

어프렌티스 프로젝트

– Introduction –

산업인공지능센터 김 재 영

Contents

I. 교과목 소개

1. 강의 계획
2. 프로젝트 파트 진행 및 과제
3. 평가
4. 출석수업 강의실
5. 강의 자료 및 과제
6. 교수 소개

II. 프로젝트 소개

III. Q&A

교과목 소개



1. 강의 계획

1. 프로젝트 파트

주별	강의 주제	강의 방법
09/01	오리엔테이션	출석
09/08	디자인 씽킹과 AI 프로젝트	원격
09/15	데이터 수집	원격
09/22	데이터 전처리	원격
09/29	데이터 분석 및 시각화	원격
10/06	모델 생성	원격
10/13	모델 테스트 및 업그레이드	출석
10/20	결과 발표	출석

2. 연구계획서 작성 파트

주별	강의 주제	강의 방법
10/27	학술 논문의 개념 및 구조	출석
11/03	학술 논문 작성 절차	출석
11/10	직무 관련 주제 발표	출석
11/17	특강 및 논문 검색 방법	원격
11/24	글쓰기 및 발표자료 작성 방법	출석
12/01	연구계획서 작성 방법	출석
12/08	연구계획서	출석

프로젝트 파트 진행 및 과제

주차	수업내용	프로젝트 진행	프로젝트 과제	비고
1	강의소개			
2	디자인 씽킹과 AI 프로젝트		프로젝트(개별프로젝트)의 문제의 식별 및 정의	• 자신의 업무와 연관된 문제 프로젝트에 대하여 가산점
3	데이터 수집	제출한 개별 프로젝트 문제의 정의 피드백	프로젝트 데이터 수집	• 상황에 따라 linear Regression 내용이 추가 될 수 있음
4	데이터 분석 및 시각화	개별 데이터 수집 결과 피드백	데이터 전처리	
5	데이터 전처리	데이터 분석 제출 결과 피드백		
6	Regression 모델	데이터 전처리 제출 결과 피드백		
7	모델훈련 및 분석	모델 훈련		
8	결과 발표	최종 결과 발표	프로젝트 결과 보고서	- 모델 테스트 결과 및 평가 - 결과 분석

***** 수업 진행 상황에 따라 수업 일정, 내용, 과제 등이 사전 공지와 함께 조정될 수 있습니다.**

평가

- 2개 파트 분리 평가
 - 프로젝트 파트 50%(김 재영교수)
 - 연구계획서 작성 파트 50%(이 광연교수)
- 프로젝트 파트 평가

평가 항목		반영 비율	
출석		10%	10%
과제		20%	20%
프로젝트	발표	30%	70%
	보고서	40%	

출석 수업 강의실

주차	날짜	담당교수	대면/비대면	평가	강의실
1	09/01(목)	김재영	Offline		B222
2	09/08(목)	김재영	online		
3	09/15(목)	김재영	online		
4	09/22(목)	김재영	online		
5	09/29(목)	김재영	online		
6	10/06(목)	김재영	online		
7	10/13(목)	김재영	Offline		B222
8	10/20(목)	김재영	Offline	평가	B224(국제회의실)
9	10/27(목)	이광연	Offline		B222
10	11/03(목)	이광연	Offline		B222
11	11/10(목)	이광연	Offline		B222
12	11/17(목)	이광연	online		
13	11/24(목)	이광연	Offline		B222
14	12/01(목)	이광연	Offline		B222
15	12/08(목)	이광연	Offline	평가	B224(국제회의실)

강의 자료 및 과제

- 강의 자료는 CBNU eCampus 강의홈 Home 강의자료 참조
- 과제는 CBNU eCampus 강의홈 과제 Assignment를 참조하고 제출

교수 소개

- 충북대학교 산업인공지능센터 초빙교수 김재영
- 연구실
 - 충북대학교 오창캠퍼스 융합기술연구원 C655호 iLab연구실
 - 28116 충북 청주시 청원구 오창읍 양정4길 45
- 연락처
 - 전 화: 043)249-1464 / 010-4952-2800
 - Email: jaykim@chungbuk.ac.kr

프로젝트 소개



What is data?

