

제 1강 빅데이터 이해하기

중앙대학교 일반대학원 문화예술경영대학원

CONTENTS .

- 01 데이터란 무엇인가
- 02 빅데이터의 개념
- 03 빅데이터의 특성
- 04 스몰데이터와 빅데이터의 비교
- 05 데이터과학자 보다 문제 해결사 되기
- 06 세상의 다양한 분석자원들
- 07 데이터 분석과제 기획

교수 소개



오인택(吳仁澤)
PH.D. IN TAEK OH

학력

- 송실대학교 공학박사(IT Policy & Management)
- 서강대 경제학석사 (기술경제학)

경력

- KTDS DX사업본부장
- KT IT부문 IT컨설팅본부장(2021 ~ 2023)
- KT IT부문 SW개발본부 IT Delivery 1PTF장 (2020)
- KT IT부문 DX혁신담당 (2018~2020)
- KT IT부문 경영플랫폼담당 (2015~ 2018)
- KT NexR 사외이사 겸직 (2016 ~ 2019)

주요 활동

- 21년 등단(나래시조시인협회), 시조시인
- 대표시집 『은꽃연가』, 좋은땅출판사
- 한국산림문학 이사
- 문경새재여름시인학교 사무국장
- 한국 문인협회 정회원
- 중앙대학교 일반대학원 문화예술경영학과 겸임교수
- 한국정보처리학회 부회장
- KAIST Fourth Revolution Intelligence Center 자문위원
- 한양대학교 소프트중심대학 교과과정 자문위원
- 한국정보산업협회 한이음 대회 심사위원
- 서울시 기술연구소 자문위원

연락처

- ohintaek@gmail.com
- 010-3010-1651(언제든지 연락을 해도 좋습니다)

소개

N | **KT오인택**

통합VIEW이미지지식iN인플루언서동영상쇼핑뉴스어학사전지도...

 www.etnews.com

[인사이트]오인택 KT 상무 "RPA로 2년간 70만 업무시간 절감, 생산성 극대화하겠다" - ...

“로봇 프로세스 자동화(RPA)로 **KT** 임직원이 업무 시간을 70만 시간 이상 절감했습니다.” **오인택 KT DX혁신담당(상무)**은 “RPA로 임직원의 일하는 방식의 DNA를 바꾸고 있다”며 이같이 말했다. 이어 “전표처리 등...

2020.11.19.



 www.etnews.com

KT "일잘하는 '전대리' '마비서' 덕에 연 70억 절감" - 전자신문

KT가 로봇틱 프로세스 자동화(RPA) 콜라봇 사업에 드라이브를 건다. 영수증 처리 등 단순반복 업무를 자동화해 전사적으로 연간 27만 시간을 절약한 경쟁력을 토대로 기업고객을 확보한다. **오인택 KT 상무**(IT기획...

2019.06.25.

 m.post.naver.com > viewer

AI 전대리-마비서 덕에 연간 70억원 절감한 사연은

영수증 처리 등 단순반복 업무를 자동화해 전사적으로 연간 27만 시간을 절약한 경쟁력을 토대로 기업고객을 확보한다.**오인택 KT 상무**(IT기획실 고객경영플랫폼 담당)는 최근 “**KT RPA 서비스**는 직원 업무를 코딩해 자동화한 형태 패턴으로 고객 맞춤형 지원이 가능하다”며 “데이터를 분석하고 처리하는 일, 영상을 편집하는 일 등 손은...



 www.dt.co.kr > contents

"디지털전환 늦으면 도태... 우리가 최선의 방향 제시하겠다" - 디지털타임스

오인택 KT IT부문 IT Delivery PTF 상무 “**KT**의 B2B 사업은 과거 통신과 회선 기반 위주에서 변하고 있습니다. 디지코(디지털기업)로 변화함에 따라 고객 관점의 DX가 중요하게 됐고, 기업은 고객에게 새로운 가치를 제공하는 방향으로 나아가고 있습니다.” **KT IT부문 IT 딜리버리 P-TF 오인택**(사진) 상무는 20일 디지털타임스와 서면 인...

2021.05.20.



 www.ddaily.co.kr > m_article

“KT IT 서비스 전문성? 수천억 내부 프로젝트로 충분히 쌓았죠” - 디지털데일리

하지만 이미 수천억원 이상 내부 프로젝트를 통해 내재된 IT 역량을 기반으로 충분히 만족할 만한 서비스를 제공하고 있습니다.” **오인택 상무 KT IT부문 IT컨설팅본부 IT컨설팅1담당 상무**<사진>는 최근 디지털데일리와 인터뷰를 통해 이같이 자신했다. 지난 2020년 ‘디지코(디지털 플랫폼 기업)’으로의 도약을 선언한 **KT**는 인공지능... 2022.09.06.



01.

데이터란 무엇인가

데이터란 무엇인가

01

데이터 분석 활동과 DIKW 피라미드(1/3)

- 데이터 분석을 활용한 다양한 문제해결 과정에서 데이터(Data) → 정보(Information) → 지식(Knowledge) → 지혜(Wisdom)가 순차적으로 나타남

단계	설명	예시
지혜 (Wisdom)	<ul style="list-style-type: none">지식의 축적과 아이디어가 결합된 창의적인 산물근본 원리에 대한 깊은 이해를 바탕으로 도출되는 창의적인 아이디어	<ul style="list-style-type: none">로션이외에 다른 상품 판매실적도 비교해봄계절과 지역에 따라, 영업직원 판매방식에 따라 실적차이가 나는지 점검
지식 (Knowledge)	<ul style="list-style-type: none">상호 연결된 정보패턴을 이용해 예측한 결과물개인의 경험 결합해 고유의 지식으로 내재화	<ul style="list-style-type: none">각 대리점별 판매실적을 기준으로 로션을 안정적으로 공급하고, 인센티브도 차등 지급함
정보 (Information)	<ul style="list-style-type: none">데이터를 통해 도출된 다양한 정보를 구조화해 유의미한 정보로 분류함데이터의 가공·처리, 데이터간 상관·연관관계 속에서 의미를 부여함	<ul style="list-style-type: none">매월 총 300개 판매실적을 보이고 있음B대리점의 판매실적이 A대리점보다 2배임
데이터 (Data)	<ul style="list-style-type: none">개별 데이터 자체는 특별한 의미부여가 안된 객관적 사실타 데이터와 상관관계가 없는 가공하기 전의 순수한 수치나 기호 그 자체	<ul style="list-style-type: none">월평균 로션판매량 조사 A대리점 : 100개 B대리점 : 200개

데이터란 무엇인가

02

데이터 분석 활동과 DIKW 피라미드(2/3)

• 데이터(Data)

- 분석가가 관심을 가진 어떤 현상이나 대상에 대한 관찰·수집한 자료로 특별한 의미가 부여되지 않은 “객관적 사실” 그 자체를 의미함
- 타 데이터와 어떤 비교하기 이전의 가공하기 전의 순수한 상태를 수치나 기호로 기록함
- 수치라는 것은 남자 1, 여자 2, 또는 판매실적을 A등급 B등급 등 정성적(qualitative)으로 기록하거나 판매실적을 100개, 100만원 등으로 정량적(quantitative)으로 기록하는 행위임
- A대리점, B대리점도 하나의 분석대상에 대한 데이터이며, 대리점에서 판매하는 여러 제품들, 이들이 가진가격, 디자인, 무게, 용량, 성분 등 여러 속성도 분석대상 데이터가 될 수 있음
- 어떠한 현상이나 대상, 또 그들이 가진 여러 속성들 중에서 나중에 의사결정에 활용할 목적으로 관찰/수집/측정되는 것이 바로 데이터임
- 하지만 의사결정에 얼마나 중요하게 활용이 될지 아직까지는 알 수 없으며, 타 데이터와의 관련성도 갖지 못하는 단순한 사실 그 자체에 불과함

• 정보(Information)

- 관찰·수집된 데이터를 적절한 “가공·처리”를 통해서 데이터 자체의 특성은 물론, 타 데이터와의 공통점·차이점 등을 “상호 비교”할 수 있도록 한 것임
- 정보도 수치와 기호를 활용해 나타내지만 데이터와는 달리 테이블이나 그래프 등을 적절히 활용해 어떠한 문제해결에 필요한 “근거자료”로서의 형식을 가지게 됨

데이터란 무엇인가

03

데이터 분석 활동과 DIKW 피라미드(3/3)

• 지식(Knowledge)

- 테이블 그래프로 가공 처리된 정보라는 근거를 관련된 "업무에 활용해 문제를 해결하고 "의사결정을 수행하는 일련의 업무수행 수단을 지식이라고 함
- 개인은 학습, 현장경험, 시행착오, 체험을 통한 지식활용과정에서 겉으로 드러나지 않는 무형의 지식이라는 암묵지(tacit/intangible knowledge)가 습득·축적됨
- 암묵지는 개인에게 축적된 내면화(Internalization)된 지식으로 체화 되어 있기 때문에 외부에 표출되어 다른 사람에게 공유되기 어려우므로 조직지식으로 공통화(Socialization) 하는 게 중요함
- 반면에 개인에게 내재된 경험을 객관적인 문서, 매체에 저장·가공·분석해 놓은 결과물 형태인 형식지(explicit/tangible knowledge)의 경우에는 구체적인 유형·대상이 있기 때문에 지식의 전달과 공유가 매우 용이함

• 지혜(Wisdom)

- 관찰/수집된 데이터로부터 적절한 가공/처리를 통해 근거를 마련하고, 이를 관련된 업무에 활용해 의사결정을 수행하는 지식까지 만들어 내는 과정을 통해 축적된 지식과 경험을 체계화하여 창의적인 아이디어까지 이끌어 내는 것을 의미함
- 창의적인 아이디어라는 것은 새로운 무엇인가를 만들어내는 것을 의미하는 것이 아니라 데이터-정보-지식을 통해 축적된 지식과 경험을 체계화하여 업무활용 수준을 높이고 타업무 분야로 확장하여 적용하는 것을 의미함

데이터란 무엇인가

04

데이터를 바라보는 위상의 변화

• 원래 데이터에 대한 인식

- 원래 데이터라는 말은 활용 목적이 정해지지 않은 있는 그대로의 자료(data)라는 말로 단순히 고려되었으며, 이를 의사결정에 활용하기 위한 목적으로 가공/처리하는 정보(information)라는 말의 중요성이 더 강조되어 왔음
- 또한 가공/처리된 정보를 토대로 관련 업무에 적절하게 활용하고 경험을 쌓아가는 지식(knowledge)으로의 전환, 나아가 축적된 지식·경험이 하나의 업무활동체계로 수립되는 지혜(wisdom)에 이르기까지 개념적인 발전을 이루어 왔음
- 데이터는 사업을 운영하는 데 있어 보조적 역할로 비쳐지거나 사용로그, 개인정보 등 상대적으로 좁은 범위에 한정된 것으로 여겨졌으나, 다음처럼 그 영향력 및 중요도, 활용가능성으로 인해 빅데이터라는 이름으로 부르게 됨

• DT 시대의 도래

- 알리바바의 마윈회장은 "지난 20년간 지속된 IT의 시대가 저물고, 앞으로 30년간 DT혁명에 기반한 새로운 인터넷 시장이 열리게 될 것이며, 이제는 방대한 고객 데이터를 활용해 개별 고객의 요구에 부응할 줄 아는 기업이 성공하는 DT시대가 될 것" 이라고 함
- IT의 주도권이 인프라, 기술, 소프트웨어 등에서 데이터로 옮겨가고 있으며, 데이터가 IT에서 분리된 독립적인 주체로 발전하고 있음
- 기존의 IT(Information Technology) 시대에서는 수작업의 자동화와 데이터의 가공·처리에 초점을 맞추었다면, DT(Data Technology) 시대에서는 데이터 자체를 활용해 제품·서비스를 기획하고 새로운 비즈니스를 창출하는 것이 중요해짐

• 21세기 원유, 빅데이터

- 데이터가 과거 원유처럼 사회 저변을 떠받치는 에너지원이자 모든 산업의 성장과 변화를 위한 필수자원으로 자리매김함
- 원유를 정제해 석유와 아스팔트 등 다양한 제품을 만들 수 있듯이, 앞으로 기업들은 빅데이터를 활용해 다양한 서비스를 창출할 수 있다는 의미임

02.

