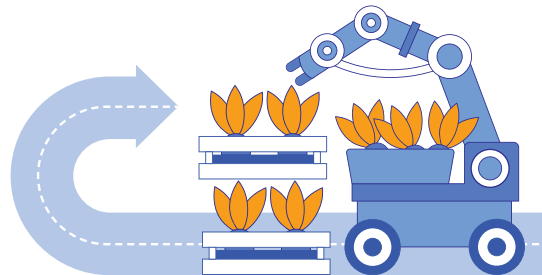




고등학교

데이터 과학





미래 사회를 정의하는 핵심 키워드를 뽑는다면, 대부분의 사람들이 주저 없이 ‘인공지능’을 선택할 것입니다. 이러한 인공지능 사회의 주인공이 되기 위해서는 어떤 능력이 필요할까요?

가장 먼저 데이터와 정보로 인한 디지털 세상의 변화를 인식하고, 정보의 사회적 가치를 탐구할 수 있어야 합니다. 왜냐하면 이를 바탕으로 디지털 세상의 주인공이 되기 위한 포용성을 갖출 수 있기 때문입니다.

또한, 정보를 처리하는 다양한 원리와 기술에 기반한 컴퓨팅 사고력을 함양하고, 실생활 및 다양한 학문 분야의 문제를 해결하는 능력과 태도를 길러야 합니다. 왜냐하면 이를 바탕으로 디지털 세상의 주인공이 되기 위한 주도성을 갖출 수 있기 때문입니다.

위와 같은 능력을 깊이 있게 갖추기 위해서는 ‘데이터 과학’ 과목을 학습해야 합니다.

‘데이터 과학’ 과목을 학습함으로써 데이터와 데이터 처리에 대한 다양한 방법론을 기반으로 통계와 기계학습 등을 활용하여 다양한 학문 분야의 문제 해결과 의사 결정에서 통찰력을 얻을 수 있습니다. 또한, 개인의 삶과 연관되어 일상과 향후의 직업에서 기술의 발전에 능동적으로 대처할 수 있는 능력을 기를 수 있습니다.

그리고 데이터 문해력을 함양함으로써 급증하는 데이터를 비판적으로 분석력을 갖추게 되며, 이를 바탕으로 미래 사회와 환경 변화 등에 대한 통찰력과 책임감 있는 자기 주도성을 갖춘 디지털 시민으로 성장할 수 있습니다. 특히, 컴퓨팅 사고력을 바탕으로 실생활 및 다양한 분야의 문제를 창의적으로 해결하는 데 필요한 능력과 태도를 함양하며, 데이터 과학의 지식이 필요한 진로와 연계된 기초 경험을 쌓는 데에도 도움이 됩니다.

이를 위해 본 교과서는 각 단원별로 다음과 같은 목표를 달성할 수 있도록 집필하였습니다.

I 단원에서는

데이터 과학의 발전에 따른 사회의 특성과 데이터의 가치를 이해하고, 데이터에 기반한 합리적인 의사 결정을 실천하는 태도를 기를 수 있는 내용과 활동으로 구성하였습니다.

II 단원에서는

데이터 분석과 관련된 효과적인 방법을 이해하고, 문제 상황에 따라 데이터의 관계를 파악하여 다양한 분석 방법을 적용할 수 있는 능력을 기를 수 있는 내용과 활동으로 구성하였습니다.

III 단원에서는

문제를 합리적으로 해결하기 위한 모델을 구성하고, 문제 해결 과정에서 발생할 수 있는 여러 쟁점을 비교하며, 분석된 결과의 의미를 찾아 비판적으로 해석하는 능력과 태도를 기를 수 있는 내용과 활동으로 구성하였습니다.

IV 단원에서는

데이터 과학을 기반으로 한 문제 해결이 합리적 의사 결정에 효과적임을 인식하고, 데이터 과학의 방법으로 문제를 해결하는 능력과 태도를 기를 수 있는 내용과 활동으로 구성하였습니다.

본 교과서를 학습한 여러분들이 컴퓨팅 사고력을 기반으로 디지털 사회에서 데이터의 역할 및 잠재적 가치와 데이터 과학에 기반한 문제 해결 과정의 중요성을 인식하며, 다양한 분야의 문제를 해결하고 합리적 의사 결정을 위한 통찰의 역량을 키울 수 있기를 바랍니다.

구성과 특징



▶ 시작하기

대단원 도입

단원의 핵심 아이디어

- ★ 데이터 분석을 위해서는 데이터
- ★ 데이터 처리는 데이터를 분석



데이터 준비와

- 1 데이터 수집 및 특성 분석
- 2 데이터 전처리
- 3 데이터 수집 간의 관계 파악
- 4 동일한 데이터의 다양한 분석 방법

데이터에서 원하는 자료를 얻거나 자료를 정렬하기 위한 분석 방법입니다. 그러나 원본데이터에 할 일은 정확하고 신뢰할 수 있는 데이터를 수집하는 것입니다. 데이터는 종종 값이 누락되기도 하고, 분석에 적합하지 않은 행을 제거하기도 합니다. 이 단계에서는 데이터를 수집하고 전처리하는 후, 다양한 분석 방법을 적용하여 분석 결과를 비교하는 과정을 통해 데이터를 알고 이해하여, 합리적인 선택과 결정의 도구를 활용할 수 있는 능력을 기를 수 있도록 합니다.

소단원 도입

2 데이터의 형태와 속성

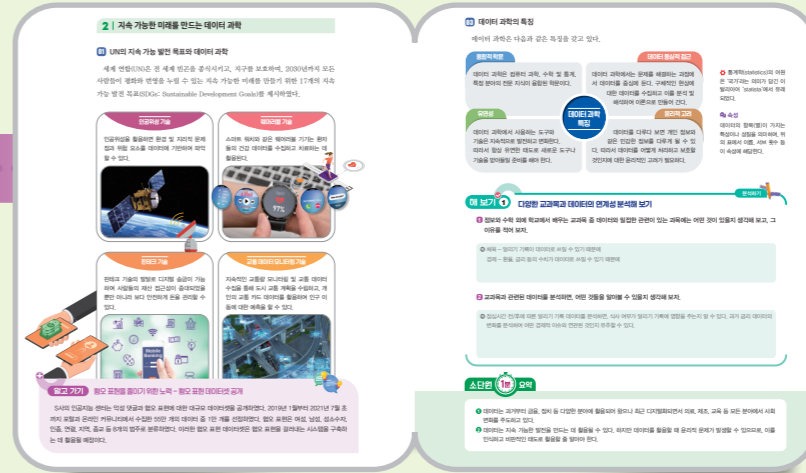
- 1 데이터의 형태
- 2 데이터의 속성



주어진 주어진 문제를 해결하기 위해 수집할 수 있는 데이터의 형태를 설명하시오.

데이터의 형태와 속성을 설명하고, 어떤 데이터를 수집할 수 있는지 설명하시오.

▶ 본문 전개하기



해 보기 1

1 정보와 수학 외에 이유를 적어 보기

알고 가기

S사의 인공지능

소단원 1분 요약

통계학(statistics)은 '국가'라는 의미가 아니라 'statista'에서 되었다.

속성

데이터의 항목

[대단원 도입] 대단원과 소단원 제목을 통해 앞으로 학습할 내용을 미리 살펴볼 수 있도록 구성했습니다. 또한 교육 과정의 핵심 아이디어를 소개하여 해당 단원에서 꼭 배워야 내용을 파악하고 이를 토대로 학습을 진행할 수 있도록 했습니다.

[소단원 도입(생각 열기)] 학습 목표와 관련된 일상적인 문제 상황이나 사례, 기사 등을 삽입하여 사 진과 함께 제시하여 학습자의 흥미와 관심을 돋우고, 질문을 통해 앞으로 배울 내용에 대해 고민해 보도록 했습니다.

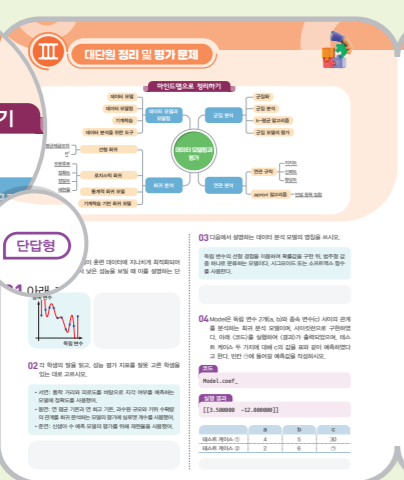
[본문] 단원에서 다루는 주요 개념, 기초 지식, 원리 등을 사진, 삽화, 통계 자료 등 다양한 시각 자료를 활용하여 구조화하여 제시함으로써 학습자의 이해도를 높이고, 자기 주도 학습을 촉진하도록 했습니다.

• 용어 설명 본문에서 다루는 중요한 용어, 개념 등에 대해 보충 설명을 제공했습니다.
• 보조 설명 또는 팁 학습 내용을 이해하는 데 도움이 되는 간단한 정보나 지식을 제공했습니다.

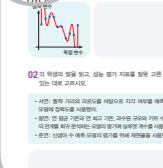
평가 문제

마인드맵으로 정리하기

데이터 모델과 모델링

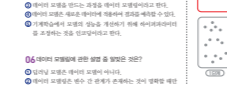


단답형



선택형

서술형



평가 문제

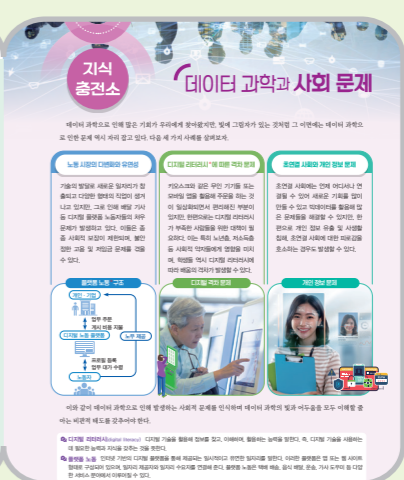
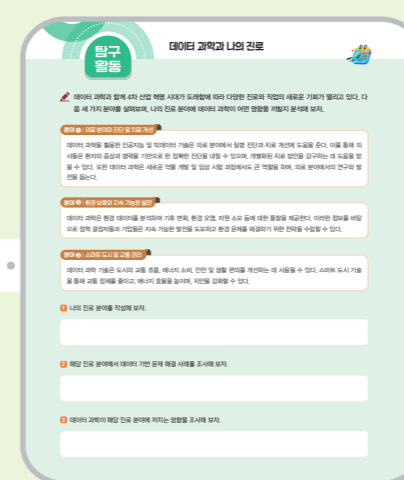
구분	종류	가치	단위	정확도
1	수치	1.5	cm	1.0
2	문자	1.5	cm	1.4
3	문자	1.5	cm	1.3
4	문자	1.5	cm	1.2
5	문자	1.5	cm	1.1
6	문자	1.5	cm	1.0

정리하기



[대단원 정리 및 평가 문제]
• 마인드맵으로 정리하기 마인드맵을 통해 학습한 핵심 개념을 상기시키고 학습자가 스스로 정리할 수 있도록 했습니다.

• 평가 문제 다양한 유형의 평가 문제(선택형, 단답형, 서술형 등)를 활용하여 학습한 내용을 한 번 더 정리하고 평가할 수 있도록 했습니다.



[탐구 활동] 소단원 내용을 정리·포괄하는 활동을 제시하여 주요 개념을 이론적으로 이해하는 것뿐만 아니라 실제 상황에서 적용 가능한 문제 해결 능력을 기를 수 있도록 했습니다.

[지식 충전소] 최신 정보와 다양한 분야의 읽기 자료를 제공하여 학습자의 흥미를 유발하고 데이터 과학 분야의 최근 동향을 파악할 수 있도록 안내했습니다.

