



Introduction To Data Analysis

데이터 분석 입문

Lecture 01. 강의 소개

인공지능소프트웨어학과 강환수 교수

강의 소개



교수 소개

강환수 교수

- ✓ 인공지능소프트웨어학과 학과장 교수
- ✓ 연락처: 02-2610-1941
- ✓ 연구실: 2호관 706호
- ✓ E-mail: hskang@dongyang.ac.kr
- ✓ Github Homepage
+ <https://github.com/ai7dnn>



DMU*ai*

동양미래대학교 인공지능소프트웨어학과

인공지능 서비스
전문 소프트웨어 개발자 인재양성

Dongyang Mirae University
Dept. Of Artificial Intelligence

DMU*Ai* 동양미래대학교 인공지능소프트웨어학과

인공지능 서비스 전문 소프트웨어 개발자 인재양성
Dongyang Mirae University Dept. Of **Artificial** Intelligence

2021.

교육부의 신산업분야
특화 선도전문대학 지원사업에 선정

2022

인공지능소프트웨어학과를 신설하여
인공지능 서비스 구현 분야의
전문기술 인재 육성

2023

인공지능소프트웨어학과와 정원 80명으로 증원해
국내 최고의 인공지능 서비스 구현의
전문기술 인재 양성 요람으로 육성

교육과정

빅데이터분석 트랙 중 초급

목표 직업군	인공지능서비스구현 & 응용SW엔지니어링					
직무 핵심역량	1학년		2학년		3학년	
	1학기	2학기	1학기	2학기	1학기	2학기
컴퓨팅사고 · 프로그래밍	컴퓨터공학기초 프로그래밍기초	데이터베이스 자바프로그래밍	자료구조 객체지향프로그래밍			
웹·모바일 서비스 구현	UI/UX 디자인 웹프로그래밍기초	웹서버프로그래밍	모바일프로그래밍 웹프로젝트	모바일프로젝트		
빅데이터 분석		데이터분석입문	빅데이터분석 프로그래밍	빅데이터응용 프로그래밍	빅데이터분석 프로젝트	
인공지능구현	인공지능		기계학습 · 딥러닝		인공지능서비스프로젝트	현장실습
	파이썬프로그래밍 인공지능개론	오픈소스소프트웨어 인공신경망	기계학습 딥러닝프로그래밍	컴퓨터비전 딥러닝응용프로그래밍	자연어처리 인공지능라이브러리 인공지능캡스톤디자인	최신인공지능 인공지능비즈니스모델 졸업작품
마이크로 전공트랙	타과 학생의 인공지능 리더러시 및 세부 심화 인공지능 기술 함양을 위해 3개의 마이크로 전공트랙 제도 운영					
	인공지능 마이크로 전공트랙	빅데이터분석 마이크로 전공트랙	기계학습 · 딥러닝 마이크로 전공트랙		전공필수교과목	파이썬프로그래밍 인공신경망 인공지능캡스톤디자인 졸업작품

빅데이터분석 관련 교과목

데이터분석입문(1학년 2학기)

- ✓ 파이썬을 활용한 데이터분석기초, csv 자료 중심
- ✓ Matplotlib 데이터 시각화 기초
- ✓ Numpy, pandas 입문

빅데이터분석프로그래밍(2학년 1학기)

- ✓ 판다스(pandas)를 활용한 데이터분석
- ✓ Matplotlib 데이터 시각화 심화

빅데이터응용프로그래밍(2학년 2학기)

- ✓ 판다스(pandas)를 활용한 시계열 데이터분석

빅데이터분석프로젝트(3학년 1학기)