다양한 답변

- 사실과 통계
- 데이터베이스의 레코드 모음
- 정보
- 컴퓨터에 저장되거나 컴퓨터에서 사용되는 사실, 수치 또는 정보
- 숫자, 소리, 이미지
- 기록 및 거래
- 보고서
- 컴퓨터가 처리하는 것

데이터 (전)처리 관점

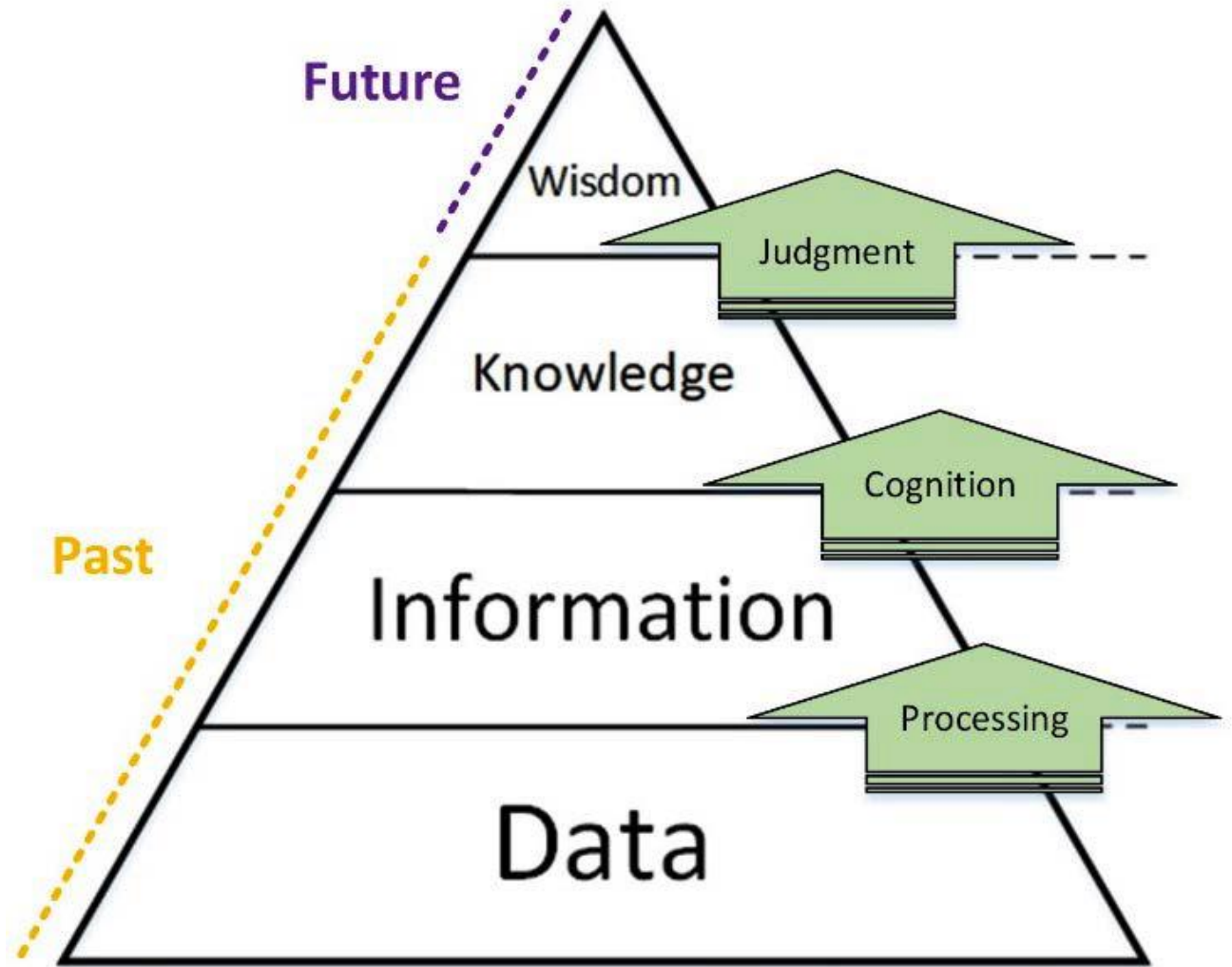
- 실제 존재하는 것의 측정 또는 모델을 나타내는 기호 또는 기호

데이터의 이해

- 고차원의 규약 및 이해(Higher-level conventions and understandings - HLCUs)
- HLCUs 가 없으면 데이터는 무의미한 기호(Symbol)와 신호(Sign)들의 파일

