

기초빅데이터프로그래밍



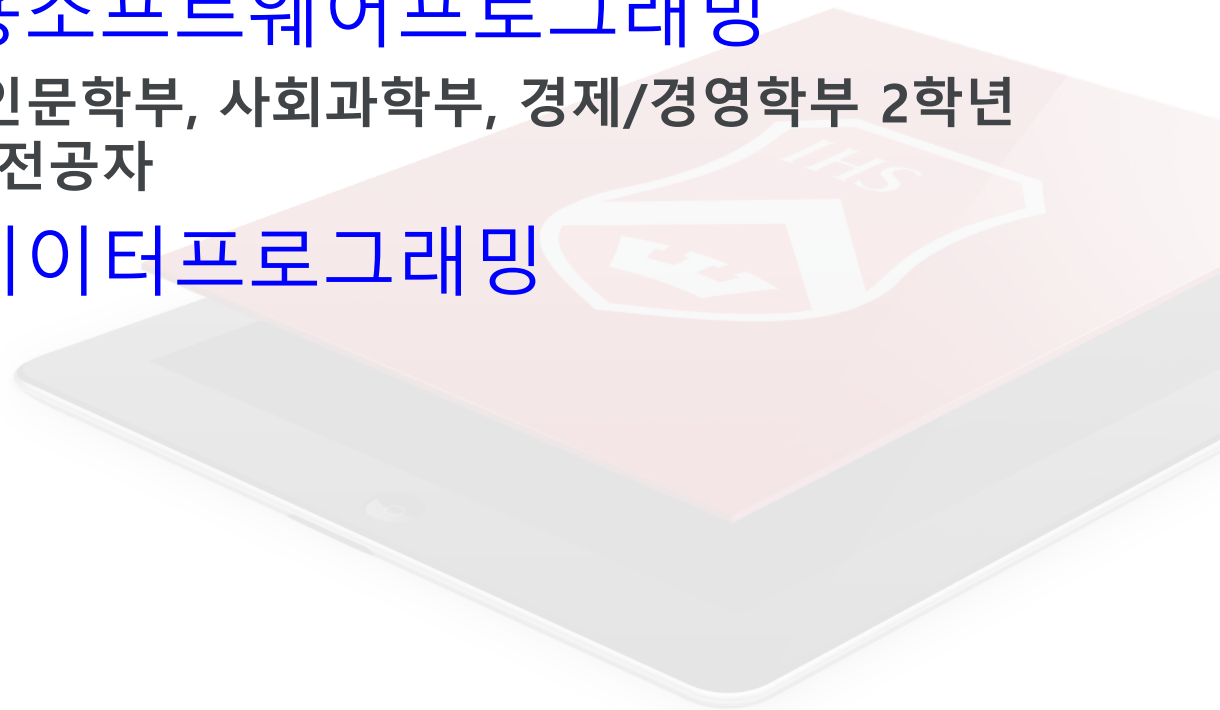
과목번호 STS2011

➤ 과목명 history

- 2017 고급응용소프트웨어프로그래밍
- 2018 기초응용소프트웨어프로그래밍
 - 수강대상: 국제인문학부, 사회과학부, 경제/경영학부 2학년
∴ 비컴퓨터공학전공자
- 2019 기초빅데이터프로그래밍

➤ 내용

- 고급 파이썬



빅데이터 사이언스 연계전공 교육 과정 (36학점)


