

Data Analysis

2019. 12.

Instructor : 전 윤 길

1. 데이터 분석의 이해

1.1 인공지능과 데이터 분석



I 1.1 인공지능 (A.I. : Artificial Intelligence)



■ Artificial Intelligence

Intelligence : the ability for problem solving

Artificial Intelligence : Intelligence demonstrated by machines

I 1.1 인공지능 (A.I. : Artificial Intelligence)

인공지능 이란?

지능(Intelligence) 이란?

- 사고능력
- 사전적 정의 intellectual[mental] faculties; mental capacity; intelligence; intellect (▶intelligence는 두뇌의 활동, intellect는 지적인 일에 흥미가 있어 두뇌를 활동시키는 힘이 있음).
- 일반적 정의
- 적응적(Adaptive), 학습능력(Learning Ability), 선행지식(Use of Prior Knowledge) 활용, 여려가지 다른 정신 과정들의 복잡한 상호과정과 조정하는 능력, 문화특수성(Cultural specificity), 비보편성, 감정과 독립하여 사고 가능 ([Source : https://ko.wikipedia.org/wiki/](https://ko.wikipedia.org/wiki/))

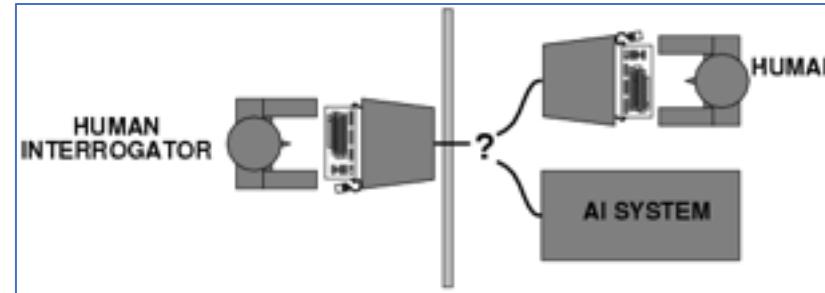
다시, 인공지능(*artificial intelligence*)이란?

- 시스템에 의해 만들어진 지능

I 1.1 인공지능 (A.I. : Artificial Intelligence)

인간처럼 행동하는가? (Turing Test)

- Turing (1950) : "Computing machinery and intelligence",
- "Can machines think? → "Can machines behave intelligently?"



인간처럼 생각하는가? (인지과학 : Cognitive Science)

- 1960's "cognitive revolution" : information-processing psychology
- 합리적으로 생각하는가? (Laws of Thought) : 규범적(Normative) 사고하는가?
- 합리적으로 행동하는가? (doing the right thing)

I 1.1 인공지능 (A.I. : Artificial Intelligence)

문제해결 (Problem solving)

문제 타입(Problem types)

- 단일 상황 문제(Single-sate problem) : 결정론적(Deterministic), 완전 관측(fully observable)
- 규격 문제(Conformant problem) : 비식별적(Non-observable)
- 상황 적응 문제(Contingency problem) : 비결정적 문제(Nondeterministic and/or problem)
- 탐험 문제(Exploration problem) : 알 수 없는 상태(Unknown state space)

Ex) Tree search algorithm

$F(x_1, x_2) = y$

$F(\text{problem}, \text{strategy}) = \text{a solution}$

`Tree_Search(problem, strategy) returns a solution, or failure`

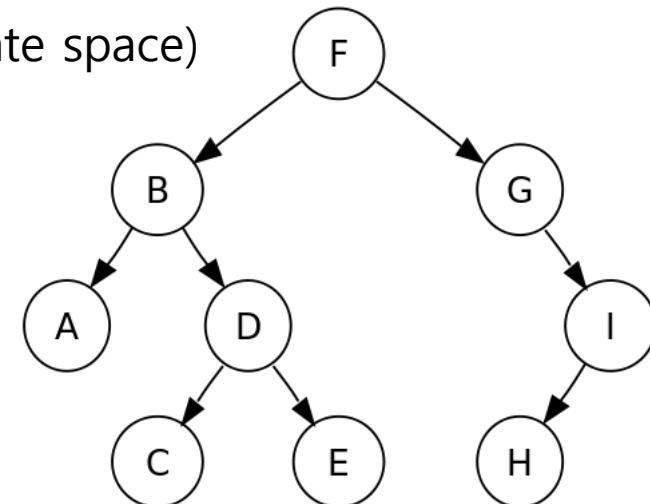
`loop do`

`if true then return`

`if`

`else`

`End`



→ 각 문제에 대한 제약조건 (Constraint)내 문제해결

I 1.1 인공지능 (A.I. : Artificial Intelligence)

