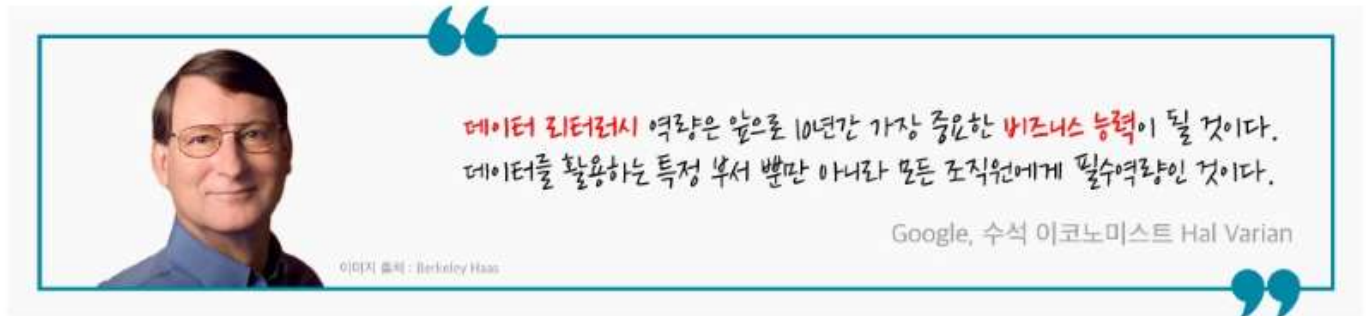


- ✓ "현대 사회에서 데이터 리터러시는 모든 조직 구성원에게 요구되는 필수적인 역량으로 자리잡고 있습니다."



[알아봅시다!](#)

✓ 데이터 리터러시의 이해

데이터 리터러시(Data Literacy)란?

Data literacy

Data literacy (데이터 리터러시)

“ 데이터를 정보로서 읽고, 이해하고, 만들고, 커뮤니케이션 하는 능력 ”

From Wikipedia, the free encyclopedia

Data literacy is the ability to read, understand, create, and communicate data as information. Much like literacy as a general concept, data literacy focuses on the competencies involved in working with data. It is, however, not similar to the ability to read text since it requires certain skills involving reading and understanding data.^[1]

- 데이터 리터러시(Data Literacy)란? **데이터 문해력**이란 뜻으로 다양한 방식으로 데이터를 읽고, 이해하고, 활용하는 능력
- 단순히 데이터 분석 도구를 다루는 기술적 능력을 넘어, 데이터를 통해 문제를 정의하고 해결책을 도출하는 종합적인 역량을 의미

데이터 리터러시의 중요성

- 데이터 기반의 의사결정 능력 향상
- 문제해결 능력 향상
- 비즈니스 성과 및 생산성 개선
- 디지털 혁신 및 경쟁력 강화
- 데이터 문화 형성

데이터 리터러시를 통해 우리는?

- 데이터를 활용해 해결할 수 있는 문제를 명확히 정의할 수 있다.
- 관련된 데이터 소스와 그 특성을 이해할 수 있다.
- 사용되는 데이터가 적합하고 충분한지 평가할 수 있다.
- 분석 결과를 해석하고 의미 있는 인사이트를 도출할 수 있다.
- 도출된 통찰을 바탕으로 의사결정을 내릴 수 있다.
- 사례를 통해 생성된 가치를 설명할 수 있다.

'어떤 문제'를 해결하기 위해 '어떤 데이터'를 '어떻게 모아', '어떤 방법(알고리즘)'으로 분석/모델링 하여 '어떤 비즈니스 가치(Value)'를 제공할 수 있다.