I 데이터 과학의 이해

- 01 | 데이터 과학과 문제 해결**
 - 1 데이터와 데이터 과학 11
 - 2 데이터 과학의 문제 해결 과정 14
- 02 | 데이터의 형태와 속성**
 - 1 데이터의 다양한 형태 19
 - 2 데이터의 속성 21
- 03 | 데이터셋과 데이터베이스**
 - 1 데이터셋 25
 - 2 데이터베이스 26
 - 3 데이터베이스의 통합적 활용 27
 - 4 데이터베이스의 생활 속 활용 29
- 04 | 세상을 바꾸는 데이터 과학**
 - 1 데이터의 역사 33
 - 2 지속 가능한 미래를 만드는 데이터 과학 36
 - 3 데이터 과학 속 데이터 윤리 38

○ 대단원 정리 및 평가 문제

II 데이터 준비와 분석

- 01 | 데이터 수집 및 특성 분석**
 - 1 데이터 수집 47
 - 2 데이터의 특성 분석 51
- 02 | 데이터 전처리**
 - 1 데이터 전처리 57
 - 2 데이터 분석을 위한 시각화 63
- 03 | 데이터 속성 간의 관계 파악**
 - 1 데이터 속성 간의 관계 분석 69
 - 2 데이터 속성 간의 관계 분석 실습하기 73
 - 3 데이터 통합하여 분석하기 78
- 04 | 동일한 데이터의 다양한 분석 · 비교**
 - 1 다양한 방법으로 데이터 분석하기 85
 - 2 토마토 재배 관련 데이터 분석 · 비교하기 86

○ 대단원 정리 및 평가 문제



III 데이터 모델링과 평가

- 01 | 데이터 모델과 모델링**
 - 1 데이터 모델과 모델링 105
 - 2 데이터 분석을 위한 도구 107
 - 3 사이킷런으로 데이터 모델링 시작하기 109
- 02 | 회귀 분석**
 - 1 선형 회귀 119
 - 2 로지스틱 회귀 135
- 03 | 군집 분석**
 - 1 군집화와 군집 분석 145
 - 2 k-평균 알고리즘으로 군집화하기 146
 - 3 군집 모델의 평가 153
- 04 | 연관 분석**
 - 1 연관 분석과 연관 규칙 157
 - 2 연관 규칙의 성능 159
 - 3 선형적(apriori) 알고리즘 161

○ 대단원 정리 및 평가 문제



IV 데이터 과학 프로젝트

- 01 | 데이터 과학 적용과 데이터 적합성**
 - 1 데이터 과학의 적용 사례 171
 - 2 데이터로 해결 가능한 주제 174
- 02 | [데이터 과학 프로젝트 ①] 슬기로운 의사 생활** 178
- 03 | [데이터 과학 프로젝트 ②] 지혜로운 어부 생활** 196

○ 대단원 정리 및 평가 문제



부록

- (데이터 분석을 위한) 코랩 활용하기 216
- (데이터 분석을 위한) 판다스 기본기 익히기 218
- (최적의 분류 기준을 찾는) 의사결정트리 모델 알아보기 220

단원의 핵심 아이디어

- ★ 디지털 사회의 시민에게는 데이터에 기반한 합리적인 의사 결정이 필요하다.
- ★ 데이터 과학에 대한 이해는 데이터를 활용하여 복잡한 문제를 해결하는 데 도움을 준다.

데이터 과학의 이해

- 01 | 데이터 과학과 문제 해결
- 02 | 데이터의 형태와 속성
- 03 | 데이터셋과 데이터베이스
- 04 | 세상을 바꾸는 데이터 과학



데이터 양의 폭발적인 증가와 데이터 처리 기술의 발전으로 인해 데이터 과학은 다양한 분야에서 더 나은 의사 결정을 도우며 우리의 삶과 사회의 변화를 이끌고 있다.

이 단원에서는 데이터 과학의 개념과 함께 데이터의 잠재적 가치를 살펴보고, 진로 및 직업과 관련한 데이터 기반의 문제 해결 사례를 분석하며 데이터의 역할 및 중요성을 파악하도록 한다.

1

데이터 과학과 문제 해결



- 학습 목표**
 - 데이터 과학의 개념과 데이터 과학의 문제 해결 단계를 설명할 수 있다.
 - 실생활의 문제 해결 사례를 데이터 기반의 의사 결정 상황에 적용할 수 있다.
- 학습 요소**
 - 데이터 과학의 개념, 데이터 과학의 문제 해결 단계



21세기의 원유, 데이터

데이터 과학은 의료, 축산, 마케팅, 안전 등 다양한 분야에서 우리 삶을 바꾸고 있다.

사례 1

환자의 질환에 따라 건강 관리 패턴을 파악해 관련 정보 제공

<당뇨병 환자에게 응급 진료가 필요한 상황>

아래의 증상이 나타난 경우 즉시 의료진과 상담해야 합니다.

8시간 이상 심한 설사	4시간 이상 구토
38.3도 이상의 체온이 24시간 이상 지속	혈당이 300mg/dL 이상 지속되며 고혈당 증상 동반
저혈당 대처 후에도 혈당이 70mg/dL 이하	발에 상처, 물집, 염증이 생긴 경우

사례 2

돼지 사료 지급 패턴을 분석해 맞춤형 사료 지급



사례 3

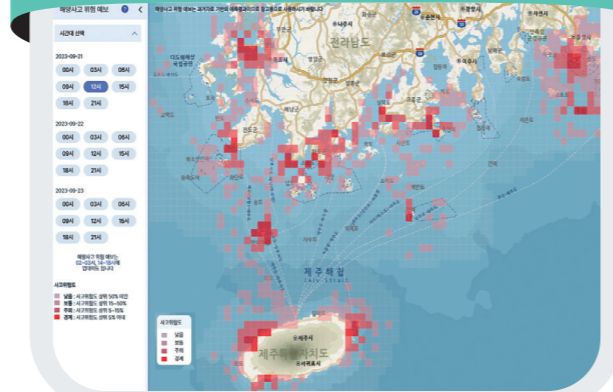
소상공인을 위한 데이터 기반 상권 분석 보고서

업소	업소수	전년대비	방범1km지역	전년대비
	21개	▼ 4.6%	137개	▲ 0.7%
매출	월평균 매출액	전년대비	월평균 매출건수	전년대비
	896만원	▼ 1%	1,183건	▼ 4.7%
인구	유통인구			
	유통인구	세대요율	세대상률	세대내비

* 업소별 월평균 매출 기준이며, 분석영역 내 업소수가 2개 이하일 경우 매출은 해당 업종 평균 데이터로 제공함

사례 4

빅데이터 기반 해양 사고 위험 예보



데이터가 우리 삶에 어떤 영향을 끼치고 있는지 이야기해 볼까?

1 | 데이터와 데이터 과학

01 데이터 과학의 개념

데이터 과학은 데이터를 기반으로 실제 현상을 이해하고 분석하기 위해 통계, 데이터 분석, 기계학습 및 관련 방법을 통합하는 학문 분야다. 특히 데이터의 관찰, 측정, 데이터 수집 및 해석 등의 단계를 거치면서 지식과 통찰*을 이끌어 내고, 과학적 접근 방법을 적용한다는 특징을 갖고 있다. 최근에는 기상청에서 날씨를 예측하거나, 스포츠 분야에서 데이터에 기반한 작전을 계획하는 데도 데이터 과학이 활용되고 있다.

통찰
사물이나 현상을 꿰뚫어 보는 능력 또는 특정 대상을 보편적인 시각 외에 다른 시각으로 볼 수 있는 능력

데이터를 분석한 결과 이번 태풍은 다행히 우리나라를 비껴갈 것 같습니다.



이번 대회에 승리 요인은 우리 선수들과 상대팀 선수들의 데이터를 분석해 전술을 구상한 결과라고 생각합니다.



02 데이터 기반 의사 결정의 중요성

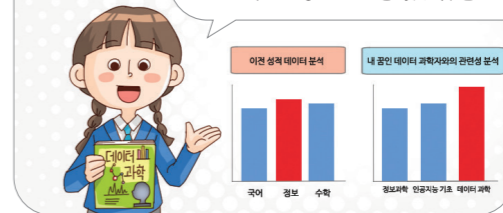
데이터를 수집·분석하여 의사 결정을 내리는 것을 '데이터 기반 의사 결정'이라고 한다. 축구 선수들의 데이터를 기반으로 선발 여부를 결정하는 것도 데이터 기반 의사 결정이다. 또한 우리 생활 속의 크고 작은 결정을 내릴 때도 데이터에 기반한 의사 결정을 하면 개인의 기억과 경험, 주관에 의해 의사 결정을 하는 방식에 비해 객관적이고 합리적인 선택을 할 수 있다.

예를 들어 학생들이 선택 과목을 결정할 때 이전 학기 성적, 과목에 대한 흥미도, 미래 산업의 발전 가능성 등 다양한 데이터를 고려하면 더 나은 선택을 할 수 있다. 또한 버스 노선을 결정할 때 승객들의 이동 패턴과 시간대별 승객 수 등의 데이터를 분석하면, 합리적인 버스 노선을 만들 수 있다.

지난 학기에 친한 친구가 선택한 과목을 듣고 마음에 들어서, 이번에도 그 친구가 선택한 과목을 선택하려고요.



제가 이전에 받은 성적과 진로 분야와의 관련성을 교과별로 수치화해 보니, 이 과목이 가장 저에게 잘 맞는 것 같아요!



경험, 주관에 의한 의사 결정과 데이터 기반 의사 결정의 사례

해 보기 2 하루 동안 생활 속에서 접하는 데이터 찾아보기

1 아침에 일어나서 밤에 잠들 때까지, 자신의 생활 속에서 어떤 데이터를 접하고 있는지 기록해 보자.

생활 속에서 접하는 데이터



	언제	어디에서	무엇을	어떻게	왜
예	아침에 옷 입기 전	집에서	기온을	스마트폰으로	어떤 옷을 입을지 결정하려고
1					
2					
3					
4					

2 생활 속에서 발생하는 데이터 중에는 어떤 목적을 위해 수집되는 경우도 있지만 그렇지 않은 경우도 있다. 만약 데이터를 추가로 수집한다면 어떤 문제를 해결할 수 있을지 생각해 보고, 친구들과 함께 이야기해 보자.

1 기존에는 수집되지 않았던 데이터 중에서 어떤 데이터를 추가로 수집할 수 있을까?

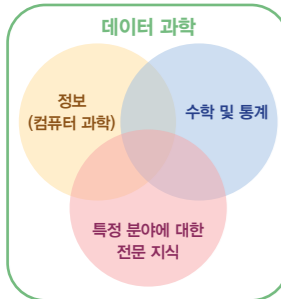
2 1에서 말한 데이터를 수집할 수 있다면 어떤 문제를 해결할 수 있을까?

3 데이터를 수집하는 과정에서 윤리적인 문제가 발생할 가능성은 없을까?

03 데이터 과학의 특징

데이터 과학은 다음과 같은 특징을 갖고 있다.

데이터 과학의 융합적 특징을 표현한 벤 다이어그램



▲ 그림 1-1 | 데이터 과학의 융합적 특징

융합적 학문

데이터 과학은 컴퓨터 과학, 수학 및 통계, 특정 분야의 전문 지식이 융합된 학문이다.

데이터 중심적 접근

데이터 과학에서는 문제를 해결하는 과정에서 데이터를 중심에 둔다. 구체적인 현상에 대한 데이터를 수집하고 이를 분석 및 해석하여 이론으로 만들어 간다.

데이터 과학 특징

유연성

데이터 과학에서 사용하는 도구와 기술은 지속적으로 발전하고 변화한다. 따라서 항상 유연한 태도로 새로운 도구나 기술을 받아들일 준비를 해야 한다.

윤리적 고려

데이터를 다루다 보면 개인 정보와 같은 민감한 정보를 다루게 될 수 있다. 따라서 데이터를 어떻게 처리하고 보호할 것인지에 대한 윤리적인 고려가 필요하다.

해 보기 1 다양한 교과목과 데이터의 연계성 분석해 보기

1 정보와 수학 외에 학교에서 배우는 교과목 중 데이터와 밀접한 관련이 있는 과목에는 어떤 것이 있을지 생각해 보고, 그 이유를 적어 보자.

- 예 체육 - 달리기 기록이 데이터로 쓰일 수 있기 때문에
- 경제 - 환율, 금리 등의 수치가 데이터로 쓰일 수 있기 때문에

2 교과목과 관련된 데이터를 분석하면, 어떤 것들을 알아볼 수 있을지 생각해 보자.