빅데이터의 개념

빅데이터의 개념

01

빅데이터 시대가 열린 근본적 이유

- 데이터 생산이 폭발적으로 증가했으며, 이에 필요한 기술적 성능의 고도화, 관련 비용의 하락에 따른 것임
- 각 분야별 빅데이터 등장 배경

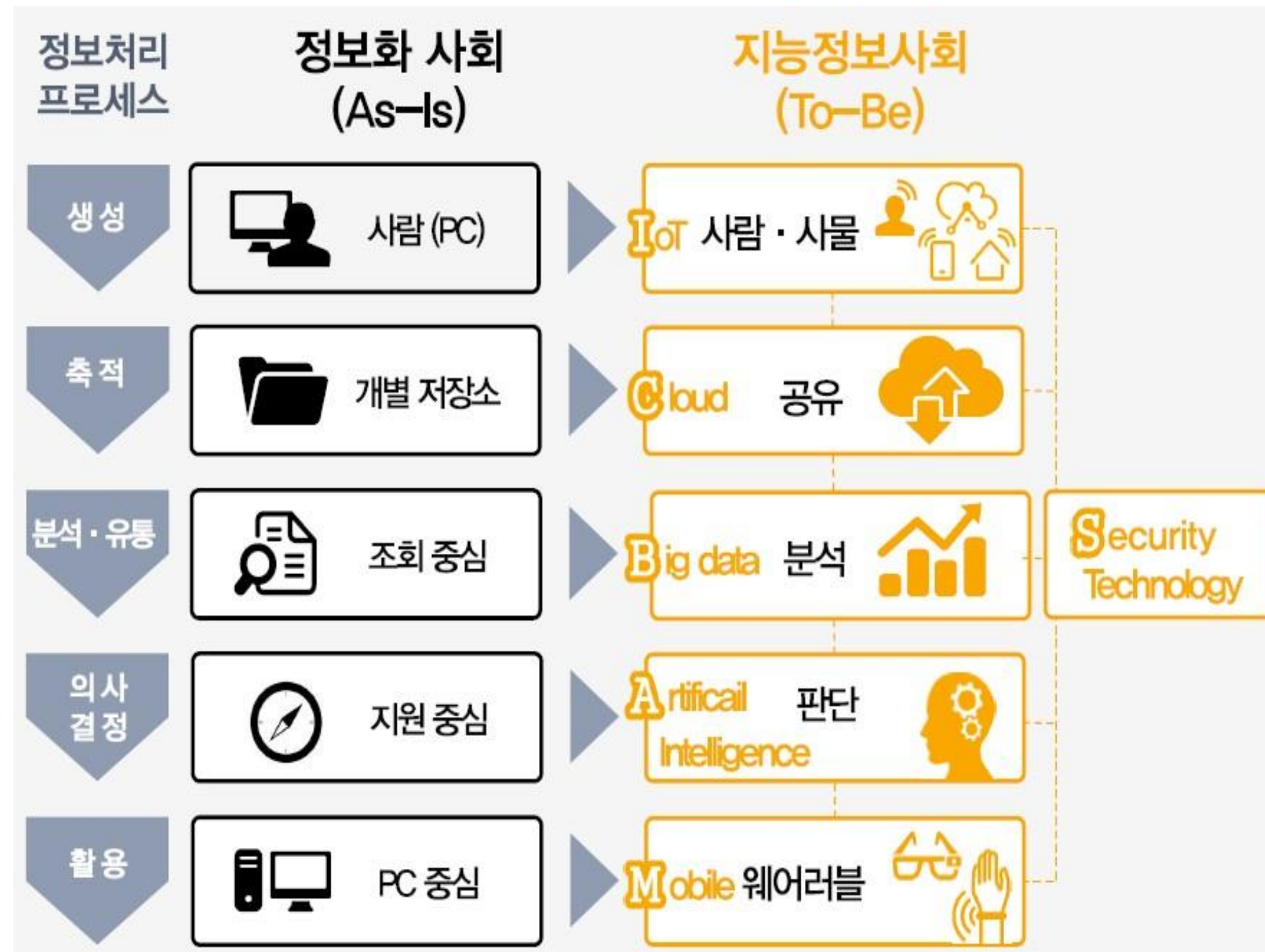
구 분		설 명
산업계 변화	고객 데이터 축적·활용 증가	<ul style="list-style-type: none">• 거대기업들이 온/오프라인에서 사용자/소비자 행태 정보를 수집/분석해 경영/경쟁 전략에 활용함• 정보가 지속적으로 축적되면서 기업들이 보유한 데이터가 거대한 가치창출이 가능할 만큼 충분한 규모에 도달함• 질적 변화를 일으킬 기술이 접목됨으로써 거대가치를 창출할수 있는 기회가 나타남• 보유 데이터에 숨어있는 가치를 발굴해 새로운 성장 동력원으로 만들어 낼 수 있음
학계 변화	거대 데이터 활용 과학 확산	<ul style="list-style-type: none">• 거대 데이터를 다루는 학문 분야가 늘어나 필요한 기술 아키텍처 및 통계도구들도 지속적으로 발전• 인간 게놈 프로젝트의 인간 유전자 정보해석이 10년에서 일주일로 가능할 만큼 기술이 발전함• 비용도 약 1만분의 1로 낮아짐• 대형 강입자충돌기의 경우 발생 데이터의 0.001%만 보관되 연구에 활용됨• NASA의 기후 시뮬레이션 센터에서 32페타바이트의 기후관찰 정보를 활용해 슈퍼컴에서 시뮬레이션함
기술 변화	디지털화 저장기술 인터넷 보급 모바일혁명 클라우드 컴퓨팅	<ul style="list-style-type: none">• 디지털화의 급진전, 저장기술의 발전과 가격하락• 인터넷 발전과 모바일 시대의 진전에 따른 클라우드 컴퓨팅 보편화• 아날로그의 디지털로의 전환은 데이터의 생산, 유통, 저장의 편리성을 획기적으로 개선함• 오리지널과 동일한 복제품을 만들어 낼 수 있음• 디지털 시대 초기에는 디지털화 비용과 저장 비용이 증가함• 그러나 압축 기술이 발전하면서 비용도 빠르게 감소하기 시작함

빅데이터의 개념

02

ICBM 시대의 근간 데이터

- 지능정보화 사회로의 촉진자



- "4차 산업혁명" 시대를 맞아 IoT, Cloud, Bigdata, Mobile, Security로 대변되는 "ICBMS"가 국가 핵심 경쟁력으로 급부상함
- ICBMS는 하나의 독립적인 정보통신기술이면서도 톱니바퀴처럼 맞물려 상호작용하는 플랫폼을 총칭하는 말로 빅데이터가 근간을 이루고 있음
- 인터넷에 연결된 수많은 IoT 기기들이 수집한 데이터는 Cloud에 모인 후 Bigdata분석을 통해 의미 있는 정보로 만들어짐
- 과거에는 알 수 없었던 새로운 정보를 언제 어디서나 Mobile로 공유하면서 새로운 가치를 창출하며, 이 모든 과정은 안전한 Security가 유지되는 상태에서 이루어 져야함을 의미함

빅데이터의 개념

03

빅데이터에 대한 다양한 개념

- 기존 데이터베이스 관리도구가 수집, 저장, 분석, 관리할 수 있는 범위를 넘어서는 데이터셋(Dataset)을 의미함
- 최근에는 대규모의 다양한 데이터에서 경제적으로 필요한 가치를 추출 가능한 차세대 기술과 아키텍처로 의미가 확장됨

초점	주체	정의내용
데이터규모 자체특성 변화에 초점	Mckinsy(2011)	• 일반적인 DB소프트웨어로 저장, 관리, 분석할 수 있는 범위 초과하는 규모의 데이터
	가트너 그룹(2012)	• 데이터의 양(Volume), 데이터 유형과 소스의 다양성(Variety), 데이터 수집/처리 속도(Velocity)의 급격한 증가로 나타난 현상
분석비용 기술적 변화에 초점	IDC(2011)	• 다양한 종류의 데이터로부터 저렴한 비용으로 가치를 추출 • 데이터의 초고속 수집/발굴/분석을 지원하도록 고안된 차세대 기술/아키텍처
인재·조직 포괄적 변화에 초점	노무라연구소	• 기본적인 데이터, 데이터 처리/저장/분석기술 이외에 의미 있는 정보도출에 필요한 인재/조직 까지 포함해야 함
	메이어-쾨베르그 & 쿠키어(2013)	• 대용량 데이터를 활용해 작은 용량에서 얻을 수 없었던 새로운 통찰/가치 추출 • 이를 활용해 시장, 기업, 시민, 정부 관계 등 많은 분야에 변화를 가져옴

👉 새롭게 나타난 것은 아니고 기존의 데이터, 처리방식, 다루는 사람/조직 차원에서 변화
- 데이터와 이를 다루는 기술차원에서 패러다임 전환

03.

빅데이터의 특성

빅데이터의 특성

01

빅데이터의 3V와 3A특성

- 폭발적인 빅데이터에 대한 관심은 3V라는 기술적인 특성이 주효했지만 분석적 관점에서 정확성과, 민첩성, 실행가능성이 보다 강화되었으며, 나아가 비즈니스 전략 및 경쟁방식에 커다란 변화를 가져옴



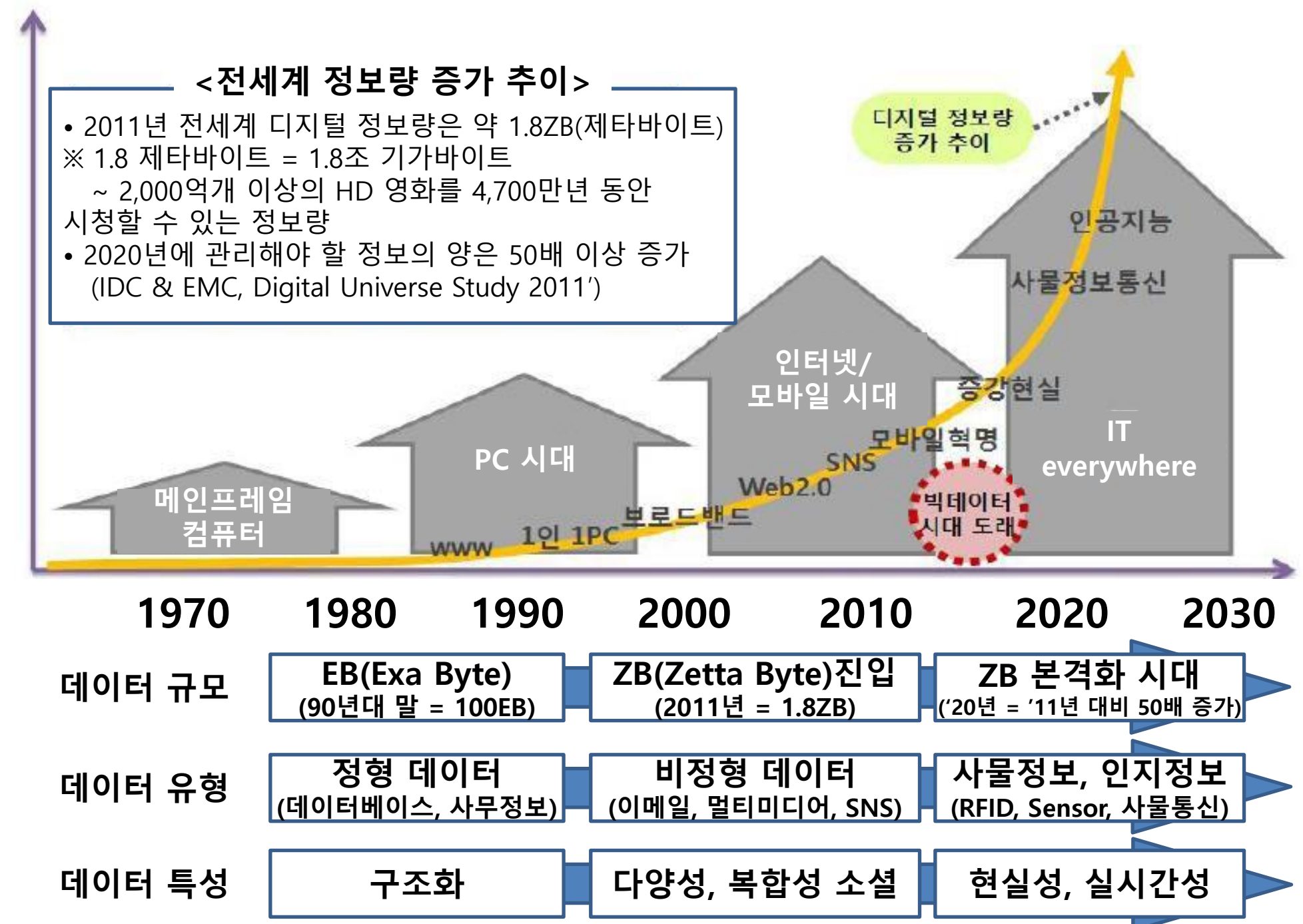
빅데이터의 특성

02

빅데이터의 물리적·기술적 특성

- IT의 일상화가 이루어지는 스마트시대에는 소셜, 사물, 라이프로그 데이터 등이 결합된 데이터의 폭발적 증가로 인한 혼돈과 잠재적인 가능성이 존재하는 빅데이터 시대가 도래함

