

비전공자를 위한

AI 데이터 분석 입문

권순동 교수

인공지능이란 무엇인가

- 인공지능 Artificial Intelligence
 - 추론이나 학습과 같은 인간 고유의 생각과 행동 능력을 모방하는 것
- 약한 인공지능(Weak AI, Narrow AI)
 - 특정 영역의 문제 해결에 최적화된 AI 기술로 알고리즘과 데이터, 규칙 등을 필요로 하는 것
 - 추천시스템, 로봇 청소기, 인공지능 비서, Virtual Influencer, RPA
 - 인간이 AI를 활용하여 능력이 더 증강된다 하여 증강지능(Intelligence Augmented)으로 불리기도 함
- 강한 인공지능 (Strong AI)
 - 인간 수준의 지능을 가지고 스스로 생각하고 일하는 AI (터미네이터 영화)

인공지능과 4차 산업혁명

대상



기술적용



큰 변화

인력

증기기관

1차 산업혁명 (기계화) 1784

생산과 동력의 비분리

전기

2차 산업혁명 (대량생산) 1870

인간의 수작업

컴퓨터

3차 산업혁명 (사무자동화) 1969

인간의 두뇌

인공지능

4차 산업혁명 현재

인공지능 활용 사례

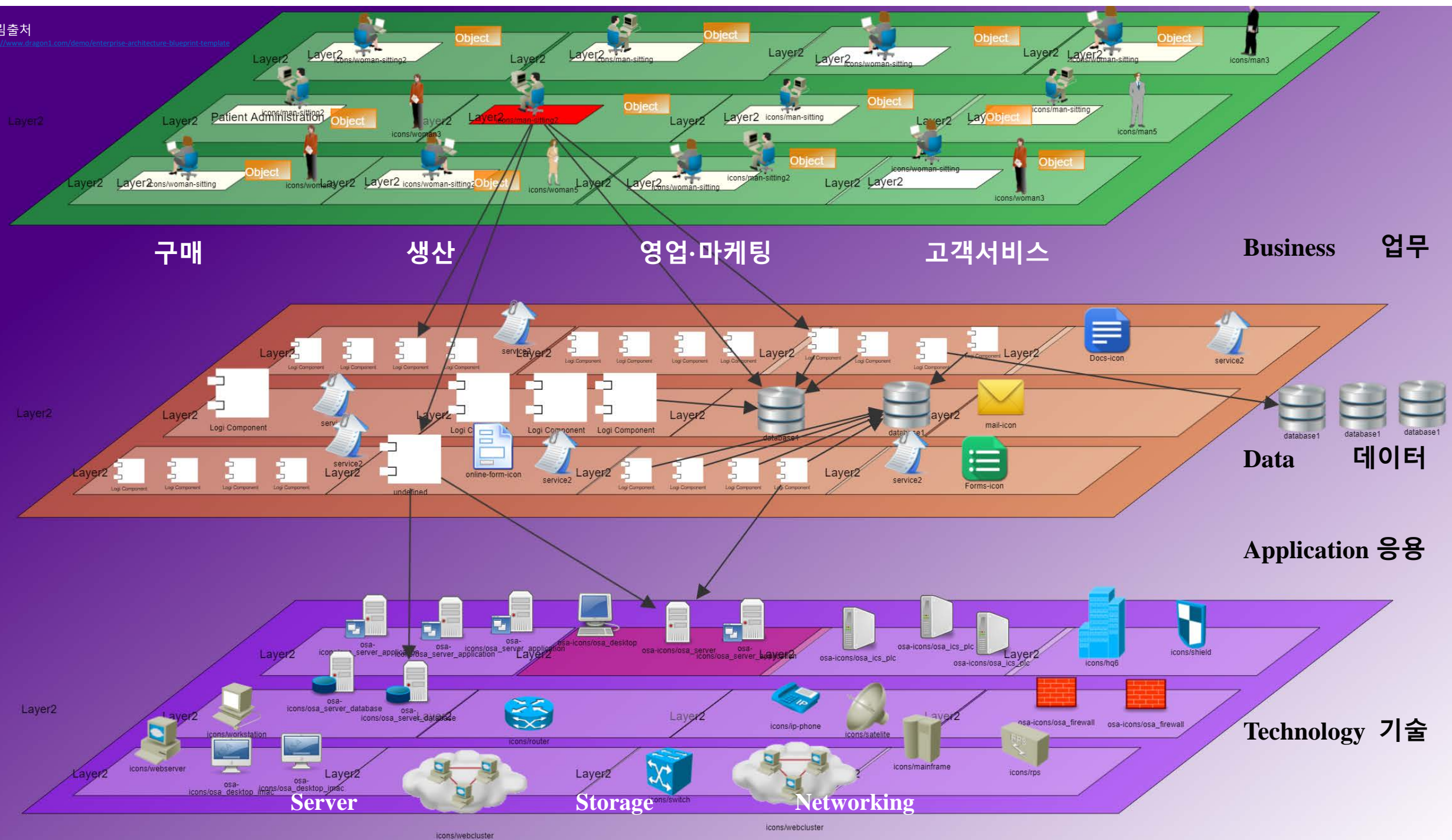
- 4차 산업혁명을 주도하며 생산성을 크게(10배) 향상시킴
- 스마트공장-맞춤형 제품을 저비용으로 생산
- 구글 인공지능 데이터 분석 - 맞춤광고, 상품맞춤추천, 동영상추천
- 신약개발의 기간 및 비용 단축
 - 코로나19 진단키트 개발
 - "기존 방법으로 100명의 전문가가 3개월 동안 할 것을
 - AI와 컴퓨터로 3시간 만에 해결할 수 있다" (씨젠 천종윤 대표)

인공지능을 이용한 데이터 분석

- AI는 컴퓨터, 소프트웨어, 통계학 전공자들이 다루는 것 아닌가?
- AI를 비전공자가 다루는 것이 의미가 있나?
- AI를 비전공자가 어느 수준까지 알면 좋은가?
- AI 기반 데이터 분석의 핵심 내용은 무엇인가?