- 예 점심시간 전/후에 따른 달리기 기록 데이터를 분석하면, 식사 여부가 달리기 기록에 영향을 주는지 알 수 있다. 과거 금리 데이터의 변화를 분석하여 어떤 경제적 이슈와 연관된 것인지 유추할 수 있다.

2 | 데이터 과학의 문제 해결 과정

데이터 과학의 문제 해결 과정



▲ 그림 1-2 | 데이터 과학의 문제 해결 과정

맥락

사물이나 대상 등이 서로 연결되어 있는 관계

데이터 과학의 문제 해결 과정은 크게 문제 정의하기, 데이터 수집하기, 전처리 및 탐색하기, 모델링하기, 모델 평가하기, 모델 활용하기의 6단계로 구분할 수 있다. 그러나 6개의 단계를 순서대로 모두 거쳐야 하는 것은 아니다. 문제 상황에 따라 일부 단계는 생략될 수도 있고, 일부 단계는 반복될 수도 있으며, 경우에 따라 이전 단계를 다시 수행할 수도 있다. 따라서 문제 해결 과정의 전체적인 흐름을 이해하고, 각 상황에 따라 유연하게 문제 해결 과정을 설계할 수 있어야 한다.

01 문제 정의하기

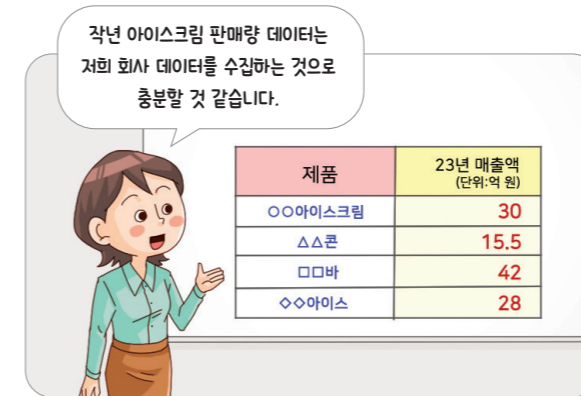
문제 정의 단계는 문제와 관련된 다양한 맥락*을 바탕으로 문제를 어떻게 정의할지 결정하고, 전체적인 계획을 수립하는 단계다.

예를 들어, 아이스크림 회사에서 '올여름 아이스크림 생산량'을 결정해야 할 경우, 기존 아이스크림 판매 데이터를 바탕으로 아이스크림 판매에 영향을 끼치는 요인들을 분석하여 문제를 정의해야 한다. 이때 기온이나 장마와 같은 날씨 요인이 아이스크림 판매에 큰 영향을 끼친다고 생각하고 문제를 정의할 수도 있고, 아이스크림 소비를 주도하는 청소년 및 청년층의 기호에 맞는 신상품을 개발하는 것이 판매에 큰 영향을 끼친다고 생각하고 문제를 정의할 수도 있다.



02 데이터 수집하기

데이터 수집 단계에서는 정의된 문제와 직접 연관된 데이터를 수집해야 한다. 만약 데이터의 양이 부족하거나 누락된 데이터나 오류가 있는 데이터가 많다면 뛰어난 분석 방법을 적용한다고 하더라도 좋은 결과가 나오기 어렵다. 따라서 데이터 수집 단계는 데이터 과학의 문제 해결 과정에서 매우 중요한 단계다. 이때 수집된 데이터의 양과 질에 따라 문제를 다시 정의해야 할 수도 있기 때문에 문제 정의와 데이터 수집 단계는 매우 밀접한 관계가 있다.



모델링

지도는 실제 지형이나 도로, 건물 등의 위치를 토대로 만들지만, 현실 세계의 모습을 단순화하고 추상화하여 표현한 것이다. 이와 같이 어떤 대상을 단순화하고 추상화하여 일반적으로 표현하는 과정을 모델링이라고 하고, 모델링된 결과를 모델이라고 한다.

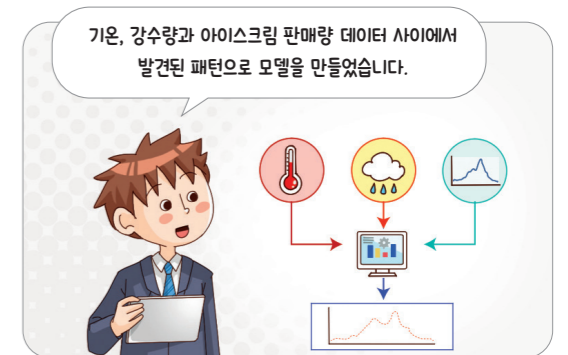
03 전처리 및 탐색하기

전처리 및 탐색 단계는 수집된 데이터에 비어 있는 값이나 정상적인 값에서 극단적으로 벗어난 값 등의 오류가 있는지 확인하고, 오류가 있을 경우 데이터를 분석하기 전에 적절한 방법으로 처리하는 단계다. 데이터에 이상이 없는지 확인하는 작업은 매우 중요한데, 데이터에 오류가 있을 경우 전체적인 문제 해결 과정에 심각한 장애를 초래하기 때문이다. 또한 데이터를 시각화하며 데이터의 다양한 특징을 탐색하는 것도 전처리 및 탐색 단계에서 해야 할 중요한 일이다.



04 모델링하기

모델링(modeling)* 단계는 데이터로부터 유용한 패턴을 추출하고, 그 패턴을 기반으로 모델을 구축하는 단계다. 모델은 데이터를 직접 분석하거나 자동화된 알고리즘을 통해 만들 수 있다. 예를 들어, 기온 및 강수량 데이터와 아이스크림 판매량 데이터 간의 패턴을 모델로 만들어서 아이스크림 판매량을 예측하는 모델을 만들 수 있을 것이다.





탐구 활동

05 모델 평가하기

모델 평가 단계는 모델링 단계를 통해 만들어진 모델이 문제를 해결하는 데 적합한 모델인지 종합적인 관점에서 평가하는 단계다. 평가 단계를 통해 모델을 만들 때 고려하지 못한 중요한 요소가 있는지 최종적으로 점검하고, 모델을 문제 해결에 적용할지 여부를 결정한다. 예를 들어, 기존에 만들었던 모델이 작년의 아이스크림 판매량을 적절하게 예측하는지 확인하여 모델의 활용 여부를 결정할 수 있다.



06 모델 활용하기

모델 활용 단계는 이전 단계에서 만든 모델을 문제 해결 및 의사 결정에 적용하는 단계다. 또한 모델을 적용하는 과정에서 모델이 문제 해결에 적합한지를 지속적으로 점검하는 것도 모델 활용 단계에서 이루어진다. 예를 들어, 올여름 아이스크림 생산량을 결정하는 의사 결정에 아이스크림 판매량 예측 모델을 활용할 수 있다. 이렇게 데이터에 기반한 판매량을 예측 모델에 활용하면 적절한 양의 생산을 통해 더 효율적인 기업 활동을 할 수 있게 된다.

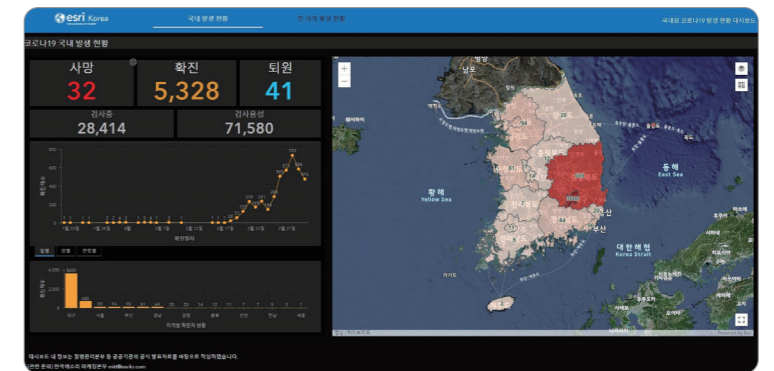


2020년에 발생한 코로나19 감염병 대응 과정은 데이터에 기반하여 문제를 해결한 대표적 사례다. 코로나19 등 감염병 방역 정책과 관련한 다음 기사를 읽고, 데이터 기반 의사 결정의 가치를 생각해 보자.

과학기술정보통신부(이하 과기정통부)와 질병관리청(이하 질병청), 정보통신산업진흥원은 코로나19 등 신종 호흡기계 감염병 대응 협력 상황을 점검하고, ‘한국형 신규 감염병 대응 시스템’ 구축 등 추진 경과를 공유했다.

현재 우리나라 대표 출연 연구 기관, 인공지능 기업, 병원 등 15개 기관이 참여해 감염병 전파 매개 변수 분석, 공간 내 감염원 전파 양상 분석, 유행 예측 모델 및 방법론 개선, 온라인 기반 유행 양상 분석, 의료 자원 관리 등 5개 과제를 중심으로 해결책을 개발 중에 있다.

질병청 위기대응분석관은 “코로나19 팬데믹과 같은 긴급 상황 대응 시, 과학적 근거에 기반한 신속한 의사 결정이 무엇보다 중요하다.”라며 “향후 질병청의 감염병 대응에 있어 과기정통부의 인공지능 기반 해법이 근거 중심 정책 수립에 큰 도움이 될 것으로 기대한다.”라고 밝혔다.



국내 코로나19 발생 현황 대시보드

출처 인공지능 · 데이터에 기반한 감염병 방역 정책 수립, ZDNET(2021. 11. 24.), <https://zdnet.co.kr/view/?no=20211124103726>

1 코로나19 감염병을 해결하는 과정에서 데이터에 기반하여 의사 결정을 내렸던 사례를 찾아보고, 데이터 기반 의사 결정의 가치를 설명해 보자.

2 자신의 관심 분야에서 데이터에 기반하여 문제를 해결한 사례를 찾아보고, 친구들에게 그 가치를 설명해 보자.

소단원 1분 요약

- 1 데이터 과학은 데이터와 빅데이터의 기반 위에 실제 현상을 이해하고 분석하기 위해 통계, 데이터 분석, 기계학습 및 관련 방법을 종합적으로 활용하는 학문 분야다.
- 2 데이터 과학의 문제 해결 과정은 크게 문제 정의하기, 데이터 수집하기, 전처리 및 탐색하기, 모델링하기, 모델 평가하기, 모델 활용하기의 6단계로 구분할 수 있다.

2

데이터의 형태와 속성



학습 목표

- 생활 속 데이터에서 정형 데이터와 비정형 데이터를 구분할 수 있다.
- 데이터 속성에서 데이터의 잠재적 가치를 판단할 수 있다.

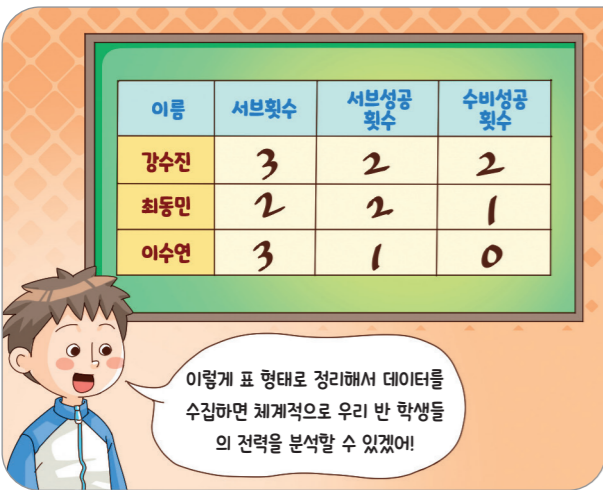
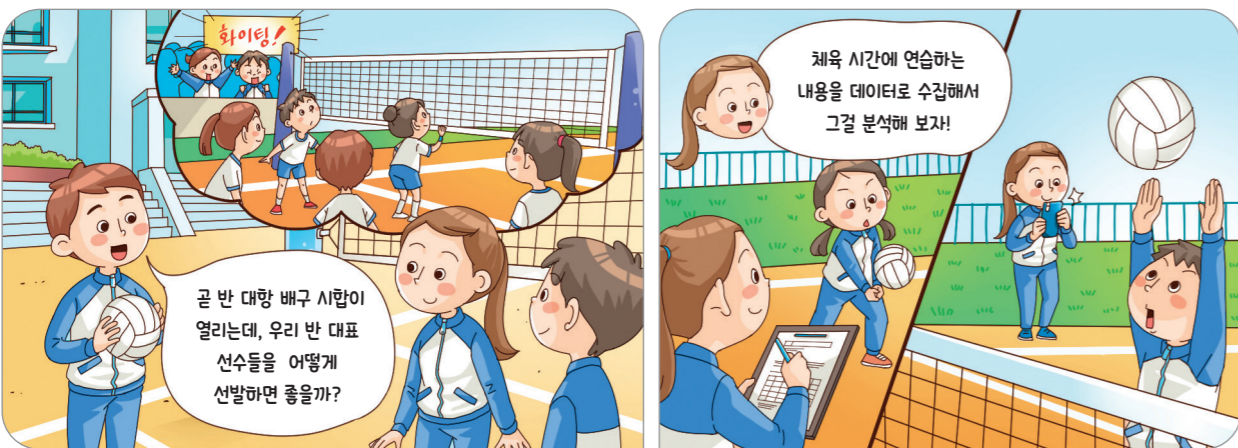
학습 요소

- 정형 데이터, 비정형 데이터, 데이터 속성, 데이터의 가치

생각 열기

데이터의 형태

우리 주변의 문제를 해결하기 위해 수집할 수 있는 데이터의 형태를 살펴보자.