- 실시간 연결과 소통의 스마트 혁명은 데이터 폭증을 발생시켰고 기존의 데이터 저장·관리·분석기법은 한계와 도전에 직면함
- 스마트 단말 확산, SNS활성화, 사물네트워크의 확산으로 데이터 폭발이 더욱 가속화되며 점차적으로 빅데이터 기반이 확대되고 있음
- 빅데이터는 고객정보와 같은 정형화된 자산정보(내부) 뿐만 아니라 외부데이터, 비정형, 소셜, 실시간 데이터 등이 복합적으로 구성됨
- 빅 데이터는 규모, 다양성, 속도의 증가 특성을 갖고 있으며, 이 요소들이 충족될수록 빅데이터에 적합함



빅데이터의 특성

03

빅데이터의 분석적·사용자적 특성(1/2)

- 빅데이터를 기술요소로만, 또는 통계전문가의 전유물로 인식하기 보다는 누구나 주어진 문제해결을 위한 도구로서 분석을 왜 하며, 분석결과를 어떻게 활용한 것인지에 더 많은 관심을 갖는 것이 필요함

스몰데이터 분석사례

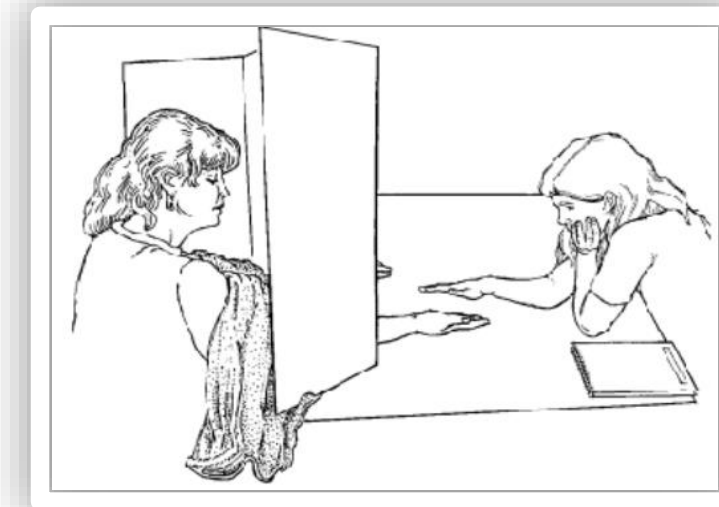
• 미국의 기치료 현황

- 환자의 에너지 장을 다스리면 병이 치료된다는 주장
- 기치료사가 환자의 몸 10센티미터 위에서 양손을 머리에서 다리쪽으로 움직이면서 나쁜기운을 제거해 치료
- 세계적으로 100여 개 간호대학 정규교육과정운영, 북미 80 개 병원에서 간호사의 정식치료서비스

- 에밀리(초3)는 기치료 비디오를 보고 과연 효과가 있을까 의문을 품음
- 만일 기치료가 실제로 무의미 하다면 상당한 교육적·사회적 불필요한 비용을 치르고 있는 심각한 상황임
- 간호사 엄마와의 공동연구진행

기치료에 대한 심층연구 (A Close Look at Therapeutic Touch)

- 관련 연구조사 : 853편 논문조사
- 성과변수설정 : 에밀리의 손에서 나오는 기를 느낀 기치료사들의 비율
- 자료수집 : 기치료사 21명 대상 총 280회 실험 실시
- 가설검증 : 손의 위치를 맞춘 기치료사는 44%에 불과해 효과 없음
- 연구성과 : 최연소 유명학술지(JAMA, 1988) 논문 게재로 기네스북에 등재됨



빅데이터의 특성

04

빅데이터의 분석적적·사용자적 특성(2/2)

- 기술발달에 따른 빅데이터의 등장으로 데이터분석을 통한 보다 정확하고, 유연하고, 근거있는 문제해결 및 액션시나리오의 기획이 가능해짐

스몰데이터에 대한 빅데이터 방식의 접목

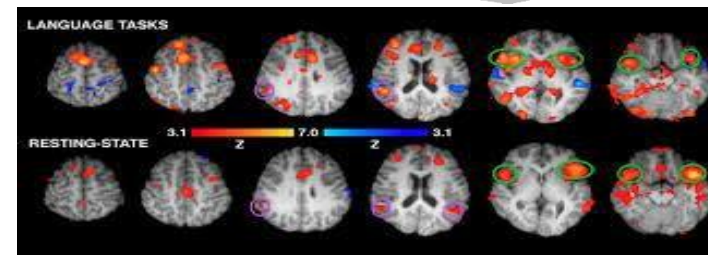
- 기존 스몰 데이터 연구사례의 한계점

- 기치료사는 환자의 나쁜 기운에 반응해 좋은 에너지를 방출하는 훈련이 되어 있음
- 맑은 초등 3년의 손에는 부정에너지가 약할 가능성
- 따라서 새로운 접근방식의 기치료 효과에 대한 검증 필요

- 기치료 무용론 프레임으로 해석과 액션
 - 미국의 기치료 교육의 실제적 효과가 없음이 검증되었으므로 현재 각급 대학의 교육과정운영과 의료기관의 관련 의료서비스의 전면 중단·폐기 근거로 활용
- 기치료 활성화 프레임으로 해석과 액션
 - 동양한의학에 비해 미국 기치료 교육과 실제 의료서비스 수준이 현저한 격차를 보이고 있으므로 이를 해소하기 위한 대규모 연구개발비용 투자의 근거로 활용

1 fMRI(Functional magnetic resonance imaging; 기능적 자기공명영상)분석

- 기치료시 나타나는 혈류나 포도당 대사의 환자 뇌기능 인지과정을 분석해 검증



2 소셜데이터 텍스트마이닝분석

- 기치료에 대한 소셜미디어 상의 토픽, 연관, 감성어 분석해 검증



3

분석을 왜 하는가?

- 문제특성은 스몰과 빅이 동일하지만 접근방법의 변화필요
- 빅데이터는 완전히 새로운 분석주제가 아니며, 어제의 문제가 오늘도, 그리고 내일도 계속해서 나타남 분석결과에 대한 해석과 액션방안이 중요함

- 기치료 무용론 프레임으로 해석과 액션
 - 미국의 기치료 교육의 실제적 효과가 없음이 검증되었으므로 현재 각급 대학의 교육과정운영과 의료기관의 관련 의료서비스의 전면 중단 폐기 근거로 활용
- 기치료 활성화 프레임으로 해석과 액션
 - 동양한의학에 비해 미국 기치료 교육과 실제 의료서비스 수준이 현저한 격차를 보이고 있으므로 이를 해소하기 위하여 대규모 연구개발비용 투자의 근거로 활용

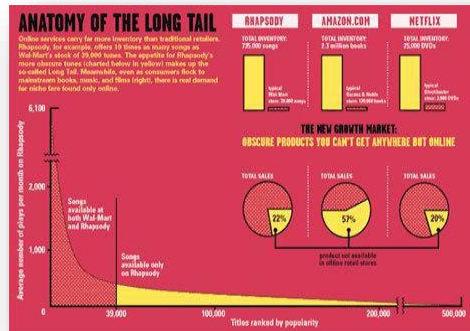
빅데이터의 특성

05

비즈니스 관점의 3V

- 빅데이터는 제품·서비스와 이용자 확대(Volume), 다양한 연관·파생 비즈니스에 적용(Variety), 트렌드세터로서 전략적 기회를 선점(Velocity)할 수 있음

아마존 롱테일효과 : 제품·서비스와 이용자 확대(Volume)



- 매장선반보다 온라인에서는 디지털화로 진열 상품가지수가 늘어남
- 검색 및 추천을 통해서 진정한 틈새상품, 과거 히트상품, 전통 경로에서 미 판매제품까지 가능
- 특색제품구비에 대한 호감도 상승, 히트상품과 틈새제품을 같이 찾는 구매행동에 변화

페어캐스트 항공권 가격추천 : 다양한 연관·파생 비즈니스에 적용(Variety)



- 승객마다 가격은 천차만별, 항공사는 예약부도 문제
- 전노선의 항공권 가격과 남은 날짜에 따라 팔린 가격분석
- 호텔룸, 콘서트·레스토랑 티켓 등 땡처리 비즈니스
- 또한 중고차 거래 등 제품차별성이 적고, 가격이 동적인 비즈니스에 확대 적용

유유제약 멍치료제 : 트렌드세터로서 전략적 기회를 선점(Velocity)



- 소셜분석 결과 사람들은 멍 치료연고 존재자체를 잘 모름
- 멍이 발생하면 소고기와 계란 등 민간요법과 "가린다"라는 행위가 연관어로 나타남
- 여름이외에 겨울에도 버즈량이 많다는 것을 발견
- 기존 어린이이용 부종완화제 제품에 대한 리포지셔닝으로 새로운 시장기회 포착

04.

