01. 4차 산업혁명의 이해

- 18세기: 증기기관을 기반으로 한 기계화 혁명의 1차 산업혁명 시대
- 19~20세기 초반: 전기를 사용한 대량 생산 혁명의 2차 산업혁명 시대
- 20세기 후반: 컴퓨터와 인터넷이 보급, 지식 정보 혁명의 3차 산업혁명 시대
- 21세기 현재: 4차 산업혁명 시대

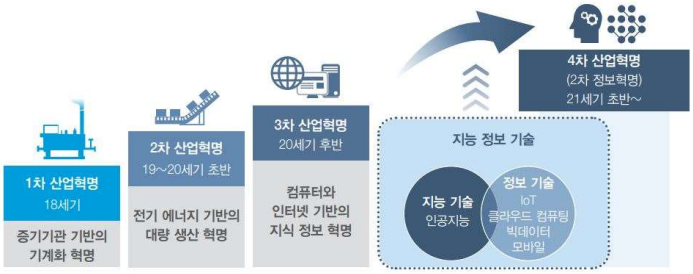


그림 1-1 **산업혁명의 흐름**(출처: 미래창조과학부 블로그)

01. 4차 산업혁명의 이해

■ 초연결

- 초연결
  - 사물인터넷의 진화와 디지털화로 인해 사물과 공간, 인터넷의 상호의존성이 증폭, 제품과 서비스의 연결성이 무한 확장
  - 사물인터넷과 5세대 통신(5G)은 초연결 사회를 실현시킨 대표적 기술
- 사물인터넷
  - 2005년 국제전기통신연합에서 처음으로 공식 정립
  - ITU에 따르면 IoT는 '언제나, 어디서나, 어느 것과도 연결될 수 있는 새로운 통신 환경
  - RFID태그를 부착하여 정보를 수집하는, 단순한 센서 네트워크(USN)에서 시작
  - 센서 네트워크는 사물과 사물간 통신을 의미하는 M2M으로 발전
  - 사람, 업무, 데이터까지 모든 것이 연결되어 상호 통신하는 만물인터넷(IoE)으로 발전할 전망

그림 1-2 **네트워크에 연결된 사물과 세계 인구수 비교**(출처: CISCO, IBSG, HP)

01. 4차 산업혁명의 이해

■ 초연결

- 5세대 통신 5G
  - 최대 20Gpbs 속도가 가능하고 일상적으로는 100Mbps 속도가 가능한 고속
  - 기존보다 1만배 이상 더 많은 트래픽을 수용하는 대용량
  - 평방 킬로 미터당 1백만 개의 기기 사용이 가능한 고밀집
  - 배터리 하나로 10년 간 구동 가능한 고에너지 효율
  - 1ms 이하의 낮은 지연시간
  - 이동 간 제로 중단을 실현하는 고안정성

  - 5G의 특징은 초고속, 초연결, 초저지연으로 요약 가능

표 1-1 5G의 특징

특징	설명
초고속	초광대역 무선통신(eMBB)enhanced Mobile BroadBand • 유선과 무선의 차이가 없는 **대용량** 및 **고속**의 데이터 이용 환경 제공 • 4G에 비해 최대 20배 더 빠른 20Gbps까지 구현 가능(일상적으로는 100Mbps 보장 목표) • 모바일로 8K 콘텐츠 송수신 가능
초연결	대규모 사물통신(mMTC)massive Machine Type Communication • 산업 또는 일반인에게 IoT 사용 환경 제공 • 현재보다 최대 500배 더 많은 기기와 **고밀집** 연결 가능 • **고에너지 효율** • 스마트폰의 인터넷, PC의 인터넷을 넘어 진정한 IoT 가능
초저지연	고신뢰/초저지연 통신(uRLLC)ultra-Reliable Low-Latency Communication • 최대 1ms까지의 **낮은 지연성**과 **고안정성**을 목표로 데이터 통신 서비스의 품질(QoS)을 제공

01. 4차 산업혁명의 이해

■ 초지능

- 초지능 - 슈퍼 인텔리전스
  - 인공지능의 지능이 인간 을 넘어서는 특이점을 통과하여 거의 모든 영역에서 인간의 인지 능력을 크게 능가 하는 경우
- 초지능 - 하이퍼 인텔리전스
  - 인간의 지능과 인공지능이 협력하여 더 스마트한 서비스를 제공하는 것
  - 인공지능 기능을 추가하여 사물을 더 스마트하게 만드는 사물의 지능화를 의미
- 초지능: 하이퍼 인텔리전스에서 슈퍼 인텔리전스로 진화하는 인공지능
- 빅데이터: 초지능의 원료 / 인공지능: 초지능을 제작