- ✓ 공공데이터 분석 프로젝트 수행

— 교과목 개요

과목 개요	<ul style="list-style-type: none"> • 본 과목에서는 우리 일상생활 속 데이터에 질문을 던지고, 이를 해결하는 형태로 수업을 진행합니다. • 기온, 인구, 대중교통 같은 실제 공공데이터를 활용한 데이터 분석 프로젝트를 진행합니다. • 데이터 다운로드부터 데이터 탐색, 문제 도출 및 문제 해결 방법 구상까지 차근차근 데이터 분석 절차를 학습하고, 파이썬(Python)을 활용하여 코드를 작성해 나가면서 데이터 분석 능력을 향상시킵니다. • matplotlib 라이브러리를 활용하여 막대그래프, 파이 차트, 버블 차트와 같은 다양한 시각화 방법을 학습하고, numpy, pandas 같은 라이브러리 활용법도 함께 학습합니다.
학습목표 및 성취수준	<ul style="list-style-type: none"> • 데이터 분석 능력을 배양할 수 있다. • 파이썬 프로그래밍 능력을 배양할 수 있다. • 주어진 데이터로부터 파이썬을 활용하여 데이터를 정제할 수 있다. • 파이썬을 활용하여 데이터를 다양한 형태로 시각화 할 수 있다. • 빅데이터의 개념 및 특징을 이해할 수 있다.

— 수강 안내

수업 시 사용 도구	<ul style="list-style-type: none">파이썬 프로그래밍 실습도구<ul style="list-style-type: none">- Jupyter Notebook- Google CoLab- Visual Studio Code(vs code)
평가방법	<ul style="list-style-type: none">출석: 20%, 과제 및 수업참여: 10%, 중간고사: 30%, 기말고사: 40%
수강안내	<p>※ 수업 구성: 일반 수업 3주(1주, 5주, 14주) + 온라인 수업 10주 + 중간고사 1주 + 기말고사 1주 = 총 15주</p> <ul style="list-style-type: none">선수과목: 파이썬타 학과 전공생의 수강 신청 환영합니다.

평가 방법

총점 100점

✓ 출석 (20%) / 과제와 수업참여 (10%) / 중간고사 (30%) / 기말고사 (40%)

평가 방식

✓ 상대평가 방식

등급	평가 비율	
A+, A0	0~35% 이내	0~75% 이내
B+, B0	-	
C+, C0	20% 이상	
D+, D0		
F		

+ 성적과 관계없는 F학점 조건
: 무단결석 4일(=12시수)

LESSON 01

데이터 분석이란?



데이터 분석(Data Analysis)

- ✓ 데이터로부터 유용한 정보를 발굴하고 결론 내용을 알리며 의사결정을 지원하는 것을 목표로 데이터를 정리, 변환, 모델링하는 과정 (from Wikipedia)
- + 특히, 오늘날 비즈니스 부문에서 데이터 분석은 의사 결정을 더 과학적으로 만들어주고 비즈니스를 더 효율적으로 운영할 수 있도록 도와주는 역할을 함



- 데이터 분석 도구 (Tool)에는 무엇이 있을까?

— 데이터 분석 도구의 종류

① 파이썬 (Python)



② R



③ 엑셀 (Excel)



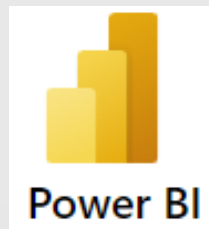
④ SQL (Structured Query Language, 구조화 질의어)



⑤ 태블로 (Tableau)

⑥ Power BI (Business Intelligence)

⑦ 구글 애널리틱스 (Google Analytics, GA)



데이터 분석을 하려면?

① 데이터에 대한 지식

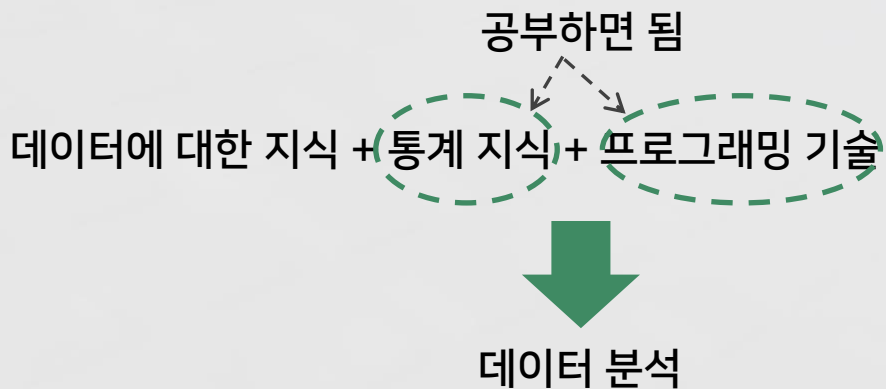
- ✓ 데이터를 제공하는 쪽에서 사전 정보를 줄 수도 있고, 안 줄 수도 있음