스몰데이터와 빅데이터의 비교

스몰데이터와 빅데이터 비교

01

공통점과 차이점

- 스몰데이터와 빅데이터간의 공통점과 차이점에 대한 이해를 토대로 분석에 대한 다양한 견해를 파악할 수 있음

구분	스몰데이터 분석	빅데이터 분석
분석 목적	<ul style="list-style-type: none">표본분석을 토대로 모집단의 특성을 추론하여 해석함비교집단간 특성차이,조치수단 간 효과차이 비교를 통해 관심대상 집단·수단을 선별하는데 중점	<ul style="list-style-type: none">대규모 데이터에 숨어있는 패턴을 발견하고 규칙을 도출함도출된 패턴 및 규칙을 이용해 개별대상별 (고객·제품 등) 액션 방안 마련에 중점
데이터 수집	<ul style="list-style-type: none">사용자를 대상으로 한 서베이나 포커스그룹 인터뷰 활용으로 상당한 조사기간이 수반됨	<ul style="list-style-type: none">대규모 내부거래처리 데이터, 음성, 동영상, 외부공공 및 소셜 데이터 활용으로 실시간에 가까운 분석
분석 기법	<ul style="list-style-type: none">기술통계 및 차이분석, 회귀분석, 구조방정식 등 추론통계 중심	<ul style="list-style-type: none">군집분석, 연관분석, 분류분석, 예측분석 등 정형 및 텍스트 마이닝 기법이 추가됨
분석 도구	<ul style="list-style-type: none">전통적 SPSS, SAS 등 상용 통계분석 패키지 활용	<ul style="list-style-type: none">R, 하둡 등의 오픈소스 및 클라우드 기반의 분석도구 활용
적용 분야	<ul style="list-style-type: none">제조, 금융, 유통, 관광, 보건, 행정, 국방 등 공공 및 민간 전 분야를 망라함	<ul style="list-style-type: none">동일한 적용분야를 가지며, 보다 다양한 데이터소스 및 분석기법 활용으로 분석근거의 정확성과 예측력이 향상됨
분석 사용자	<ul style="list-style-type: none">통계학 및 마이닝 전문가에게 의뢰를 통한 분석: Analysis	<ul style="list-style-type: none">업무실무자 스스로 셀프분석을 통한 의사결정에 활용: Analytics

스몰데이터와 빅데이터 비교

02

빅데이터의 본질적·근본적 변화(1/2)

- 기존의 스몰데이터 환경에서의 데이터분석 특성 대비 빅데이터 환경에서는 다음과 같이 4가지 측면의 근본적인 변화가 나타남

1

사전처리(pre-processing)

- 쏟아지는 정보의 양이 정보관리 시스템의 한계를 넘어섬
- 필요한 정보만 수집하고 불필요 정보를 버려 시스템 효율성 강화
- 가치가 있을 것이라고 정해진 특정한 정보만 모아서 처리함
- 사전에 표준화 포맷으로 인쇄된 문서를 통해 필요한 정보만 수집
- 개개인의 특수상황을 반영하는 세부 정보수집은 포기해 관리비용을 줄임
- 사용자 로그데이터에 관심을 가지지 않음

사후처리(pro-processing)

- 가능한 많은 데이터를 모으고 그 데이터를 다양한 방식으로 조합해 숨은 정보 찾아냄
- 구글은 로그 데이터를 분석해 광고 매칭해 세계적 기업으로 성장함

2

표본조사(sampling survey)

- 통계학의 발전목표가 적은 데이터로 보다 풍부한 결과를 얻기 위한 것이었음
- 데이터 수집비용, 대용량 데이터 처리도구 비용 등에 부담이 됨
- 분석목표에 따라 필요한 정보를 수집하므로 질문이 바뀌면 데이터를 다시 수집해야 함
- 데이터 활용 융통성이 매우 떨어짐

전수조사(population census)

- 인터넷 로그 등을 이용해 특정 사이트를 이용하는 사용자 전수조사가 가능함
- 샘플링이 주지 못하는 패턴이나 정보를 다양하게 제공해 줌
- 모든 데이터를 모아서 상황에 필요한 질문을 통해 다양한 방식으로 데이터를 재가공이 가능해짐
- 데이터 활용 융통성이 매우 높음

스몰데이터와 빅데이터 비교

03

빅데이터의 본질적·근본적 변화(2/2)

- 기존의 스몰데이터 환경에서의 데이터분석 특성 대비 빅데이터 환경에서는 다음과 같이 4가지 측면의 근본적인 변화가 나타남

3

질(quality)

- 정확하고 모든 조건을 충족시키는 샘플/사례만 분석하게 되는데, 이처럼 모든 조건을 충족하는 데이터세트는 상당부분의 사례를 제외시키는 결과를 낳음
- IBM은 고도로 정교하게 번역된 말뭉치(corpus)만을 이용해 수백만건 수준에 불과함

양(quantity)

- 엄청난 데이터의 양이 질적으로 전환되는 과정에서 인사이트가 나타남
- 제외되는 사례 없이 다른 변수들도 모두 분석에 활용해 더 많은 가치를 추출하게 됨
- 데이터 수가 증가함에 따라 사소한 몇 개의 오류는 대세에 영향을 주지 못함
- 지속적으로 추가될 때 양질의 정보가 오류 정보보다 많기에 전체적으로 좋은 결과산출에 기여함
- 번역품질의 데이터베이스의 양에 따라 성패가 갈림
- 구글은 번역에 참고할 말뭉치 데이터베이스로 잘된 것, 오역 있는 것을 다 수용해 수십억 건에 이름

4

인과관계(causation)

- 데이터 수집비용이 매우 비싸 이론에 기초한 변인결정, 엄격한 실험을 통한 정제된 데이터 획득, 정교한 인과관계(causation)를 찾음
- 특정 현상이 일어나는 이유와 과정을 설명하려고 함
- 이론의 종말처럼 인과관계가 완전히 불필요한 것은 아님
- 변인들 간의 인과관계를 많이 알수록 현상에 대한 이해와 폭이 깊어짐

상관관계(correlation)

- 데이터획득비용이 기하급수적으로 감소, 모든 곳에서 데이터가 넘쳐남
- 인과관계 규명(이유를 몰라도) 없이 상관관계로 비즈니스를 성공적으로 이끌 수 있음
- 상관관계 분석이 주는 인사이트가 인과관계에 의한 미래예측을 점점 더 압도하게 됨
- 신속한 의사결정을 원하는 비즈니스에서 실시간 상관관계 분석에서 도출된 인사이트 바탕으로 수익창출가능
- 독감확산 경로예측에 해당 키워드의 검색빈도 증감을 살펴보는 정도로 충분
- 특정 지표의 변화가 주가와 밀접한 상관관계가 있으면, 신속히 거래해 이익을 취하는게 중요함
- 인과관계 분석은 타이밍을 놓치게 되어 주식거래 목적에 부합하지 않음

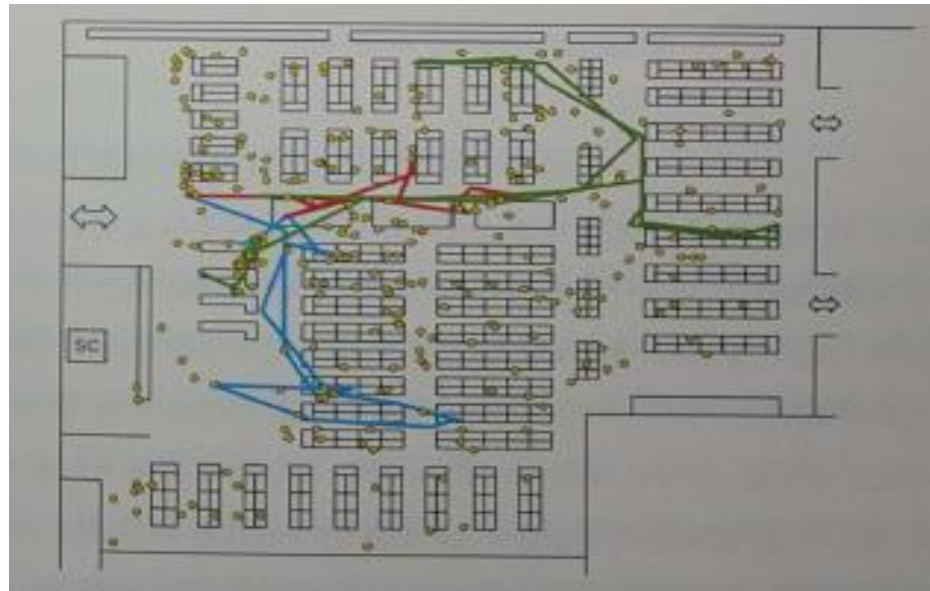
스몰데이터와 빅데이터 비교

04

데이터분석의 어제와 오늘

- 기존의 스몰데이터에 기반한 데이터분석 제약조건에서 벗어나 다양한 내·외부, 정형·비정형 빅데이터 획득을 통해 새로운 통찰과 가치를 창출함

대형매장 고객동선 추적



- (스몰) 조사요원이 표본고객을 따라다니면서 동선과 구매활동을 관찰 · 기록
- (빅) 이동카트 및 선반, 물품, CCTV 등에 센서를 부착한 IoT를 이용해 고객의 매장내 구매활동을 입체적으로 센싱
- 고객별 이동순서, 탐색시간, 미구매 상품 등 파악

외국인 관광객 만족도조사



- (스몰) 공항에서 출국하는 외국인 관광객 표본을 대상으로 연차보고서 분석
- (빅) 외국관광객이 주로 사용하는 SNS 및 해당국가 언어를 직접분석
- 방문전의 기대사항 및 방문과정의 체험, 방문후의 느낌을 종합적으로 분석

질병치료법 검색



- (스몰) 매일 3천여편의 의학논문이 게재, 의사들의 논문 리딩 시간은 일주일에 몇시간에 불과해 특정증상에 대한 치료법 습득 어려움
- (빅) 3초안에 증상과 관련한 원인 및 치료법 등을 제시함
- 연관증상에 대한 추가정보 제공으로 종합적인 치료 가능

스몰데이터와 빅데이터 비교

05

빅데이터의 현주소

- 빅데이터에 대한 공급자관점의 마케팅 열풍이 절정으로 치닫고 있으며, 데이터분석의 본질적 가치인 문제해결에 대한 관심이 증대되고 있음
- e-비즈니스 열풍의 교훈
 - 인터넷이 등장했던 1990년대 중반 기존의 전통적 오프라인(offline; click and mortar) 기반의 비즈니스를 인터넷 기반의 온라인(online; dotcom) 변환해야 한다는 e-비즈니스(electronic business)에 대한 열풍이 일어남
 - 이후 M(모바일)-비즈니스, U(유비쿼터스)-비즈니스, 최근의 O2O(online to offline) 서비스에 이르기까지 온라인과 오프라인에 대한 접목을 강조하고 있음
 - 이제는 e라는 표현이나 온라인이란 표현 없이 "비즈니스"는 당연히 온오프 구별 없이 상호 접목이 되어야 하는 것이 하나의 원칙이 되고 있음
- 마케팅적 용어표현 보다는 빅데이터는 결국 문제해결이 핵심
 - 빅데이터도 기존의 스몰데이터와의 물리적·기술적 차이를 강조하는 마케팅적 표현이 초창기에 과열되었음
 - 그러나 데이터와 인프라 측면의 규모, 다양성, 속도라는 물리적·기술적 차이만을 강조하다보니 데이터분석의 본질적 가치에 대한 맹신과 과신, 회의와 불신이라는 부작용을 낳게 됨
 - 빅데이터라는 용어도 데이터마이닝, 머신러닝, 딥러닝, 인공지능 등 다양한 기술적, 방법론적 용어들로 확대 재생산 되고 있음
 - 이제는 스몰데이터 vs. 빅데이터라는 용어보다는 주어진 문제해결에 전문가의 직관과 업무경험에 더해 데이터를 기반으로한 근거를 접목하는 균형감 있는 의사결정이라는 문화적 패러다임이 더 중요하다고 판단됨



비즈니스는
온 오프 구별 없이
처리되어야 함

Big Data
vs. Small Data



05.

데이터과학자 보다 문제 해결사 되기

데이터과학자 보다 문제 해결사 되기

01

분석대가들이 진정으로 우리에게 원하는 것

- 수백년간의 유산을 보유한 통계학과 머신러닝이 진정으로 전해주는 메시지는 통계·머신러닝 전문가가 아닌 다양한 현안들에 분석기법을 적절하게 활용할 줄 아는 문제해결사가 되라는 것임

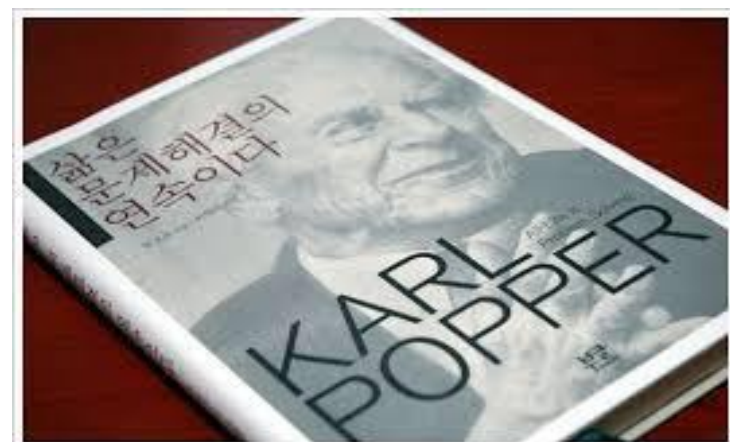
- 영어회화를 익혀나가는데 있어 영문학과 전공자처럼 동일한 커리큘럼과 범위, 수준, 일정으로 학습하는 것이 상당히 비효율적이고 결코 그럴 이유도 없으며, 커뮤니케이션을 위한 하나의 수단으로 접근하는 것이 필요함
- 현재 국내외에서 진행되는 빅데이터 관련 커리큘럼은 통계학이나 머신러닝 전공자의 수업과 동일한 커리큘럼과 수준으로 진행되고 있어서 비전공자들에게 학습로드가 상당한 부담이 됨
- 수학적으로 생경한 개념과 용어들, 알 수 없는 수식으로 가득찬 지문들, 왜 보다는 어떻게에 치우쳐 있는 설명들로 인해 분석자체가 하나의 숙제가 되버림
- 자기 업무에 필요한 문제들 먼저 리스트업하고 이를 해결하기 위한 하나의 해결도구로서 분석적 방안을 활용중심으로 하나·둘씩 익혀나가는 접근방법이 요구됨

데이터 사이언티스트



- 통계학을 전공해야 한다?
- 머신러닝 전문가가 되어야 한다?
- 데이터분석을 해야 업무가 혁신된다?
- 분석과제를 기획해야 한다?