✓ 데이터 리터러시 핵심 역량



[Communication](#)

✓ 데이터와 인공지능의 개념

✓ 데이터의 정의 및 종류

데이터의 개념

이론을 세우는데 기초가 되는 사실이나 자료

- 기호화, 숫자화한 자료
- 존재적 특성 : 객관적인 사실
- 당위적 특성 : 추론이나 예측을 위한 근거

데이터의 유형

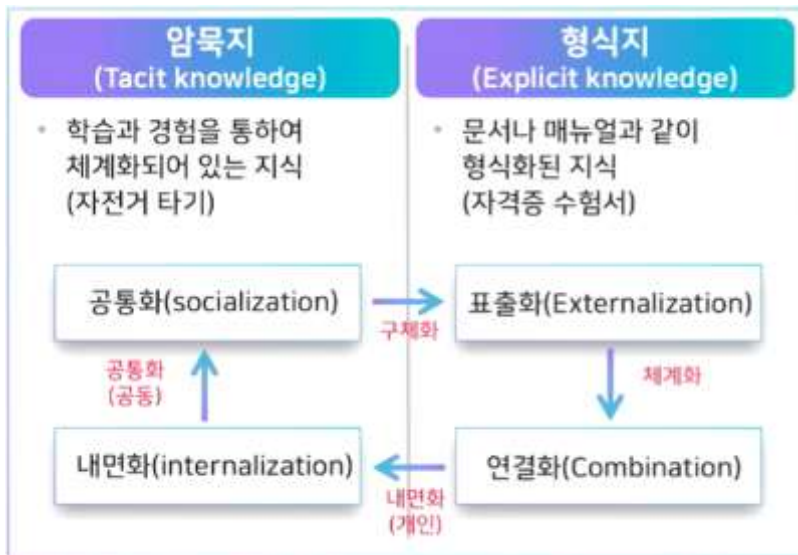


- 정량적 데이터 : 수치, 도형, 기호로 표현된 데이터 (예: 온도, 강우량, 전공 등)
- 정성적 데이터 : 언어나 문자로 표현된 데이터 (예: 리뷰, 주관식 설문조사 등)

데이터와 정보



지식의 상호작용



데이터의 가치와 데이터 기반 의사결정의 중요성

- 투명성 제고로 연구개발 및 관리 효율성 제고
- 시뮬레이션을 통한 수요 포착 및 주요 변수 탐색으로 경쟁력 강화
- 고객 세분화 및 맞춤 서비스 제공
- 알고리즘을 활용한 의사결정 보조 혹은 대체
- 비즈니스 모델과 제품, 서비스의 혁신
- 사회가 원하는 변화와 상황의 환경 탐색 및 미래 대응

✓ 인공지능 기본 개념

- 인간의 지능을 갖고 있는 기능을 갖춘 컴퓨터 시스템
- 인간의 지능을 기계 등에 인공적으로 시연(구현)한 것
- 범용 컴퓨터와 같은 지능을 만들 수 있는 방법론이나 실현 가능성 등을 연구하는 과학 분야를 지칭

데이터와 인공지능의 관계

- 데이터는 인공지능의 원료이며, 인공지능은 머신러닝과 딥러닝을 포괄하는 개념
- 인공지능(Artificial Intelligence, AI)은 인간의 지능적 행동을 모방하는 컴퓨터 시스템 또는 기계를 의미
- 인공지능은 문제를 해결하기 위해 특정 방법론을 선택하며 입력 데이터와 출력 결과 간의 관계를 학습하는 역할을 함
- 데이터 중심 최적 행동 결정 : 일정한 패턴을 학습하여 의사결정 규칙을 생성
- 알고리즘 중심 최적 행동 결정 : 알고리즘을 사용하여 입력 데이터를 학습하여 의사결정 기준을 생성

인공지능의 발전 단계



인공지능의 종류



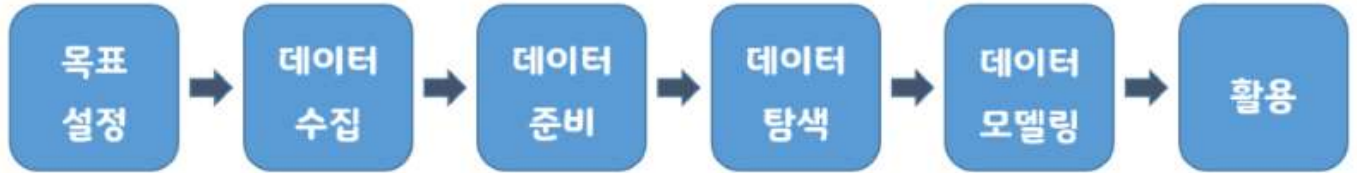
✓ 데이터 분석 방법론

데이터 분석 시 품질 확보를 위하여 단계별로 수행해야 하는 활동, 작업, 산출물을 정의한 것으로, 프로젝트의 성격과 요구사항, 환경에 따라 조합하거나 변경해서 사용함