- **전공 예비** (이수 학점에 포함되지 않음 - 거의 모든 학생이 필수로 수강, 아닐 경우 수강)
 - 컴퓨팅 사고력, 미적분학II(SS2006) 또는 대학수학 또는 경제수리기초(ECO2003) 중 택 1
- **빅데이터 기초 (9학점(3과목) - 필수)**
 - 컴퓨터 (3학점): 기초응용소프트웨어프로그래밍(STS2011) -> 기초 빅데이터 프로그래밍 과목명 변경
 - 수학 (3학점) : 선형대수학(MAT2110)
 - 통계학 (3학점): 통계학입문(MAT3020), 경영통계학(MGT2002), 경제통계학(ECO2004) 중 택 1
- **빅 데이터 심화 (24학점(8과목) - 선택 : 각 트랙에서 최소 2과목 이상 선택)**
 - 데이터 엔지니어 트랙 (9과목):
 - IT (7과목): 기초 C 언어(CSW2010), 기초 JAVA 언어(CSW2020), 자료구조 입문(CSW2030), 알고리즘 입문(CSW3010), 데이터베이스 입문(CSW3030), **빅데이터 컴퓨팅(신규)**, **머신러닝의기초 및 프로그래밍(AI연계전공과 공유)** - 컴퓨터공학과에서 동일(유사) 과목 수강 시 인정
 - 수학 (2과목): 산업수학 종합설계(MAT4331), 이산수학(MAT3110)
 - 데이터 분석 트랙 (11과목)
 - 경영학 (4과목): 통계자료분석(MGT4202), 응용경영통계(MGT4207), 빅데이터와 비즈니스 인텔리전스(MGT4226), **웹데이터 분석과 응용(AI연계전공과 공유)**
 - 경제학 (4과목): 계량경제학I(ECO2009), 경제수리통계학(ECO4004), 금융계량경제학(ECO4003), 금융시장행태주의(ECO4034)
 - 융합SW(인문컨텐츠 융합) (3과목): 코퍼스언어학(CSW2050), 빅데이터학(CSW3060), 정보처리 및 자연언어처리(CSW4020)
- **빅데이터 종합설계 (캡스톤) (3학점(1과목) - 필수)**
 - 빅데이터 종합설계(캡스톤)(신규)

과목개요

- 본 과목은 고급 파이썬 과정으로서, 문제 해결을 위해 **기초적인 파이썬 프로그래밍이 가능한 학생들을 대상으로 객체지향프로그래밍, 표준 Library, 예외처리, 정규식 등을 포함하는 파이썬 언어의 문법을 학습하고, Anaconda 개발환경에서 실습과 프로젝트를 통해 고급 파이썬 언어 활용능력을** 배양한다.
- 또한, Matplotlib, NumPy, Pandas 등과 같은 데이터 과학 영역에서 많이 사용되는 패키지들을 학습한다.

소프트웨어란?

소프트웨어란
무엇인가 ?



```
g[] sett
settings[0].compare
if (name.compareTo("") != 0) {
    name += " ";
}
name += etr.getString(settings[1]);
else if (settings[0].compareTo("d") == 0) {
    if (name.compareTo("") != 0) {
        name += " ";
    }
    name += DateUtils.format(etr.getDate(settings[1])
} else if (settings[0].compareTo("n") == 0) {
    if (name.compareTo("") != 0) {
        name += " ";
    }
    comSysNumber = etr.getDouble
    f = NumberFormat.g
    d(false);
    sysNum
```

상상을 현실로 만들어 주는 소프트웨어

소프트웨어의 정의

- **컴퓨터 Hardware** = 계산과 논리연산(비교)만 수행하는 단순한 기계
- **알고리즘** = 문제를 계산과 비교의 순차적 수행을 통하여 해결하는 방법
- **컴퓨터 프로그램** = 알고리즘을 컴퓨터가 이해할 수 있는 언어로 표현(코딩)한 것
- **소프트웨어** = 컴퓨터 프로그램 + 데이터 + 관련 문서

소프트웨어의 본질

- 문제를 해결하는 **솔루션**
- 지식과 경험을 코딩한 **지식재산**

왜 소프트웨어가 강력한가?

- **점진적 개선** 가능
- **누적하여 성능 추가** 가능
- 다수를 **집합하여 문제 풀이** 가능

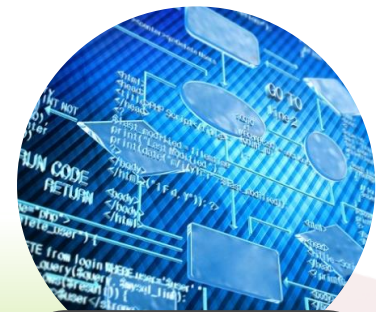
무엇이 소프트웨어 혁명을 가능하게 했는가?