문제 해결사



- 절대로 분석전공자처럼 될 수 없다.
- 결코 분석전공자와 경쟁할 이유가 없다.
- 혁신을 위한 방법은 얼마든지 있다.
- 누구든지 적절한 수준에서 즐겁게 활용할 수 있다.

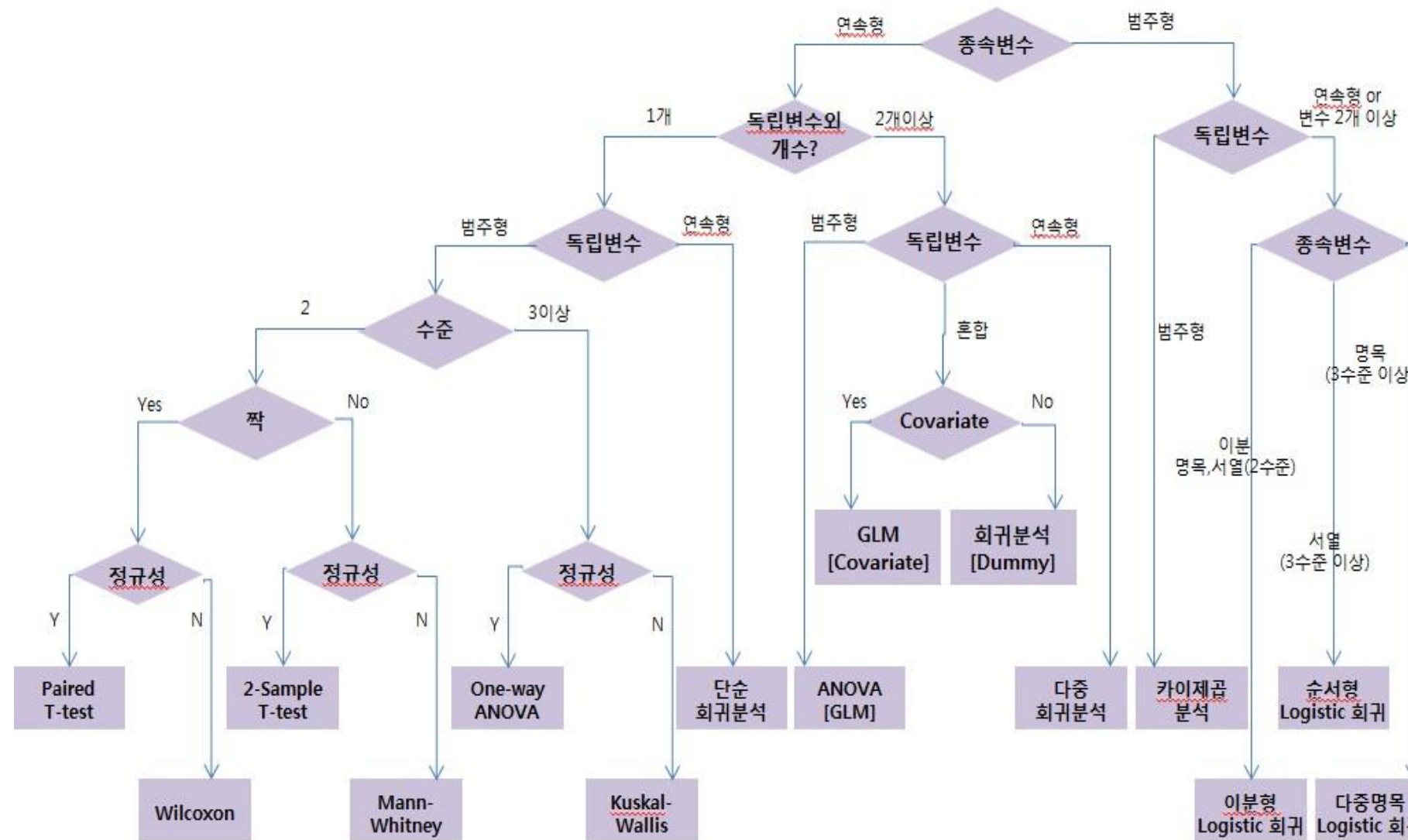
데이터과학자 보다 문제 해결사 되기

02

통계·머신러닝 이론의 블랙박스화

- 빅데이터는 전통적인 통계학의 프레임, 수학적 알고리즘에서 벗어나 문제해결 중심의 재해석을 통한 적용·활용방안 중심의 사고전환이 필요함

현행 데이터분석 접근방식



- 공급자 관점의 분석기법
 - 분석기법들의 명칭이 학자와 분야에 따라 달라 혼란유발
- 수단과 목적이 뒤바뀜
 - 고수준의 분석기법 사용을 위해 기법에 문제특성을 끼워 맞춤
- 전공학습 프레임에 갇힘
 - 통계학·머신러닝 전공자들과 동일한 방식·내용의 학습을 당연시함

- 업무활용 관점에서 재해석 필요
 - 통계학·머신러닝의 난해한 이론체계를 블랙박스로 가정하고 이용·활용에 초점
 - 자기의 문제특성에 맞는 적절한 분석기법·원리의 선별능력중요

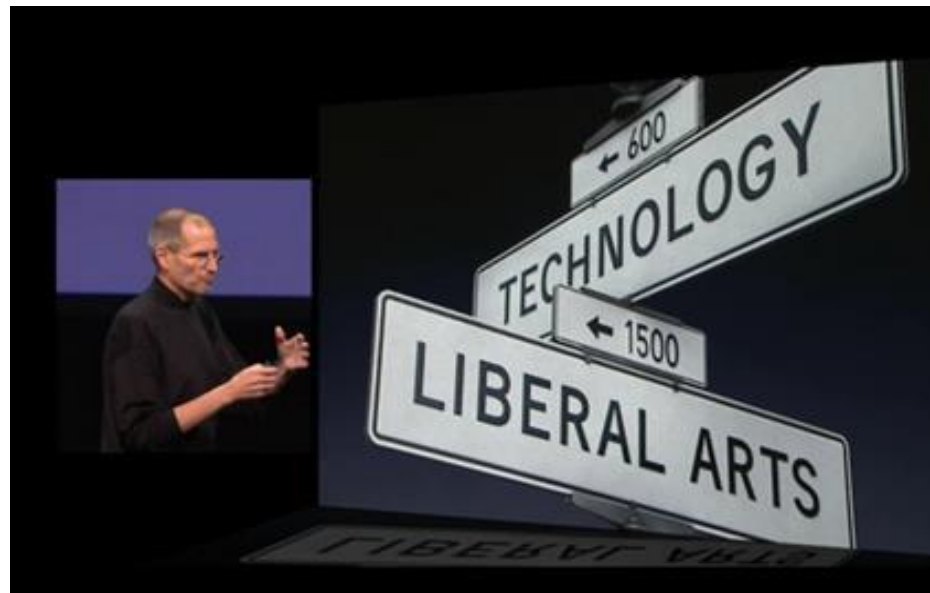
데이터과학자 보다 문제 해결사 되기

03

데이터분석의 과학상자화·레고박스화

- 누구든지 자신의 관심과 경험을 바탕으로 분석대가들이 만들어 놓은 분석기법을 선택·활용하여 자신의 업무에 대한 문제해결사(Problem Solver)가 될 수 있음

기술과 인문학의 교차점



- 아이폰에 들어가는 각종 전자·센서 부품의 기술적 작동원리를 난해한 공학이론, 수학기초 수준이 아닌 개념적으로 이해
- 사용자들의 생활양식, 업무환경 속에 ICT 기능특성이 어떻게 적용·활용될 수 있는지에 대한 탁월한 아이디어를 제품·서비스로 만들어 냄

과학상자 놀이



- 아이들이 과학상자를 가지고 놀 때 각 부품들의 규격, 동작원리, 성분, 주의사항 등에는 관심이 없음
- 자동차, 비행기, 배, 장비 등 만들고 싶어하는 구체적인 대상(목표)을 정하고 각 부속품을 자유롭게 활용하는 데 중점을 둠

온도 데이터 분석



- 온도를 측정하는 도구로 온도 변화시 액체물질의 열팽창성, 열기전성, 전기저항성, 방사열발생성을 측정하는 개념적 원리정도 이해하면 됨
- 업무에서 온도가 관련된 분석 대상·주제는 무엇인지, 온도의 변화에 영향을 미치는 선행변수와 결과변수에 대한 학습이 더 중요함

데이터과학자 보다 문제 해결사 되기

04

데이터분석을 학습하기 위한 3가지 방법

- 데이터분석을 익히고 활용하기 위한 접근방식에는 데이터분석 기법중심, 분석도구 중심, 문제해결 중심의 3가지가 있으며 개인적 성향 및 업무여건에 따라 고려가 필요함

데이터분석 기법 중심



- 가장 일반적인 접근방법
- 기초통계학, 확률분포, 데이터마이닝, 텍스트마이닝, 최적화 등 분석기법을 하나씩 학습함
- 통계학, 산업공학, 경영공학 등 전공특성에 따라 사용하는 분석기법 종류에 차이가 있음
- 전통적 통계학·마이닝 접근법

분석도구 중심



- R, SAS, SPSS, Matlab 등 분석도구의 기능코드나 메뉴를 하나씩 익혀나감
- 도구마다 특화된 학문분야 및 산업분야가 있음
- 추론통계 및 마이닝 등 기법별로 개별 소프트웨어를 익혀야 함
- 개발자 중심 접근법

문제 해결 중심

보조 활동	하부구조	하부지원 : 기획, 재무, MIS, 법무, 총무				
	기술개발	연구, 설계, 개발, 디자인				
	인적자원	직무관리, 보상관리, 평가관리, 조직관리				
주요 활동	기술획득	제품설계	제조	마케팅	유통유통	서비스
	·원천기술 ·기술정보화 ·지적재산권 ·생산공정	·제품기획 ·특성 ·디자인 ·품질	·입지 ·공수 ·원재료조달 ·부품/조립	·상품, 가격 ·광고, 홍보 ·판매촉진/판매원 ·포장, 배달	·수송 및 저장 ·유통 ·통합	·보통제도 ·직접/특립 ·속도 ·가격

Source: Michael E. Porter, Competitive Advantage, 1985.

- 내가 속해 있는 산업·업종, 기능부서 등의 고유한 이슈 중심
- 전통적인 중시되어오는 핵심 성과지표의 현황 및 목표를 통해 분석과제 접근이 가능
- 또는 경쟁사·벤치마킹을 통해 새로운 KPI도입가능
- 문제해결사 중심 접근법

06.

