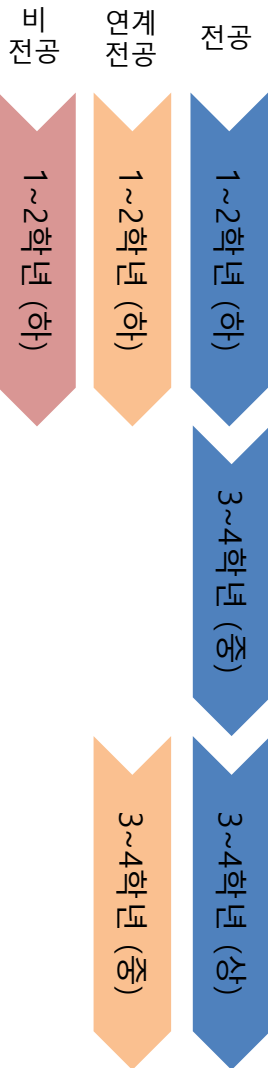


기계학습

데이터 이해와 데이터 시각화

AI 리터러시 교육 모델



1단계: 데이터 이해 및 데이터 시각화 교육 (No-code AI)

*데이터 리터러시: 데이터를 목적에 맞게 활용하는 데이터 해석 능력

2단계: 툴을 활용한 데이터 분석 기초 (No-code AI)

*프로그램이 아닌 툴을 이용한 데이터 분석

3단계: 파이썬을 활용한 데이터 분석 기초 (Code AI)

4단계: 머신러닝 기초 (Code AI)

*머신러닝 프로세스 & 기본 알고리즘 이해 / 공개 라이브러리 사용

5단계: 공공데이터를 활용하여 실 생활 문제를 해결하기 (Code/No-code AI)

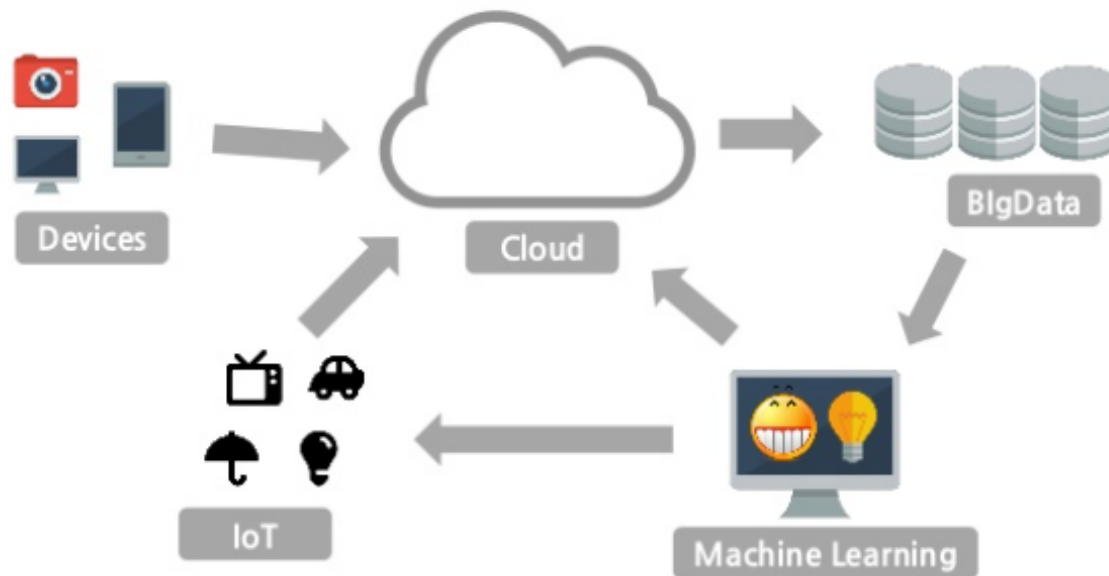
6단계: 전공 기반 데이터를 이용하여 **분야별** 문제 해결하기 (Code/No-code AI)