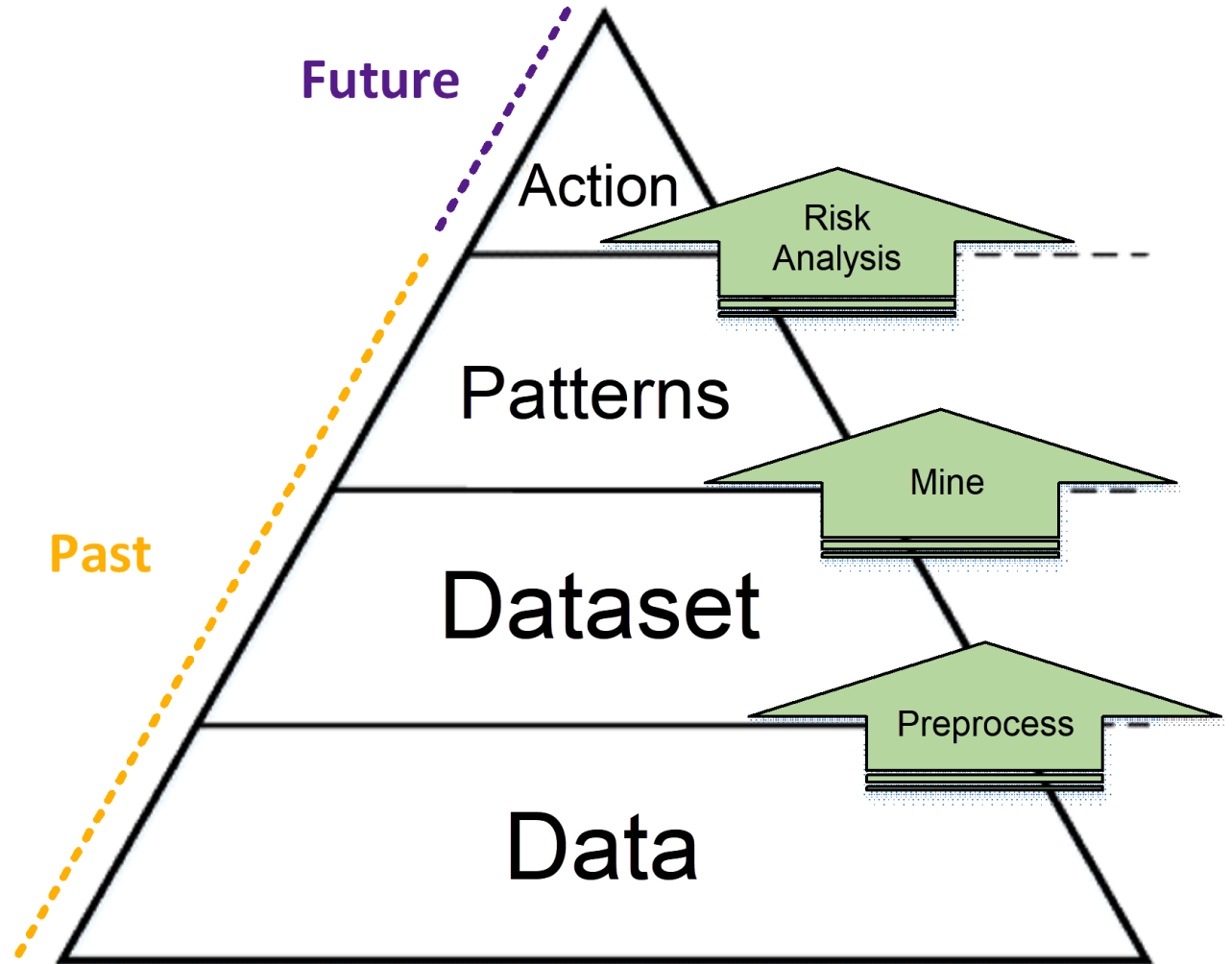
DIKW pyramid

HLCUs - 인간의 언어와 이해



DDAP pyramid

HLCUs - 머신 러닝(ML) 및
인공 지능

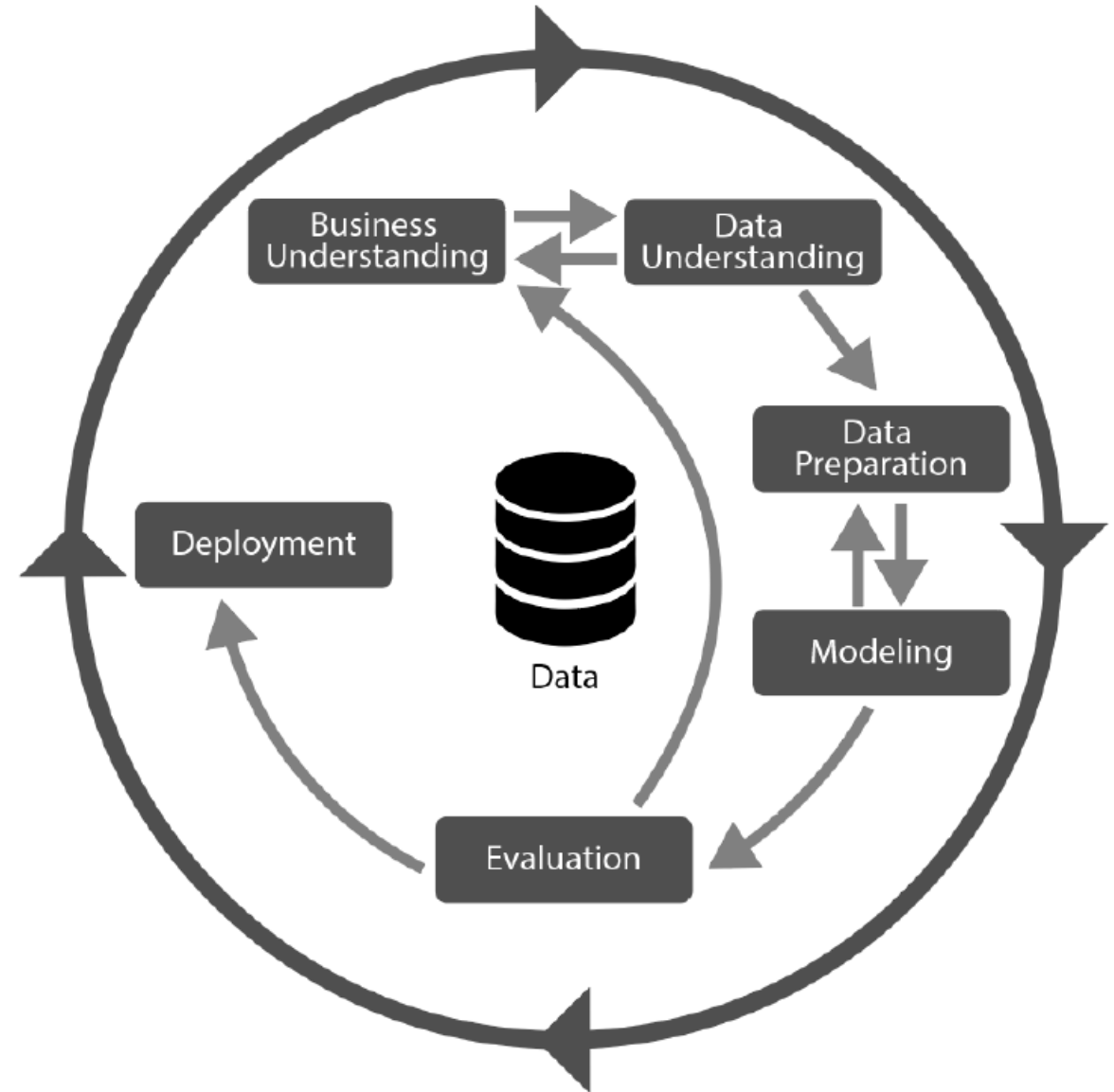


데이터 분석

- 통계학, 기계학습, 데이터 시각화를 포함한 다양한 분석 방법이나 도구를 이용하여 데이터분석 결과를 비즈니스에 유용한 정보로 전환하는 과정
- 데이터 분석 방법론: 데이터 분석을 효과적으로 활용하기 위해서 체계화된 절차와 방법
 - KDD(Knowledge Discovery in Database) 분석 방법론
 - CRISP-DM(Cross Industry Standard Process for Data Mining) 분석 방법론
 - SEMMA(Sample, Explore, Modify, Model and Assess) 분석 방법론

