

The top of the slide features several white, paper-cut style clouds of various sizes and a small white airplane, all hanging from thin vertical lines against a dark grey background.

ADsP 1. 데이터의 이해

키워드 살펴보기



1과목 키워드 살펴보기



데이터 정의	데이터 유형 정성적, 정량적	지식의 차원 암묵지, 형식지	암묵지와 형식지 상호작용 공통화-표출화-연결화-내면화	데이터 계층 구조 데이터 - 정보 - 지식 - 지혜
데이터베이스 특징	통합, 저장, 공유, 변화되는 데이터	DBMS 정의, 종류 RDBMS(관계), ODBMS(객체)	데이터베이스 관련 용어 DD(데이터 사전), ERD, SQL, 메타데이터	
데이터베이스 설계 절차	요구조건 분석 - 개념적 - 논리적 - 물리적 설계	NoSQL 특징, 종류 RDBMS보다 덜 제한적, 확장성, 세세한 통제, MongoDB, HBase, Redis		
기업 내부 데이터베이스 솔루션 1	OLTP, OLAP, CRM, SCM, Data Warehouse, Data Mart, ERP	데이터웨어하우스의 4대 특징 데이터 통합, 시계열성, 주제 지향적, 비소멸(비휘발)성		

1과목 키워드 살펴보기



기업 내부 데이터베이스 솔루션 2

BI, BA, 블록체인, KMS, RFID,

빅데이터 정의, 4V

Volume, Variety, Velocity + Value
투자비용 요소, 비즈니스 효과 요소

빅데이터 출현 배경

클라우드 컴퓨팅 → 경제적 효과 제공
양질 전환, 비정형 데이터, 처리기술

빅데이터 역할

석탄/철, 원유, 렌즈, 플랫폼
(공동 활용 목적, 페이스북, API..)

빅데이터 가치 산정 - 어려움

데이터를 언제, 어디서, 누가 활용
기존에 없던 가치 창출, 분석 기술 발달

빅데이터가 ... 본질적인 변화

사후처리, 전수조사,
양(Quantity), 상관관계

빅데이터 활용 기법

연관규칙학습, 유형분석(분류), 유전 알고리즘, 기계학습, 회귀 분석, 감정분석, 소셜 네트워크 분석(=사회 관계망 분석)

빅데이터 활용을 위한 3대 요소

자원, 기술, 인력

1과목 키워드 살펴보기



빅데이터 위기요인, 통제방안

사생활 침해, 책임 원칙의 훼손, 데이터의 오용
알고리즘미스트 등장!

개인정보 비식별화 기법

데이터 마스킹, 데이터 범주화, 가명, 잡음
첨가, 총계, 평균값 대체, 데이터 값 삭제

분석 애플리케이션 사례

에너지 - 트레이딩, 수요예측

데이터 사이언스의 정의, 다른 학문과 차이점

다양한 유형의 데이터, 분석+구현+전달=포괄적
총체적 접근법, 과학과 인문학의 교차로

데이터 사이언티스트의 역량

가트너 - 데이터관리, 분석 모델링, 비즈니스 분석, 소프트 스킬
하드 스킬, 소프트 스킬 (세부 항목도!)

정보 vs 통찰력

통찰력 : 모델링, 실험설계, 권고, 예측, 최적화

인문학 열풍

복잡한 세계, 비즈니스 중심 → 서비스, 시장창조

의사 결정 오류

로직 오류, 프로세스 오류

가치 패러다임 변화

Digitalization - Connection - Agency

데이터 사이언스의 한계와 인문학

분석은 가정에 근거 → 의구심, 가정과 현실
불일치에 대해 계속 고찰, 예측 위험 살피기



데이터의 크기

K-M-G-T-**Peta-Exa-Zeta-Yota**

구글의 실시간 자동 번역시스템

관련 빅데이터의 특징 : Volume

SQL 함수 중 '그룹' '조건'

SELECT 컬럼 FROM 테이블 WHERE 조건식 GROUP BY 그룹화할 컬럼 **HAVING 조건식** ORDER BY 정렬 컬럼

미래사회의 특성, 빅데이터의 역할

불확실성 - 통찰력, 리스크 - 대응력, **스마트 - 경쟁력**, 융합 - 창조력

2021년 : 키워드 살펴보기에서 8 ~ 9개 풀기 가능 했음

The top of the slide features several white paper-cut clouds of various sizes and a small white paper airplane, all hanging from thin vertical lines against a dark grey background.