위의 대표 선수 선발을 위해, 어떤 데이터를 추가로 수집할 수 있을까?

1 | 데이터의 다양한 형태

데이터를 구조적인 특성에 따라 분류하면 정형 데이터와 비정형 데이터로 나눌 수 있다.

01 정형 데이터

정형 데이터(structured data)는 데이터의 속성이나 구조가 명확하고 체계화된 데이터를 말한다. 정형 데이터의 대표적인 예는 행과 열로 이루어진 표 형태의 데이터다. 이때 데이터의 각 열은 해당 데이터의 특정 속성을 나타내고, 각 행은 다양한 속성으로 이루어진 구체적인 값을 갖는다. 예를 들어 오른쪽 표에서 서브 횟수 열을 보면 각 학생들이 서브를 시도한 횟수를 알 수 있고, 각 행에서는 학생별 이름, 서브 횟수, 서브 성공 횟수를 확인할 수 있다.

번호	이름	열(column)	
		서브 횟수	서브 성공 횟수
1	고지성	10	4
2	김지혁	8	3
3	나승빈	11	6
...
18	최지만	12	5
19	추성준	9	6
20	하승수	7	5

▲ 표 I-1 | 학생들의 서브 횟수를 나타낸 정형 데이터의 예

속성

데이터의 항목(열)이 가지는 특성이나 성질을 의미하며, 위의 표에서 이름, 서브 횟수 등이 속성에 해당한다.

이러한 정형 데이터는 데이터의 속성*을 파악하기 쉽고, 데이터의 수집·관리가 비교적 쉽다는 장점이 있어서 기업이나 공공 기관 등에서 많이 활용되고 있다.

정형 데이터

특징

- 구조가 명확하다.
- 속성을 파악하기 쉽다.
- 수집·관리가 비교적 쉽다.

예 일별 박스 오피스* 데이터

2023년 09월 20일(수)

순위	영화명	개봉일	매출액	매출액 점유율	매출액 증감(전일 대비)
1	잠	2023-09-06	258,641,035	26.5%	-16,578,784
2	그란 투리스모	2023-09-20	138,453,229	14.2%	136,869,229
3	베니스 유령 살인사건	2023-09-13	66,520,686	6.8%	-17,448,636
4	거미집	2023-09-27	59,490,000	6.1%	46,917,000
5	달짝지근해: 7510	2023-08-15	55,471,640	5.7%	-806,623
6	오픈하이머	2023-08-15	58,469,141	6.0%	-4,484,437
7	콘크리트 유토피아	2023-08-09	30,456,828	3.1%	-3,791,277

* 박스 오피스(box office) 연극, 영화, 공연 등에서 흥행 결과를 알 수 있는 총수입 금액

▲ 표 I-2 | 일별 박스 오피스 데이터

해 보기 1 정형 데이터 찾아보기

우리 학급과 관련된 정형 데이터에는 어떤 것들이 있을지 생각해 보자.


예 수학여행 참가 희망 여부 조사 등

02 비정형 데이터


비정형 데이터(unstructured data)는 구조를 명확하게 정의하기 어려운 데이터를 말한다. 비정형 데이터는 정형 데이터처럼 명확한 속성이 정의되어 있지 않지만, 데이터를 가공하여 속성을 추출할 수 있다. 예를 들어,

영화에 대한 사람들의 댓글 리뷰 데이터를 숫자로 변환하면, 이를 영화의 새로운 속성으로 사용할 수 있다. 대표적인 비정형 데이터로는 이미지, 텍스트, 음성 등 우리 생활에서 자주 접하는 데이터가 있다. 비정형 데이터는 구조에 제약을 받지 않고 자유롭게 표현할 수 있다는 장점이 있다.


비정형 데이터




이미지



텍스트



음성



영상

- 자유로운 구조를 가질 수 있다.
- 우리 생활에서 자주 접하는 데이터다.

알고 가기 반정형 데이터

정형 데이터와 비정형 데이터 외에도 반정형 데이터(semi-structured data) 형태의 데이터도 최근 데이터 과학에서 많이 활용되고 있다. 반정형 데이터는 일부분이 정형화된 데이터로, 구조가 존재하지만 구조를 바꿀 수도 있는 유연한 형태의 데이터를 말한다. 대표적인 반정형 데이터로는 XML(eXtensible Markup Language)과 JSON(JavaScript Object Notation)이 있다. 이러한 반정형 데이터는 데이터 구조를 쉽게 변경할 수 있고, 필요한 경우 표 형태로도 쉽게 변환할 수 있어 데이터 공유에 많이 활용된다.

공공행정 공공기관
미리보기

XML 한국소비자원_품목별 피해구제 사례

한국소비자원에서 제공하는 품목별 피해구제 사례입니다. 품목, 출처, 제목(비정형), 질문(비정형), 답변(비정형) 등의 정보를 포함하고 있습니다.※ 주의 : 답변 데이터는 답변 일자의 범...

제공기관 한국소비자원 수정일 2022-08-30 조회수 1546 다운로드 577 키워드 피해구제,사례,Q&A 다운로드

공공행정 국가행정기관
미리보기

XML 공정거래위원회_결정문 파일데이터

2008년도 이후의 공정거래위원회 공개 결정문(의결서, 결정서, 재결서, 시정권고서 등) 파일데이터를 제공함

제공기관 공정거래위원회 수정일 2022-08-05 조회수 716 다운로드 270 키워드 결정문,의결서,시정권고서 바로가기

국토관리 공공기관
미리보기

JSON 한국수자원공사_우량수위 관측정보

□ 사용방법 : 검색창에서 "우량수위 관측정보" 검색 -> "Sheet" 또는 "API" 라벨 메뉴 선택 -> 서비스목록에서 "우량수위 관측정보(DATA 조회)" 항목 선택 □ 개 요 : K-water에서 관리중...

제공기관 한국수자원공사 수정일 2022-07-19 조회수 6587 다운로드 543 키워드 시간별수위,우량,누적우량 바로가기

환경기상 자치행정기관
미리보기

XML 경기도 의정부시_공원수목_이미지정보

경기도 의정부시의 공원 내 수목의 이미지정보로 시도명, 시군구명, 관리번호, 파일명, 가로크기(픽셀), 세로크기(픽셀) 등의 항목으로 구성되어 있습니다.

제공기관 경기도 의정부시 수정일 2023-06-13 조회수 763 다운로드 310 키워드 공원,수목,이미지 다운로드

▲ 공공데이터포털에서 활용되는 반정형 데이터(XML, JSON)

2 | 데이터의 속성

다음은 어느 지역의 도서관 관련 데이터를 표로 나타낸 것이다. 표와 같은 정형 데이터는 데이터의 속성을 파악하기 쉽기 때문에 별도의 설명이 없어도 다양한 사실을 알 수 있다. 예를 들어, '시작시각' 속성을 통해 도서관이 언제 시작하는지 알 수 있고, '휴관일' 속성을 통해 어떤 요일에 휴관하는지 알 수 있다. 또 '위도'와 '경도' 속성을 활용하면 도서관의 위치까지 알 수 있다. 이와 같이 행과 열로 이루어진 표 형태의 정형 데이터는 구조가 명확하여 데이터의 속성을 쉽게 파악할 수 있다.

도서관명	도서관 유형	휴관일	평일운영 시작시각	평일운영 종료시각	토요일 운영 시작시각	토요일 운영 종료시각	열람 좌석수	자료수 (도서)	대출가능 권수	대출가능 일수	위도	경도
A 도서관	공공 도서관	일	8:00	22:00	8:00	22:00	92	60080	5	14	36.34926753	127.330255
B 도서관	공공 도서관	월	8:00	22:00	8:00	22:00	212	90102	5	14	36.33771587	127.3374864
C 도서관	공공 도서관	월	8:00	22:00	8:00	22:00	98	48726	5	14	36.38210426	127.3204213
F 어린이 도서관	공공 도서관	금	9:00	18:00	9:00	18:00	150	65915	10	14	36.30504773	127.3684557
G 교육청 도서관	공공 도서관	둘째, 넷째주 월	7:30	22:00	7:30	22:00	1600	304305	10	15	34.80731478	126.3761686
H 정보도서관	공공 도서관	수	9:00	23:00	9:00	22:00	136	66137	5	14	37.486252	126.959668
J 주민 자치 도서관	작은 도서관	토·일	9:00	18:00	0:00	0:00	20	3440	5	20	36.29561259	127.1412686
K 도서관	공공 도서관	월	8:00	21:00	9:00	18:00	115	68037	10	14	36.12620527	127.0973608
L 도서관	공공 도서관	월	9:00	18:00	9:00	18:00	158	55110	3	14	35.45273826	128.5305123

▲ 표 I-3 | 도서관 관련 데이터

한편 데이터의 속성은 문제 상황에 따라 그 중요성이 달라질 수 있다. 데이터를 활용해 우리 학교에서 가장 가까운 도서관이 어디인지 확인할 때는 위도와 경도 속성이 중요한 속성이 되고, 도서관에서 과제에 필요한 책을 언제 찾으러 갈 것인지 결정할 때는 휴관일과 시작시각이 중요한 속성이 될 수 있다. 또한 문제 해결에 필요한 새로운 속성이 추가되면 데이터의 가치가 더 높아질 수도 있기 때문에 데이터의 속성은 데이터의 가치를 결정하는 데 중요한 역할을 한다.

알고 가기 파이썬 언어의 다양한 데이터형

데이터의 각 속성에는 숫자, 문자 등 다양한 형태의 데이터가 저장된다. 이 중 파이썬(python) 프로그래밍 언어를 기반으로 데이터 분석에서 사용되는 대표적인 데이터형(data type)은 다음과 같다.

데이터형 이름	데이터의 형태	예시
object	텍스트(문자열)	'공공도서관', '작은 도서관'
int64	정수	92, 60080
float64	실수	36.34926753, 127.330255
bool	참(True) 또는 거짓(False)	True, False
datetime64	날짜와 시간	2025-03-01 13:30:03

해 보기 2 도서관 데이터의 다양한 속성이 가진 가치 살펴보기

1 21쪽에 제시된 도서관 데이터를 사용하여 다음 문제 상황을 해결하고자 할 때, 문제 상황에 따라 필요한 속성이 어떻게 달라질지 생각해 보자.

문제 상황	우리 집이랑 가장 가까운 도서관은 어디일까?	빌려야 할 책이 많다면 어떤 도서관으로 가는 것이 좋을까?	다양한 책을 찾고 싶다면 어떤 도서관으로 가는 것이 좋을까?
필요한 속성	위도, 경도		

2 제시된 속성 외에 어떤 속성이 있으면 도서관 데이터의 가치를 높일 수 있을지 생각해 보자.

