

데이터분석전문가(ADsP) Part 1

데이터 이해

01

1장 데이터의 이해

2장 데이터의 가치와 미래

3장 가치 창조를 위한 사이언스와 인사이트



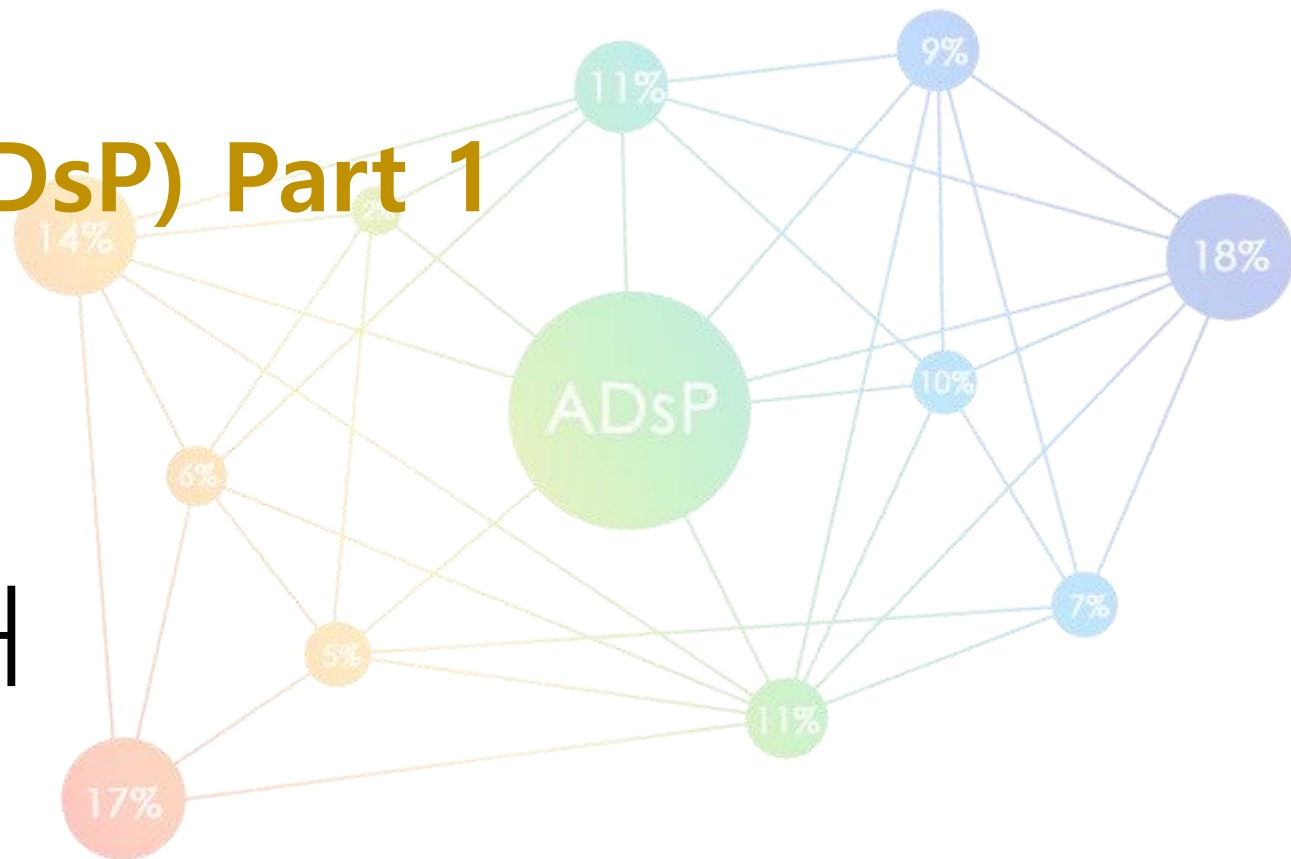
데이터분석준전문가(ADsP) Part 1

데이터 이해

01

1장 데이터의 이해

1. 데이터와 정보
2. 데이터베이스 정의와 특징
3. 데이터 베이스 활용



1. 데이터의 이해



1장 데이터와 정보

01 데이터 정의?

데이터는 객관적 사실 → 존재적 특성 동시에 추론, 예측, 전망, 추정을 위한 당위적 특성(~해야만 하는)을 가짐.
이는 다른 객체와의 상호관계 속에서 가치를 갖는 것을 의미함.

02 데이터 유형

정성적 데이터: 언어, 문자 등으로 기술(비정형 데이터 형태로 저장, 분석에 시간과 비용이 필요)

정량적 데이터: 수치, 기호, 도형으로 표시(데이터 양이 증가하더라도 저장, 분석이 용이)

03 지식경영

지식경영 = 암묵지+형식지의 상호작용 속에서 지식이 공유되며 생성이 된다.

암묵지: 학습과 체험을 통해 개인에게 습득, 시행착오와 오랜 경험을 통해 개인에게 습득된 무형 지식으로 외부와 공유되기 어려움. ex) '김장김치담그기', '자전거타기'

형식지: 교과서, 메뉴얼, 비디오, DB 형상화된 지식을 의미. 지식의 공유가 용이

1. 데이터의 이해



04

암묵지+형식지의 4단계 지식전환 모드

1단계: **공통화** (암묵지 지식 노하우를 다른 사람에게 알려주는 것)

2단계: **표출화** (암묵적 지식 노하우를 책이나 교본 등 형식지로 만드는 것)

3단계: **연결화** (책이나 교본(형식지)에 자신이 알고 있는 새로운 지식(형식지)를 추가하는 것)

4단계: **내면화** (만들어진 책이나 교본(형식지)을 보고 다른 직원들이 암묵적 지식(노하우)을 습득)

→ 이것을 SECI 모델(Socialization-Externalization-Combination-Internalization Model)