세상의 다양한 분석자원들

세상의 다양한 분석 자원들

01

[A]분석대상·객체(1/2)

- 데이터분석 대상인 세상에 존재하는 다양한 유·무형의 객체마다의 고유한 내재속성과 행동속성, 이들간의 관계 속에 모든 문제와 해답이 들어 있음

사람·내외부 고객



- 보는 비즈니스의 중심에는 고객이 있으며, 고객이 제품·서비스를 구매하고, 불만을 제기하는 주체라는 측면에서 프로필, 라이프스타일, 구매이력 등이 분석대상임

제품·서비스



- 기업이 만들어 내는 제품 서비스가 과연 유행을 선도하고 고객의 선택을 받을 수 있는지 측면에서 제품명, 가격, 무게, 색상, 기능, 디자인, 내구성 등이 분석대상임

온·오프채널



- 고객이 원하는 제품·서비스 기업이 어떤 경로로 전달한 것인지 측면에서 채널의 메뉴구성, 이용단계, 이동경로, 이용빈도, 구매율, 편의성, 결재수단, 정보구조 등이 분석대상임

시설·장비



- 제품·서비스 생산과 제조 위한 시설·설비,장비의 운영·관리 측면에서 온도, 회전수, 에러율, 소음, 진동, 색상, 동작횟수, 사용연한 등이 분석대상임

세상의 다양한 분석 자원들

02

[A]분석대상·객체(2/2)

- 고객이 제품서비스를 이용하는 환경적 측면에서 영향을 미치는 시공간 요소와 자연조건, 정책여건에 대해서도 분석모델에 포함해야 함

시간·요일·기간·트렌드



- 24시간 365일 비즈니스가 일관되게 진행되지 않는다는 측면에서 고객이 제품서비스를 언제, 어느 기간동안, 어느 추세로 이용하는지 등이 분석대상임

지역·장소·공간·위치



- 지대와 장소에 따라 고유한 문화적 특성이 존재하며, 지리적인 접근성, 유동성, 이용조건이 다르다는 측면에서 지역, 장소, 공간, 위치 등이 분석대상임

날씨·기상·자연



- 가장 변화무쌍하게 비즈니스에 영향을 미치는 외부환경적 요인이라는 측면에서 기온, 습도, 풍향, 풍속, 강수량, 강수확률, 대기질, 기상특보, 재난, 이상기후, 이동경로, 발생위치 등이 분석대상임

정책·법률·규제·제도



- 각종 규제, 정책, 법률은 기업에게 코스트 상승요인이자 기업발전의 저해 또는 발전을 위한 기반이라는 측면에서 공정거래, 건강, 안전, 환경 및 보안 준수 등이 분석대상임

세상의 다양한 분석 자원들

03

[B] 산업 & 업종 이슈

- 데이터분석은 지구상에 존재하는 산업·업종의 문제해결을 위한 하나의 도구이므로 각 산업과 업종에서 중요한 비중을 차지하는 고유한 문제와 최근에 부각되는 이슈에 집중해 데이터분석을 접목해야 함

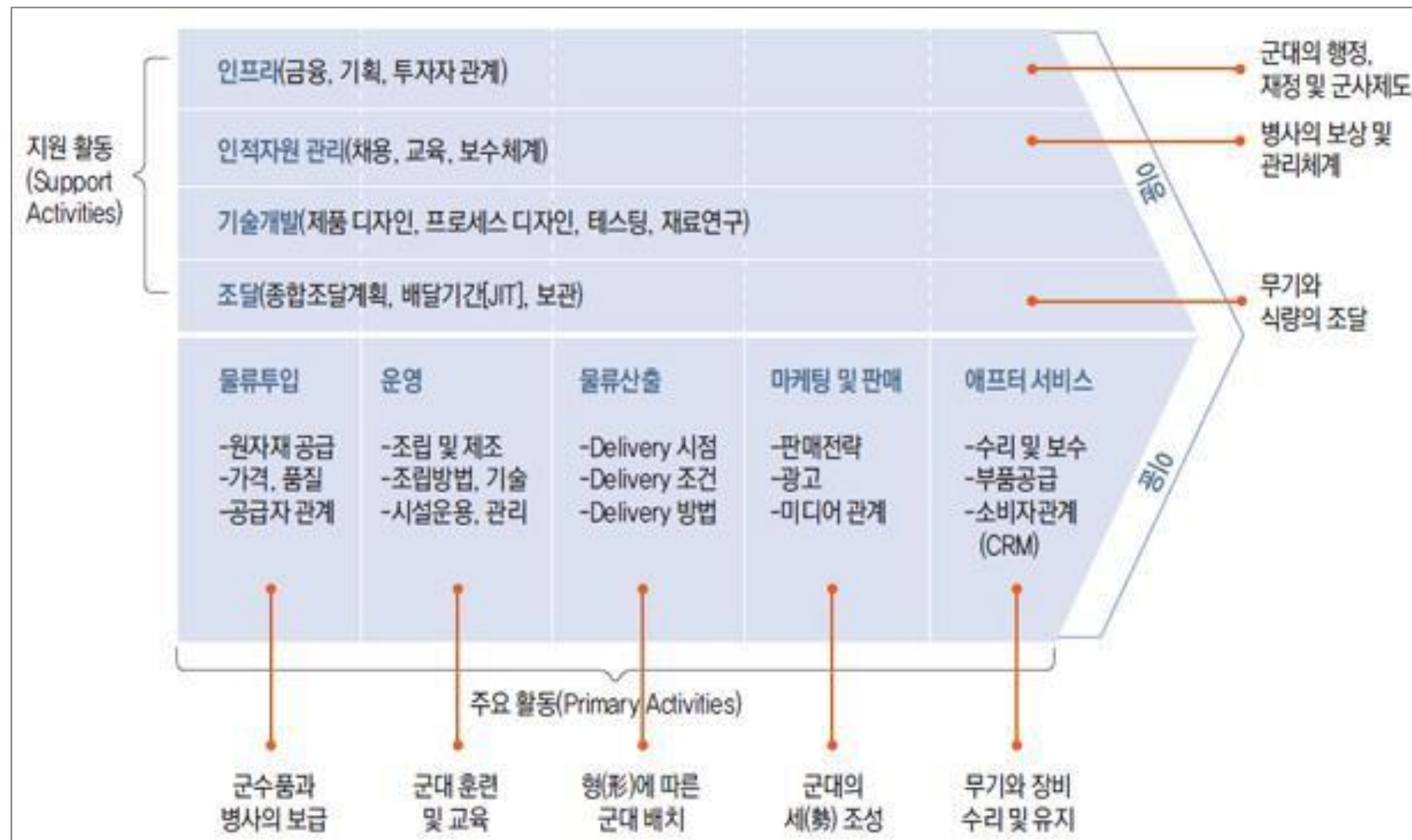
산업분류	업종
농업·임업·어업	• 농업(작물재배, 축산), 임업(임업·벌목, 채취), 어업(어로, 양식)
광업	• 석탄·원유·천연가스, 철·비철금속, 비금속광물(건설자재 채취·석쇠)
생산·제조	• 식료품, 음료, 담배, 섬유, 의복·액세서리, 잡화, 화학물질, 의약품, 전기·전자, 의료·정밀, 기계·장비, 자동차, 가구, 기계, 악기·귀금속 등
수도·하수·폐기물·재생	• 수도, 하수·폐수·분뇨처리, 폐기물 수집·운반·처리·원료재생, 환경정화·복원
건설	• 종합건설(건물·주택·토목·지반·시설·조경), 전문직공사(건축·해체, 설치·장식, 통신·전기)
도소매·유통	• 자동차·부품판매, 도매·상품중개, 소매(백화점, 대형마트, 전자상거래, 통신·방문판매)
운송·물류·창고	• 육상운송, 수상운송, 항공운송, 창고·보관, 물류·택배
숙박·음식점	• 숙박(일반, 호텔·콘도, 기숙·고시원), 음식점·주점(한식·외국식, 간이·배달, 주점, 커피)
정보통신	• 출판(서적, SW·게임 개발·공급), 영상·오디오 제작·배급, 방송, 우편·통신, 컴퓨터 프로그래밍, 시스템통합·관리, 정보서비스(포털, 자료처리, 뉴스제공)
금융·보험	• 금융(은행, 조합, 신탁·투자, 카드·할부, 여신), 보험·연금, 금융서비스(증권, 선물, 연금)
부동산	• 부동산·건물 개발·임대·공급, 부동산 중개·자문·감정평가
전문과학기술서비스	• 연구개발업, 전문서비스(법·회계·광고·조사·컨설팅), 건축·엔지니어링, 전문과학기술서비스(수의업, 디자인, 사진, 통번역)
사업지원서비스	• 사업시설(유지·관리, 청소·방제, 조경), 사업지원(고용알선·인력공급, 여행사, 경호·경비, 콜센터), 임대업(운송장비, 용품, 콘텐츠,)
공공행정·국방	• 정부행정, 국방행정, 사회보장행정
교육서비스	• 유아·초등·중등·고등대학, 학원·수련시설, 사회교육·직업훈련, 온라인학원
보건·사회복지	• 보건(병원, 요양원, 한의원), 복지·보육시설·복지관, 상담서비스,
예술·스포츠·여가	• 창작예술·여가관련 서비스, 스포츠·오락관련 서비스, 유원지·테마파크, 복권·사행시설

세상의 다양한 분석 자원들

04

[C] 가치체인 업무활동 이슈(1/2)

- 일련의 활동을 통해 가치를 창출하는 가치체인(Value Chain)에서도 데이터분석을 통해 보다 생산적·효과적으로 만들어 나갈 수 있음



- 가치사슬은 한 기업의 활동을 연관성이 있는 몇 개의 활동들로 나누어 경쟁우위의 원천인 저원가 또는 차별화가 가능한 부분을 찾아내기 위한 프레임워크임
- 주요활동(primary activities)은 직접적으로 고객에게 부가가치 창출에 기여하는 활동으로서 물류투입, 운영, 물류산출, 마케팅과 판매, 애프터서비스 등 다섯 가지 활동으로 구성됨
- 지원활동(support activities)은 직접적으로 부가가치를 창출하지는 않지만 다른 활동을 보조해주는 기능을 하는 조달, 기술개발, 인적자원관리, 인프라(금융, 기획, 투자자관계) 등 네 가지 활동으로 구성됨

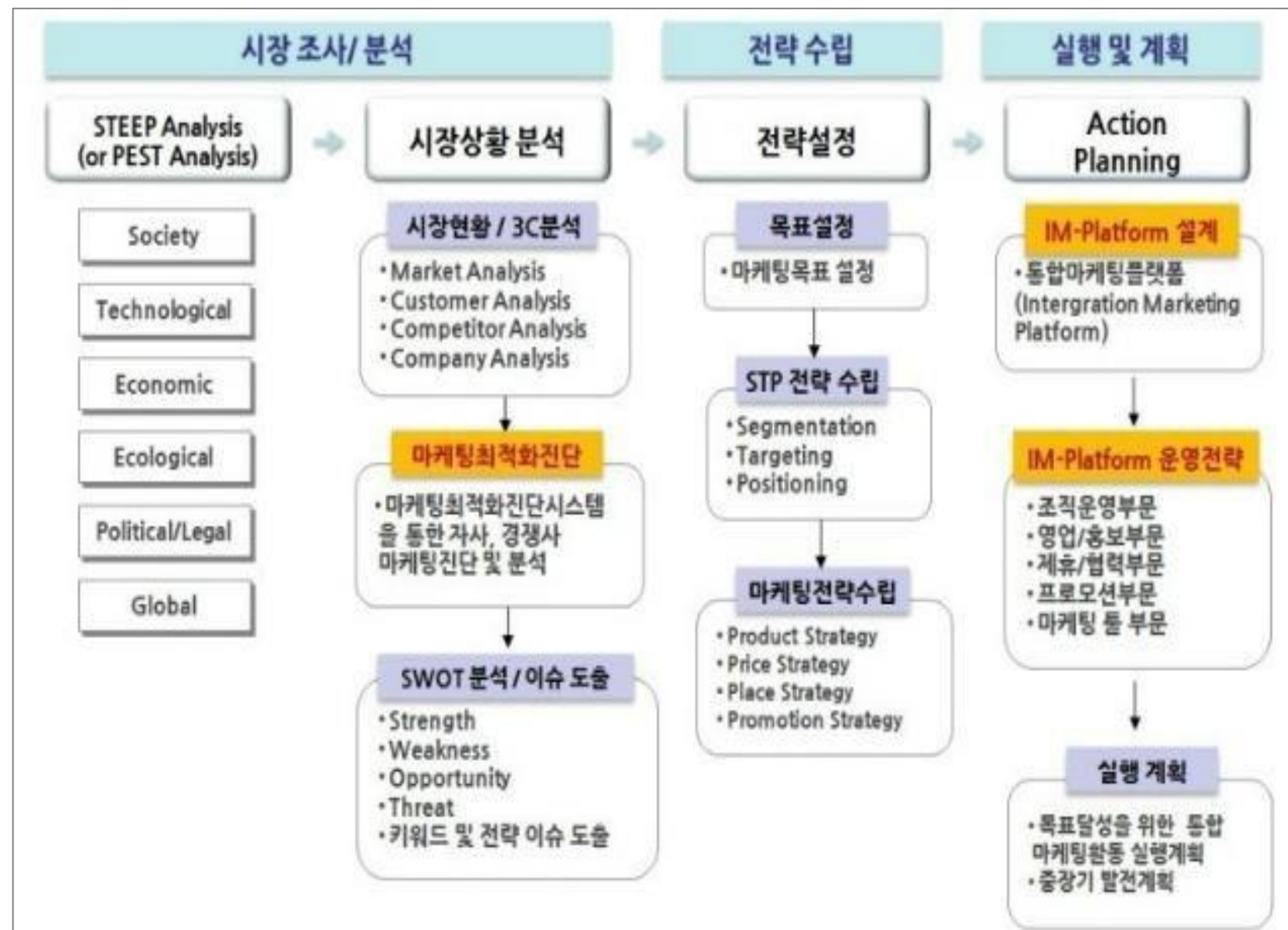
기업 자체기준이나 경쟁자보다 미흡한 부분, 고객 요구사항 미충족 요소를 찾아내고 이를 개선하는데 데이터분석을 접목해야 함