예 내가 관심 있는 분야의 책이 몇 권씩 소장되어 있는지 기록된 속성

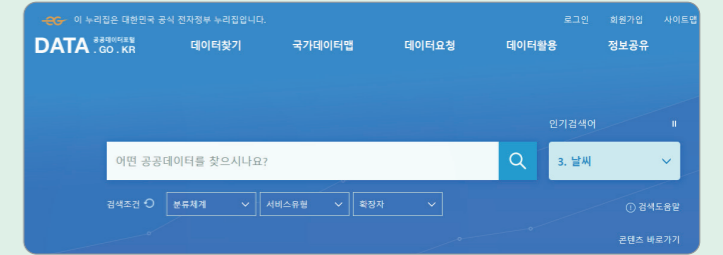
소단원 1분 요약

- 1 데이터를 구조적인 특성에 따라 분류하면 구조가 명확하게 정의된 정형 데이터와 구조를 명확하게 정의하기 어려운 비정형 데이터로 나눌 수 있다.
- 2 데이터의 속성은 데이터의 가치를 결정하는 데 중요한 역할을 하며, 문제 상황에 따라 그 중요성이 달라질 수 있다.

탐구 활동

'공공데이터포털'에서 다양한 형태의 데이터 수집하기

1 '공공데이터포털(<https://www.data.go.kr>)' 사이트는 공공 기관에서 제공하는 데이터를 쉽고 편리하게 사용할 수 있도록 정부에서 운영하는 곳이다. 이곳에서는 다양한 분야의 데이터가 정형, 비정형, 반정형 형태로 제공되고 있다.



공공데이터포털 화면

공공 데이터의 예

	A	B	C	D	E	F
1	사고유형대분류	사고유형중분	사고유형	사고건수	사망자수	중상자수
2	차대사람	횡단중	횡단중	12767	445	599
3	차대사람	차도통행중	차도통행중	3628	146	143
4	차대사람	길가장자리구	길가장자리구	1973	33	56
5	차대사람	보도통행중	보도통행중	2226	21	74
6	차대사람	기타	기타	14675	333	488
7	차대차	정면충돌	정면충돌	7892	205	353
8	차대차	측면충돌	측면충돌	71687	404	1775
9	차대차	후진충돌	후진충돌	3514	2	30
10	차대차	추돌	추돌	32717	379	756
11	차대차	기타	기타	44222	248	999
12	차량단독	전도	전도	956	82	38
13	차량단독	전복	전복	165	36	8
14	차량단독	공작물충돌	공작물충돌	2742	340	125
15	차량단독	주/정차차량	주/정차차량	23	3	
16	차량단독	도로이탈	도로이탈 추돌	385	73	19
17	차량단독	도로이탈	도로이탈 기타	156	18	7
18	차량단독	기타	기타	3401	148	111
19	철길건널목	철길건널목	철길건널목	1	0	

사고 유형별 교통사고 통계(도로교통공단)



독도의 사계절(외교부)

1 위에서 예시로 제시된 두 가지 데이터는 어떤 형태의 데이터인지 구분해 보자.

예시 데이터	정형 데이터	비정형 데이터
사고 유형별 교통사고 통계 데이터		
독도의 사계절 데이터		

2 '공공데이터포털' 사이트에서 관심 분야의 데이터를 내려받고, 어떤 형태의 데이터인지 구분해 보자.

내가 수집한 관심 분야의 데이터	정형 데이터	비정형 데이터

3 관심 분야의 데이터로 어떤 문제를 해결할 수 있을지 생각해 보고, 데이터의 잠재적 가치에 대해 이야기해 보자.

3

데이터셋과 데이터베이스



- 학습 목표**
 - 데이터셋과 데이터베이스를 설명할 수 있다.
 - 서로 다른 데이터셋을 통합하는 데이터베이스의 필요성을 설명할 수 있다.
- 학습 요소**
 - 데이터셋, 데이터베이스



데이터 통합의 중요성

다음 사례를 통해 서로 다른 데이터들을 통합하는 것이 왜 중요한지 생각해 보자.

학생들이 잔반을 너무 많이 남기는데, 이 문제를 어떻게 해결할 수 있을까?

아 그래, 데이터를 활용해 볼까?

먼저 월별 식단 데이터가 있으니 이걸로 메뉴와 재료를 확인해 볼 수 있겠네.

급식 일자	메뉴	칼로리
2023 0927	오곡밥, 토란국, 갈비찜, 콩나물무침	833kcal
2023 0928	잡곡밥, 무채국, 깍두기, 고등어구이	871kcal
2023 0929	참쌀밥, 만둣국, 오이냉채, 호박볶음	781kcal

그리고 날짜별 잔반량 데이터가 있으니, 이 둘을 분석해 보면 잔반을 줄일 수 있는 방법을 찾을 수 있을 것 같아!

급식 일자	메뉴	칼로리	잔반량
2023 0927	오곡밥, 토란국, 갈비찜, 콩나물무침	833kcal	25kg
2023 0928	잡곡밥, 무채국, 깍두기, 고등어구이	871kcal	19kg
2023 0929	참쌀밥, 만둣국, 오이냉채, 호박볶음	781kcal	32kg

선생님, 어떻게 이렇게 잔반량을 줄이셨나요?

서로 다른 데이터들을 함께 통합하며 분석하니 이전에는 분석할 수 없었던 잔반 문제의 원인을 찾을 수 있었어요!

하하

Q 위 사례에서 서로 다른 데이터를 통합했을 때 기존 데이터에서는 알 수 없었던 어떤 새로운 사실을 알아낼 수 있었을까?

1 | 데이터셋

데이터셋(dataset)이란 하나의 주제에 대한 여러 데이터들이 모인 집합(set)을 말한다. 예를 들어 급식 메뉴에 대한 여러 데이터를 하나로 모은 표는 급식 데이터셋이라 할 수 있고, 생활 속의 사물이나 동물에 대한 이미지 데이터를 하나로 모은 것은 생활 이미지 데이터셋이라 할 수 있다.

데이터와 데이터셋

〈급식 데이터셋〉 급식 메뉴에 대한 여러 데이터가 모여서 급식 데이터의 집합을 이룬다

학교명	급식 일자	급식 인원 수	요리명	칼로리 정보
A중학교	20230927	513	오곡밥 알토란국 매운돈사태갈비찜 삼색나물 배추김치 전통식혜 송편	833.3 kcal
B중학교	20230927	170	다시마무채국 깍두기 요구르트 열대과일샐러드 옛날도시락비빔밥 스마일감자&케첩	871.0 kcal
C고등학교	20230927	600	참쌀밥 쇠고기토란국 매운돼지갈비찜 호박볶음 배추겉절이 미나리약과&송편 육전&불고기산적&양파소스	781.4 kcal
D고등학교	20230927	1051	혼합잡곡밥 시금치된장국 매운돼지갈비찜 배추김치 송편 달밤라테 모듬전&초간장	1252.7 kcal
E고등학교	20230927	150	혼합잡곡밥 굴림고기만두국 계맛살오이냉채 스펀구이&케첩 배추김치 레드자몽에이드 짬뽕무침	1439.8 kcal

▲ 표 1-4 | 급식 데이터셋

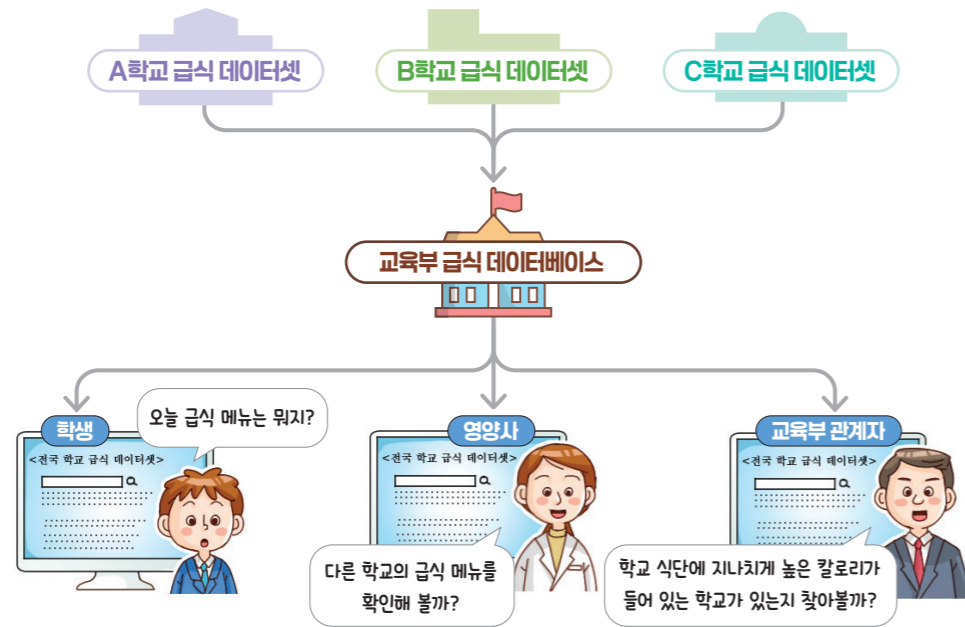
〈생활 이미지 데이터 및 데이터셋〉 CIFAR-10

- 비행기
- 자동차
- 새
- 고양이
- 사슴
- 개
- 개구리
- 말
- 배
- 트럭

+ CIFAR-10 데이터셋은 10개의 주제(비행기, 자동차 등)에 대한 이미지가 각각 6,000개로 구성된 이미지 데이터셋으로, 2009년에 알렉스 크리체프스키(Alex Krizhevsky), 제프리 힌튼(Geoffrey Hinton) 등이 수집 및 공개했다.

2 | 데이터베이스

데이터베이스(database)는 여러 사람이 공유하여 사용하기 위해 통합 관리되는 데이터의 집합으로, 데이터베이스 시스템을 사용하면 대규모의 데이터를 효율적으로 저장하고 검색하며, 업데이트하고 관리할 수 있다. 여러 사람이 대규모의 데이터셋을 통합적으로 관리하고 사용하려면 데이터베이스가 필요하다. 예를 들어, 데이터베이스를 활용하면 전국 모든 학교의 급식 데이터셋을 통합하여 관리하고 다른 학교의 데이터셋을 열람할 수 있다.



데이터베이스를 사용하여 데이터를 통합적으로 관리하면 다음과 같은 이점이 있다.

편리한 버전 관리

데이터베이스를 활용하면 언제, 누가, 어떤 데이터를 수정했는지 기록을 유지할 수 있다. 따라서 데이터에 오류가 발생했을 경우 원인을 찾아 그 이전 버전으로 쉽게 복구할 수 있다. 예를 들어, 누군가가 실수로 중요한 데이터를 삭제했다고 하더라도 그 이전 버전으로 복구할 수 있다.

데이터 무결성 유지

데이터의 제약 조건을 설정하여 데이터 입력, 수정, 삭제 시 발생할 수 있는 오류를 줄일 수 있다. 예를 들어, 식단의 칼로리 항목에 -1000과 같은 값이 입력되거나 같은 내용이 중복되어 입력될 경우 입력되지 않도록 할 수 있다.

데이터 보안 강화

데이터베이스는 사용자별로 데이터 접근 권한을 설정할 수 있다. 이를 통해 민감한 정보를 안전하게 보호하고, 각 사용자에게 필요한 데이터만 제공할 수 있다. 예를 들어, 다른 학교의 급식 데이터를 수정하지 못하게 접근 권한을 설정할 수 있다.

체계적인 데이터 관리

데이터베이스를 사용하면 데이터를 체계적으로 분류, 저장, 검색할 수 있다. 예를 들어, 급식 메뉴, 음식의 재료, 알레르기 정보 등 다양한 데이터를 체계적으로 저장할 수 있다. 또한, 학생이나 학부모들이 특정 날짜의 급식 정보를 알고 싶을 때, 데이터베이스를 통해 실시간으로 데이터를 검색할 수 있다.

3 | 데이터베이스의 통합적 활용

인터넷 게시판, 블로그, 메신저 등 대부분의 서비스는 데이터베이스를 사용한다. 그리고 데이터베이스에는 여러 종류의 데이터셋이 들어가는데, 이때 각각의 데이터셋은 표 형태로 저장되며 이를 테이블(table)이라고 한다.