CRISP-DM Workflow

Open industry standard
(Cross-industry Standard
Process for Data Mining)



Business Understanding



- 데이터를 분석하는 이유를 이해
- 데이터 분석 모델과 결론이 비즈니스에 미치는 영향을 이해

Data Understanding



- 데이터 수집
- 문제 해결에 필요한 데이터를 선택할 수준의 충분한 이해
- 사용 가능한 데이터의 결정하고 데이터 인벤토리 작성
- 데이터 조사 및 평가
- 이 단계의 주요 결과는 데이터를 통해 발견된 내용 및 예상되는 작업을 설명하는 보고서

Data Preparation



- 모델링에 사용하는 데 필요한 정리 및 변환을 적용
- 데이터 타입, 수집 및 정제/정리 프로세스의 중요성을 잘 이해하는 것이 필요
- Garbage in, garbage out(가비지 인 가비지 아웃, GIGO, 기고)
- 데이터의 추가적인 속성(attribute) 찾고 모델링을 위한 데이터 세트 설정 과정에 추가적인 시각화 필요

Modeling



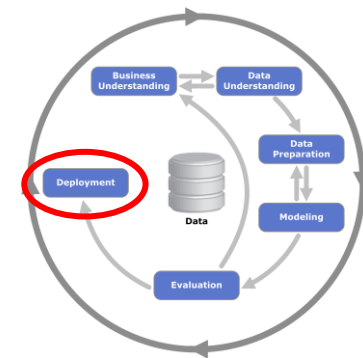
- Model architecture specification
 - 모델 유형 선택(선형 회귀, 인공 신경망, 그래디언트 부스팅 등)
 - 모델 학습 및 예측 기능 비교
- Training
 - Training 프로세스 구성
 - Training 프로세스 실행

Evaluation



- 최종 성능 측정기준으로 모델 평가
- 측정기준에 적합한 샘플 데이터를 사용하여 모델의 성능에 대한 일관되고 독립적인 측정 필요

Deployment



- 대부분의 모델은 일단 개발완료 후에 서비스 시스템에 적용
- 전자 상거래 사이트에서 모델이 추천 시스템의 기초가 되는 경우 시스템 적용과정이 매우 중요
- 모델 및 시스템의 버전 업그레이드에 대한 고려 필요