세상의 다양한 분석 자원들

05

[C] 가치체인 업무활동 이슈(2/2)

- 가치사슬은 단계(Phase)-활동(Activity)-과업(Task) 프로세스라는 계층구조 속에서 데이터분석을 활용할 수 있는 비즈니스 문제들이 존재함



- 마케팅 업무활동의 경우 시장조사분석전략수립 실행 및 계획이라는 단계에 따라 활동 및 세부과업을 진행하게 되며, 이에 필요한 근거자료의 수집·가공, 분석·해석, 업무활용이라는 데이터분석활동이 수반됨
- 마케팅 4P인 신제품기획(Product), 가격책정(Price), 판매채널(Place), 촉진(Promotion)과 고객관계관리(CRM: Customer Relationship Management)를 위한 잠재고객발굴, 가망고객전환, 업셀링, 크로스셀링, 휴면고객활성화, 고객이탈방지 등을 위한 업무계획수립 및 실행결과 분석이 데이터분석의 주요 주제가 될 수 있음.

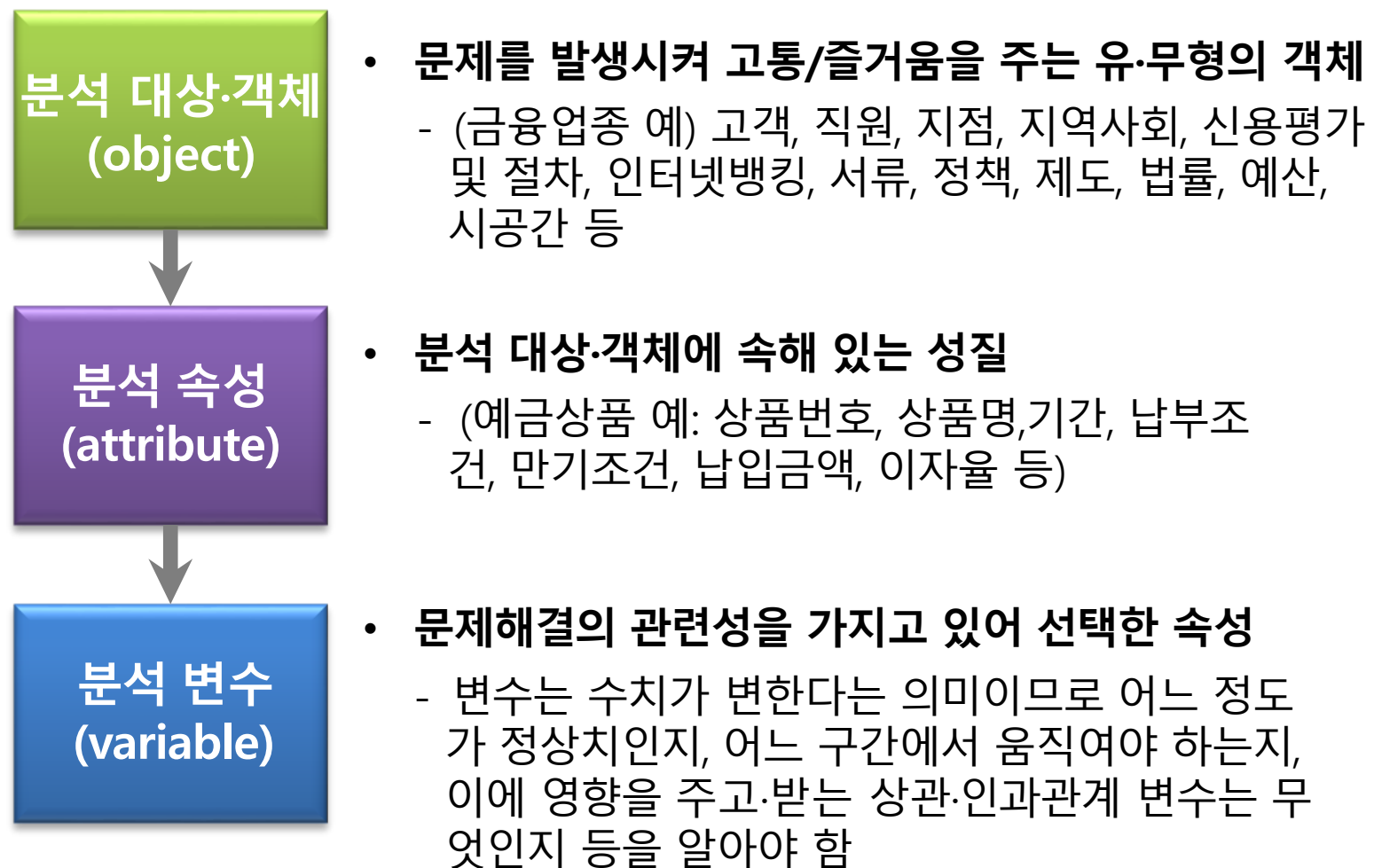
세상의 다양한 분석 자원들

06

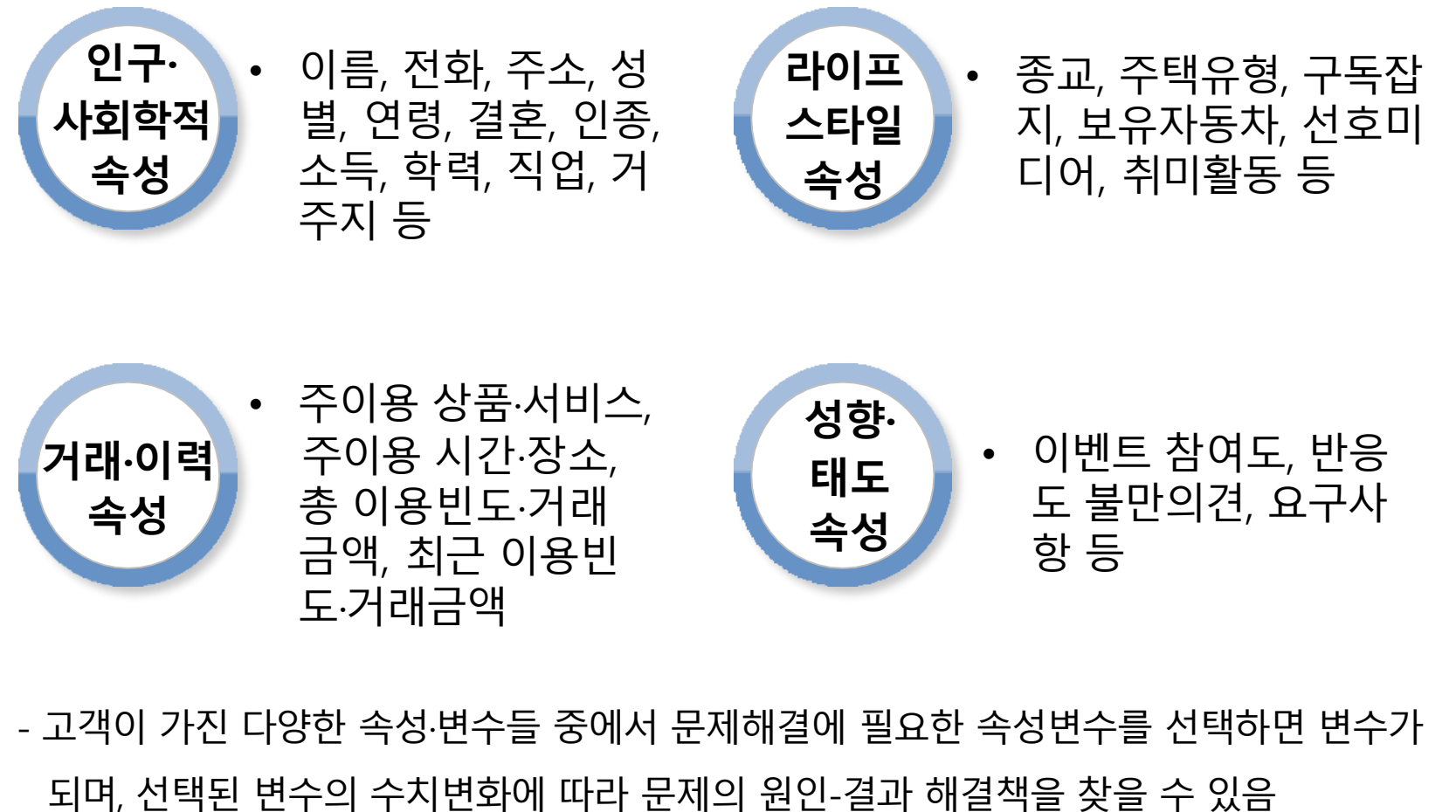
분석대상·객체에서 다양한 분석자원의 추출

- 분석 대상·객체, 속성, 변수로 구성된 데이터(분석문제)자원 Pool, 변수간 관계 및 판단근거 학습이 분석방법론의 핵심이 되어야 함

분석대상·객체로부터 분석에 필요한 속성·변수의 추출



속성·변수 리스트 예 : 고객



세상의 다양한 분석 자원들

07

온도속성을 이용한 분석사례

- 온도라는 속성이 문제 상황별로 어떠한 영향을 미치는 자를 통해서 주어진 문제해결에 필요한 실마리를 얻을 수 있음

체온과 면역력



- 모든 질병과 노화는 체온저하와 밀접한 관련성이 있음
- 체온 1도를 올리면 면역력이 5배 정도 높아짐
- 36도 이상이면 효소 활동이 활발해 면역력이 강화
- 낮으면 자율신경에 이상이 있고, 암세포 활동이 나타남

데이터센터 온도와 서버안정성



- 서버실은 22도라는 게 알려진 관행이었음
- 구글에서 온도와 서버의 성능, 수명과의 관계분석결과 27도에서 서버가 안정적임
- 온도 5도 중대로 냉각에너지를 반 이상 절감
- 페이스북은 연평균기온이 낮은 스웨덴 북부에 IDC 예정

시설원에 온도와 작물생육



- 작물생육에 영향을 주는 요소 중의 하나는 온도임
- 작물의 종류에 따라, 계절에 따라, 성장단계에 따라 적정유지 온도범위가 존재함
- 저온장해시 탄수화물이나 원형질 합성에 어려우며, 고온장해시 기공폐쇄로 광합성 작용이 감소함

07.