고성능 컴퓨터 칩



초고속 네트워크

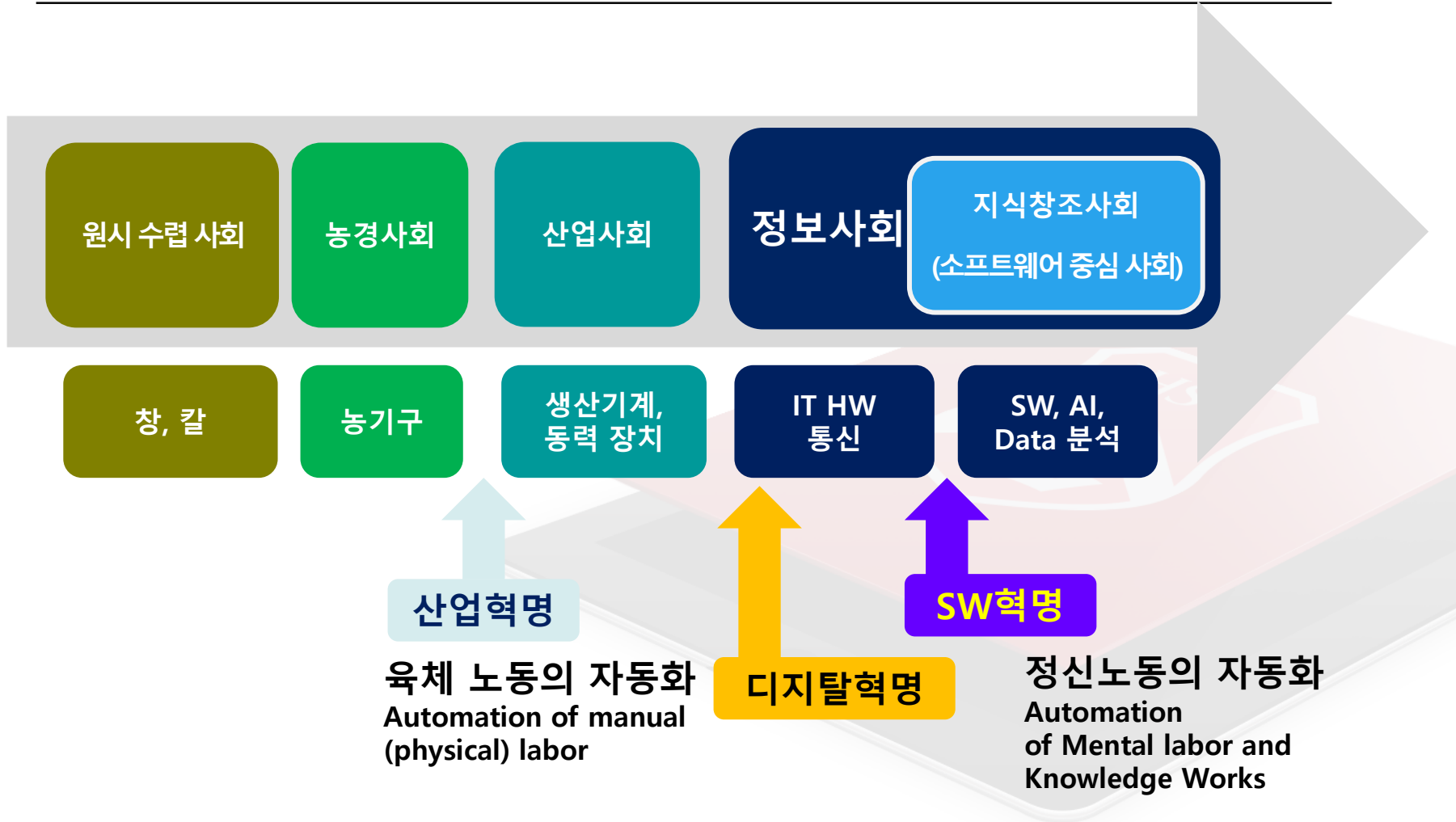


SW·인공지능

이 세가지 기술의 결합인 디지털기술이
인류 역사 이래 최고의 메타기술
(Meta-Technology: 기술혁신을 일으키는 기술)