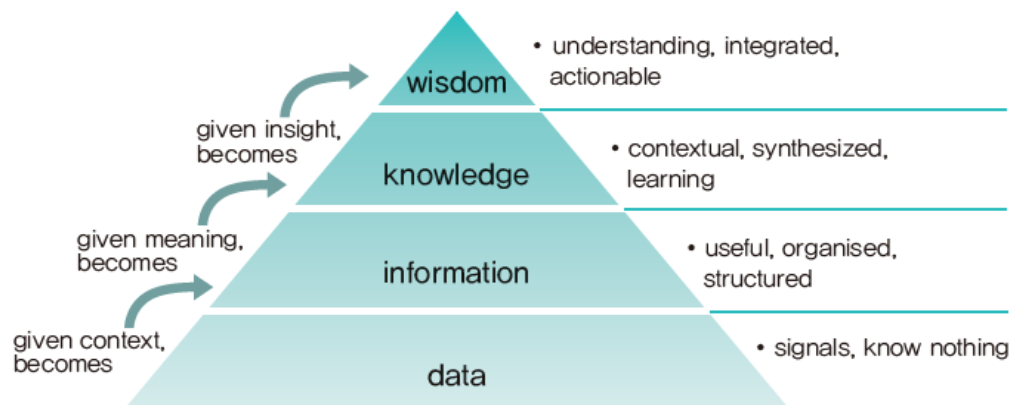
1.데이터의 이해



05

데이터와 정보와의 관계

Data → Information → Knowledge → Wisdom 계층구조는 데이터의 당위적 특성과 관련



Data: 존재 형식을 불문하고, 타 데이터와의 상관관계가 없는 가공하기 전의 순수한 수치나 기호

→ A 마트 100원, B 마트는 200원 연필을 판매

Information: 데이터의 가공 및 상관관계간 이해를 통해 패턴을 인식하고 의미 부여

→ A 마트의 연필가격이 더 싸다.

Knowledge: 상호 연결된 정보 패턴을 이해하여 이를 토대로 예측한 결과물

→ 상대적으로 저렴한 A마트에서 연필을 사야겠다.

Wisdom: 근본 원리에 대한 깊은 이해를 바탕으로 도출되는 아이디어

→ A 마트의 다른 상품들도 B마트보다 쌀 것이라고 판단

1. 데이터의 이해



06 데이터베이스란?

데이터베이스는 “동시에 복수의 적용 업무를 지원할 수 있도록 복수 이용자의 요구에 대응해서 데이터를 받아들이고 저장, 공급하기 위하여 일정한 구조에 따라서 편성된 데이터의 집합”, “관련된 레코드의 집합, 소프트웨어로는 데이터베이스 관리시스템(DBMS)을 의미

07 관계형 데이터베이스 관리시스템(RDBMS) VS 객체 지향 데이터베이스 관리시스템(ODBMS)

관계형 데이터베이스 관리시스템(RDBMS)

일련의 정형화된 테이블로 구성된 데이터 항목들의 집합체로서, 그 데이터들은 데이터베이스 테이블을 재구성하지 않더라도 다양한 방법으로 접근하거나 조합이 가능

객체 지향 데이터베이스 관리시스템(ODBMS)

객체들을 생성하여 계층에서 체계적으로 정리하고, 다시 계층들을 하위 계층이 상위 계층으로부터 속성과 방법들을 물려받을 수 있는 계승 가능한 구조로 구성됨

1. 데이터의 이해



08

데이터베이스 특징

- ① **통합된 데이터(integrated data)**: 데이터베이스에서 동일한 내용의 데이터가 중복되어 있지 않다는 것을 의미
- ② **저장된 데이터(stored data)**: 자기 디스크나 자기 테이프 등과 같이 컴퓨터가 접근할 수 있는 저장매체에 저장되는 것을 의미
- ③ **공용 데이터(shared data)**: 여러 사용자가 서로 다른 목적으로 데이터베이스의 데이터를 공동 이용
- ④ **변화되는 데이터(Changed data)**: 새로운 데이터의 추가, 기존 데이터의 삭제, 갱신으로 항상 변화하면서도 항상 현재의 정확한 데이터를 유지해야 한다는 의미

1. 데이터의 이해

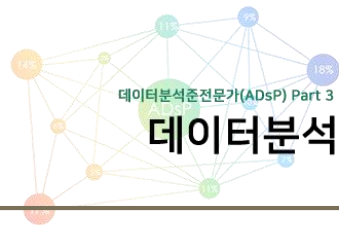


09

데이터베이스 특성

- ① **정보의 축적 및 전달 측면** → 대량의 정보를 정보처리기가 읽고 쓸 수 있는 기계가 독성과 필요한 정보를 검색할 수 있는 검색 가능성, 정보통신망을 이용하여 원거리에서도 온라인으로 이용할 수 있는 원격 조작성을 갖는다.
- ② **정보 이용 측면** → 이용자의 정보 요구에 따라 다양한 정보를 신속하게 획득하고, 원하는 정보를 경제적으로 찾아 낼 수 있다.
- ③ **정보 관리 측면** → 방대한 양의 정보를 체계적으로 축적하고 새로운 내용 추가나 갱신이 용이
- ④ **정보기술발전 측면** → 데이터베이스는 정보처리, 검색, 관리 소프트웨어 등 네트워크 발전 기술을 견인할 수 있다.
- ⑤ **경제·산업적 측면** → 데이터베이스는 인프라로서 특성을 가지고 있어 경제, 산업, 사회 활동의 효율성을 제고하고 국민의 편익을 증진하는 수단으로 의미를 가진다.

