

# 데이터 분석 기초

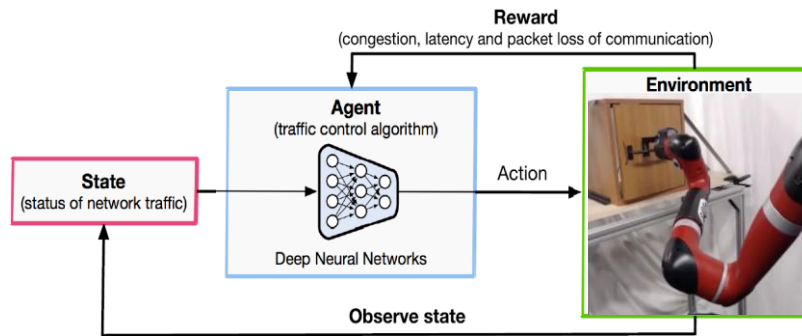


김유성 (yskim525@skku.edu)

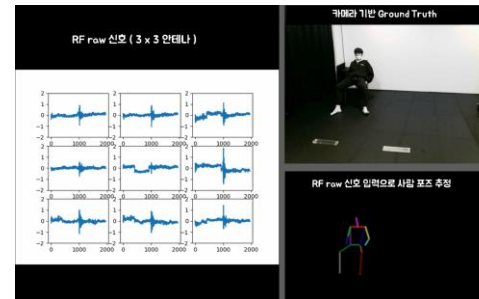
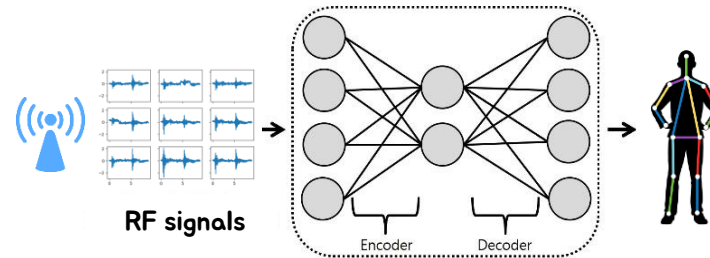
Computer System and Intelligence Lab.