인공지능 역사

1950 ~ 1970년대 태동기

튜링의 사고기계 제안(1950)
인공지능의 탄생(1952-1956)
퍼셉트론(Perceptron), 신경망(neural networks)(1957)
IJCAI학회 창립(1969)

AI 첫번째 인공지능의 겨울(1974 ~1980)

Minsky & Papert의 신경망의 한계(1969)

1970 ~ 1990년대 제1차 AI 산업화

AAAI학회 창립(1980)
지식기반 전문가 시스템(1975-1985)의 산업화

AI 두번째 인공지능의 겨울(1987 ~ 1993)
자연어(natural language)에서의 실패(Pinker & Prince, 1988)

1990 ~ 2010년대 과학적방법론의 도입

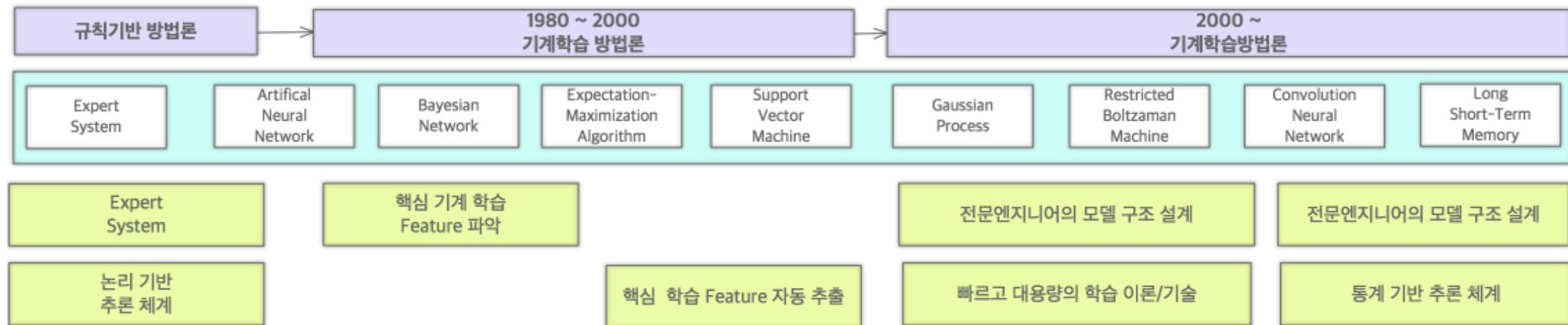
신경망과 머신러닝 연구
LeCun의 컨볼루션 신경망 개발(1989)
베이지안넷(1988)과 확률적 추론연구

2010 ~ 현재 : 제2차 AI 산업화

빅데이터와 머신러닝의 산업화
자율주행차, 인공지능 스피커, 인공지능 어드바이저
딥러닝 기술의 산업화(2012- 현재)

I 1.1 인공지능 (A.I. : Artificial Intelligence)

인공지능 방법론



유행과 사고의 변화 (Trend and change of Thoughts)

- 유행의 변화(Change of trends) :** **유행에는 기준이 없다 (NO Standing Trend)**
- 사고의 변화(Change of Thoughts) :** 요소의 중요성에서 **학습하는 구조의 중요성으로의 변화 (from importance of Feature to importance of Learning Structure)**

Domain Knowledge is very important on traditional methodologies of AI.