1.데이터의 이해



10

기업내부 데이터베이스 솔루션

구분	주요 특징
1980년대 기업 내부 데이터베이스	<ul style="list-style-type: none">① OLTP(On-line Transaction Processing) 온라인 거래처리: 주 컴퓨터와 통신회선으로 접속되어 있는 복수의 사용자 단말에서 발생한 트랜잭션을 주 컴퓨터에서 처리하여 그 결과를 즉석에서 사용자 되돌려보내 주는 처리 형태. 여러 과정이 하나의 단위 프로세스로 실행되도록 하는 프로세스② OLAP(On-Line Analytical Processing) 온라인 분석처리: 다차원으로 이루어진 데이터로부터 통계적인 요약 정보를 제공할 수 있는 기술
2000년대 기업 내부 데이터베이스	<ul style="list-style-type: none">① CRM (Customer Relationship Management): 선별된 고객으로부터 수익을 창출하고 장기적인 고객관계를 가능케함으로써 보다 높은 이익을 창출할 수 있는 솔루션을 말한다.② SCM (Supply Chain Management): SCM이란 제조, 물류, 유통업체 등 유통공급망에 참여하는 모든 업체들이 협력을 바탕으로 정보기술(Information Technology)을 활용, 재고를 최적화하기 위한 솔루션이다.

1. 데이터의 이해



11

분야별 기업내부 데이터 베이스 솔루션

분야	주요 솔루션
제조부문	<ul style="list-style-type: none">① DW(Data warehouse): 정보 검색을 목적으로 구축된 데이터베이스이다. 데이터 웨어하우스가 전사적인 규모의 시스템이라 한다면 데이터 마트는 사업부 단위의 소규모 데이터 웨어하우스라 할 수 있다.② ERP(Enterprise Resource Planning): 제조업을 포함한 다양한 비즈니스 분야에서 생산, 구매, 재고, 주문, 공급자와의 거래, 고객서비스 제공 등 주요 프로세스 관리를 돕는 여러 모듈로 구성된 통합 애플리케이션 소프트웨어 패키지를 의미③ BI(Business Intelligence): 기업의 DW(Data Warehouse)에 저장된 데이터에 접근해 경영의사결정에 필요한 정보를 획득하고 이를 경영활동에 활용하는 것을 말함④ CRM(Customer Relationship Management): 선별된 고객으로부터 수익을 창출하고 장기적인 고객관계를 가능케함으로써 보다 높은 이익을 창출할 수 있는 솔루션을 말함

1.데이터의 이해



실시간 기업(Real Time Enterprise)

->RTE는 기업의 주요 업무 프로세스에서 발생하는 최신의 정보를 필요한 사람에게 실시간으로 전달하여 즉각적인 모니터링과 신속한 대응이 가능한 기업이다. 정보기술의 전략적 활용이라는 측면에서 '02년부터 가트너 그룹이 공식적으로 사용하기 시작한 개념이다.

사례) 월마트는 QR(Quick Response) 시스템 구축을 통해 제품의 판매 현황을 매시간 수집하여 고객이 원하는 제품을 원하는 시점에 구입할 수 있도록 하고 있다.

1.데이터의 이해



금융부문

- ① EAI는 Enterprise Architecture Integration의 약자로 기업 애플리케이션 통합을 의미. 기업 내의 ERP(전사적자원관리), CRM(고객관계관리), SCM(공급망계획)시스템이나 인트라넷 등의 시스템 간에 상호 연동이 가능하도록 통합하는 솔루션
- ② EDW(Enterprise Data Warehouse)는 기존 DW(Data Warehouse)를 전사적으로 확장한 모델인 동시에 BPR과 CRM, BSC 같은 다양한 분석 애플리케이션들을 위한 원천이 된다. 따라서 EDW를 구축하는 것은 단순히 정보를 빠르게 전달하는 대형 시스템을 도입한다는 의미가 아니라 기업 리소스의 유기적 통합, 다원화된 관리 체계 정비, 데이터의 중복 방지 등을 위해 시스템을 재설계하는 것임
- ③ **블록체인(Blockchain)**: 데이터 분산 처리 기술. 네트워크 참여하는 모든 사용자가 모든 거래내역 등의 데이터를 분산, 저장하는 기술을 말함. 블록들을 체인 형태로 묶는 형태이기 때문에 블록체인이라는 명칭이 생겨남. 기존거래 방식에서 데이터를 위, 변조하기 위해서는 은행의 중앙서버를 공격하면 가능했으나 최근 몇몇 은행 전산망 해킹 사건이 발생이 했으나 블록체인인 경우 사실상 해킹이 불가능함
- ④ 이밖에 ERP, e-CRM 등 이용

1. 데이터의 이해



유통부문

- ① KMS(Knowledge Management System) 지식관리시스템의 약자. 조직 내의 지식을 체계적으로 관리하는 시스템을 의미, 예전에는 대부분의 기업이 물품을 생산하던 환경이었지만 요즘에는 지적 재산이 매우 중요해짐에 따라 기업을 관리하는 시스템이 등장함
- ② RFID는 무선주파수(RF, Radio Frequency)를 이용하여 대상(물건, 사람 등)을 식별할 수 있는 기술로서, 안테나와 칩으로 구성된 RF 태그에 사용 목적에 알맞은 정보를 저장하여 적용 대상에 부착한 후 판독기에 해당되는 RFID 리더를 통하여 정보를 인식함
- ③ 이 외에도 CRM, SCM 등 이용

1.데이터의 이해



12

BI와 BA 차이

구분	BI(Business Intelligence)	BA(Business Analytics)
목적	과거의 성과를 측정하고 향후 비즈니스를 계획	데이터와 통계를 기반으로 성과에 대한 이해와 비즈니스 통찰력에 초점을 둔 분석 방법
응용	데이터 기반의 의사결정	<ul style="list-style-type: none">• 사전에 예측하고 최적화하기 위함• BI보다 진보된 형태

2. 기출문제



Q1

다음은 데이터 정의에 관한 설명이다. 가장 부적절한 것은??

- ① 객관적 사실이다.
- ② 데이터는 추론과 추정의 근거를 이루는 사실이다.
- ③ 개별 데이터 자체로는 의미가 중요한 객관적 사실이다.
- ④ 단순한 객체로서의 가치와 다른 객체와의 상호 관계 속에서 가치를 갖는다.

2. 기출문제



Q2

객체들을 생성하여 계층에서 체계적으로 정리하고, 다시 계층들을 하위 계층이 상위계층으로부터 속성과 방법들을 물려받을 수 있는 복잡한 데이터 구조를 관리하는 DBMS은?