② 통계 지식

- ✓ 데이터로부터 어떤 정보를 추출해 낼 것인지? 의사결정에 도움이 되는 정보인지 판단

③ 프로그래밍 기술

- ✓ 분석하고자 하는 내용을 프로그램으로 구현하는데 필요

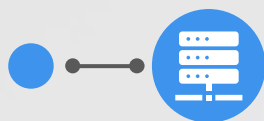


LESSON 02

빅데이터와 분석 방법

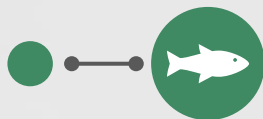


빅데이터 개요



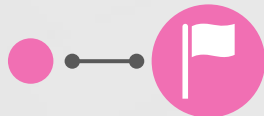
기존의 데이터 분석역량을 넘어서는 방대한 분량의 데이터

- 디지털 기기 및 센서 등의 보급화를 통해 실생활에서 규모를 가늠하기 힘든 수준의 많은 정보와 데이터 생성
- 데이터 규모의 방대함, 데이터의 종류와 속성 다양, 매우 빠른 주기로 생성



디지털기기와 센서를 이용한 사람 및 기계간의 정보를 주고받는 환경의 확산으로 점차 가속화

- IoT나 M2M 환경의 확산을 통한 막대한 정보 생성
- 사용자가 직접 제작하는 동영상 및 SNS 등을 통한 데이터 증가



빅데이터 분석을 통해 새로운 가치 창출

- 생성된 빅데이터를 필요한 목적에 맞게 가공하고 분석하여 새로운 결론을 얻고, 이를 통한 최적의 답안 제시
- 빅데이터에서 얻을 수 있는 기존의 패턴 분석으로 향후에 일어날 현상이나 상태를 예측하고 대응

→ 데이터 분석 과정

- ✓ 문제의 정의
- ✓ 수집 및 분석할 데이터의 정의
- ✓ 분석계획/표본데이터의 수집
- ✓ 데이터의 취득
- ✓ 데이터 전처리 및 정제
- ✓ 탐색적 데이터 분석
- ✓ 기계학습 등을 이용한 분류 또는 예측 모델링
- ✓ 보고서 작성

데이터 분석 방법

Raw데이터 수집

데이터 전처리

데이터 분석

데이터 시각화

시간 분석 방법

- 시간을 기반으로 데이터를 분석하는 방법
- 예측 기법에 많이 적용되는 방법 (주식시장, 날씨 변화)

분포 분석 방법

- 데이터 분포(산점도, 상관관계, 히스토그램)를 기반으로 분석하는 방법
- 산점도 및 히스토그램을 이용해 밀집도를 확인해 분석

지도 분석 방법

- 지도 데이터를 이용해 분석하는 방법
- GPS, 위/경도를 이용해 데이터를 분석

통계적 분석 방법

- 확률, 분산, 표준편차 등을 이용해 분석하는 기법
- P-value(유의확률), T검정(집단 분석)등의 분석 방법이 대표적

기계학습을 이용한 분석 방법

- 데이터를 기반으로 한 지도학습, 비 지도학습 분석
- 대표적으로 지지벡터머신, 뉴럴넷 등의 분석 방법이 대표적

ETC..