Even in the Cutting edge technology(**Deep Neural Network** in Machine Learning)

“understanding domain knowledge is the key to find solution”

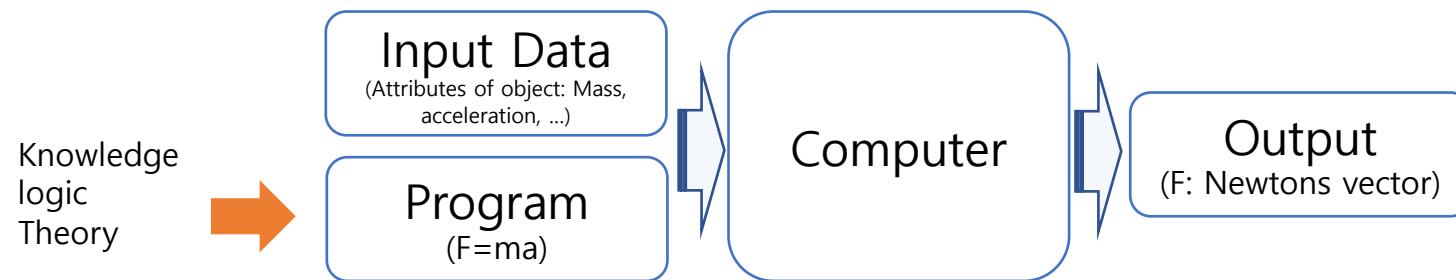
I 1.1 인공지능과 데이터 분석

데이터 분석 (Knowledges based Data Analysis)

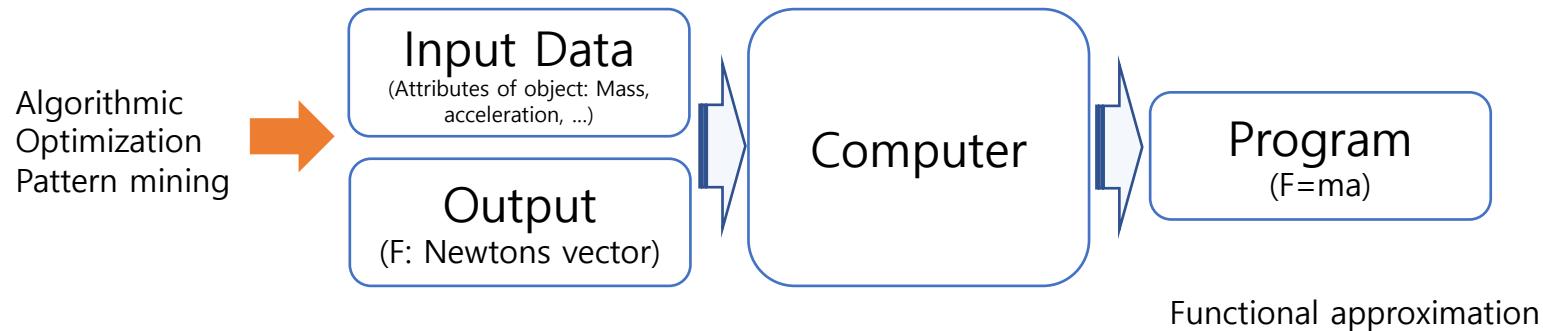
- 티코 브라헤(Tycho Brahe, 1546~1601) : 정확한 천체 관측 → **Big Data**
- 요한스 케플러(Johannes Kepler, 1571~1630): 데이터부터 이론 도출(law of planetary motion) → **Analysis**
- 아이작 뉴턴(Issac Newton, 1643~1727): 진리에 도달(universal gravitation) → **Truth**



구조적 접근(Structural Approach)



알고리즘 접근(Algorithmic Approach)



머신러닝(Machine Learning)은 알고리즘적 접근
: 함수 근사(Functional approximation)



I 1.1 인공지능과 데이터 분석

데이터 분석 (Data Analysis)

데이터(Data) 란?