- ① 객체지향 데이터베이스 관리시스템(ODBMS)
- ② Mysql
- ③ 관계형 데이터베이스 관리시스템(RDBMS)
- ④ Operating System

2. 기출문제



Q3

다음은 어떤 데이터의 유형인가?

지역별 온도·풍속·강우량과 같이 수치로 명확하게 표현되는 이것은 데이터의 양이 크게 증가하더라도 이를 관리하는 시스템에 저장·검색·분석하여 활용하기가 매우 용이하다.

2. 기출문제



Q3

다음은 어떤 기업내부데이터 솔루션에 대한 설명인가?

물류, 유통업체 등 유통 공급망에 참여하는 모든 업체들이 협력을 바탕으로 정보 기술(Information Technology)을 활용, 재고를 최적화 하기 위한 솔루션이다.

2. 기출문제



Q4

데이터의 가공 및 처리와 데이터간 연관관계 속에서 의미가
도출된 것을 무엇이라 하는가?

2. 기출문제



Q5

내재된 경험을 문서나 매체로 저장하는 것은?

- ① 표출화
- ② 공통화
- ③ 내면화
- ④ 연결화

2. 기출문제



Q6

DIKW 내용 중 다른 하나는??

- ① 오늘의 날씨는 21도이다
- ② A마트의 쇠고기는 1킬로 2만 원이다
- ③ B회사의 매출액은 3천만 원으로 예상된다
- ④ C마트는 100원에 D마트는 200원에 연필을 판다.

2. 기출문제



Q7

다음 중 암묵지가 아닌 것은?

- ① 김장김치 담그기의 노하우
- ② 암묵지는 개인에게 체화되기 되어 있기 때문에 공유하기 어렵다
- ③ 현장작업과 같은 경험을 통해 획득할 수 있는 지식
- ④ 회계, 재무 관련 대차대조표에 요구되는 지식의 매뉴얼 등이 암묵지이다.

2. 기출문제



Q8

데이터에 대한 설명으로 부적절한 것은?

- ① 데이터는 지식경영의 핵심 이슈인 암묵지와 형식지의 상호작용을 한다
- ② 형식지는 문서나 매뉴얼처럼 외부로 표출되어 여러 사람이 공유할 수 있는 지식이다
- ③ 형식지란 개인에 체화된 비밀스러운 지식이다.
- ④ 지식의 차원에 대해 가장 널리 알려진 것은 Polany가 두가지 차원으로 구성된 암묵지와 형식지이다

2. 기출문제



Q09

다음 중 정량적 데이터가 아닌 것은?

- ① 도형
- ② 그림
- ③ 숫자
- ④ 문자

2. 기출문제



Q10

()란 데이터로부터 의미 있는 정보를 추출해내는 학문이며, 정형 또는 비정형 막론하고 인터넷, 휴대전화, 감시용 카메라 등에서 생성되는 숫자와 문자, 영상 정보 등 다양한 유형의 데이터를 대상으로 하며, 분석뿐 아니라 이를 효과적으로 구현하고 전달하는 과정까지 포함한 포괄적 개념이다?