* 증기기관, 전기와 같이 기술혁신과 경제발전을 가속화함으로써 사회·경제 변혁의 동인이 되는 기술

소프트웨어 중심사회의 도래



MWC 2017, MARS

MWC에 전시된 최신 IT 기술

혼합현실(MR)

SK텔레콤 텔레프레존스

중국 스타트업 비저너테크

삼성전자 C랩 모니터러스



'인텔' 부스에 마련된 MR가인 홀로렌즈로 전시관을 바라본 모습

인공지능(AI)

LG전자 G5, 노키아3·5의 구글 어시스턴트

IBM 왓슨

삼성SDS의 인공지능 매장관리 비서

화장품 브랜드 올레이의 인공지능 기반 앱



IBM 인공지능 '왓슨'의 조언으로 완성된 패션 브랜드 '마르체사'의 드레스

로봇(Robot)

일본 소프트뱅크의 사람을 닮은 로봇 페퍼

소니 인공지능 비서 로봇

엑스페리아 에이전트

중국 스타트업 포캅엑스 가정용 비서 로봇



영국 '우분투' 부스에 전시된 로봇

5G(5세대 이동통신)

SK텔레콤·KT·NTT도코모(일본)·텔레포니카(스페인)·우레두(카타르) 등 글로벌 통신업체의 5G 통신망 구축

중국 ZTE의 세계 최초 5G 기반 스마트폰



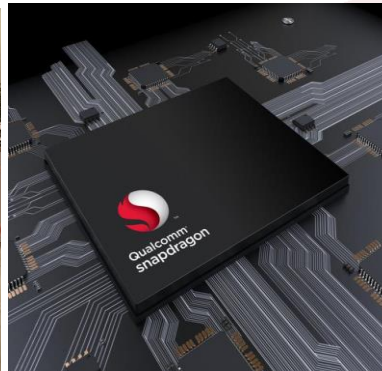
KT가 5G 기술을 토대로 전시한 스카이다이빙 가상현실 체험 장비

MWC 2018

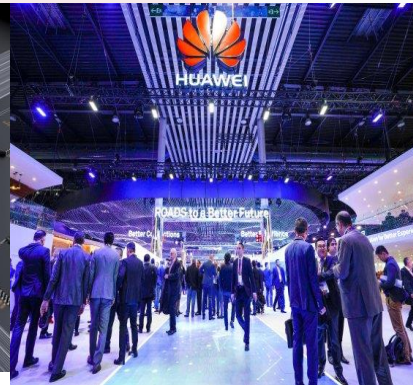
- Mobile World Congress 2018의 화두: 5G 이동통신
- 현재 4세대(4G) 이동통신보다 20배 정도 속도가 빨라지기 때문에 끊임없이 고품질 데이터를 주고 받아야 하는 자율주행차, 스마트시티 등의 서비스에 반드시 필요.
- 자율주행차, 스마트폰 등 5G 기기의 두뇌를 만드는 칩셋 제조사,
- 5G 이동통신망을 구축해야 하는 이동통신사,
- 5G 시대 허브 역할을 할 스마트폰을 만들 제조사



Seat: Leon Cristobal



퀄컴 스냅드래곤 x50



화웨이: 5G 통신장비



LG전자 V30싱큐

MWC 2019

GSMA Innovation City 2019 Partners

Together with the GSMA, companies in the Innovation City showcased a spectacular array of innovative, cutting edge products and services that demonstrate how mobile is creating a better future.