- 추리, 토론, 계산의 근거
- 측정 및 통계량으로의 사실 정보 ("factual information")

데이터 형태(Type of Data)

- Quantitative data : expressed numerical terms
- Qualitative data : verbal or narrative format
- Mixed Data : Surveys

분석(Analysis) 이란?:

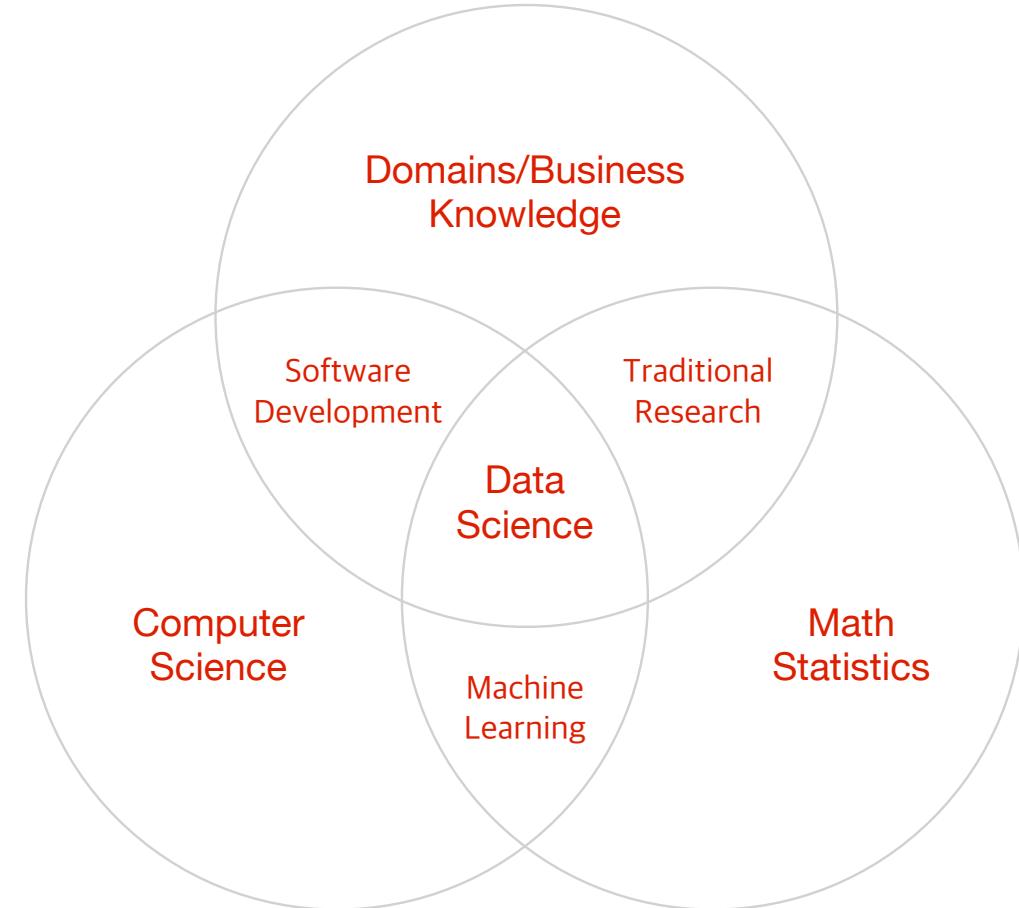
- 정보에 대한 해석 및 의미 파악
- 데이터 실험을 통해 패턴의 파악

데이터 분석의 목적:

- 유용한 정보의 발견
- 결과에 대한 예측
- 의사결정의 근거 활용

데이터 분석 과정(Process of Data Analysis) :