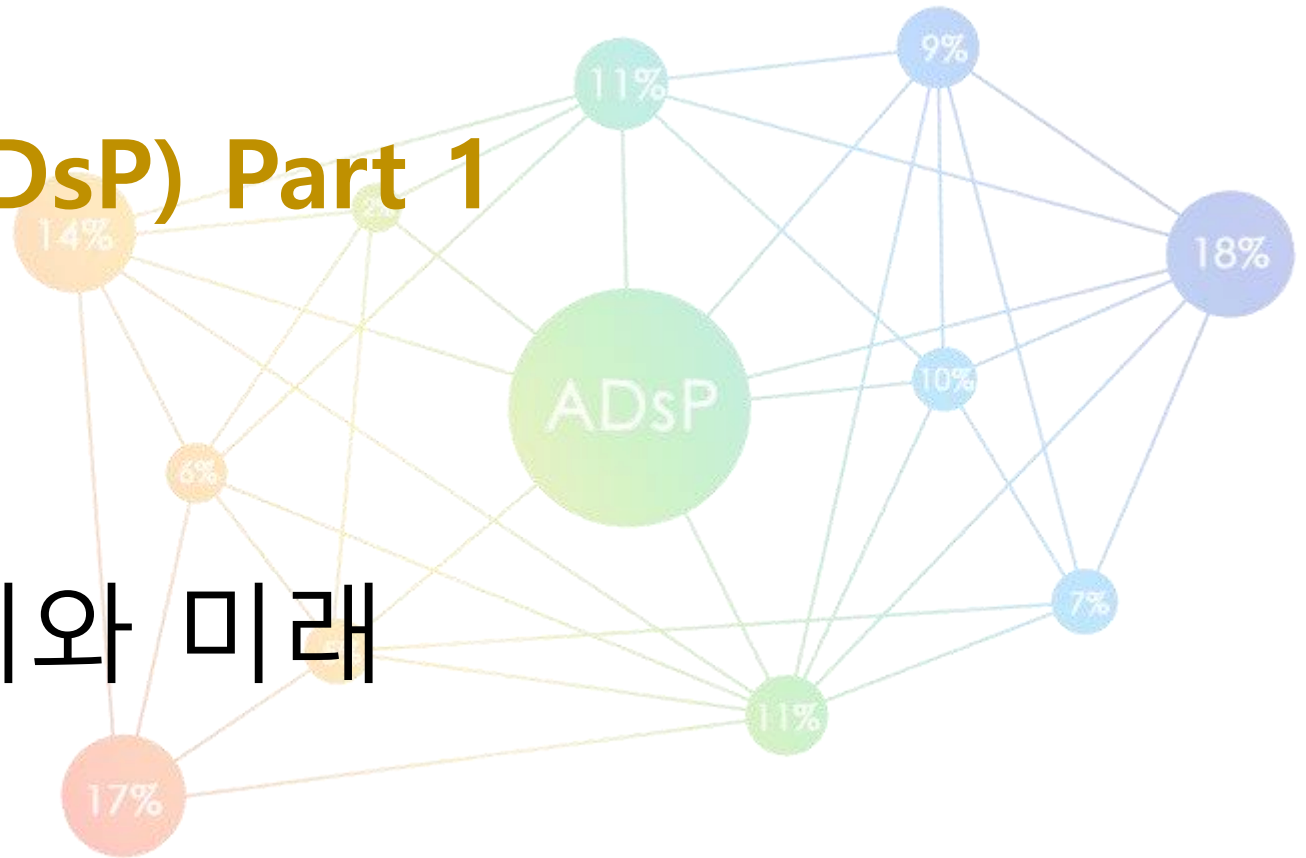
데이터분석전문가(ADsP) Part 1

데이터 이해

01

2장 데이터의 가치와 미래

1. 빅 데이터의 이해
2. 빅 데이터의 가치와 영향
3. 비즈니스모델
4. 위기요인과 통제 방안
5. 미래의 빅 데이터



2.데이터의 가치와 미래



13

빅데이터란?

가트너 그룹 더그래니가 언급한 빅데이터 정의

- ① **Volume**(데이터의 크기): 생성되는 모든 데이터를 수집
예) 트위터 매일 12테라 생성
- ② **Variety**(데이터의 다양성): 정형화된 데이터를 넘어 텍스트/오디오/비디오 등 모든 유형의 데이터를 분석 대상으로 한다.
예) 구글은 2,300억 단어분석 음성인식 엔진 개발
- ③ **Velocity**(데이터의 속도): 두 가지 관점의 속도를 의미함 즉 사용자 원하는 시간 내 데이터분석 결과 제공과 데이터의 업데이트되는 속도가 매우 빨라짐
예) 브리티시 텔레콤: 1초 60기가의 데이터 전송

2.데이터의 가치와 미래



14

빅데이터 역할

① 빅데이터는 산업혁명의 석탄, 철

빅데이터는 석탄과 철이 산업혁명에서 했던 역할을 차세대 산업혁명에서 해낼 것으로 기대된다. 빅데이터 역시 지금의 제조업뿐 아니라 서비스 분야의 생산성을 획기적으로 끌어올려 혁명적 변화를 가져올 것으로 기대되고 있다.

② 빅데이터는 원유

빅데이터는 원유처럼 각종 비즈니스, 공공기관 대국민 서비스, 경제 성장에 필요한 '정보'를 제공함으로써 산업 전반의 생산성을 한 단계 향상시킬 것으로 기대가 된다.

③ 빅데이터는 렌즈

현미경이 생물학 발전에 미쳤던 영향만큼 나아가 데이터가 산업 전반에 영향을 미칠 것으로 기대된다.

ex) 구글 'Ngram Viewer'를 통해 수천만 권의 책을 디지털화

④ 빅데이터는 플랫폼

플랫폼이란 다양한 차원에서 활용되는 개념이지만, 비즈니스 측면에서는 일반적으로 '공동 활용의 목적으로 구축된 유무형의 구조물'을 의미한다.

ex) 페이스북에서 앱

2.데이터의 가치와 미래



15

빅데이터가 만들어내는 본질적인 변화

- ① 사전처리에서 사후처리 시대로
- ② 표본조사에서 전수조사로
- ③ 질보다 양으로
- ④ 인과관계에서 상관관계로

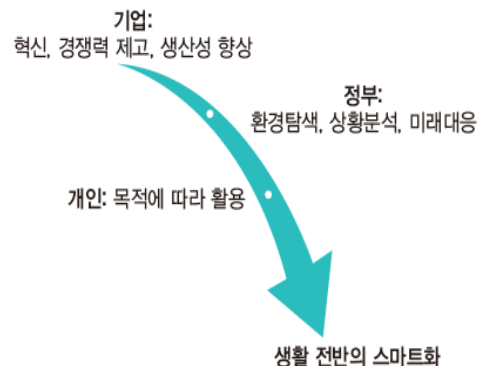
2.데이터의 가치와 미래



16 빅데이터의 가치 산정이 어려운 이유

- ① 데이터의 활용 방식: 재사용이나 재조합(mashup), 다목적용 데이터 개발 등이 일반화되면서 특정 데이터를 언제·어디서·누가 활용할지 알 수 없다.
- ② 새로운 가치 창출: 빅데이터 시대에는 기존에 없던 가치를 창출함에 따라 그 가치를 측정하기 어렵다.
- ③ 분석기술의 발달: 지금은 가치 없는 데이터도 새로운 분석 기법의 등장으로 거대한 가치를 만들어내는 재료가 될 가능성이 있다.

17 빅데이터의 영향



2.데이터의 가치와 미래



18

빅데이터 활용 사례

- ① 구글의 검색엔진, 월마트의 구매패턴 분석, IBM 왓슨 의료분야에 활용
- ② 정부의 실시간 교통정보 활용, CCTV 국가안전에 활용
- ③ 정치인의 사회관계망분석을 통한 유세, 가수의 팬 음악청취 기록 분석 활용 등
- ④ 아마존의 킨들에 쌓이는 전자책 읽기 관련 데이터 분석해 저자들에게 제공

2.데이터의 가치와 미래



19

빅데이터 활용 테크닉

① 연관규칙학습(Association rule learning)

어떤 변수 간에 주목할만한 상관관계가 있는지를 찾아내는 방법이다.

예) 슈퍼마켓에서 상관관계가 높은 상품을 함께 진열(우유와 기저귀)

② 유형분석(Classification tree analysis)

‘사용자가 어떤 특성을 가진 집단에 속하는가?’와 같은 문제를 해결하고자 할 때 사용

예) 온라인 수강생들을 특성에 따라 분류

③ 유전 알고리즘(Genetic algorithms)

‘최대의 시청률을 얻으려면 어떤 프로그램을 어떤 시간대에 방송해야 하는가?’와 같은 문제를 해결할 때 사용한다. 최적화의 메커니즘을 찾아가는 방법이다.

예) 연료 효율적인 차를 개발하기 위해 어떻게 원자재와 엔지니어링을 결합해야 하는가?