AI 기반 데이터 분석의 핵심 내용

- Business 업무
- Data 데이터
- Application 응용 .변수지정 .모델선택 .성능확인
- Technology 기술



조직 아키텍처

Enterprise Architecture, EA

- 엔터프라이즈 아키텍처
- 전사 아키텍처
- 정보기술 아키텍처
- 소프트웨어 아키텍처
- 조직 전체 조감도(가시화), 설계도(계획), 청사진, 구조도, 프레임워크
- 데이터 분석 아키텍처(절차, 구조, 방법)

조직 아키텍처

Enterprise Architecture, EA

• Business 업무: 구매 생산 판매 인적자원

• Data 데이터:

• Application 응용:

• Technology 기술:

컴퓨터, 인터넷

Business 업무

1. **지도**에 무인교통단속카메라 위치 표시하기
2. **주가 예측**: 주가 데이터를 확보하여 다음주 주가를 예측해보기
3. **매출액 예측**: 내일 기온이 36도라면 아이스크림이 얼마나 판매될 것인가를 예측해보기
4. **통신사 고객이탈 예측**: 월통신요금이 80달러이고 가입기간이 12개월일 때, 이 고객은 다른 회사로 전환할 것인가?
5. **신입사원 이직원인 분석**: 신입사원 이직원인을 찾아보고, 대처방안 수립해보기

Data 데이터

- 데이터 유형
 - 숫자, 문자, 사진, 음성, 동영상 데이터
- 정형 데이터
 - 연속형 데이터, 이산형 데이터
- 비정형 데이터
 - 블로그, 페이스북, 인스타그램, 뉴스, 음성, 동영상

Application 응용 (데이터 분석 유형)

- 추정estimation: 연속형 데이터
- 분류classification: 범주형 데이터
- 군집화clustering: 시장 세분화
- 유사성 집단화affinity grouping: 장바구니 분석

추정 분석

알려져 있지 않은 연속형 변수의 동작 값,
즉, 미래 값을 추정하는 것이다.

유사성 집단화 분석

변수들 간의 관계를 알려주는데,
관계의 성격과 빈도수를 함께 알려준다.

군집 분석

데이터셋을 상호 배타적인 집단으로 구분하는 것으로,
집단 내 구성원 간에는 가능한 서로 가깝고,
집단이 다르면 가능한 멀어지도록
분할하는 기법이다.

분류 분석

데이터를 가장 효과적이고 효율적으로
사용하기 위해 범주나 집단으로
데이터를 조직화하는 과정이다.

Application 응용 (분석 기법)

- Machine Learning

- LinearRegression
- DecisionTreeRegressor
- XGBRegressor

- Deep Learning (인공신경망 Artificial Neural Network)

- keras Sequential
- keras Dense

Application 응용 (절차)

- 변수지정: 독립변수 x 와 종속변수 y
 - 데이터 표준화
 - train data, test data
- 모델선택
 - LogisticRegression
 - DecisionTreeClassifier
 - XGBClassifier
 - keras Sequential
 - `model.fit(x, y)`
- 성능확인 : R^2 , accuracy

Technology 기술

- Google colab 클라우드

- <https://colab.research.google.com>

- 강의자료: https://colab.research.google.com/drive/1YsYu_cq2oHGKpgjA9MnDgmLvERASGgRc?usp=sharing

- 딥러닝 PC

- jupyter notebook <https://www.anaconda.com>

AI 데이터 분석의 차이점

- 데이터 분석과 빅데이터의 차이점
- Excel과 빅데이터 분석의 차이점
 - Excel은 마우스 클릭하여 또는 명령어를 입력하여 실행
 - Pandas는 명령어를 입력하여 실행

AI 데이터 분석의 학문적 특징

- 업무 지식 (domain knowledge) 사회과학, 의학, 약학, 식품, 패션, 경제학, 행정학
- 공 학 컴퓨터공학, 소프트웨어공학, 데이터 처리, python
- 자연과학 통계학, 수학

AI를 더 잘 활용하려면

- Business – Domain knowledge 에 대한 이해
- Data – 데이터에 대한 이해
- Application – 예측 성능을 더 높여야 하는가
- Technology – 분석용 컴퓨터를 장만해야 하는가, Cloud를 이용해야 하는가?

- 전공자
 - 더 잘 개발함
- 비전공자
 - 더 일을 잘함

데이터 확보 방법

- 공공데이터 포털 데이터
 - 공공데이터 포털
 - 주가 데이터
- 기관별 공개 데이터
 - 한국노동연구원의 한국노동패널조사 데이터 분석
 - 한국고용정보원 (청년패널조사, 대졸자직업이동경로조사, 고령화연구패널조사)
 - 정보통신정책연구원(KISDI) 한국미디어패널조사

데이터 분석 도구

- Python
 - 소프트웨어 개발 언어
 - Python 언어의 일부가 데이터 분석에 사용됨
- R
 - 통계분석 언어 (tool, 도구)