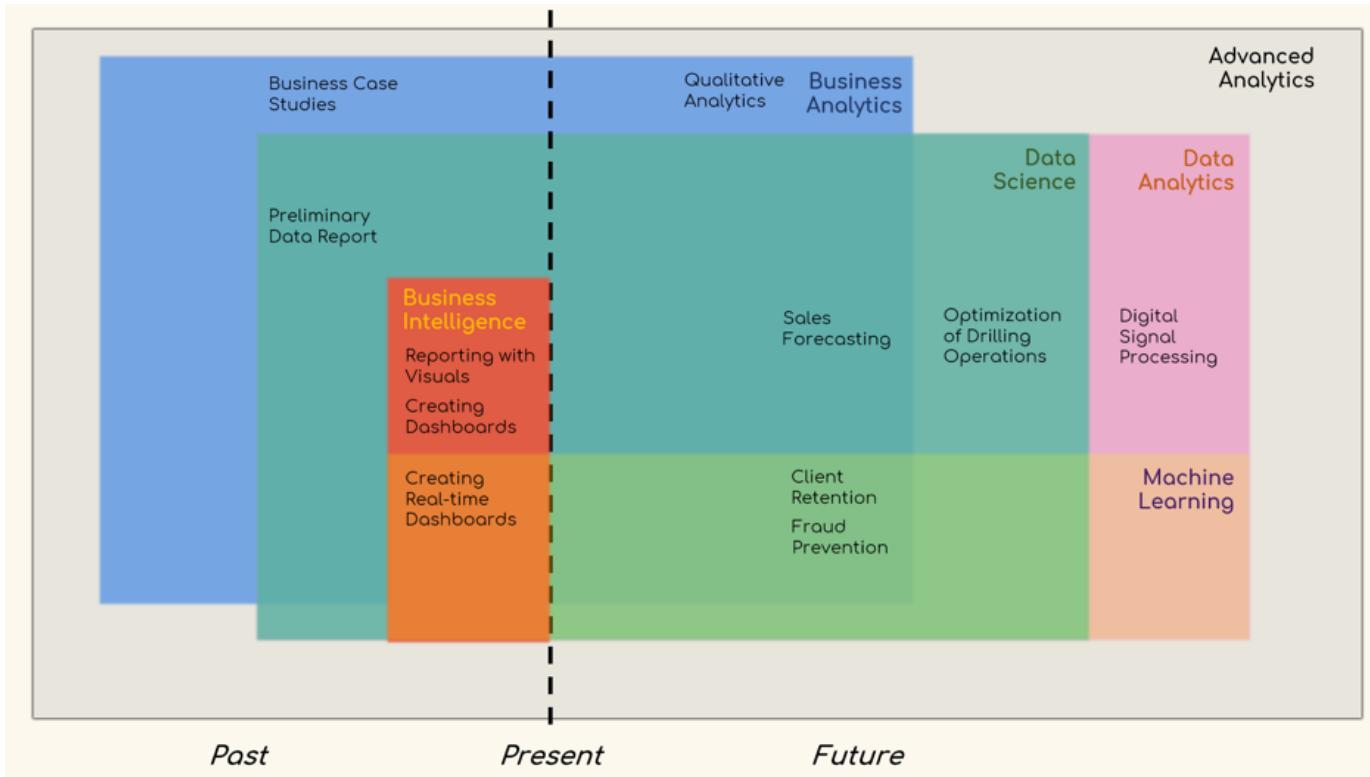
1. 조사(Inspecting)
2. 정제(Cleansing)
3. 변환(Transforming)
4. 모델링(Modeling)