④ 기계 학습(Machine learning)

기존의 시청 기록을 바탕으로 시청자가 현재 보유한 영화 중에서 어떤 것을 가장 보고 싶어 할까와 같은 문제

2.데이터의 가치와 미래



를 해결할 때 사용된다. 기계학습은 훈련 데이터로부터 학습한 알려진 특성을 활용해 '예측'하는 일에 초점을 맞춘다.

예) 넷플릭스 영화추천 시스템

⑤ 회귀분석(Regression Analysis)

'구매자의 나이가 구매 차량의 타입에 어떤 영향을 미치는가?'와 같은 질문에 답할 때 사용된다.

⑥ 감정분석(Sentiment Analysis)

'새로운 환불 정책에 대한 고객의 평가는 어떤가?'를 알고 싶을 때 활용한다.

예) 소셜미디어에 나타난 의견을 바탕으로 고객이 원하는 것을 찾아 낼 때 사용된다.

⑦ 소셜 네트워크 분석(Social network analysis)=사회관계망분석(SNA)

영향력 있는 사람을 찾아낼 수 있으면, 고객들 간 소셜 관계를 파악할 수 있다.

2.데이터의 가치와 미래



20

빅데이터 위기요인과 통제방안

① 사생활 침해

→ **(위기요인)** 빅데이터 시대가 본격화되면서 우리를 둘러싼 정보 수집 센서들의 수가 점점 늘어나고 있고, 특정 데이터가 본래 목적 외에 가공처리돼 2차·3차적 목적으로 활용될 가능성이 증가하면서 사생활 침해를 넘어 사회·경제적 위협으로 변형될 수 있음.

익명화(Anonymization): 사생활 침해를 방지하기 위해 데이터에 포함된 개인 식별 정보를 삭제하거나 알아볼 수 없는 형태로 변환하는 것을 말한다.

→ **(통제방안)** 동의에서 책임으로 - 개인정보의 활용에 대한 개인이 매번 동의하는 것은 경제적으로도 매우 비효율적이다. 따라서 사생활침해 문제를 개인정보 제공자의 동의를 통해 해결하기보다는 개인정보 사용자에게 책임을 지움으로써 개인정보 사용 주체가 보다 적극적인 보호 장치를 강구하게 하는 효과가 발생할 것으로 기대된다.

② 책임 원칙의 훼손

→ **(위기요인)** 빅데이터 기반분석과 예측 기술이 발전하면서 정확도가 증가한 만큼, 분석 대상이 되는 사람들은 예측 알고리즘의 희생양이 될 가능성이 증가한다. 그러나 잠재적 위험 사항에 대해서도 책임을 추궁하는 사회로 변질될 가능성이 높아 민주주의 사회 원칙을 크게 훼손 할 수 있다.

→ **(통제방안)** 기존의 책임원칙을 강화할 수밖에 없다.

③ 데이터의 오용

→ **(위기요인)** 빅데이터는 일어난 일에 대한 데이터에 의존한다. 그것을 바탕으로 미래 예측하는 것은 적지 않은 정확도를 가질 수 있지만 항상 맞을 수는 없다. 주어진 데이터에 잘못된 인사이트를 얻어 비즈니스에 직접 손실을 불러올 수 있다.

→ **(통제방안)** 데이터 알고리즘에 대한 접근권 허용 및 객관적 인증방안을 도입 필요성 제기

2.데이터의 가치와 미래



21. 빅 데이터 활용 3요소

- ① 데이터: 모든 것을 데이터화하는 추세는 빅 데이터 시대의 피할 수 없다.
- ② 기술: 빅 데이터 분석 알고리즘의 진화가 가속화 될 것이다.
- ③ 인력: 데이터 사이언티스트와 알고리즘미스트의 역할이 중요해질 것으로 전망된다.

2. 기출문제



Q1

다음 중 빅데이터 활용에 필요한 3요소는 무엇인가?

- ① 데이터, 기술, 인력
- ② 프로세스, 기술, 인력
- ③ 데이터, 프로세스, 인력
- ④ 인력, 데이터, 알고리즘.

2. 기출문제



Q2

다음 중 빅데이터 특징으로 옳지 않은 것은?

- ① 데이터 분석의 발달로 표본조사로서 새로운 가치나 지식을 발견하게 되었다.
- ② 데이터의 질보다 양을 강조한다
- ③ 비즈니스 상황에서는 인과관계를 모르고 상관관계 분석만으로 충분한 경우가 많 다
- ④ 사전처리에서 사후처리 시대로 변화하였고, 사전 처리의 대표적인 예로는
표준화 된 문서 포맷을 들 수 있다

2. 기출문제



Q3

빅데이터의 활용テクニック 중 보기는 무엇에 해당되는가?

- ① 유전알고리즘
- ② 감정분석
- ③ 연관 규칙 학습
- ④ 회귀분석

우유구매자가 기저귀도 같이 구매하는가 또는
기저귀 구매자가 맥주도 같이 구매하는가를 알아본다

2. 기출문제



Q4

빅데이터시대가 도래하면서 발생하는 사생활침해를 막기 위해서데이터에 포함된 개인 식별 정보를 삭제하거나 알아볼 수 없는 형태로 변환 하는 포괄적 기술을 무엇이 라 하는가?

- ① 익명화(Anonymity)
- ② 일반화(Generalization)
- ③ 가명(Pseudonym)
- ④ 치환(Permutation)

2. 기출문제



Q5

다음 중 빅데이터 위기 요인과 해결 방안을 잘못 연결된 것을 고르시오?

- ① 가, 나
- ② 가, 다
- ③ 나, 다
- ④ 가, 나, 다

가. 사생활 침해 → 동의제를 책임제로 전환
나. 책임훼손의 원칙 → 알고리즘 접근성 허용
다. 데이터의 오용 → 결과 기반 책임 원칙

2. 기출문제



Q6

빅데이터 위기 요인 중 책임 훼손의 사례로 맞는 것은?