ADsP 2. 데이터 분석 기획

키워드 살펴보기





분석 기획 유형 4가지

what(분석대상), how(분석방법)
최적화, 솔루션, 통찰, 발견

과제단위(당면한 분석 주제 해결), 마스터플랜 단위(지속적 분석문화 내재화)

과제단위 : Speed & Test, Quick-Win, Problem Solving
마스터플랜 단위 : Accuracy & Deploy, Long Term View, Problem Definition

분석 기획 시 고려사항

가용한 데이터, 적절한 유스케이스 탐색,
장애요소들에 대한 사전 계획 수립

데이터 유형 분류

정형, 반정형, 비정형
데이터 분류

데이터 저장 방식

RDBMS : Oracle, MSSQL, MySQL
NoSQL : MongoDB, HBase, Redis, Cassandra
분산파일시스템 : HDFS

분석방법론의 구성요소

상세한 절차, 방법, 도구와 기법
템플릿과 산출물

기업의 합리적 의사결정 장애요소

고정관념, 편향된 생각, 프레임링 효과



분석 방법론의 모델 3가지

폭포수 모델 - 하향식 접근 방법
 나선형 모델 - 점증적 개발
 프로토타입 모델 - 상향식 접근 방법

KDD 분석 방법론 (5단계)

데이터셋 선택 - 데이터 전처리 - 데이터 변환 - 데이터 마이닝 - 데이터 마이닝 결과평가

잡음, 이상치, 결측치 식별/제거

변수 선택, 차원 축소, 데이터셋 변경 작업

CRISP-DM 분석 방법론 (6단계) - 영문도!

업무(Business) 이해 - 데이터 이해 - 데이터 준비 - 모델링 - 평가 - 전개

모델 평가

분석 결과 평가
 모델링 과정 평가
 모델 적용성 평가

KDD

데이터셋 선택
 데이터 전처리

데이터 변환

데이터 마이닝

CRISP-DM

데이터 이해

데이터 준비

모델링

CRISP-DM : Business Understanding - Data Understanding - Data Preparation - Modeling - Evaluation - Deployment

CRISP-DM 분석 방법론 - 업무 이해 순서

업무 목적 파악 - 상황 파악 - 데이터 마이닝 목표 설정 - 프로젝트 계획 수립

2과목 키워드 살펴보기



빅데이터 분석 방법론

분석 기획

비즈니스 이해 및 프로젝트 범위 설정 - 프로젝트 정의 및 수행 계획 수립 - 프로젝트 **위험 계획** 수립

데이터 준비

필요 데이터 정의 - 데이터 스토어 설계 - 데이터 수집 및 정합성 점검

데이터 분석

분석용 데이터 준비 - 텍스트 분석 - 탐색적 분석 - 모델링 - 모델 평가 및 검증

추가적 데이터 확보가 필요한 경우
반복적인 피드백을 수행하는 구간

위험 대응 방법

회피(Avoid)
전이(Transfer)
완화(Mitigate)
수용(Accept)

시스템 구현

평가 및 전개



분석 과제 도출 방법

하향식 접근 방법

문제가 확실할 때 사용

상향식 접근 방법

문제 정의 자체가 어려운 경우 사용

디자인 싱킹(Thinking)

중요 의사결정시 상향식(발산)과 하향식(수렴)을 반복적 사용

하향식 접근 방법 (Top-Down Approach) 의 데이터 분석기획 단계

Problem Discovery - Problem Definition - Solution Search - Feasibility Study

문제 탐색 - 문제 정의 - 해결 방안 탐색 - 타당성 검토

상향식 접근 방법

비지도 학습, 디자인 싱킹의 발산 단계,
반복적인 시행착오를 통해 수정하며
문제 도출

비즈니스 모델 기반 문제 탐색 : 비즈니스 모델 캔버스, 5가지 영역 : **업무, 제품, 고객, 지원 인프라, 규제와 감사**

분석 기회 발굴의 범위 확장 : 거시적 관심 요인, **경쟁자 확대**, 시장의 니즈 탐색, 역량의 재해석 관점

외부 참조 모델 기반 문제 탐색 : 유사/동종 사례 벤치마킹 (Quick & Easy 방식)

분석 유즈 케이스 : 향후 데이터 분석 문제로의 전환 및 적합성 평가에 활용

2과목 키워드 살펴보기



분석 프로젝트 특징

분석 과제 주요 특성

Data Size, Data Complexity, Speed, **Analytic Complexity**, **Accuracy & Precision**