데이터 사이언티스트(Data Scientists):
데이터를 다루고 분석하는 사람

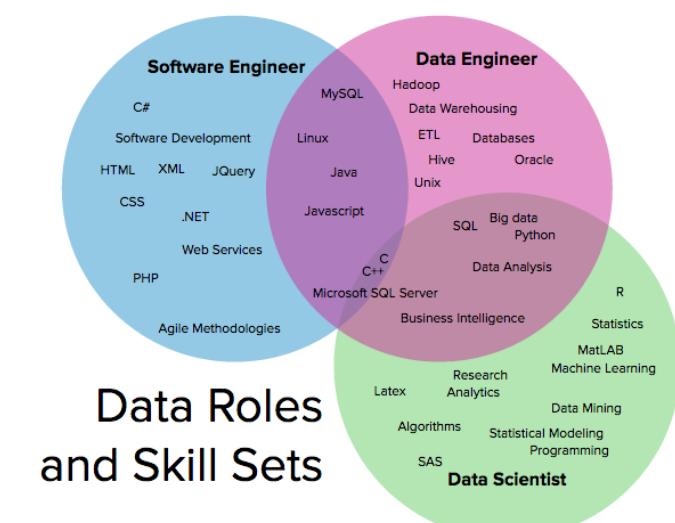
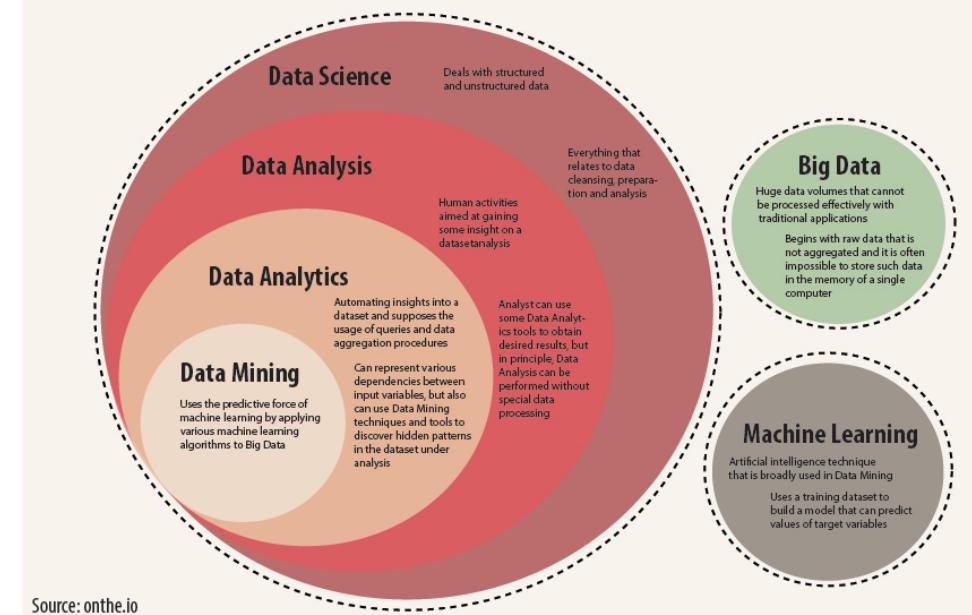
1.1 인공지능과 데이터 분석

데이터 분석관련 구분(예시)



Source : Iliya Valchanov, 'Analytics, Data Science, Business Intelligence, Machine Learning, and AI'
<https://www.datascience.com/blog/data-science-vs-business-intelligence-machine-learning-ai>

What is the difference between Data Science, Data Analysis, Big Data, Data Analytics, Data Mining and Machine Learning?



Source : Difference Between Data Scientists, Data Engineers, and Software Engineers - According To LinkedIn
<https://www.analytikus.com>

1.2 데이터 분석 적용사례



| 데이터 분석의 적용 사례 (일반)

1. 헬스케어(Healthcare)

- 1) Supporting diagnosis by detecting variations in patient data
- 2) Early identification of potential pandemics
- 3) Imaging diagnostics

2. 자동차(Automotive)

- 1) Autonomous fleets for ride sharing
- 2) Semi-autonomous features such as driver assist
- 3) Engine monitoring and predictive, autonomous maintenance

3. 금융서비스(Financial services)

- 1) Personalized financial planning
- 2) Fraud detection and anti-money laundering
- 3) Automation of customer operations

4. 운송 및 유통(Transportation and logistics)

- 1) Autonomous trucking and delivery
- 2) Traffic control and reduced congestion
- 3) Enhanced security

5. 테크놀로지, 미디어, 통신(Technology, media, and telecommunications)

- 1) Media archiving, search, and recommendations
- 2) Customized content creation
- 3) Personalized marketing and advertising

6. 소매 및 소비자 판매(Retail and consumer)

- 1) Personalized design and production
- 2) Anticipating customer demand
- 3) Inventory and delivery management

7. 에너지(Energy)

- 1) Smart metering
- 2) More efficient grid operation and storage
- 3) Predictive infrastructure maintenance

8. 제조(Manufacturing)

- 1) Enhance monitoring and auto-correction of processes
- 2) Supply chain and production optimization
- 3) On-demand production

| 데이터 분석의 적용 사례 (Manufacturing : 제조)