Google

 HUAWEI

kt

 SIERRA
WIRELESS

lifecell  TURKCELL



Pavilions

Zones and pavilions representing all sectors of the mobile industry



Innovation City

Presenting the fusion of 5G, AI and IoT



MWC 2020


코로나19 확산 우려 ... 취소



24-27 February 2020



MWC 2021



[Registration](#)
[Login](#)
[Press Zone](#)
[Exhibitor](#)

[ATTEND](#)
[AGENDA](#)
[EXHIBITORS](#)
[MEET & NETWORK](#)

#MWC21

Agenda

From the most inspiring speakers, to participating in debates about the hottest topics in technology – find out how to plan and make the most of your MWC experience.

FILTER BY TOPICS

[4YFN](#)
[Conference](#)

[GSMA Ministerial Programme](#)

[Partner Programmes](#)

[Specialist Workshops](#)

[Virtual Exhibition Tours](#)

FILTER BY SESSION TYPE


[Virtual](#)

[Physical](#)

Learn straight from the brands

PRE Event	MON Jun 28	TUE Jun 29	WED Jun 30	THU Jul 1
--------------	---------------	---------------	---------------	--------------

08:00 (CEST)




PARTNER PROGRAMMES · 08:00 - 09:00 [VIRTUAL](#)

Getting the Edge on Revenue Opportunities

[HALL VIRTUAL](#)
[MY MWC ONLINE](#)

09:00 (CEST)



VIRTUAL EXHIBITION TOURS · 09:00 - 10:00 [VIRTUAL](#)

5G Tours - Live Streaming

[HALL VIRTUAL](#)
[MY MWC ONLINE](#)

MWC 2022

MWC Barcelona 2022 Themes



5G Connect

[LEARN MORE](#)

Event Theme Partner



Advancing AI

[LEARN MORE](#)



CloudNet

[LEARN MORE](#)



FinTech

[LEARN MORE](#)



**Internet of
Everything**

[LEARN MORE](#)



Tech Horizon

[LEARN MORE](#)



Data 에 길이 있다

Big Data (빅데이터)

자율주행 자동차

IoT (사물인터넷)

인공지능

Retailer(인터넷 상거래)



데이터의 정의

- 데이터(data) : 현실 세계에서 측정된 객관적 사실
- 정보(information) : 데이터를 목적에 맞게 가공해서 의미와 가치 부여

데이터의 예	정보의 예
<ul style="list-style-type: none">- 오늘은 떡볶이를 30개 팔았다- 오늘은 하루 종일 비가 내렸다	<ul style="list-style-type: none">- 떡볶이 매출이 매 월 2%씩 상승- 하루 종일 비가 내리는 날은 떡볶이 구매 고객이 평소보다 15% 감소

빅데이터

빅 데이터는 통상적으로 사용되는 데이터 수집, 관리 및 처리 소프트웨어의 수용 한계를 넘어서는 크기의 데이터를 말한다. 빅 데이터의 사이즈는 단일 데이터 집합의 크기가 수십 테라바이트에서 수 페타바이트에 이르며, 그 크기가 끊임없이 변화하는 것이 특징이다.

- 대량의 복잡한 데이터 집합



기존 데이터베이스 관리도구로 데이터를 수집, 저장, 관리, 분석할 수 있는 역량을 넘어서는 대량의 정형 또는 비정형 데이터 집합 및 이러한 데이터로부터 가치를 추출하고 결과를 분석하는 기술

정형? 반정형? 비정형? 데이터

정형 데이터(Structured Data)

- 형태(고정된 필드)가 있으며, 연산이 가능, 주로 관계형 데이터 베이스
- ex) 회사 매출 이나 직원의 데이터(excel 파일)

반정형 데이터(Semi-Structured Data)

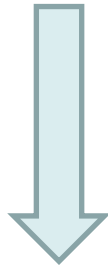
- 형태가 있지만 연산이 불가능
- ex) XML, HTML 등의 웹 문서

비정형 데이터(Unstructured Data)