- 하향식 접근법 (Top-Down Approach) : 문제가 주어지고 이에 대한 해법을 찾기 위해 데이터를 수집하고 분석하는 방식
- 상향식 접근법 (Bottom Up Approach) : 문제 정의가 어려운 경우 데이터를 기반으로 문제의 재정의 및 해결방안을 탐색하는 방식
- 프로토타이핑 접근법 : 먼저 분석을 시도해보고 그 결과를 확인하면서 반복적으로 개선해 나가는 방식

✓ 빅데이터 과학 방법론

데이터 과학은 빅데이터를 다루고 그 안에서 가치를 도출하는 과정으로, 일반적으로 6단계 구성으로 하향식 접근법과 프로토타이핑 접근법을 융합하여 체계적인 단계 수행과 반복적인 모델을 구축할 수 있도록 함



1. 목표 설정(문제 정의) : 데이터 분석의 목표를 설정하는 단계

데이터 분석을 통해 얻고 싶은 결과가 무엇인지 설정하고, 분석의 목적, 이유, 비즈니스에 미치는 영향, 구체적인 설계와 지표, 일정과 예상 결과 등에 대한 상세 요구사항 설정

ex) 2025년 수요 예측, 농작물 생산량 증가 방법, 영화 추천 등

2. 데이터 수집 : 분석에 필요한 데이터를 확보하는 단계

직접 데이터를 수집, 기존에 있는 데이터를 활용, 어떠한 방법으로 데이터를 확보하고 어떠한 형식으로 데이터를 저장하고 관리할지 결정

ex) 기존 데이터 활용(30년간 판매량, 기후 데이터 등), 직접 수집(센서 데이터, 크롤링 등)

※ 데이터가 무료로 공개되어 있어 이를 활용하여 다양한 데이터 수집 가능

- 공공 데이터 포털 (data.go.kr)
- 국가 통계 포털 (kosis.kr)
- Google Dataset Search (datasetsearch.research.google.com)

이외에도 다양한 공개 데이터 저장소가 있음

3. 데이터 준비 : 데이터 분석이 편리하도록 데이터를 처리하는 단계

- 데이터 여과 : 데이터 보정, 삭제, 중복성 확인
- 데이터 정제(Data Cleaning) : 누락된 값(결측치), 잘못된 값(이상치) 등 목적에 맞지 않는 데이터를 수정하거나 제거
- 데이터 통합 및 축소 : 필요한 데이터 통합 및 불필요한 항목 제거
- 데이터 변환(Data Tidying) : 데이터 분석을 위해 데이터를 알아보기 쉽고 처리하기 편리하도록 데이터 추출 또는 변환 (Feature Engineering)

ex) 결측치를 평균값 또는 0으로 수정, 이상치 샘플 제거, 온도와 습도 데이터, 평점만 추출, 카테고리화, 정규화 등

4. 데이터 탐색 : 데이터와 변수간의 관계를 이해하는 단계

탐색적 데이터 분석(EDA, Exploratory Data Analysis)을 통해 데이터의 변수 간 관련성 파악, 데이터 분포 확인 및 편차 및 패턴 존재 여부를 확인