- 1) 프로세스 모니터링 및 자동수정 기능 향상 (Enhance monitoring and auto-correction of processes)
- 2) 공급망 및 생산 최적화(Supply chain and production optimization)
- 3) 주문형 생산(On-demand production)

Top 8 Data Science Use cases

- 예측 분석(Predictive analytics)
- 고장 예측 및 예방 정비(Fault prediction and preventive maintenance)
- 수요 예측 및 재고 관리(Demand forecasting and inventory management)
- 가격 최적화(Price optimization)
- 보증 분석(Warranty analysis)
- 로봇화(Robotization)
- 제품 개발(Product development)
- 공급망 위험 관리(Managing supply chain risk)
- 컴퓨터 비전 어플리케이션(Computer vision applications)
 - 고품질 제어 개선 (Improved high-quality control)
 - 인건비 감소(Decrease in labor cost)
 - 고속처리기능(High-speed processing capability)
 - 지속적 운영(Continuous operability)

Source : <https://activewizards.com/blog/top-8-data-science-use-cases-in-manufacturing/>

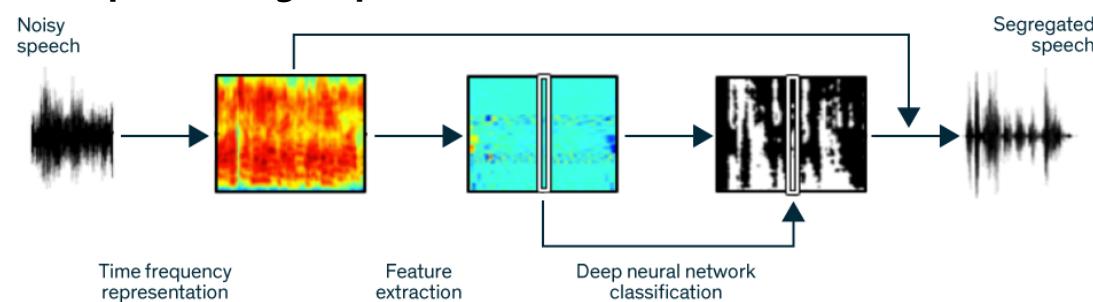
| 데이터 분석의 적용 사례(일반)

강화학습: 자율주행 자동차
(Reinforcement Learning : Self driving car)



Source : SAE

딥러닝: 음성인식
(Deep Learning : Speech detection)



Source : <https://news.developer.nvidia.com/reinventing-the-hearing-aid-with-deep-learning/>

비전 : 얼굴 인식
(Computer vision : facial recognition)



Source : https://www.washingtonpost.com/news/world/wp/2018/01/07/feature/in-china-facial-recognition-is-sharp-end-of-a-drive-for-total-surveillance/?noredirect=on&utm_term=.6d7bb68f9e5d

A CCTV display using the facial-recognition system Face in Beijing.

1.3 데이터 분석 프로젝트 프로세스



| 데이터 분석의 프로젝트 프로세스

데이터 전략(Data Strategies) : 양적 및 질적 분석(quantitative and qualitative analyses)

전략1(Strategy1) : 데이터 시각화(Visualizing the Data)

→ 리포팅(reporting), 발표(presentation of findings)

전략2(Strategy2) : 탐색적 분석(Exploratory Analyses)

→ 미래의 분석에 대한 기본 베이스라인(baseline) 활용

전략3(Strategy3) : 트랜드 분석(Trend Analysis)

→ 문제 정의와 해석을 바탕으로 해당 산업을 바탕으로 데이터 분석 모델 적용

전략4(Strategy4) : 평가(Estimation)