<http://csi.skku.edu>

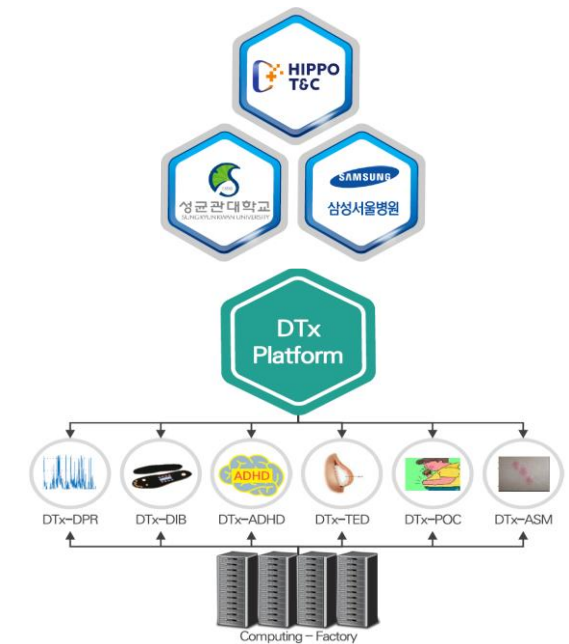
## Intelligent Systems



## RF based Learning Systems



## Digital Therapeutics

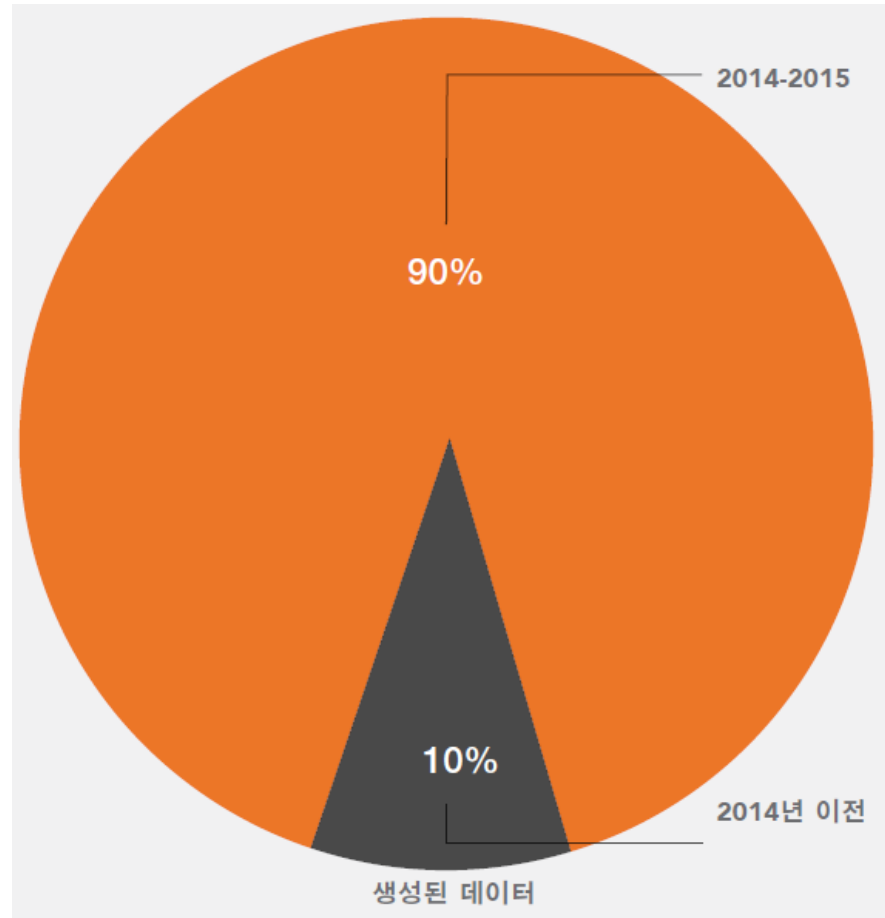


# 데이터 분석?

- 데이터를 정제 및 가공 해서 의미 있는 정보를 추출
- 추출된 정보를 중요한 의사 결정에 활용
  - ◎ 본 강의에서는 데이터 분석에 대해 전반적인 이해와
  - ◎ 실습을 통해 다양한 데이터 정제, 가공, 가시화 기술들을 배우며
  - ◎ 새로운 가치를 찾아내는 경험을 축적



# 데이터 성장



- 전 세계 데이터의 90% 가 지난 2년 동안 새롭게 생성 됨
  - 초연결 사물 인터넷 시대
  - 수천대의 배송 차량 동선 최적화
  - 병원의 디지털 정보 (영상 데이터)
  - 항공기 센서 데이터 (매일 25,000 대)
  - 금융 거래 실시간 분석 및 이상 거래 활동 감지
  - 기상 측정 인공위성, 기상 관측소 시간당 15회 (22억 5000만 개 이상의 데이터 요소)
  - 유럽입자 물리연구소, 입자충돌 실험 1초마다 40 TB

# 세상을 바꾸는 데이터

모바일 데이터를 사용하여 전염병 감염 패턴을  
파악하고 집중 방역활동 장소를 선정

범죄가 발생하기 쉬운 장소를 예측하여, CCTV 설치 및  
순찰을 강화 하여 절도 및 폭력 범죄를 감소시킴

관광객의 이동 패턴을 파악하여 공용 자전거 서비스,  
숙박, 쇼핑과 같은 시설 투자 및 마케팅에 활용

농기구 센서 정보를 사용하여 재배할 작물, 파종 시기,  
식품의 신선도 추적, 변화하는 기후에 적응하는 방법을 결정

# 커리큘럼

1주차 : 과목 소개

2주차 : 파이선 기초 - 선택문 반복문

3주차 : 파이선 기초 - 자료구조, 리스트, 사전

4주차 : 파이선 기초 - 함수

5주차 : 넘파이

6주차 : 판다스1

7주차 : 판다스2

8주차 : 중간 시험

9주차 : 가시화

10주차 : 데이터 탐색1

11주차 : 데이터 탐색2

12주차 : 데이터 분석

13주차 : 캐글 문제 도전1

14주차 : 캐글 문제 도전2

15주차 : 기말 시험

# Grading

- Lecture Attendance : 10%
- Practice Attendance : 15%
- Individual Assignment : 25%
- Midterm Exam : 25%
- Final Exam : 25%