만약 도서관에 데이터베이스가 없다면, 내가 원하는 책이 도서관에 있는지 확인하기 어려울 뿐 아니라, 대출 여부를 확인하는 것은 더욱 어렵다. 하지만 우리는 도서관 홈페이지를 통해 책이 도서관에 구비되어 있는지 쉽게 확인할 수 있고, 만약 누군가가 대출해 갔다면 반납 예정일이 언제인지도 알 수 있다. 또한, 여러 도서관 중에서 내가 찾는 책을 빌릴 수 있는 도서관은 어디인지도 쉽게 검색할 수 있는데, 이를 위해 데이터베이스는 여러 데이터를 통합하여 관리한다.



예를 들어 도서관 데이터베이스에는 도서관에서 소장하고 있는 책의 저자나 출판사 등이 저장된 책 목록 테이블과 회원의 이름과 회원 번호, 연락처 등이 저장된 명단 테이블, 그리고 누가 언제부터 언제까지 책을 대출했는지가 저장된 대출 기록 테이블이 있다.

데이터베이스의 각 테이블에 저장된 데이터들은 공통적인 속성을 기반으로 서로 연결될 수 있다. 예를 들어, 책 목록 테이블에 책의 대출 여부를 나타내는 속성이 없을 경우, 해당 책의 도서관관리번호를 대출 기록 테이블에서 확인하여 대출 여부를 확인할 수 있다.

해 보기 1

데이터베이스의 필요성 생각해 보기

모둠 활동 또는 과제를 수행했던 경험을 떠올려 보며, 데이터베이스의 네 가지 이점과 관련하여 어려움을 겪었던 경험을 모둠원들과 이야기해 보자.



1 책 목록 테이블에서 제목 검색

도서관리번호	제목	지은이	출판사	구입연도
354-2316	오만과 편견	제인 오스틴	민음사	2021
354-5672	위대한 개츠비	F. 스콧 피츠제럴드	문예출판사	2023
354-1249	앵무새 죽이기	하퍼 리	민음사	1998
354-9423	1984	조지 오웰	민음사	2014
354-7754	모비 딕	허먼 멜빌	문학동네	2024

2 책 목록 테이블에 있는 도서관리번호로 대출 기록 테이블에서 대출 여부 확인

도서관리번호	Member ID	대출일	반납일
354-2316	001	2023-10-01	2023-10-15
354-9423	002	2023-10-05	2023-10-19
354-1249	003	2023-11-01	2023-11-15

Member ID	이름	연락처	이메일 주소
001	홍길동	030-1234-5678	honggildong@example.com
002	강감찬	030-8765-4321	kanggamchan@example.com
003	이순신	030-5678-1234	yisunshin@example.com

표 I-5 | 책 목록 테이블

표 I-6(위) | 대출 기록 테이블, 표 I-7(아래) | 회원 명단 테이블

또한 각 도서관의 데이터베이스를 통합하여 전국 단위의 도서관 데이터베이스를 운영한다면 누구나 찾고자 하는 책이 어떤 도서관에 있는지 쉽게 검색할 수 있게 된다.



전국 도서관 데이터베이스

A 도서관 DB B 도서관 DB C 도서관 DB D 도서관 DB

4 | 데이터베이스의 생활 속 활용

데이터베이스는 기본적으로 추가 및 생성, 읽기, 갱신, 삭제와 같은 기능을 갖는다. 도서관의 사례를 통해 우리 생활에서 데이터베이스가 어떻게 활용되는지 간략히 살펴보자.

01 데이터 생성하기

데이터베이스에 새로운 데이터를 생성하거나 추가할 수 있다. 도서관에 새로운 책이 들어온 경우, 책 목록 테이블에 책에 대한 정보를 입력하면 데이터베이스에 새로운 책에 대한 데이터가 생성된다. 또한 새로운 테이블을 생성하거나 새로운 데이터베이스를 생성할 수도 있다.

도서관리번호	제목	지은이
354-2316	오만과 편견	제인 오스틴
354-5672	위대한 개츠비	F. 스콧 피츠제럴드
354-1249	앵무새 죽이기	하퍼 리
354-9423	1984	조지 오웰
354-7754	모비 딕	허먼 멜빌



새로운 책이 입고되면 해당 책에 대한 데이터를 생성한다.

표 I-8 | 데이터 생성

02 데이터 읽기

데이터베이스에 저장된 데이터를 읽어올 수 있다. 도서관 데이터베이스에 접근하면 책의 제목을 기준으로 조회할 수도 있고, 지은이 이름, 도서관에 등록된 연도 등 다양한 기준으로 조회할 수도 있다.

통합체육관광부 국립중앙도서관

로그인 회원가입 내서재 바꾸니 ENGLISH 홈페이지내 검색

자료검색 디지털컬렉션 신청참여 도서관 이용 도서관 소개

결과 내 재검색

전체 모비딕 다국어입력기 상세검색 유형별 저자별 주제별

"전체: 모비딕"에 해당하는 자료의 검색결과 총 350 건 입니다. 목차/초록으로도 검색해보시겠습니까? 검색 히스토리

유형전체 도서 고문헌 학위논문 잡지/학술지 신문 기사 멀티미디어 장애인자료 웹사이트 해외한국관련자료 외부연계자료 기타

자료형태 목록형 표지형 한자>한글 한글>한자 더보기

· 온라인자료 · 오프라인자료

도서 (162)건

1. 도서 모비딕

허먼 멜빌 원작/알렉스 니노 그림/김민구 역 | 교육출판공사 | 2005 | 비지일: 20050829 | 청구기호: 아808.3-5-107-42 |

자료이용장소: 어린이청소년도서관(역삼동) 연구자료실(멀티미디어, 3층)

도서 (162) 원문이용방법: [국립중앙도서관, 중앙도서관]-영양, 인쇄시, 과금

원문보기



03 데이터 갱신하기

데이터베이스에 저장된 데이터를 수정할 수도 있다.
만약 도서관에 책을 반납했을 경우 대출 기록 테이블의 데이터를 업데이트한다. 게시판의 경우 글의 내용을 수정할 때도 데이터 갱신이 이루어진다.



도서관리번호	Member ID	대출일	반납일
354-2316			
354-9423	002	2023-10-05	2023-10-19
354-1249	003	2023-11-01	2023-11-15

● 대출했던 도서를 반납하면 대출 현황 데이터베이스가 갱신된다.

▶ 표 I-9 | 데이터 갱신

04 데이터 삭제하기

데이터베이스에 저장된 데이터를 삭제할 수도 있다.
만약 도서관에서 오래된 책을 정리하여 버릴 때는 책 목록 테이블에서 해당 데이터를 삭제한다.



도서관리번호	제목	지은이	출판사	구입연도
354-2316	오만과 편견	제인 오스틴	민음사	2021
354-5672	위대한 개츠비	F. 스콧 피츠제럴드	문예출판사	2023
354-1249	영무새 죽이기	하퍼 리	민음사	1998
354-9423	1984	조지 오웰	민음사	2014
354-7754	모비 딕	허먼 멜빌	문학동네	2024

● 오래된 도서를 처리하거나 도서가 분실된 경우 해당 데이터를 삭제한다.

▶ 표 I-10 | 데이터 삭제

소단원 1분 요약

- 1 하나의 주제에 대한 데이터들의 집합을 데이터셋이라고 하며, 대규모의 데이터셋들을 효율적으로 통합하고 관리하기 위해 데이터베이스를 사용한다.
- 2 생활 속에서 이용하는 블로그, SNS, 온라인 쇼핑몰 등 대규모 데이터를 여러 사람이 공유하는 서비스들은 데이터베이스를 활용하여 데이터를 추가, 조회, 수정, 삭제하는 등의 작업을 처리한다.

만약 인터넷 쇼핑몰을 만든다면, 쇼핑몰의 회원 데이터를 관리하기 위한 데이터베이스를 어떻게 설계하면 좋을지 모둠을 구성하여 논의해 보자.

1 [쇼핑몰 사용자 테이블]을 설계해 보자.

	속성명(영문)	설명	데이터형 (정수/실수/텍스트/ 이미지/영상 등)	조건
0	uid	회원들을 구분할 수 있는 사용자 아이디	텍스트	중복되면 안 됨. 8자 이상, 32자 이하
1	passwd	사용자 비밀번호	텍스트	숫자, 특수 문자, 영문자 조합 10글자 이상
2	name	사용자 이름	텍스트	6자 이상, 30자 이하
3	email	사용자 이메일 주소	텍스트	이메일 형식에 맞아야 함. 중복되면 안 됨.
4	point	쇼핑몰 포인트	정수	0 이상이어야 함.

2 [쇼핑몰 제품 테이블]을 설계해 보자.

	속성명(영문)	설명	데이터형 (정수/실수/텍스트/ 이미지/영상 등)	조건
0	pid	쇼핑몰에 등록된 물품에 대한 아이디	정수	중복되면 안 됨.
1	like	사람들이 '좋아요'를 누른 숫자	정수	0 이상이어야 함.
2	pname	상품명	텍스트	6자 이상, 100자 이하
3	price	상품 가격	정수	0 이상이어야 함.
4	image	제품 이미지	이미지	이미지 형식의 확장자

3 [장바구니 테이블]을 설계해 보자.

	속성명(영문)	설명	데이터형 (정수/실수/텍스트/ 이미지/영상 등)	조건
0	uid	회원들을 구분할 수 있는 사용자 아이디	텍스트	[쇼핑몰 사용자] 테이블에서 불러올 것
1	pid	쇼핑몰에 등록된 물품에 대한 아이디	정수	[쇼핑몰 제품] 테이블에서 불러올 것
2	history	이전에 구입한 횟수	정수	0 이상이어야 함.
3				
4				

4

세상을 바꾸는 데이터 과학



- 학습 목표**
 - 데이터로 인한 사회 변화를 인식하고 설명할 수 있다.
 - 나의 진로 및 관심 분야와 관련된 데이터 기반 문제 해결 사례를 분석할 수 있다.
- 학습 요소**
 - 다양한 분야 속 데이터 과학, 미래 사회와 데이터 과학



데이터와 우리 생활

우리는 이미 데이터와 떼려야 뗄 수 없는 데이터 시대를 살고 있다. 데이터를 활용하여 많은 문제를 해결할 수 있지만, 아직 해결되지 않은 문제도 여전히 남아 있다. 다음 내용을 보고 질문에 답해 보자.

데이터 전송 문제

선생님! 와이파이 연결이 잘 안 돼요!

데이터 기반의 추천 문제

음 이 앱은 내가 원하는 노래를 너무 잘 추천해 준단 말이야. 칭찬해!

데이터 기반의 도시 설계

나는 안전하고 편리한 도시를 설계하는 사람이 되고 싶어!

데이터 오남용 문제(데이터 윤리 문제)

다양한 센서로 데이터가 이렇게 많이 수집되고 있는데, 혹시 개인 정보가 잘못 사용되는 않을지 잘 살펴봐야겠는데.



지금까지 해결되지 않은 문제 중에서 앞으로 데이터로 해결할 수 있는 문제에는 어떤 것들이 있을까?

1 | 데이터의 역사

데이터의 역사는 날짜를 기록하기 위해 벽이나 막대기에 줄을 긋는 행위에서부터 시작되었다. 그 이후 글자와 숫자가 발명되면서 본격적으로 데이터가 기록되기 시작했다. 기원전 3,000년경 고대 메소포타미아 문명에서 만들어진 썩기(설형) 문자는 상업 분야의 거래 내역을 기록하는 데 사용되었고, 이것이 오늘날 금융 데이터의 시초라고 할 수 있다. 최근에는 금융과 기술의 합성어인 '핀테크(fintech)'의 등장으로 다양한 금융 서비스에 혁신을 이루어 내고 있다.