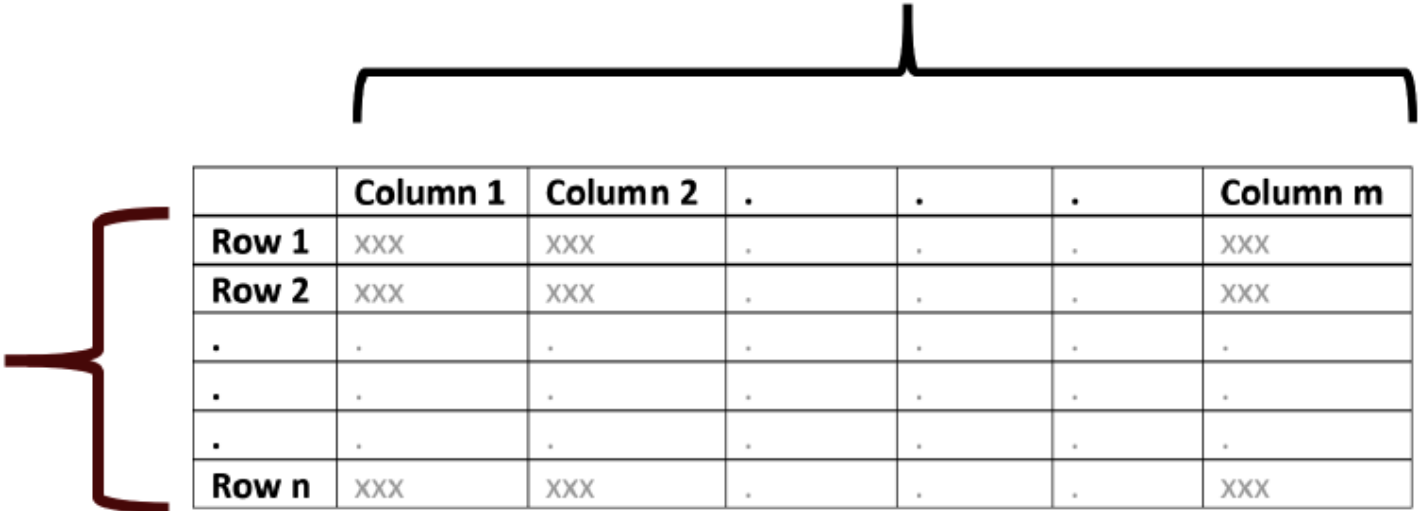
Data preprocessing

- 모델 학습에 사용되는 데이터 형태로 데이터를 가공하는 과정(in Machine learning)
- 처리 단계별 데이터
 - Raw data - 초기에 수집된 원본 데이터
 - Prepared data - 학습을 위한 데이터만 추출한 서브셋 데이터
 - 비어 있는 데이터나 불필요한 행이나 열이 삭제된 상태의 정돈된 데이터
 - Engineered feature - 머신러닝 학습과 서빙에 적절한 형태로 재가공 된 데이터
- Data preprocessing 작업
 - 오류 데이터 정제
 - Reindexing and reformatting
 - Missing values
 - Data fusion, data integration, data reduction, data transformation

The most universal data structure – a table

Data Attributes

Data Objects



	Column 1	Column 2	.	.	.	Column m
Row 1	xxx	xxx	.	.	.	xxx
Row 2	xxx	xxx	.	.	.	xxx
.
.
.
Row n	xxx	xxx	.	.	.	xxx

About the project

- 프로젝트는 개별적으로 진행
- 프로젝트 진행과 관련된 수업은 온라인으로 진행
 - 필요한 스터디와 프로젝트 수행을 위한 자기 주도적 학습 필요
- 프로젝트 진행 단계별로 과제 제출
- 프로젝트 과정 alc 결과에 대한 분석 및 평가 보고서 제출 및 발표

프로젝트 기초지식

- 통계
 - 평균(mean), 분산, 표준분산, 피어슨 상관 계수(Pearson Correlation Coefficient) 등에 대한 개념적인 이해
- Programming Language
 - Python
 - Numpy, Pandas
 - Matplotlib
- 데이터 분석(Machine Learning)
 - Linear Regression

마무리



자기 소개 및 Q&A