데이터 이해 및 데이터 시각화 교육

빅데이터 출현의 배경

■ 기술의 하모니

- 어디서 발생하는 데이터라도 쏘아 올릴 수 있는 자유롭고 **빠른 통신**
- 그렇게 발생한 많은 데이터를 받아들일 수 있는 충분한 **클라우드**
- 더 많은 곳에서, 데이터가 나오지 않던 곳에서 데이터가 나오게 하는 **IoT**
- 쌓여서 거대해진 데이터에서 숨은 패턴을 찾아내려 하는 **데이터 분석**
- 지능을 모사해서 데이터 분석의 능력과 수준을 끌어 올리기 위한 **인공지능**



빅데이터 출현의 배경

- Everything is connected

- 모두가 연결된 세상 → 데이터 축적 → 데이터 기반 의사결정의 시발점
- IoT 장비의 활성화가 결국 데이터 양의 대 폭발을 가져 온 것



facebook

YouTube

Instagram

빅데이터를 통한 세상의 변화

- 패러다임의 변화

기술 → 가격 → 기회
Data



데이터 분석을 통한
가치의 재 생산

데이터 분석

넘치는 데이터. 어떻게 감당할 것인가?



데이터 분석



Machine Learning

산업혁명 - **물건** 생산을 위해 **육체노동**을 기계로 자동화하자

머신러닝 - **정보** 추출을 위해 **정신노동**을 기계로 자동화하자
(특히 deep learning)

데이터 어떻게 생겼나?

- 배추 값 예측을 위한 데이터

- 연월일, 평균 기온, 최저 기온, 최대 기온, 강우량, 배추평균가격
 - [연월일]은 계절을 간접적으로 포함하는 정보

총 2922개 항목 중 1~10 필터					
year	avgTemp	minTemp	maxTemp	rainFall	avgPrice
20100101	-4.9	-11	0.9	0	2123
20100102	-3.1	-5.5	5.5	0.8	2123
20100103	-2.9	-6.9	1.4	0	2123
20100104	-1.8	-5.1	2.2	5.9	2020
20100105	-5.2	-8.7	-1.8	0.7	2060
20100106	-7.3	-11.4	-2.5	0.3	2060
20100107	-6.7	-11.2	-1.2	0	2140
20100108	-5.6	-11.4	1.4	0	2140
20100109	-3.1	-8.8	1.8	0.1	2140
20100110	-1.3	-5	2.8	0	2140

데이터 어떻게 생겼나?

- 남녀 분류를 위한 신체 데이터
 - 성별, 신장, 무게, 발의 크기
 - 신체적 조건은 평균적으로 남녀가 다르다는 가정

성별	신장 (feet)	무게 (lbs)	발의 크기(inches)
남성	6	180	12
남성	5.92 (5'11")	190	11
남성	5.58 (5'7")	170	12
남성	5.92 (5'11")	165	10
여성	5	100	6
여성	5.5 (5'6")	150	8
여성	5.42 (5'5")	130	7
여성	5.75 (5'9")	150	9

데이터 어떻게 생겼나?

■ 통신사 고객 이탈 예측

```
In [2]: # Show the features (=columns) and first data entries of the data frame.  
df.head(5)
```

Out[2]:

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineS
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No

- customerID — 고객 ID
- 성별 — 고객이 남성이든 여성이든
- SeniorCitizen — 고객이 노인인지 여부 (1, 0)
- 파트너 — 고객에게 파트너가 있는지 여부 (예, 아니요)
- 부양 가족 — 고객에게 부양 가족이 있는지 여부 (예, 아니요)