▶ 최초 문자인 썩기 문자가 새겨진 점토판(맥아와 보릿가루의 수렴 내역을 적은 장부)



인구 데이터 역시 데이터의 역사와 함께 해 온 중요한 데이터다. 현재까지 알려진 가장 오래된 인구 데이터는 기원전 3,000년경 이집트에서 만들어졌다. 우리나라에서는 삼국 시대, 고려 시대, 조선 시대에 이르기까지 호구 조사란 명칭으로 인구 조사가 실시되었다. 최근에는 5년마다 전 국민의 일부를 대상으로 인구주택총조사가 이루어지는데, 이렇게 수집된 데이터는 국가 정책 수립에 중요하게 사용되며 누구나 다양하게 활용할 수 있도록 개방하고 있다.



▶ 조선 시대 전국의 호수와 인구수를 기록한 책



▶ 5년마다 전 국민의 일부를 대상으로 이루어지는 인구주택총조사

🔴 통계학(statistics)의 어원은 '국가'라는 의미가 담긴 이탈리아어 'statista'에서 유래되었다.

다양한 분야 속 데이터 과학

과거 손으로 기록되던 아날로그 데이터와 달리 현재의 데이터는 대부분 디지털 데이터라는 특징이 있다. 이로 인해 전 세계의 데이터가 실시간으로 연결되고, 고성능의 컴퓨팅 자원을 활용하여 예전에는 해결하지 못했던 다양한 문제를 해결할 수 있게 되었다. 다음은 데이터 과학이 다양한 분야의 문제 해결에 적용되고 있는 사례들이다.

의료 분야



환자 치료를 위한 예측 분석

환자의 의료 기록, 유전자 데이터, 생활 습관 등의 건강 관련 데이터를 분석하여 질병의 징후를 예측하고, 이를 바탕으로 적절한 예방 조치를 취할 수 있다.

신약 개발

생물 정보학을 활용하여 유전자 데이터, 단백질 구조, 대사 경로 등의 분자 구조 데이터셋을 분석하고 이를 활용해 새로운 약물 후보를 발견하거나 기존 약물의 새로운 용도를 찾을 수 있다.

원격 의료 및 원격 환자 모니터링

웨어러블 장치 및 모바일 건강 앱에서 수집한 데이터를 활용하여 개인화된 의료 서비스를 제공할 수 있으며, 거리와 시간의 제약 없이 환자의 상태를 지속적으로 모니터링할 수 있다.

제조 분야



장비 고장 예측

센서로부터 수집한 데이터를 실시간으로 분석하여 장비의 성능을 모니터링하고, 예상되는 고장을 미리 예측하여 생산 과정의 중단을 최소화할 수 있다.

공급망 개선

공급업체, 제조업체 및 유통업체의 데이터를 분석하여 재고 수준을 최적화하고, 수요 예측을 개선하여 효율적인 물류 관리를 할 수 있다.

품질 관리 및 결함 감지

기계학습 기반의 컴퓨터 비전 기술 등을 통해 제품의 품질을 자동으로 관리하고, 결함을 신속하게 식별하여 제조 과정의 효율성을 높일 수 있다.



교육 분야



맞춤형 학습 시스템

학생 개인의 학습 내역, 학습 스타일 등을 고려하여 맞춤형 교육을 제공할 수 있다. 이를 통해 학생들이 보다 효율적으로 학습할 수 있다.

학습 과정 분석

학생의 성과, 참여도, 행동 패턴 등에 대한 데이터를 수집하고 분석하여 학습 과정을 개선할 수 있다. 이를 통해 학생들의 학습 경험을 최적화하고 학업 성취도를 높일 수 있다.

학업 문제 조기 발견 및 지원

데이터 분석을 통해 학업에 어려움을 겪고 있는 학생이나 학습 동기가 떨어지는 학생 등을 조기에 발견할 수 있다. 이를 통해 학생의 문제가 더 심각해지기 전에 적절한 지원 및 개입을 할 수 있다.



해 보기 1

관심 분야에서 데이터 과학이 접목된 사례 찾기

조사하기

본문에 제시된 분야 외에 자신의 관심 분야 중 데이터 과학이 접목된 사례를 찾아보자.

관심 분야	데이터 과학이 접목된 사례

2 | 지속 가능한 미래를 만드는 데이터 과학

01 UN의 지속 가능 발전 목표와 데이터 과학

세계 연합(UN)은 전 세계 빈곤을 종식시키고, 지구를 보호하며, 2030년까지 모든 사람들이 평화와 번영을 누릴 수 있는 지속 가능한 미래를 만들기 위한 17개의 지속 가능 발전 목표(SDGs: Sustainable Development Goals)를 제시하였다. 이러한 지속 가능 발전 목표를 달성하는 데 데이터 과학은 핵심적인 역할을 한다. 데이터 과학을 활용하면 각 목표에 대한 정확한 모니터링과 평가가 가능하며, 이를 통해 국가와 국제 사회가 자원을 효율적으로 배분하고 전략을 수립할 수 있다. 보다 구체적으로 살펴보면 데이터 과학은 다음과 같은 다양한 기술과 접목되어 전 세계의 지속 가능한 발전에 도움을 주고 있다.

인공위성 기술

인공위성을 활용하면 환경 및 지리적 문제 점과 위험 요소를 데이터에 기반하여 파악할 수 있다.



웨어러블 기술

스마트 워치와 같은 웨어러블 기기는 환자들의 건강 데이터를 수집하고 치료하는 데 활용된다.



핀테크 기술

핀테크 기술의 발달로 디지털 송금이 가능하여 사람들의 재산 접근성이 증대되었을 뿐만 아니라 보다 안전하게 돈을 관리할 수 있다.



교통 데이터 모니터링 기술

지속적인 교통량 모니터링 및 교통 데이터 수집을 통해 도시 교통 계획을 수립하고, 개인의 교통 카드 데이터를 활용하여 인구 이동에 대한 예측을 할 수 있다.



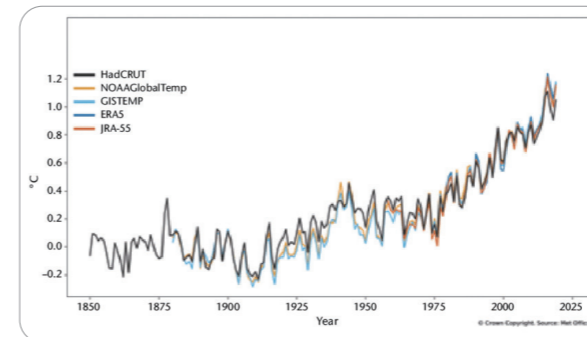
02 국가 지속 가능 발전 목표와 데이터 과학

우리나라에서도 국제 사회의 공동의 목표 달성에 기여하고 우리 사회가 처한 여러 문제들을 해결하기 위해 한국형 지속 가능 발전 목표인 국가 지속 가능 발전 목표(K-SDGs: Korean Sustainable Development Goals)를 수립했다. 여기에는 모두가 사람답게 살 수 있는 포용 사회 구현, 모든 세대가 누리는 깨끗한 환경 보전, 삶의 질을 향상시키는 경제 성장, 인권 보호와 남북 평화 구축, 지구촌 협력과 같은 5대 전략과 이를 실천하기 위한 17개 목표가 제시되었다.



▲ 국가 지속 가능 발전 목표

이러한 국가 지속 가능 발전 목표를 달성하는 데 데이터 과학은 다양한 영역에서 활용되고 있다. 예를 들어 데이터 분석을 통해 기후 위기를 진단하고 기후 위기에 대응하기 위한 구체적인 전략을 제시하고 있고, 데이터를 기반으로 작물을 재배하는 스마트팜을 통해 지속 가능한 식량 생산 체계를 구현하고 있다.



▲ 데이터 분석을 통한 기후 위기 진단: 대기질 및 폐기물 관리 등 도시가 갖는 부정적인 환경 영향을 감소시킨다(K-SDGs 11-6).



▲ 데이터 기반으로 작물을 재배하는 스마트팜: 지속 가능한 식량 생산 체계를 구축한다(K-SDGs 2-3).

해 보기 2 | 지속 가능한 미래를 만들기 위한 사례 조사하기 조사하기

본문에 제시된 사례 외에 인류의 지속 가능한 미래를 만들기 위해 데이터 과학이 활용되는 사례 또는 활용될 수 있는 분야에 대해 조사해 보자.

예) 지진, 산불, 감염병, 해양 선박 사고 등 재난 관련 데이터로 사고 예방하기

3 | 데이터 과학 속 데이터 윤리

데이터로 인해 세상의 많은 부분이 편리하게 변화하고 있지만, 데이터를 활용할 때 윤리적인 문제가 발생하기도 한다. 특히 데이터를 기반으로 한 인공지능의 발전에 따라 데이터 윤리는 더욱 중요해지고 있다.

01 데이터 편향성 문제

수집된 데이터의 분포가 고르지 못하고, 특정한 지역, 연령, 성별 등에 치우쳐 있을 경우 데이터에 편향성이 있다고 한다. 실제 현실을 대표하지 못하고 특정한 방향으로 치우쳐 있는 데이터를 가지고 모델을 만들면, 특정 대상에게 유리하거나 불리한 모델이 만들어질 수 있다.



▲ A사에서 채용 심사 시스템을 구축할 때 남성 데이터를 중심으로 모델을 학습한 결과, 여성에게 불리하게 적용되었다는 사실이 밝혀져 결국 시스템을 폐기하였다.

▶ 편향(偏向)
한쪽으로 치우침

데이터 편향성에 대한 자세한 설명은 47쪽을 참고한다.

이를 예방하기 위해서는 데이터 수집 단계에서부터 데이터가 편향되지 않았는지 검토하고, 데이터 분석의 전 과정에서 편향*을 최소화하기 위해 노력해야 한다.

02 데이터 속 혐오 표현 문제

데이터 편향성 문제가 수집된 데이터의 분포와 관련된 문제라면, 데이터의 질과 관련된 대표적인 문제로는 인공지능의 혐오 표현 문제가 있다. 과거 인공지능 챗봇이 잘못된 데이터를 학습하여 유대인 학살을 옹호하는 발언을 하여 문제가 된 것도 혐오 표현의 사례다. 이러한 문제가 발생하지 않도록 수집된 데이터를 필터링하는 과정도 필요하지만, 이는 결국 평소 사람들이 사용하는 데이터로 인해 발생한 문제이기 때문에 디지털 세상에서 바른 언어를 사용하는 습관을 가져야 한다.

알고 가기 혐오 표현을 줄이기 위한 노력 - 혐오 표현 데이터셋 공개

S사의 인공지능 센터는 악성 댓글과 혐오 표현에 대한 대규모 데이터셋을 공개하였다. 2019년 1월부터 2021년 7월 초까지 포털과 온라인 커뮤니티에서 수집한 55만 개의 데이터 중 1만 개를 선정하였다. 혐오 표현은 여성, 남성, 성소수자, 인종, 연령, 지역, 종교 등 8개의 범주로 분류하였다. 이러한 혐오 표현 데이터셋은 혐오 표현을 걸러내는 시스템을 구축하는 데 활용될 예정이다.



03 데이터 소유권 문제

학교 수업 중에 생성된 데이터는 학생의 것일까? 학교의 것일까? 서비스를 만든 기업의 것일까? 데이터가 중요한 자원이 됨에 따라 데이터의 소유권자를 누구로 볼 것인지 합리적으로 결정하는 것도 중요한 문제 중 하나다. 이러한 쟁점에 대해 충분히 논의하고 사회적 합의를 이끌어 내기 위해서는 사회 구성원들의 적극적인 참여와 관심이 필요하다.



▶ 데이터 소유권 문제

해 보기 3 데이터 윤리 문제와 관련된 사례 조사하기

▶ 데이터 윤리 문제와 관련하여 직접 경험했거나 간접적으로 들었던 사례가 있다면 소개해 보자.

소단원 1분 요약

- 1 데이터는 과거부터 금융, 정치 등 다양한 분야에 활용되어 왔으나 최근 디지털화되면서 의료, 제조, 교육 등 모든 분야에서 사회 변화를 주도하고 있다.
- 2 데이터는 지속 가능한 발전을 만드는 데 활용될 수 있다. 하지만 데이터를 활용할 때 윤리적 문제가 발생할 수 있으므로, 이를 인식하고 비판적인 태도로 활용할 줄 알아야 한다.