- 구조가 일정하지 않은 데이터, 규격화 되지 않음
- ex) 소셜데이터(트위터, 페이스북), 영상, 이미지, 음성, 텍스트 등

빅데이터의 출현 원인

1. CPU, 스토리지, 메모리 등 하드웨어의 가격 인하와 클라우드 사용
보편화(정보의 양은 3개월에 두 배씩 늘어나는 것으로 추정)



2. Hadoop, tensorflow 및 R 과 같은 오픈 소스의 기술 발전

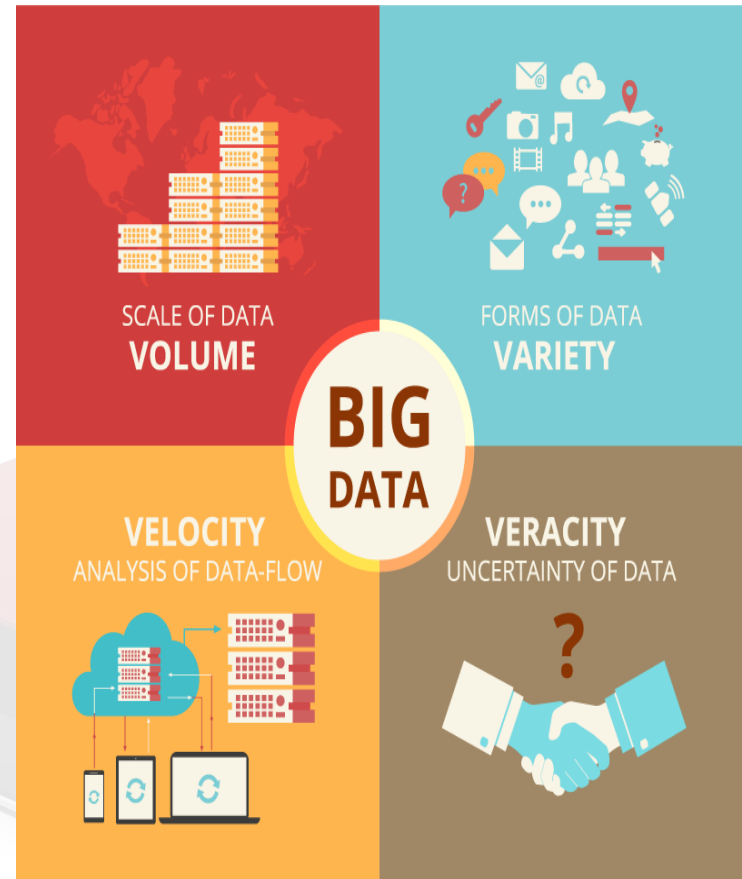
3. 기존에 존재했지만 관리되지 않고 버려지는 데이터를
활용하여 비즈니스적 가치를 찾으려는 노력이 있었기 때문.

빅데이터의 특징

빅데이터의 3V

3V는 데이터의 크기(**Volume**), 데이터의 속도(**Velocity**), 데이터의 다양성(**Variety**)을 나타내며 이러한 세 가지 요소의 측면에서 기존의 데이터베이스와 차별화된다.

추가적으로 가치(**Value**)와 진실성(**Veracity**), 시각화(**Visualization**), 가변성(**Variability**) 등이 있다.