- PhoneService — 고객에게 전화 서비스가 있는지 여부 (예, 아니요)
- MultipleLines — 고객에게 여러 회선이 있는지 여부 (예, 아니요, 전화 서비스 없음)
- InternetService — 고객의 인터넷 서비스 제공 업체 (DSL, 광섬유, 아니요)
- OnlineSecurity — 고객의 온라인 보안 여부 (예, 아니요, 인터넷 서비스 없음)
- OnlineBackup — 고객의 온라인 백업 여부 (예, 아니요, 인터넷 서비스 없음)
- DeviceProtection — 고객의 기기 보호 여부 (예, 아니요, 인터넷 서비스 없음)
- TechSupport — 고객의 기술 지원 여부 (예, 아니요, 인터넷 서비스 없음)
- StreamingTV — 고객이 스트리밍 TV를 가지고 있는지 여부 (예, 아니요, 인터넷 서비스 없음)
- StreamingMovies — 고객이 스트리밍 영화를 가지고 있는지 여부 (예, 아니요, 인터넷 서비스 없음)

데이터 이해 및 데이터 시각화 교육

데이터 시각화의 중요성

■ 데이터 시각화 정의

- 광범위하게 분산된 방대한 양의 자료를 분석해 한눈에 볼 수 있도록 도표나 차트 등으로 정리하는 것



■ 데이터 시각화 효과

- 자료로부터 정보를 습득하는 시간 절감으로 즉각적인 상황 판단이 가능
- 자료를 습득하는 사람들의 흥미를 유발하고 정보의 빠른 확산을 촉진
- 자료를 기억하는데 도움

■ 데이터 시각화 목적

- 그래픽 의미를 이용해 명확하고 효과적으로 정보를 커뮤니케이션 하기 위함

데이터 시각화의 중요성

■ 데이터

- 데이터는 연구나 조사, 발견, 수집의 결과인 일정의 기초 자료로서, **정보를 만들기 위한 원자재와 같은 것**
- 데이터는 정보 자체가 아니므로 정보로서의 가치가 부족하여 분석의 대상임

■ 정보

- 데이터와 달리 그 자체만으로 의미가 있음
- 정보는 서로 다른 데이터 간의 일정한 패턴을 가시화하여 정보를 보는 사람에게 **데이터가 내포하는 의미를 전달**
- 데이터가 정보로서 가치를 갖기 위해서는 조직화 되고 변형되어야 하며, 의미를 전달하기 위한 형태로 표현되어야 함

■ 인포그래픽

- 중요한 정보를 한장의 그래픽으로 표현해 이를 보는 사람들이 손쉽게 해당 정보를 이해할 수 있도록 만드는 그래픽 메세지

데이터 시각화의 영역

- 데이터시각화는 데이터를 기반으로 객관적 표현에 더 초점을 맞춤
 - 데이터시각화는 정보형 메시지를 전달하기 위한 작업
 - 인포그래픽은 해석된 의미-설득형 메시지 전달하기 위한 작업