- 분석 방법은 어떤 데이터를 분석하느냐에 따라 분석 방법이 달라 질 수 있음

LESSON 03

정보의 진화



데이터

- ☑ 정형화되고 기록할 만한 가치가 있다고 판단되는 어떤 현상이나 사건, 아이디어에 대한 묘사



- + 이름 : 홍길동
- + 키 : 170cm
- + 몸무게 : 70kg
- + 결혼 : 미혼
- + 특기 : 무술

정보

- ✓ 데이터는 사실들 그 자체에 대한 일차적인 표현
- ✓ 사실들과 이들로부터 유도될 수 있는 유추된 사실들
- ✓ 관련성과 목적을 갖춘 데이터(Peter Drucker)
- ✓ 정보 = 데이터 + 의미 (Thomas Davenport, Laurence Prusak)

1학기	2.5
2학기	3.0
3학기	3.0
4학기	3.3
5학기	2.9
6학기	2.7
7학기	2.5
8학기	2.3

데이터

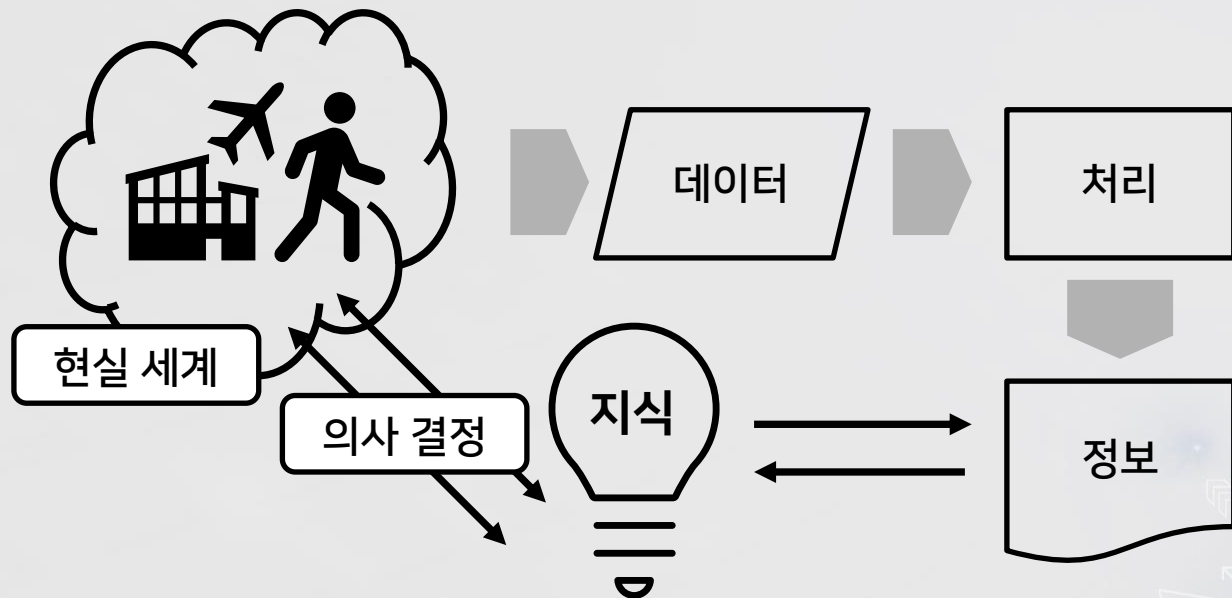


정보

지식

데이터와 정보에 비해 좀 더 상위 수준의 개념

- ✓ 정보(Information)에 개인적 경험, 아이디어 등이 더해지면 지식(Knowledge)



