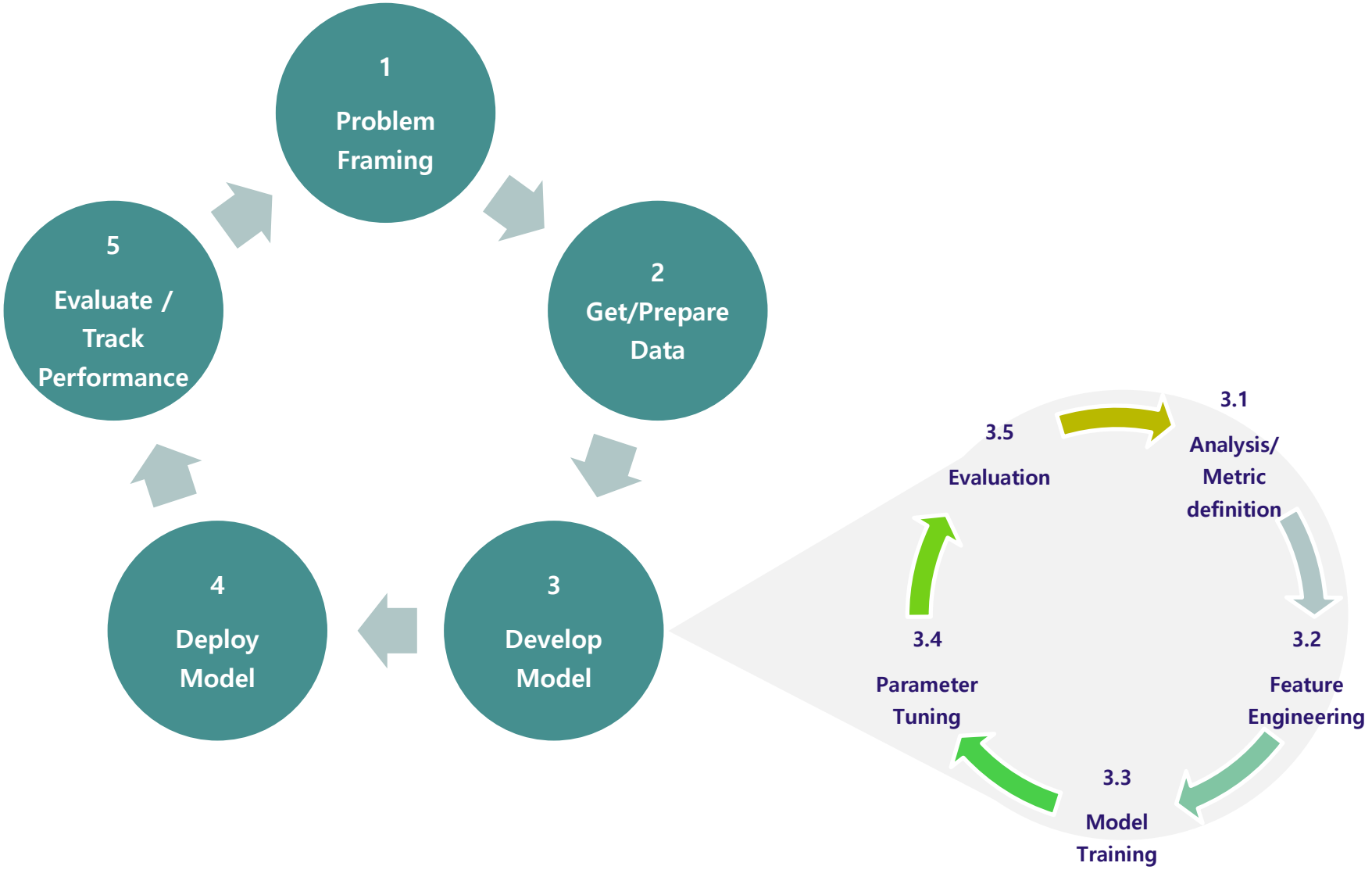
Adjust
Parameters

머신러닝 모델 : Classification

Model's Performance



머신러닝 솔루션 구현 프로세스



머신러닝 적용 분야

Finance and risk	Sales and marketing	Customer and channel	Operations and workforce
$$$ Revenue Forecasting	Sales forecasting	User segmentation	Agent allocation
Portfolio optimization	Demand forecasting	Personalized offers	Warehouse efficiency
$$$ Investment modelling	Sales lead scoring	Product recommendation	Smart buildings
Fraud detection	Marketing mix optimization		Predictive maintenance
Risk management			Supply chain optimization

과정을 마무리 하며...

- **세상에 더 좋은 머신러닝은 없다.**
  - 더 적합한 머신러닝만 있을 뿐...
- **잘 아는 것부터 점진적으로 접근해 나가라.**
  - 블랙박스에 맡기는 것이 아니라, 하나씩 처방해 나가는 것이다.
- **머신러닝은 성능점수를 최적화 하는 것이다.**
  - 즉, 성능측정기준을 무엇으로 하느냐가 중요하다.
  - 공부(운동) 잘했어? → 공부(운동)에 최적화 된 아이로 자란다.
  - 단순히 돈 벌고 싶다가 아니라,
    어떤 고객을 대상으로 어떤 상품을 얼마만큼 팔 수 있는가를 검증
- **딥러닝이 좋은 경우**
  - 내가 세상의 모든 데이터를 다 가지고 있을 때
  - 내가 가지고 있는 지식이나 능력으로 해결되지 않을 때
- **데이터사이언스는 "프로그래밍"이 아니라 "디버깅"이다.**
  - 데이터를 넣어보고, 왜 안 되는지를 끊임없이 고민
  - 머신러닝 = 러닝머신 ^^

감사합니다!

감사합니다!