정확도, 복잡도는 트레이드 오프 관계
기준점을 사전에 정의해 두어야 함

Accuracy : 분석의 활용적 측면
Precision : 분석의 안정성 측면
트레이드 오프 관계

분석 프로젝트의 관리 영역

시간, 범위, 품질, 통합, 이해관계자, 자원, 원가, 리스크, 조달, 의사소통

2과목 키워드 살펴보기



분석 마스터플랜 수립

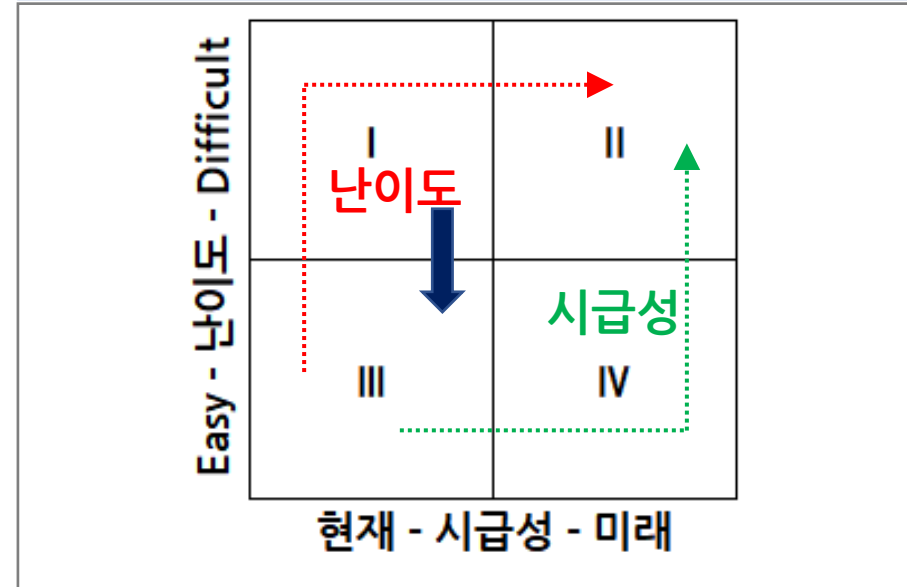
우선순위 고려 요소

전략적 중요도
ROI(투자자본 수익률)
실행 용이성

적용 범위/방식 고려요소

업무 내재화 적용 수준
분석 데이터 적용 수준
기술 적용 수준

포트폴리오 사분면을 통한 우선순위 선정





분석 거버넌스 체계 수립

- 기업에서 데이터가 어떻게 관리, 유지, 규제되는지에 대한 내부적인 관리 방식이나 프로세스
- 5가지 분석 거버넌스 체계 구성 요소 : 과제 기획/운영 프로세스, IT기술/프로그램, 분석 기획/관리 및 추진 조직, 분석교육, 데이터 거버넌스 (분석 비용 및 예산 없음)

1. 거버넌스 체계 개요

거버넌스, **분석 거버넌스**, 데이터 거버넌스

2. 데이터 분석 준비도

분석 업무 파악, 이력 및 조직, 분석 기법, 분석 데이터, 분석 문화, IT 인프라

3. 분석 성숙도 모델

비즈니스 부문, 조직/역량 부문, IT 부문을 대상

4. 분석 수준 진단 결과

도입단계, 활용 단계, 확산 단계, 최적화 단계로 구분하여 살펴봄

- CMMI (능력 성숙도 통합 모델) : ~ 성숙도를 평가하기 위한 모델

5. 분석 지원 인프라 방안 수립

데이터 거버넌스 체계요소 (세부적인 내용 꼭 알아 둘 것!)

- **데이터 표준화, 데이터 관리체계, 데이터 저장소 관리, 표준화 활동**

6. 데이터 거버넌스 체계 수립

데이터 거버넌스 구성 요소 : 원칙, 조직, 프로세스

7. 데이터 조직 및 인력 방안 수립

데이터 분석을 위한 조직 구조 : **집중형, 기능중심, 분산 (각 특징 중요!)**

8. 분석 과제 관리 프로세스 수립

과제 발굴 : 분석 아이디어 발굴, 분석 과제 후보 제안, 분석 과제 확정

과제 수행 : 팀 구성, 분석 과제 실행, 분석 과제 진행 관리, 결과 공유/개선

9. 분석 교육 및 변화 관리

모든 구성원이 데이터를 분석, 활용할 수 있도록 분석 문화를 정착, 변화시도



빅데이터 거버넌스 특징

- 분석 대상 및 목적 명확히 정의
- 품질 관리도 중요하지만, 데이터 수명주기 관리방안을 수립하지 않으면 ... 가용성, 관리비용 증대 문제
- ERD는 계속해서 변경사항을 관리해야 함
- 산업 분야, 데이터 유형, 정보 거버넌스 요소별로 구분하여 작성
- 분석 조식 및 인력에 대해 지속적인 교육과 훈련을 실시
- 개인정보보호 및 보안에 대한 방법 마련