- ① 범죄 예측 프로그램을 통해 범죄 전 체포
- ② 데이터의 본래 목적 외에 가공 처리되어 사생활이 침해되는 경우
- ③ 잘못된 지표를 사용하는 경우
- ④ 미국 NSA가 매일 17억 개의 이메일 ,전화통화 내역 수집 저장의 경우

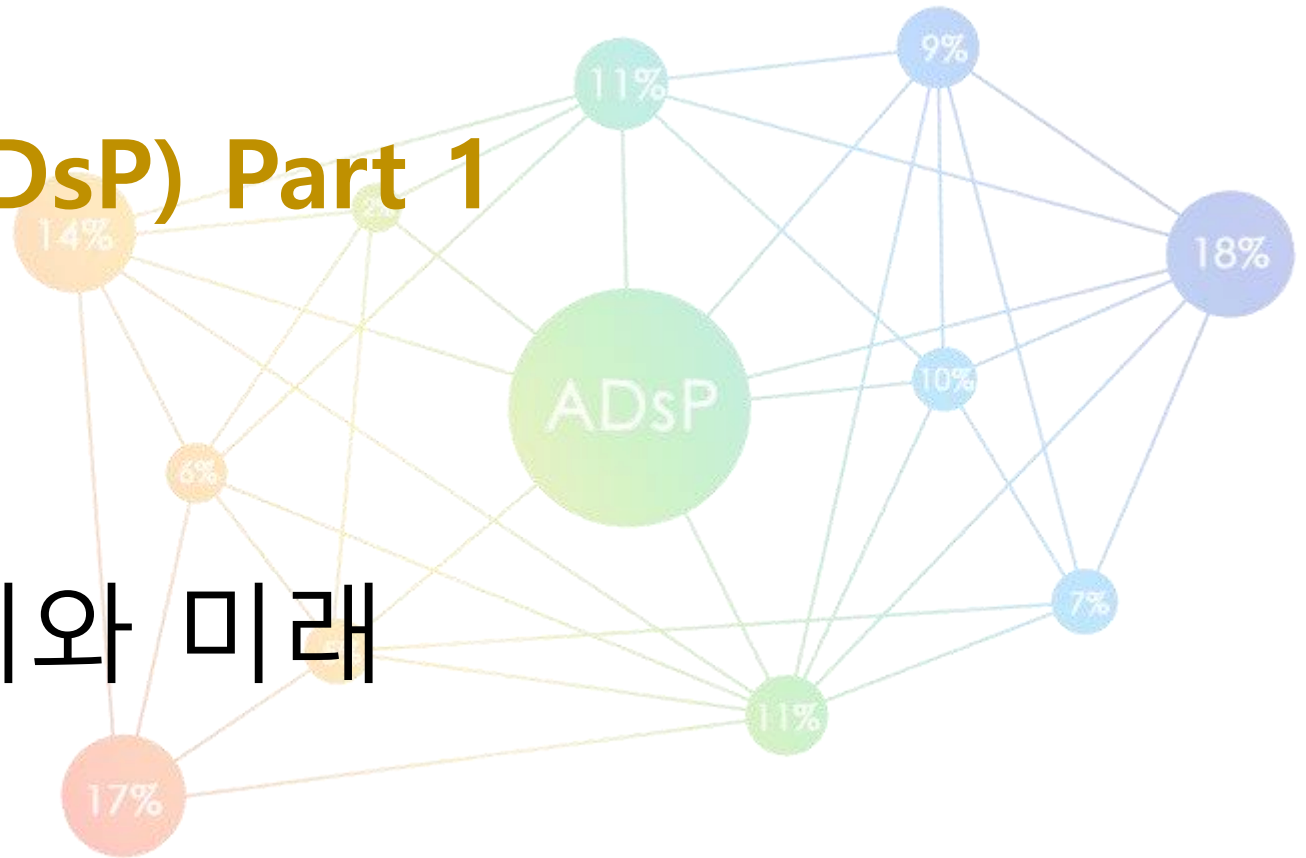
데이터분석전문가(ADsP) Part 1

데이터 이해

01

3장 데이터의 가치와 미래

1. 빅 데이터 분석과 전략 인사이트
2. 전략 인사이트 도출을 위한 필요역량
3. 빅 데이터 그리고 데이터 사이언스의 미래



3.가치 창조를 위한 데이터 사이언스와 전략 인사이드



22

빅데이터 분석, 'Big'이 핵심이 아니다.

데이터의 양이 아니라 유형의 **다양성**과 관련이 있다. 빅데이터가 가져다 주는 기회는 데이터의 크기에 있다가보다는 음성, 텍스트, 이미지, 비디오 같은 새롭고 **다양한 정보 원천의 활용**에 있다.

23

전략적 통찰이 없는 분석의 함정

토마스 데이븐포트의 분석의 기술에 의하면 **성과가 우수한 기업들도** 가치 분석적 통찰력을 갖췄다고 대답한 비율이 **매우 낮다는 사실이다**. 그만큼 기업의 핵심 가치와 관련해 전략적 통찰력을 가져다 주는 데이터 분석을 내재화하는 것이 어렵다는 것이다.

ex) 아메리카 항공사의 복잡화, 최적화가 오히려 비즈니스에 마이너스

3.가치 창조를 위한 데이터 사이언스와 전략 인사이트



24

일차적인 분석 경험이 누적되면 활용 범위를 넓혀 전략적으로 변화시켜야 한다.