ex) 상관분석, boxplot 등

5. 데이터 모델링 : 처리된 데이터 분석을 위한 모델을 구성하고 실행, 평가하는 단계

분석 목적에 따라 다양한 기법으로 데이터를 분석하는 것으로 통계 분석, 머신러닝, 딥러닝의 분석 기법을 활용하여 분석 결과를 도출

ex) 시계열 분석, 회귀 분석, 군집 분석 등

6. 활용 (결과 공유 및 자동화) : 도출된 결과에서 의미(인사이트)를 발견하고 이를 자동화하는 단계

분석결과가 예상하였던 의미나 정보가 도출되었는지 확인하고 검증, 잘못된 부분이 있을 경우 수정하여 재실행할 수 있도록 함

의미 있는 결과가 도출되었다면 이를 비즈니스에 반영할 수 있도록 보고서를 작성하거나 시스템에 반영

ex) 2025년 생산계획 수립, 농장의 온도 유지 장치 설정, 고객 맞춤 영화 추천 등

※ 시각화

데이터 또는 분석된 결과를 다양한 형태로 직관적이면서도 보기 편하게 표현하는 과정

- 전처리/탐색 : 데이터를 시각화하여 이상치 유무, 데이터 분포 확인
- 분석 : 분석된 결과를 시각화하여 의사결정의 근거 및 보고서에 활용

✓ 데이터 엔지니어링 개념

- 데이터 리터러시 역량 중 **manage**에 해당
- 수집된 데이터를 저장하고 정제하기 위하여 **데이터베이스 구축**
- 데이터베이스 내의 데이터를 활용하기 위한 **데이터 처리기술 필요**

✓ 데이터베이스 개념 및 주요 용어

데이터베이스 (Database)

여러 사람이 공유하여 사용할 목적으로 체계화해 통합, 관리하는 데이터/레코드의 집합

테이블명 : 데이터과목 학생별 성적

학번(PK)	학생명	학과번호 (FK)	학과명	성별	학점
202295153	김철수	10	컴퓨터공학과	남	3.2
202370132	최영희	06	수학과	여	3.8
202285435	홍길동	05	사회과학과	남	4.0
202425686	이몽룡	01	경영학과	남	3.6
202343287	성춘향	09	정보시스템공학과	여	3.6
202374126	박흥부	12	통계학과	남	4.3
202208048	박놀부	02	경제학과	남	3.1

DBMS (Database Management System)

이용자가 쉽게 데이터베이스를 구축하고 유지할 수 있도록 하는 소프트웨어

- 관계형 DBMS (RDBMS): **가장 널리 사용되는 유형**으로, 테이블 형식의 데이터 구조 사용, SQL(Structured Query Language)을 통해 데이터를 활용
- 계층형 DBMS: 트리 구조
- 네트워크형 DBMS: 데이터를 네트워크 구조로 표현
- 객체지향형 DBMS: 객체 개념을 데이터베이스 적용
- NoSQL(Not only SQL) : SQL만을 사용하지 않고 키/값, 문서, 그래프 등 다양한 데이터 유형의 데이터베이스를 사용하는 시스템



SQL(Structured Query Language)

관계형 데이터베이스에서 데이터를 관리하고 조작하는 데 사용되는 표준화된 프로그래밍 언어

- 데이터 정의 언어 (DDL, Data Definition Language) : 관계형 데이터베이스의 구조를 정의함
썬, 속성, 관계 인덱스 파일 위치 등 데이터베이스 고유의 특성을 포함함
create, alter, drop, truncate
- 데이터 조작 언어 (DML, Data Manipulation Language) : 데이터베이스 검색, 등록, 삭제, 갱신을 하기 위해 사용하는 데이터베이스 언어
데이터베이스의 검색 및 업데이트 등 데이터 조작을 위해 사용
select, insert, update, delete
- 데이터 제어 언어 (DCL, Data Control Language) : 데이터베이스에서 데이터에 대한 액세스를 제어하기 위한 데이터베이스 언어 또는 데이터베이스 언어 요소
박탈, 연결, 권한 부여, 질의, 자료 삽입, 갱신, 삭제 등
grant, revoke, commit, rollback

데이터 처리 기술과 프로세스

데이터 엔지니어가 하는 업무 중 가장 핵심은 데이터의 수집, 저장, 가공(ETL)으로, 이 업무를 하나로 합쳐서 데이터 파이프라인을 구축 및 관리함

- **E (Extract)** : 데이터를 추출(수집)하는 과정
기본적인 데이터 형태(Excel, CSV, JSON) 등을 직접 받아 오거나, 웹(Web), 앱(Application)의 로그(log)를 추출
(*로그(log)란 온라인 상에서 발생하는 모든 상황(클릭시간, 접속정보, 접속한 콘텐츠 등)에 대한 기록)

- **T (Transform)** : 데이터를 분석하기 쉬운 데이터로 변환하는 과정
Extract의 과정에서 다양한 형태(JSON, CSV 등)로 수집된 데이터들을 바로 저장하거나 분석하기에는 적합한 형태가 아닐 수 있기 때문에 row와 column의 형태로 변환해서 SQL, NoSQL 등으로 활용할 수 있도록 변환하는 과정
 - **L (Load)** : 데이터를 저장 또는 불러오는 과정
변환한 데이터를 데이터 레이크(Data Lake) 또는 웨어하우스(Warehouse)에 저장하는 과정
-

✓ 데이터 분석 보고서 작성 방법

데이터 분석 보고서란?