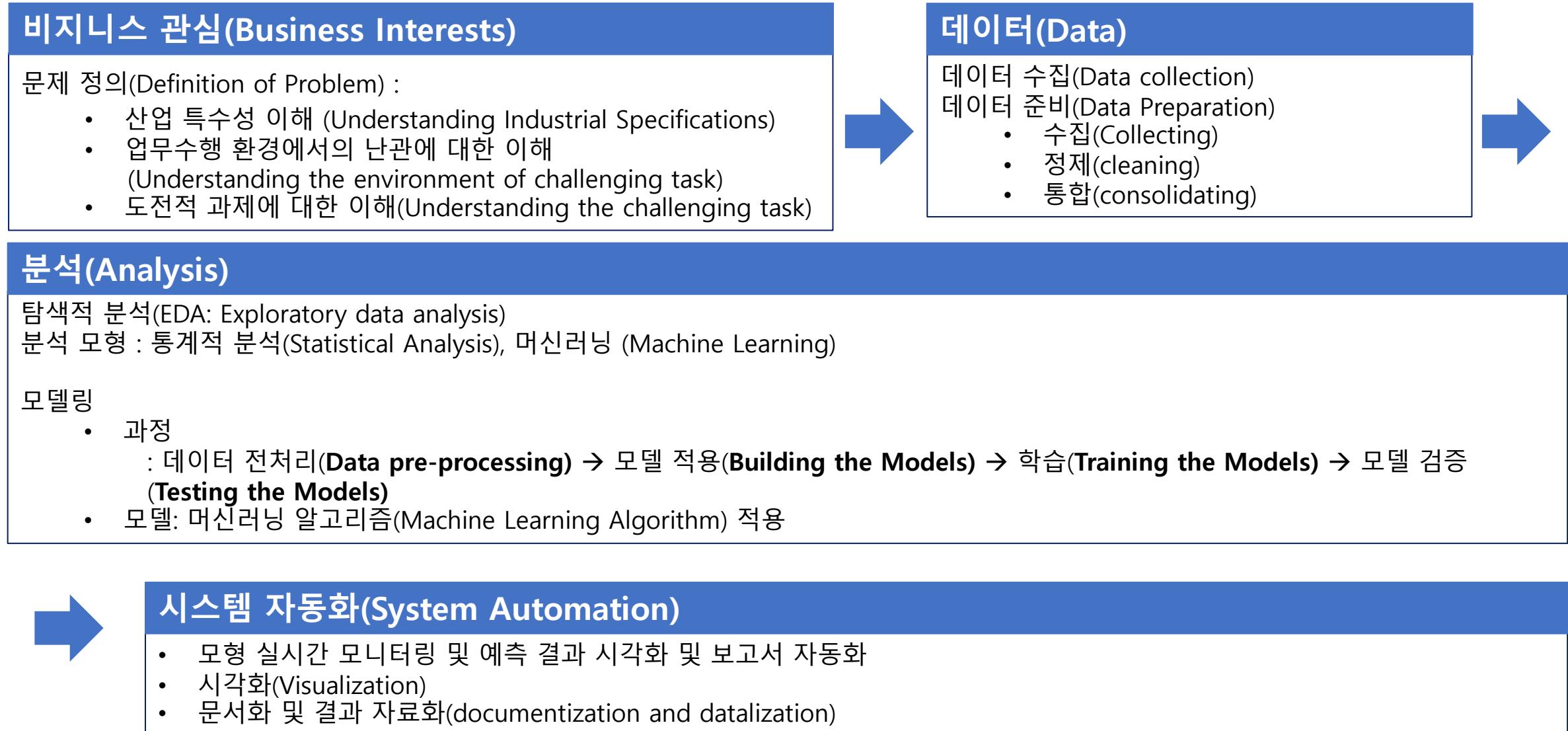
→ 예측 모형의 경우 미래에 대한 결과 확인 및 향후 조정

비지니스 문제와 데이터 분석(The Business Problem with Data Analysis)

목적(Purpose) → 질문(Questions) → 데이터 수집(Data Collection) → 데이터 분석 절차 및 방법(Data Analysis Procedures and Methods) → 결론 도출 및 해석(Interpretation/Identification of Findings) → 보고서 작성 및 향후 분석 참고 자료 활용(Writing, Reporting, and Dissemination and Evaluation)

| 데이터 분석의 프로젝트 프로세스

■ 데이터 분석 프로세스(Data Analysis Process)



| 데이터 분석의 프로젝트 프로세스

데이터 분석팀 구성(예시)