빅데이터 이용사례



사용자에 따라 다른 추천항목

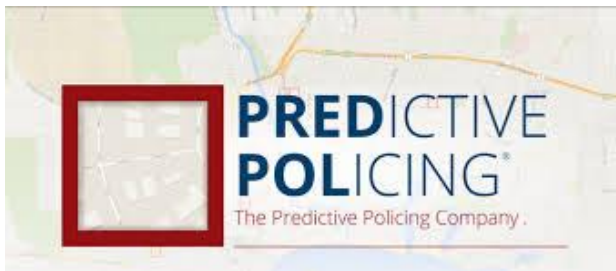


구매패턴 파악을 통한 이익극대화 방식



NETFLIX

시청자의 성향과 선호도를 수집
하고 분석하여 고객에게 개인 맞춤형
컨텐츠를 제공



Amazon의 Cloud service

- 미국의 뉴욕타임스(NYT)는 온라인 신문보기(PDF) 서비스 구축을 위해 그동안 연간 8만 달러를 사용했지만, 2007년 클라우드 컴퓨팅을 도입한 후 관리비용이 연간 840달러로 줄어들면서 **약 1/100**의 획기적인 비용절감 효과를 거뒀다.
- NYT는 **1851년부터 1980년까지** 1100만개에 이르는 TIFF의 파일형태 신문 기사를 **PDF로 전환하는 작업을** 클라우드 컴퓨팅을 활용, 단 며칠 만에 1000달러의 비용을 들여 해결했다고 밝혔다.
 - 자체 서버 이용시 14년 소요예상, 아마존의 클라우드 서비스 이용하여 24시간 내에 240달러로 완료

수업내용: 파이썬

- 객체지향 프로그래밍
- 파이썬 클래스
- 정규식
- 예외처리
- 파이썬 라이브러리를 활용한 데이터 처리
 - NumPy, matplotlib, pandas등 이용

파이썬 라이브러리

- **Numpy** : MATLAB과 유사, matrix연산을 쉽게 해 줌
 - **Pandas** : R의 dataframe과 유사하게 구현되어 있어 데이터를 손쉽게 처리할 수 있음
 - 이러한 라이브러리는 모두 Numpy를 기반으로 돼 있기 때문에 이를 사용하는 다른 라이브러리와 호환이 잘되며, 기능 확장을 하거나 통합하는데 문제가 되지 않는다.
 - **Matplotlib** :데이터 시각화 도구, 무료
- ❖ **Anaconda** : 기존의 파이썬 + matplotlib + 수학, 과학을 위한 라이브러리들을 묶어 배포

기계학습

- 우리가 가지고 있는 데이터를 기계에 학습시킨 후, 학습한 데이터와 유사한 데이터를 입력하면 기계는 우리가 알고 싶어하는 정보를 예측한다.
- 예: 이전에 온 스팸 이메일 정보를 근거로 새로 온 이메일이 스팸인지 아닌지 분류한다.
- 기계 학습은 기존 데이터를 사용하여 새로운 데이터의 정보를 유추할 수 있도록 한다.
- 가능하다면... 기계학습 알고리즘을 학습
 - Scikit-learn을 사용한 Machine Learning

Check Python

- `a = [1,2,3,4,5]` 일 때 아래 함수 중 실제 swap이 일어나는 함수는?

```
def swap_value (x, y):
```

```
    temp = x
```

```
    x = y
```

```
    y = temp
```

```
def swap_offset (offset_x, offset_y):
```

```
    temp = a[offset_x]
```

```
    a[offset_x] = a[offset_y]
```

```
    a[offset_y] = temp
```

```
def swap_reference (list, offset_x, offset_y):
```

```
    temp = list[offset_x]
```

```
    list[offset_x] = list[offset_y]
```

```
    list[offset_y] = temp
```

```
def swap_value (x, y):
```

```
    temp = x
```

```
    x = y
```

```
    y = temp
```

```
def swap_offset (offset_x, offset_y):
```

```
    temp = a[offset_x]
```

```
    a[offset_x] = a[offset_y]
```

```
    a[offset_y] = temp
```

```
def swap_reference (list, offset_x, offset_y):
```

```
    temp = list[offset_x]
```

```
    list[offset_x] = list[offset_y]
```

```
    list[offset_y] = temp
```

```
if __name__ == "__main__":
```

```
    a=[1,2,3,4,5]
```

```
    swap_value(a[1], a[2])
```

```
    print(a)
```

```
    swap_offset(1,2)
```

```
    print(a)
```

```
    swap_reference(a, 1, 2)
```

```
    print(a)
```



Python 3.6.2 Shell

File Edit Shell Debug O

Python 3.6.2 (v3.6.2:5

Type "copyright", "cr

>>>

[1, 2, 3, 4, 5]

[1, 3, 2, 4, 5]

[1, 2, 3, 4, 5]

>>>