데이터 시각화 프로세스

■ 엑셀

- 누구나 쉽게 그래프를 만들 수 있음 (일반적인 차트 가능)
- 심도 있는 분석이나 그래픽이 필요할 땐 사용하지 않음



데이터 시각화 프로세스

■ 구글 차트

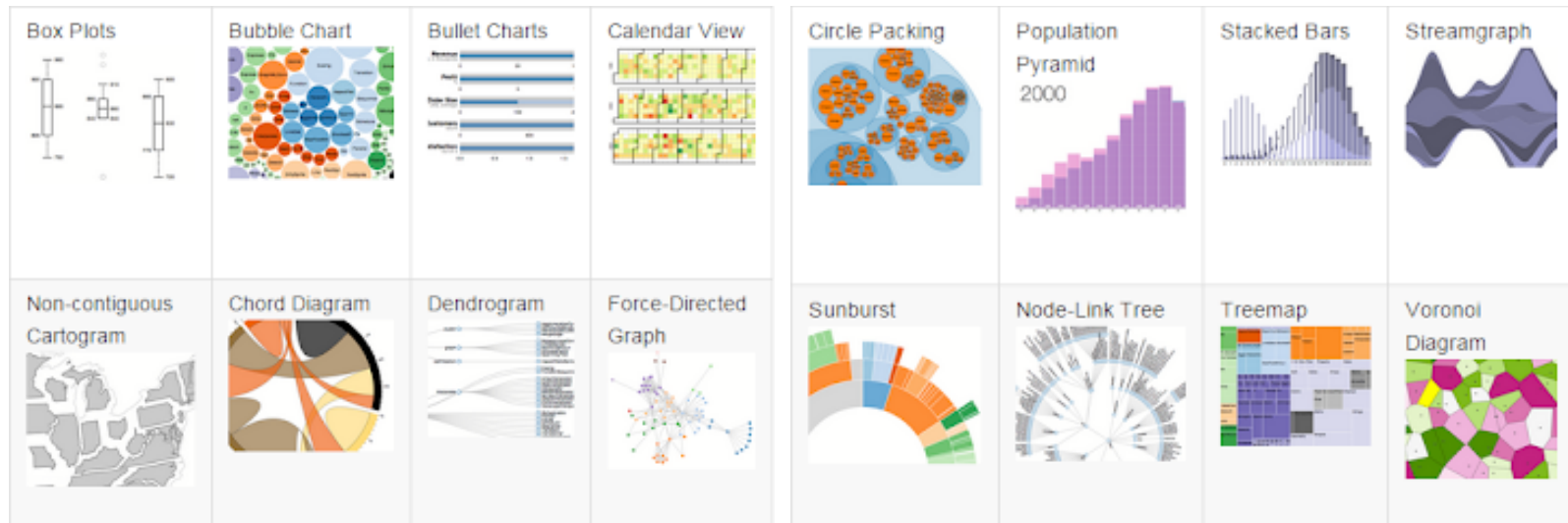
- 속도가 빠르며, 모바일에서도 잘 동작함
- 잘 모르는 초보자를 위한 위자드 툴도 제공
- 스프레드시트와 차트API 함께 사용 가능



데이터 시각화 프로세스

■ D3.js

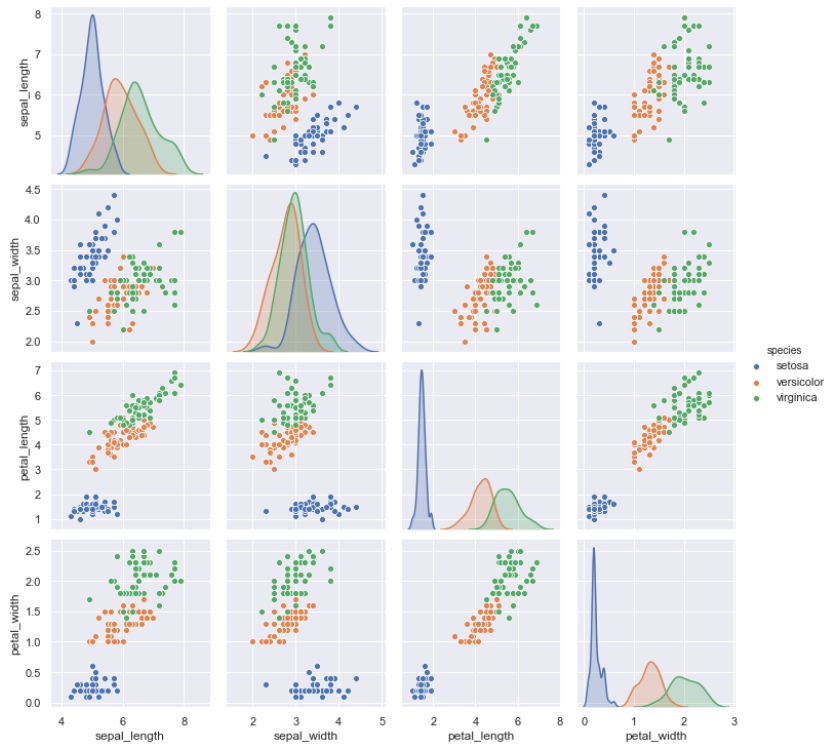
- 데이터 시각화 프레임워크, 자바스크립트로 개발
- HTML, SVG 그리고 CSS를 이용하여 시각화 요소를 만들어 냄