데이터를 활용 분석하여 사업기획, 창업, 정책 분석, 정책 제안, 현상 연구, 기타 아이디어를 보고서 형식으로 정리한 것으로, 데이터 분석을 통해 해결 방법을 찾고, 결과 및 활용 방안에 대한 보고서

- 주제를 정하고 주제에 포함된 문제에 대하여 데이터 분석을 통해 해결 방법을 찾고, 결과 및 활용 방안에 대하여 기술함
- 주제, 연구 배경, 문제점, 접근 방법, 통계분석결과, 문제해결방안, 결론(제안), 기대효과, 향후 과제(아이디어)의 내용으로 구성함
- 데이터분석 보고서는 뚜렷한 목표와 분석결과물과 결론이 있어야 함

보고서 작성 항목

- (주제 선정의 독창성) 주제 선정 및 배경
주제 선정 사유(동기, 문제점)를 구체적으로 기입
- (자료분석의 우수성, 데이터 활용성) 분석 내용·결과
 - (자료 활용 우수성) 연구 접근 방법에 대하여 간단히 기술
. 활용데이터를 요약적으로 제시
. 활용 데이터명 및 시계열 기입
. 분석 시 추가 활용한 공공·민간데이터 출처 기재(시스템, 공공데이터명 등)
 - (자료분석의 차별성) 분석 결과를 도출하기 위한 분석방법 기재
. 데이터 분석 결과를 자세히 기입
. 데이터 전처리를 하였다면, 내용 기입
. 도표, 이미지 활용 가능
- 분석 결과 및 활용 방향
 - (결과의 독창성·차별성) 분석을 통하여 얻어진 결과에 대하여 구체적인 핵심내용 기술
분석 결과를 기술함에 있어 시사점/정책 제안/창업 아이디어와 같은 핵심내용 중심으로 기술

- (연구 활용성) 연구 결과의 실현 가능성
연구결과를 활용한 정책 수립·개선, 사업가능성이나 연구 결과 활용 방향에 대해 구체적으로 작성
- 기대효과
시사점 제시, 아이디어의 파급성에 대해 구체적으로 작성 향후 연구 방향이나 제안

✓ 보고서 작성 전략

도입부, 방법론, 결과, 결론 및 권장 사항의 구조를 통해 명확한 흐름을 제공하고 정보의 전달을 최적화함

분석 기획서 작성 Tip

- 가용 데이터 고려 : 데이터 확보 가능 여부 및 데이터 유형 분석
- 적절한 활용방안, 유스케이스 탐색 : 기존의 솔루션 활용 및 과거 유사 분석 사례 참조
유사 시나리오를 통해 사용자의 공감대를 얻고 원활한 분석 수행 가능
- 장애요소에 대한 사전계획 수립 : 비용, 단순화, 실행환경, 성능 등 문제가 발생할 수 있는 부분에 대한 고려

분석 보고서 작성 Tip

- 고급 시각화 활용: 고급 시각화 도구와 기법을 사용하여 복잡한 데이터와 분석 결과를 명확하고 직관적으로 전달
대시보드, 인터랙티브 차트 등이 포함될 수 있음
- 기술적 깊이와 명확성: 분석 방법론과 결과 해석을 전문적이면서도 이해하기 쉽게 설명
전문 용어 사용 시, 그 의미를 명확히 정의하고 설명
- 실행 가능한 권장 사항: 분석 결과를 바탕으로 구체적이고 실행 가능한 권장 사항 제시
이는 조직의 전략적 목표 달성에 직접적으로 기여해야 함
- 프로젝트의 영향 평가: 분석이 조직에 미칠 장기적인 영향을 평가하고, 이를 보고서에 포함

[분석 기획서 작성](#)