데이터 분석과제 기획

데이터 분석과제 기획

01

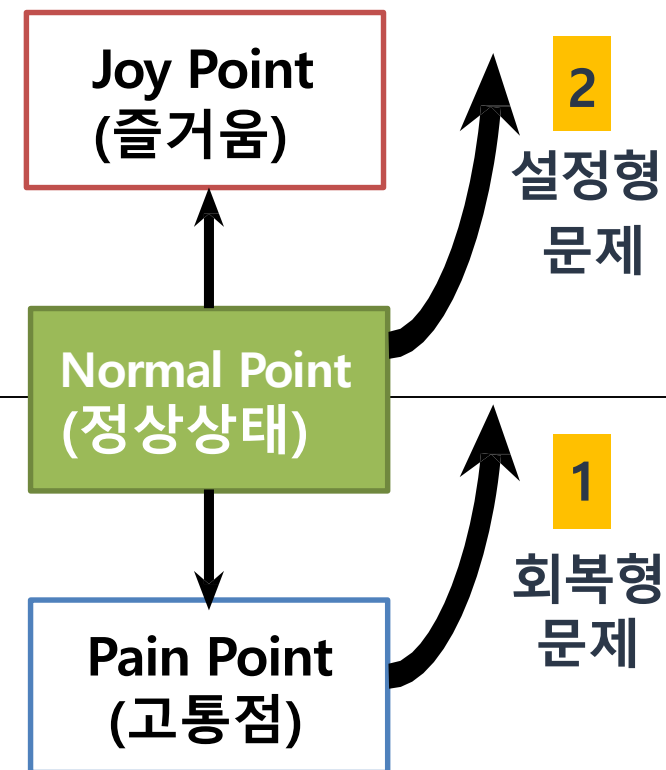
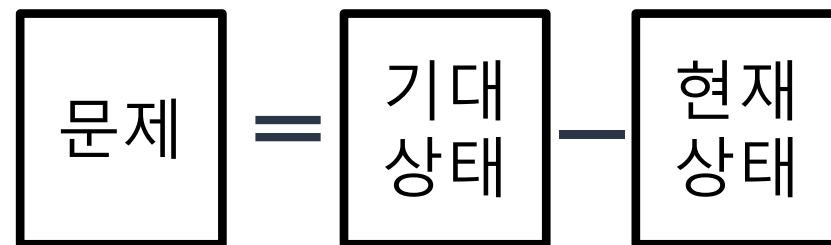
문제와 데이터분석의 관계(1/2)

- 데이터분석은 기대상태와 현재상태의 차이에서 발생하는 고통과 즐거움에 대한 현상분석을 통해 원인과 결과를 규명하고 이를 해결하기 위한 액션시나리오를 만드는 문제해결기법임

분석과제란 무엇인가

"분석과제"란 처리하거나
해결해야 할 이슈를 의미

기대상태와 현재상태의
차이에서 발생하는
"고통과 즐거움"



- 데이터분석의 주제가 되는 분석과제는 문제, 이슈, 프로젝트 등 다양한 용어로 부름
- 문제는 어떠한 기대하는 상태가 있는데, 현재상태가 그 기대수준에 미치지 못하거나 충족하지 못 하는데서 발생하게 됨
- 보통 평상시 관리하는 어떠한 기준(변수)가 정상상태에 있어야 하는데, 이 수준에서 벗어나게 되면서 고통점과 즐거움이라는 차원의 문제가 발생하게 됨
- 보통 기대상태에 비해 현재상태가 부정적인 측면에서 나쁜 쪽으로 나타났을 때 이를 원래대로 만회하기 위한 요인을 찾고 개선방안을 찾는 것이 바로 회복형 문제임
- 회복형 문제는 문제를 일으킨 원인인 공통점을 찾고 이를 약화 또는 제거함으로써 문제를 해결함
- 반면에 기대상태에 비해 현재상태를 긍정적인 측면에서 공격적으로 상향조정했을 때 이를 달성하기 위한 요인을 찾고 실현방안을 찾는 것이 바로 설정형 문제임
- 설정형 문제는 긍정적인 상황이 일회성이 아닌 지속적으로 나타날 수 있도록 촉진요인(즐거움)을 찾고 이를 강화함으로써 문제를 해결함

데이터 분석과제 기획

02

문제와 데이터분석의 관계(2/2)

- 데이터분석을 통해 주어진 문제해결을 위해 탐색·기술적 분석, 설명·인과적 분석, 최적·처방적 분석을 수행함

데이터분석이란 무엇인가

현상의 이해를 통해 미래를 예측하는 "문제해결기법"

기대상태와 현재상태의 수준을 변화시켜 문제의 크기를 '0'으로 만드는 과정

Descriptive Analysis (탐색·기술분석)

- 어떤 문제가 일어났는지
/일어나고 있는지
/일어날 것인지?
- 그 문제가 정말로 살펴
보아야할 문제인지?

문제의 중요도 인식
(Big Problem)

Predictive Analysis (설명·인과분석)

- 왜, 어떻게 문제가 일
어났는지
/일어나고 있는지
/일어날 것인지?
- 그 문제의 원인-결과?

문제 전개과정의 규명
(Mechanism)

Prescriptive Analysis (최적·처방분석)

- 어떻게 문제를 변화시
킬 것인지?
- 문제를 원하는 결과로
변화시키는 조건·조치
는 무엇인지?

문제 처리방안의 도출
(Action Scenario)

- 데이터분석이란 당면한 문제가 정말 중요한 문제인지(Big Problem)를 파악하고, 이에 대한 원인-결과 전개과정(Mechanism)에 대한 이해를 통해서 적절한 조치방안(Action Scenario)을 마련하는 하나의 문제해결 기법(Problem Solution)으로 보는 것이 본질임
- 데이터분석을 통해 현재 주어진 문제를 이해하고, 이에 대한 조치방안이 적절하게 검증되었을 때, 앞으로도 유사한 문제에 대해 이러한 조치방안이 효과적으로 사용이 될 수 있을 것이라는 믿음을 통해 미래를 예측함
- 문제의 해결을 위해서는 현재상태를 기대수준에 충족할 수 있도록 변화시켜야 하며, 또한 기대수준도 달성 가능한 수준으로 적절하게 설정하는 것이 필요함
- 데이터분석은 데이터, 통계공식, 머신러닝 알고리즘이 초점이 아니며, 또한 탐색·기술적 분석이나 설명·인과적 분석, 최적·처방적 분석들이 별도로 존재하는 것이 아닌 어떠한 분석을 하더라도 정도의 차이는 있으나 이 3가지 관점에서 분석을 수행하는 것임

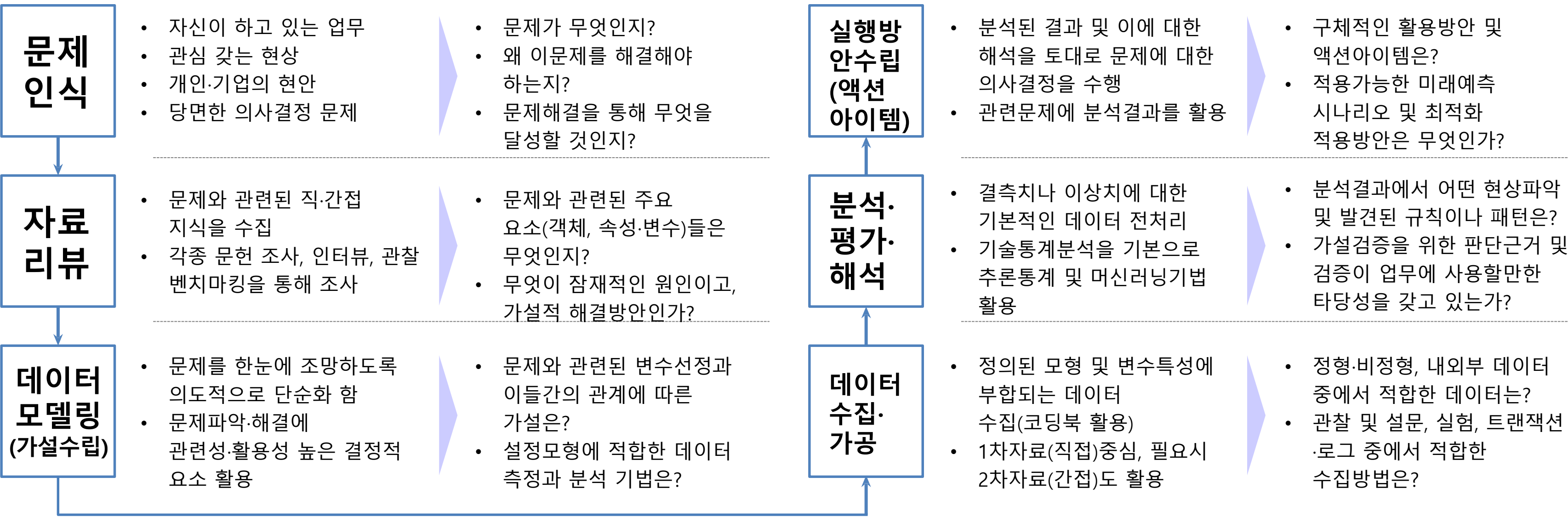
데이터 분석과제 기획

03

결과보다는 과정에 집중하는 데이터분석

- 데이터분석은 분석활동의 대상으로서 해결해야 할 문제(이슈)에 대한 현상과 원인, 해결방안에 대한 자료를 수집·가공하고 분석·평가·해석을 통해 의사결정에 활용하는 실행방안(액션시나리오)을 수립하는 "과정"임

데이터분석 수행절차



데이터 분석과제 기획

04

분석과제 기획방향(1/2)

- 수요기반 분석과제는 비즈니스 이슈 및 문제로부터 출발을 하는 과제기획방법론임

분석대상·객체로부터 분석에 필요한 속성·변수의 추출



개 념

- 주어진 문제/이슈에 대한 잠재적 원인 및 가설적 해결방안을 선행적으로 도출
- 이슈해결에 밀접한 근본적 원인에 대한 근거를 제시하고, 잠재적으로 도출된 해결방안이 실제로 실현될 수 있는지에 대한 가능성을 관련된 내/외부 데이터의 수집/가공/분석을 통하여 수행하는 접근방법

특 징

- 비즈니스 상의 이슈나 문제들은 조직의 핵심성과지표(KPI)의 수준을 바람직한 방향으로 강화시키거나 또는 부정적으로 악화시키는 요인임
- 데이터 종류나 분석기법에 문제상황을 억지로 끼워 맞추는 것이 아니라 선행적으로 문제해결 시나리오를 먼저 정의하고, 이에 적합한 데이터 및 분석기법을 찾아서 활용
- 활용되는 데이터 및 분석기법은 비즈니스 상의 이슈나 문제에 대한 해결이 가능하도록 해주는 하나의 "실행동인 (enabler)"으로서의 역할을 담당

데이터 분석과제 기획

05

분석과제 기획방향(2/2)

- 데이터주도 분석과제는 다양한 유형 및 속성을 가진 데이터로부터 출발을 하는 과제기획방법론임

데이터주도 분석과제 도출



개 념

- 다양한 데이터 원천의 조합 및 통합적/시각화 분석으로 의미 있는 패턴을 파악
- 패턴을 비즈니스 상에 활용했을 때 기존의 업무수행방식에 대한 이해를 돕고, 새로운 시각에서 이슈/문제에 대한 해결이 가능하도록 해주는 접근방법

특 징

- 전통적으로 비즈니스 운영상에서 발생하는 데이터들은 거래처리 데이터 및 각종 실적 데이터 등과 같은 정형데이터임
- 최근에는 다양한 멀티미디어 기반의 데이터 및 소셜기반의 소셜데이터, 센서와 위치기반의 사물인터넷 등의 정형/비정형 데이터가 폭넓게 생성됨
- 숨어 있는 패턴/정보/통찰을 추출함으로써 데이터가 비즈니스 변화를 리드한다는 점에서 "추진동인(driver)"의 역할을 담당

THANK YOU.