대표적인 일차적인 분석 애플리케이션 사례

산업	분석 애플리케이션
금융서비스	신용점수 산정, 사기 탐지, 고객 수익성 분석
소매업	재고 보충, 수요 예측
제조업	맞춤형 상품 개발, 신상품 개발
에너지	트레이딩, 공급, 수요 예측
온라인	웹 매트릭스, 사이트 설계, 고객 추천

3.가치 창조를 위한 데이터 사이언스와 전략 인사이트



25 데이터 사이언스가 다른 학문과의 차이점

	데이터 사이언스	통계학	데이터마이닝
분석대상	정형,비정형,반정형	정형화된 데이터	
분석방법	분석+시각화+전달 포함한 포괄적개념		분석에 초점
학문접근	종합적 학문 또는 총제적 접근법		

3.가치 창조를 위한 데이터 사이언스와 전략 인사이트



26 데이터 사이언스의 구성요소

- ① IT(Data Management)
- ② Analytics(분석적 영역)
- ③ 비즈니스 분석

27 Gartner 데이터 사이언티스트의 역량

- ① 데이터 관리
 - ② 분석 모델링
 - ③ 비즈니스 분석
 - ④ 소프트 스킬
- 공통점은 호기심에 시작

3.가치 창조를 위한 데이터 사이언스와 전략 인사이트



28

데이터 사이언티스트의 역량(소프트 스킬 VS 하드 스킬)

Hard Skill

- ① 빅데이터에 대한 이론적 지사: 관련 기법에 대한 이해와 방법론 습득
- ② 분석 기술에 대한 숙련: 최적의 분석 설계 및 노하우 축적

Soft Skill

- ① 통찰력 있는 분석: 창의적 사고, 호기심, 논리적 비판
- ② 설득력 있는 전달: 스토리텔링, Visualization
- ③ 다분야간 협력: Communication

3.가치 창조를 위한 데이터 사이언스와 전략 인사이드



29. 왜? 데이터 사이언스가 과학과 인문학의 교차로 있는가?

데이터 사이언스는 과학과 인문의 교차로에 서 있다고 할 수 있다. 그래서 세계적인 데이터 사이언스 전문가들이 이구동성으로 데이터 사이언티스트에게 스토리텔링, 커뮤니케이션, 창의력, 열정, 직관력, 비판적 시각, 글쓰기 능력, 대화 능력 등이 필요하다고 강조한다. 그리고 이러한 능력들은 대부분 인문학의 주요 주제들이다.

3.가치 창조를 위한 데이터 사이언스와 전략 인사이트



30

최근의 사회경제적 환경의 변화

- ① 단순 세계화에서 복잡한 세계화로의 변화이다. 다양성과 각 사회의 정체성과 그 맥락, 관계, 연결성, 창조성 등이 키워드로 대두되게 되었다.
- ② 비즈니스의 중심이 제품생산에서 서비스로 이동이다. 제품이 고장이 나더라도 오히려 얼마나 뛰어난 고객 서비스를 제공해 주느냐가 더 중요하게 되었다.
- ③ 경제와 산업의 논리가 생산에서 시장창조로 바뀌었다. 지금의 핵심은 새로운 현지화 패러다임에 근거한 시장창조로 이동하였다.

3.가치 창조를 위한 데이터 사이언스와 전략 인사이트



31

정보(Information) vs 통찰력(Insight)의 차이

	과거	현재	미래
정보 (information)	무슨 일이 일어났는가? ex) 리포팅(보고서)	무슨 일이 일어나고 있는가? ex) 경고	무슨 일이 일어날 것인 것인가? ex) 추출
통찰력 (insight)	어떻게, 왜 일어났는가? ex) 모델링	차선 행동은 무엇인가? ex) 권고	최악, 최선의 상황은? ex) 예측, 최적화

3.가치 창조를 위한 데이터 사이언스와 전략 인사이트



32

빅데이터 회의를 넘어: 가치 패러다임의 변화

가치 패러다임 변화는 크게 3 단계로 구분해 볼 수 있다.

- ① **‘디지털화(Digitalization)’**이다. 아날로그의 세상을 어떻게 효과적으로 디지털화 하는가가 이 시대의 가치를 창출해 내는 원천이었다. 그 대표 주자가 바로 빌 게이츠였다.
- ② 다음 단계의 가치 패러다임은 빌 게이츠가 제대로 보지 못한 **‘연결(Connection)’**이란 것이다. 새로운 시대에 서 디지털화된 정보와 대상들은 이제 서로 연결되기 시작했다. 이 연결을 얼마나 효과적이고 효율적으로 제공해 주느냐가 이 시대의 성패를 가름한다. 지금도 인터넷에는 사람, 기기 등 상당히 많은 대상이 서로 연결됐지만 ‘사물 인터넷(Internet of Things)’의 성숙과 함께 앞으로는 연결이 더 증가하고 극도로 복잡해질 것이다.
- ③ 향후에는 ‘복잡한 연결을 얼마나 효과적이고 믿을 만하게 관리해주는가’의 이슈인 **‘에이전시(Agency)’**가 키워드로 등장할 확률이 매우 높다.

3. 기출문제



Q1

데이터 사이언스 관련 내용 중 틀린 것은?

- ① 데이터 사이언스가 기존의 통계학과 다른 점은 데이터 사이언스는 총체적 접근법을 사용한다는 것이다
- ② 강력한 호기심이야말로 데이터 사이언티스트의 중요한 특징이라고 할 수 있다
- ③ 데이터 사이언스란 데이터로부터 의미 있는 정보를 추출해내는 학문이다.
- ④ 통계와 데이터 마이닝을 융합한 새로운 학문이다.

3. 기출문제



Q2

데이터 사이언스에 대한 설명이 부적절한 것은?

- ① 외국의 각 전문가들은 호기심이야말로 데이터 사이언티스트의 중요한 특징이라고 생각한다.
- ② 데이터 사이언스는 과학과 인문학의 교차로에 서 있다고 할 수 있다
- ③ 통계학은 정형 또는 비정형을 막론하고 다양한 유형의 데이터를 대상으로 한다.
- ④ 데이터 사이언스의 핵심 구성요소로는 IT 영역, 분석적 영역, 비즈니스 컨설팅 영역이 있다.

3.기출문제



Q3

다음은 어느 산업의 데이터의 일차원적인 분석
애플리케이션 사례인가?

- ① 소매업
- ② 제조업
- ③ 에너지
- ④ 운송업

트레이딩, 공급, 수요예측

3. 기출문제



Q4

통찰력을 제공하는 분석 기술이 아닌 것은?

- ① 추출
- ② 최적화
- ③ 모델링
- ④ 예측