데이터 과학과 함께 4차 산업 혁명 시대가 도래함에 따라 다양한 진로와 직업의 새로운 기회가 열리고 있다. 다음 세 가지 분야를 살펴봄, 나의 진로 분야에 데이터 과학이 어떤 영향을 끼칠지 분석해 보자.

분야 ① | 의료 분야의 진단 및 치료 개선

데이터 과학을 활용한 인공지능 및 빅데이터 기술은 의료 분야에서 질병 진단과 치료 개선에 도움을 준다. 이를 통해 의사들은 환자의 증상과 병력을 기반으로 한 정확한 진단을 내릴 수 있으며, 개별화된 치료 방안을 강구하는 데 도움을 받을 수 있다. 또한 데이터 과학은 새로운 약물 개발 및 임상 시험 과정에서도 큰 역할을 하며, 의료 분야에서의 연구와 발전을 돕는다.

분야 ② | 환경 보호와 지속 가능한 발전

데이터 과학은 환경 데이터를 분석하여 기후 변화, 환경 오염, 자원 소모 등에 대한 통찰을 제공한다. 이러한 정보를 바탕으로 정책 결정자들과 기업들은 지속 가능한 발전을 도모하고 환경 문제를 해결하기 위한 전략을 수립할 수 있다.

분야 ③ | 스마트 도시 및 교통 관리

데이터 과학 기술은 도시의 교통 흐름, 에너지 소비, 안전 및 생활 편의를 개선하는 데 사용될 수 있다. 스마트 도시 기술을 통해 교통 정체를 줄이고, 에너지 효율을 높이며, 치안을 강화할 수 있다.

1 나의 진로 분야를 작성해 보자.

2 해당 진로 분야에서 데이터 기반 문제 해결 사례를 조사해 보자.

- 문제:
- 해결책:

3 데이터 과학이 해당 진로 분야에 끼치는 영향을 조사해 보자.

지식
충전소

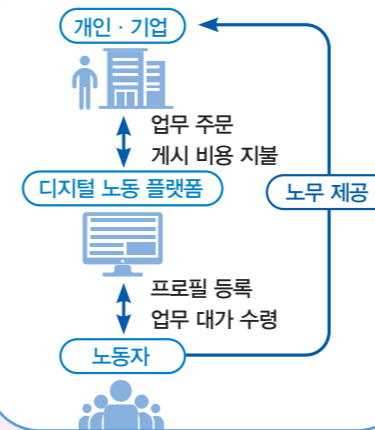
데이터 과학과 사회 문제

데이터 과학으로 인해 많은 기회가 우리에게 찾아왔지만, 빛에 그림자가 있는 것처럼 그 이면에는 데이터 과학으로 인한 문제 역시 자리 잡고 있다. 다음 세 가지 사례를 살펴보자.

노동 시장의 다변화와 유연성

기술의 발달로 새로운 일자리가 창출되고 다양한 형태의 직업이 생겨나고 있지만, 그로 인해 배달 기사 등 디지털 플랫폼 노동자들의 처우 문제가 발생하고 있다. 이들은 종종 사회적 보장이 제한되며, 불안정한 고용 및 저임금 문제를 겪을 수 있다.

플랫폼 노동 구조



디지털 리터러시*에 따른 격차 문제

키오스크와 같은 무인 기기들 또는 모바일 앱을 활용해 주문을 하는 것이 일상화되면서 편리해진 부분이 있지만, 한편으로는 디지털 리터러시가 부족한 사람들을 위한 대책이 필요하다. 이는 특히 노년층, 저소득층 등 사회적 약자들에게 영향을 미치며, 학생들 역시 디지털 리터러시에 따라 배움의 격차가 발생할 수 있다.

디지털 격차 문제



초연결 사회와 개인 정보 문제

초연결 사회에는 언제 어디서나 연결될 수 있어 새로운 기회를 많이 만들 수 있고 빅데이터를 활용해 많은 문제들을 해결할 수 있지만, 한편으로 개인 정보 유출 및 사생활 침해, 초연결 사회에 대한 피로감을 호소하는 경우도 발생할 수 있다.

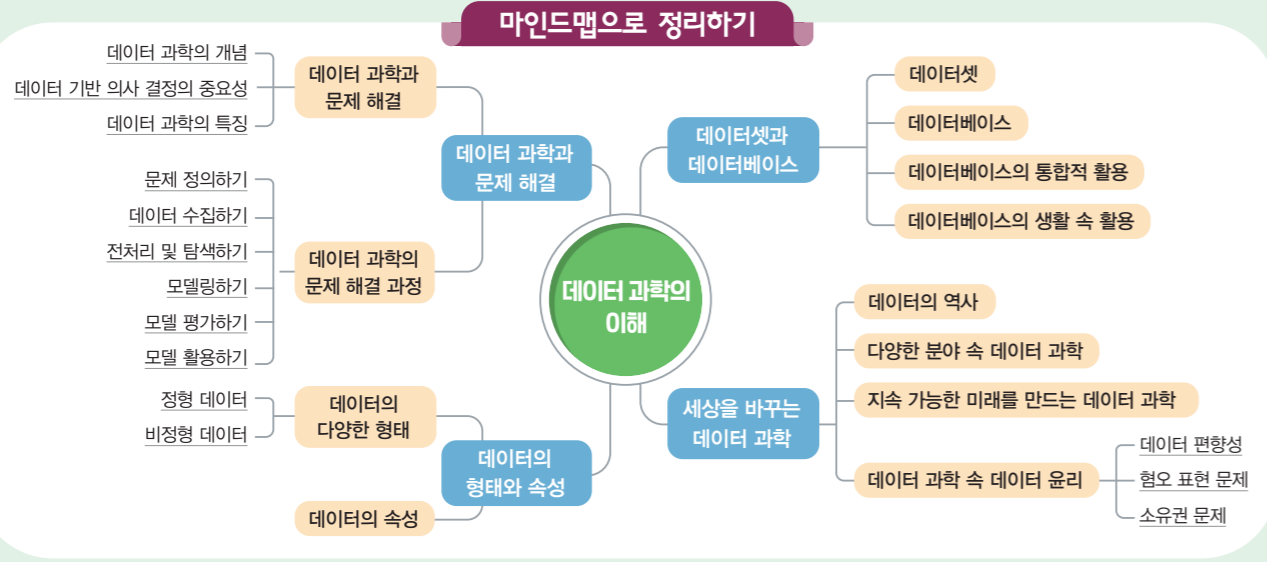
개인 정보 문제



이와 같이 데이터 과학으로 인해 발생하는 사회적 문제를 인식하며 데이터 과학의 빛과 어두움을 모두 이해할 줄 아는 비판적 태도를 갖추어야 한다.

📖 **디지털 리터러시(digital literacy)** 디지털 기술을 활용해 정보를 찾고, 이해하며, 활용하는 능력을 말한다. 즉, 디지털 기술을 사용하는데 필요한 능력과 지식을 갖추는 것을 뜻한다.

📖 **플랫폼 노동** 인터넷 기반의 디지털 플랫폼을 통해 제공되는 일시적이고 유연한 일자리를 말한다. 이러한 플랫폼은 앱 또는 웹 사이트 형태로 구성되어 있으며, 일자리 제공자와 일자리 수요자를 연결해 준다. 플랫폼 노동은 택배 배송, 음식 배달, 운송, 가사 도우미 등 다양한 서비스 분야에서 이루어질 수 있다.



선택형

01 다음에서 설명하는 개념으로 가장 적절한 것은?

데이터의 기반 위에 실제 현상을 이해하고 분석하기 위해 통계, 데이터 분석, 기계 학습 및 관련 방법을 통합하는 학문 분야이다. 특히 데이터의 관찰, 측정, 데이터 수집 및 해석 등의 단계를 통해 지식과 통찰을 이끌어 내고, 과학적 접근 방법을 적용한다는 특징을 갖고 있다.

- ① 확률과 통계
- ② 과학 데이터
- ③ 인공지능
- ④ 데이터 과학
- ⑤ 사회 과학

02 데이터 과학의 문제 해결 과정을 바르게 나열한 것은?

- ㉠ 모델 활용하기
- ㉡ 모델링하기
- ㉢ 전처리 및 탐색하기
- ㉣ 문제 정의하기
- ㉤ 데이터 수집하기
- ㉥ 모델 평가하기

- ① ㉠ - ㉡ - ㉢ - ㉣ - ㉤ - ㉥
- ② ㉥ - ㉤ - ㉡ - ㉠ - ㉣ - ㉢
- ③ ㉣ - ㉤ - ㉡ - ㉢ - ㉠ - ㉥
- ④ ㉥ - ㉣ - ㉤ - ㉢ - ㉡ - ㉠
- ⑤ ㉣ - ㉤ - ㉢ - ㉡ - ㉥ - ㉠

03 데이터베이스를 사용할 경우, 얻을 수 있는 이점으로 적절하지 않은 것은?

- ① 편리한 버전 관리
- ② 데이터 보안 강화
- ③ 데이터 무결성 유지
- ④ 체계적인 데이터 관리
- ⑤ 데이터를 활용한 미래 예측

04 데이터 과학의 문제 해결 과정 중 다음 설명에 해당하는 단계는?

수집된 데이터에 비어 있는 값이나 정상적인 값에서 극단적으로 벗어난 값 등의 오류가 있는지 확인하고, 오류가 있을 경우 데이터를 분석하기 전에 적절한 방법으로 처리하는 단계이다.

- ① 모델링하기
- ② 모델 활용하기
- ③ 모델 평가하기
- ④ 문제 정의하기
- ⑤ 전처리 및 탐색하기

05 다음 중 비정형 데이터가 아닌 것은?

- ① 우리 반 학생들의 출석 현황 데이터
- ② 1년 동안 친구들의 모습이 담긴 영상 데이터
- ③ 친구들과 스마트폰으로 함께 찍은 이미지 데이터
- ④ 친구들과 함께 합창 대회에 참여했던 음성 데이터
- ⑤ 1년 동안 우리 반의 학급 일기를 담은 텍스트 데이터

06 다음과 같은 형태의 데이터의 특징으로 알맞은 것은?

	번호	이름	서브 횟수	서브 성공 횟수
행(row)	1	고지성	10	4
	2	김지혁	8	3
	3	나승빈	11	6

	18	최지만	12	5
	19	추성준	9	6
	20	하승수	7	5

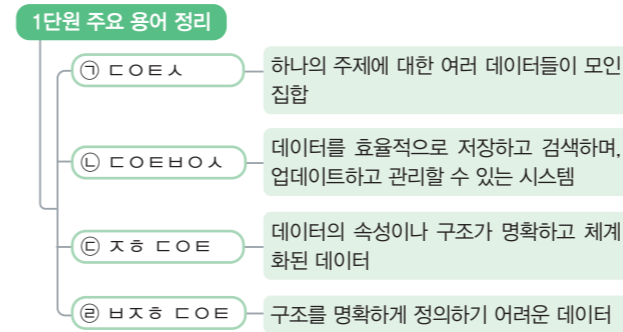
- ① 데이터의 행은 속성을 나타낸다.
- ② 데이터의 속성을 파악하기 쉽다.
- ③ 데이터의 속성이나 구조가 명확하지 않다.
- ④ 데이터의 수집, 관리가 비교적 어려운 편이다.
- ⑤ 기업이나 공공 기관에서는 많이 활용되지 않는다.

07 다음 중 데이터베이스의 기본 기능이 아닌 것은?

- ① 데이터 읽기
- ② 데이터 생성하기
- ③ 데이터 갱신하기
- ④ 데이터 삭제하기
- ⑤ 데이터 시각화하기

단답형

08 다음 초성 힌트를 보고, 1단원에서 배운 용어에 대한 마인드맵을 완성하시오.



- ㉠ _____
- ㉡ _____
- ㉢ _____
- ㉣ _____

서술형

09 교육 분야에서 데이터 과학이 활용될 수 있는 방법을 두 가지 이상 서술하시오.

교육 분야에서 데이터 과학이 활용된다면, 지금은 해결하지 못한 여러 문제들을 해결할 수 있어요! 예를 들면,



10 데이터 과학 속에서 발생할 수 있는 윤리적 문제를 두 가지 이상 서술하시오.