지식 = 내면화된 정보(개인적 경험과 아이디어) + 정보 사용 능력

지혜

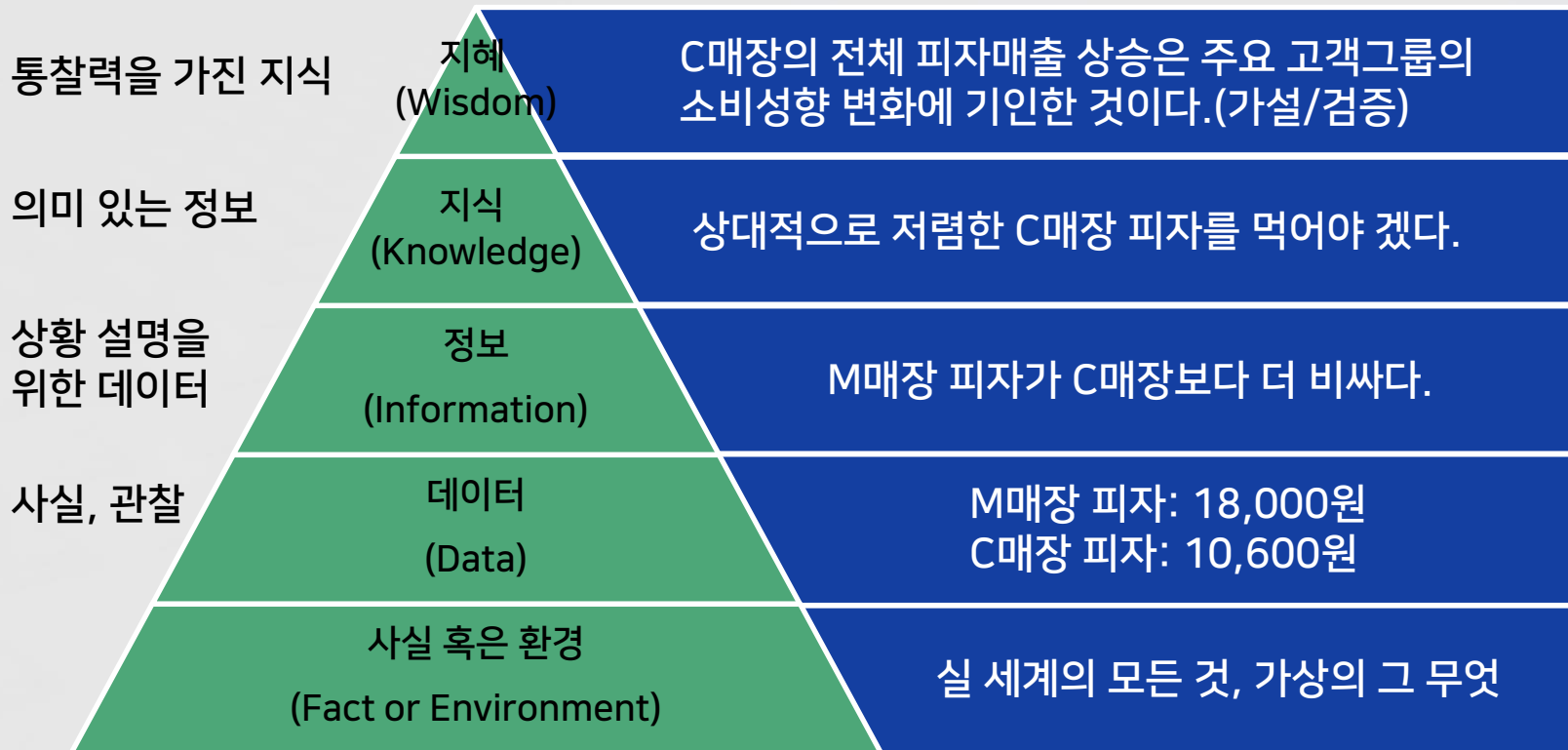
종합된 지식, 패턴화된 지식을 바탕으로 올바른 의사결정을 수행하는 능력

- ✓ 지혜(Wisdom)는 축적된 지식을 종합적으로 활용하여 앞으로의 상황을 예측하거나, 현상에 대한 관계를 파악해낼 수 있는 단계

사전적 의미

- ✓ 사물의 이치를 빨리 깨닫고 사물을 정확하게 처리하는 정신적 능력

— 정보의 진화 단계



THANK YOU!

Q & A

- + Name: 강환수
- + Office: 동양미래대학교 2호관 706호 (02-2610-1941)
- + E-mail: hskang@dongyang.ac.kr
- + Homepage: <https://github.com/ai7dnn>