데이터 시각화 프로세스

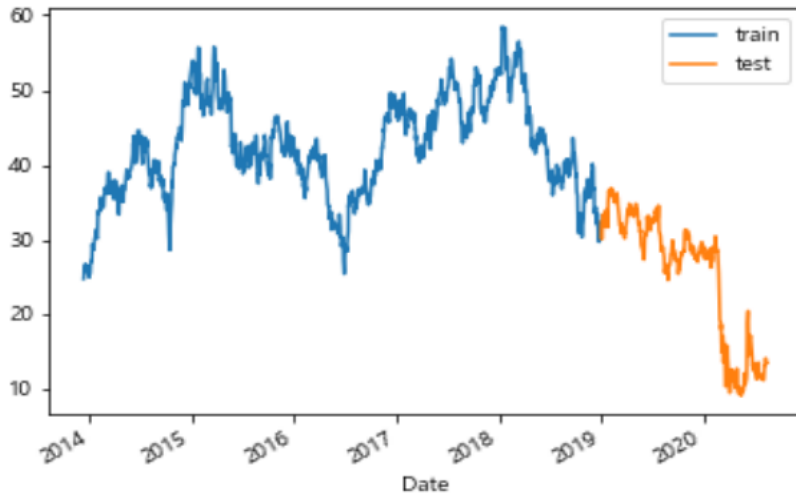
■ 파이썬

- 대규모 데이터를 무리없이 다룰 수 있어 큰 규모의 계산에 쓰임
- 적은 코드만으로 많은 기능을 만듦
- 다양한 라이브러리를 이용할 수 있음
- 미학적으로는 훌륭하지 않지만 데이터 탐색 단계에서 활용하기 좋음

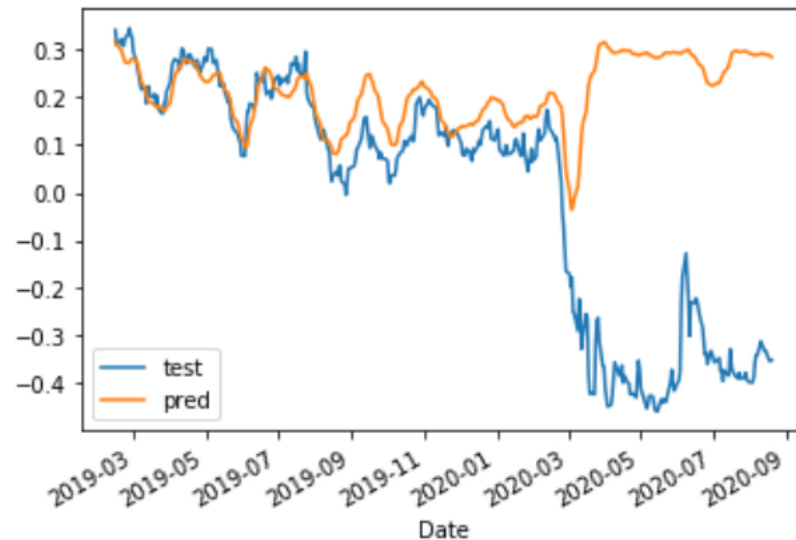


기계학습: 주가 예측 모델 시각화

■ 파이썬 예시



훈련용 데이터



코로나 이후 예측 범주를 벗어나는
주가 예측 곡선

참고자료

- 하용호, “빅데이터, 클라우드, IoT, 머신러닝. 왜 이렇게 많은 것들이 나타날까?”
- 토닥토닥 파이썬 - 머신 러닝 추가문제
- 핸즈온-머신러닝: 파이썬 - 예측모델(LSTM 모델 추가예측)