관련 용어

- Servitization : 제조업과 서비스업의 융합을 나타내는 용어
- CoE(Center of Excellence) : 조직 내 분석 전문조직을 말함
- ISP(정보전략계획) : 기업의 경영목표 달성에 필요한 전략적 주요 정보를 포착하고, 주요 정보를 지원하기 위해 전사적 관점의 정보 구조를 도출하며, 이를 수행하기 위한 전략 및 실행 계획을 수립하는 전사적인 종합추진 계획
- Sandbox : 보안모델, 외부 접근 및 영향을 차단하여 제한된 영역 내에서만 프로그램을 동작시키는 것



SQL 구문 - DDL, DML, DCL

DDL : CREATE, ALTER, DROP, TRUNCATE
DML : DELETE, INSERT, UPDATE, SELECT
DCL : GRANT, REVOKE, COMMIT, ROLLBACK

분석 과제 관리 프로세스

발굴 단계, 수행 단계의 종류, 분석과제 중 발생한 시사점과 분석 결과물은 풀(Pool)로 관리, 분석 과제로 확정된 분석과제를 풀로 관리 X

분석 마스터플랜과 ISP의 관계

분석 마스터 플랜

- 일반적인 ISP 방법론을 활용하되 데이터 분석 기획의 특성을 고려하여 수행,
 - 기업에 필요한 데이터 분석 과제를 빠짐없이 도출한 후 과제의 우선순위를 결정하고 단기 및 중,장기로 나누어 계획을 수립함
- ISP - 기업 및 공공기관에서는 시스템의 중장기 로드맵을 정의하기 위한 ISP를 수행한다 (중장기 마스터플랜을 수립하는 절차)

분석 성숙도 모델 단계

도입/활용/**확산**/최적화 중
어떤 것인가?

분석 준비도 영역

분석 업무파악, 인력 및 조직,
분석 기법, 분석 데이터, 분석
문화, **IT 인프라** 중 무엇?

빅데이터 7V

Volume, Variety, Velocity, Value
Veracity, Validity, Volatility

2021년 : 키워드 살펴보기에서 6 ~ 8 개 풀기 가능 했음

A decorative header featuring stylized white clouds and a small white airplane flying across a dark gray background.

ADsP 3. 데이터 분석

R 관련 - keyword, 총정리





R 관련 강의

3 - 10, 11, 17, 18 필수

summary 함수

연속형, 범주형 변수의 해석

- 연속형 : 최솟값, 1사분위수, 중간값, 평균, 3사분위수, 최댓값
- 범주형 : 범주, 범주별 데이터 개수가 콜론으로 구분되어 표시

결측치 대처법

단순 대처법: 완전 응답 개체분석, 평균대치법, 단순확률 대처법
다중 대처법: 추정량 표준오차의 과소추정, 계산의 난해성 문제

이상값(outlier)

- 평균 - 3*표준편차, 평균 + 3*표준편차 밖의 값
- $Q1 - IQR * 1.5$, $Q3 + IQR * 1.5$ 밖의 값
- $IQR = Q3 - Q1$
- 분석 대상이 될 수 있어 무조건 삭제 X

2021년 R :

vector 1개 (평균, 중앙값, 최빈값, 범위 구하기-손으로), matrix 1개(교재 - 266)

summary 해석 2개 (객관식1, 주관식1, 교재 - 267, 268, 269)

box-plot 2개 (교재-270, 271, 273 ~277 문제 풀이), 이상치 1개 (교재 - 302)



과목	QnA 링크
데이터 이해	https://colab.research.google.com/drive/1cCk43pbnapr0mx0SiC8ssN3ya5TJwSqG
데이터 분석 기획	https://colab.research.google.com/drive/1J7U19W-bVobaAob0wQKIsho3Uo8HkeVq
R	https://colab.research.google.com/drive/1VmixW_RYpn8_XycGAjggSCZ00sR-_8Ke
데이터 분석(필독)	https://colab.research.google.com/drive/1QDuCKk86lKTP8ox0Tw0o1D2935Tu-5-e
계속 업데이트 되고 있습니다. 좋은 질문들을 모아보았으며, 데이터 분석에 대한 것에서 시계열의 정상성, 잔차, 주성분분석, 최단연결법 군집 나누기, 덴드로그램 해석, 신뢰 구간 등의 내용을 확인해 보시기 바랍니다. *^^*	

과목	요약 링크
데이터 이해	https://colab.research.google.com/drive/1Pd-At18xeHJ39udWpMpV4MPmPnR25KnG
데이터 분석 기획	https://colab.research.google.com/drive/1_wCPXu4EdmW2Baq92mprZ5Rblriyr19l
R	https://colab.research.google.com/drive/1NlLoFbRgr4_pGDQG5iQ3hpVW40xWQsNi
데이터 분석-통계분석	https://colab.research.google.com/drive/1STSINoLgwmQNVaALizJN1BYUzlrqzqCD
데이터 분석-정형데이터	https://colab.research.google.com/drive/1C-6zmHW49XmscbBopfm3PDDMZCxbt5f1