01. 4차 산업혁명의 이해

■ 초지능

- 빅데이터
  - 디지털 환경에서 발생하는 모든 데이터를 의미
  - 4차 산업혁명의 서비스는 이전에는 사용하지 못했던 빅데이터를 사용하게 되면서 가능

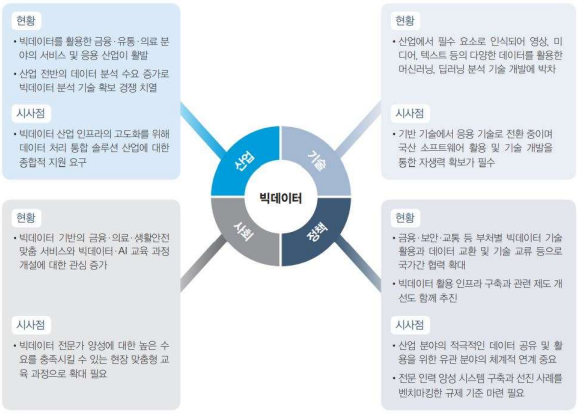


그림 1-3 온라인 뉴스의 빅데이터 키워드 분석 – 분야별 현황과 시사점

01. 4차 산업혁명의 이해

■ 초지능

- 인공지능(AI)
  - 1950년 앨런 튜링의 튜링 기계와 이미테이션 게임에서 컴퓨터 기계와 지능에 대한 논의가 시작
  - 1956년에는 다트머스대학교의 하계 컨퍼런스에서 Artificial Intelligence라는 용어가 처음 사용(1차 전성기)
  - 1970년에 들어서자 컴퓨터의 계산 기능과 논리 체계의 한계로 인공지능 이론 구현에 실패(1차 인공지능 겨울)
  - 1980년대에는 신경망 다층 퍼셉트론이 개발(2차 전성기)
  - 신경망의 성능을 높이기 위해 필요한 데이터가 부족하고 프로세서의 계산 능력이 한계에 도달(2차 인공지능 겨울)
  - 2000년대에 들어서면서 메모리, CPU, GPU 등의 하드웨어와 네트워크의 성능 향상으로 신경망 연구가 다시 활발해짐
  - 빅데이터가 출현하면서 딥러닝의 성능 향상이 가속, 구글 딥마인드의 알파고가 바둑대회에서 우승(3차 전성기)

표 1-2 AI의 주요 기술 분야

분야	설명
추론	인간의 사고 능력을 모방하는 기술
지식 표현 및 언어 지능	사람이 사용하는 자연어 이해를 기반으로 사람과 상호작용하는 기술
청각 지능	음성, 음향, 음악을 분석, 인식, 합성, 검색하는 기술
시각 지능	사물의 위치, 종류, 움직임, 주변과의 관계 등 시각 이해를 기반으로 지능화된 기능을 제공하는 기술
복합 지능	시공간, 촉각, 후각 등 주변의 상황을 인지 및 예측하고 상황에 적합한 대응을 제공하는 기술
지능형 에이전트	개인 비서, 챗봇 등 가상 환경에 위치하여 특별한 응용 프로그램을 다루는 사용자를 도울 목적으로 반복적인 작업을 자동화시켜 주는 기술
인간과 기계의 협업	인간의 감성이나 의도를 이해하기 위해 인간의 뇌 활동에 기계가 연동되어 작동하는 기술
AI 기반 하드웨어	초고속 지능 정보 처리를 구현하게 지원하는 하드웨어 기술

01. 4차 산업혁명의 이해

■ 초융합

- 디지털 트랜스포메이션
  - 디지털 기술을 활용하여 기존 산업의 운영 및 생산의 효율성과 경쟁력을 높이는 프로세스의 변화를 의미
  - 기업은 디지털 기술을 활용하여 다양한 산업 분야에서 지속적인 혁신을 추진, 특히 제조업에 주목
  - 기존 비즈니스 모델뿐만 아니라 고객의 경험을 변화시키고 추가 수익 흐름을 창출

- 로봇 프로세스 자동화(RPA)
  - 업무 과정에서 발생하는 데이터를 정형화하고 논리적으로 자동 수행하게 하는 기술
  - RPA를 통해 단순하고 반복적인 업무를 감소시키고 오류를 줄여 업무 생산성 향상을 기대할 수 있음
  - 금융 분야에서 비중이 가장 크지만 제조, 서비스 등의 다양한 분야로 확산 중

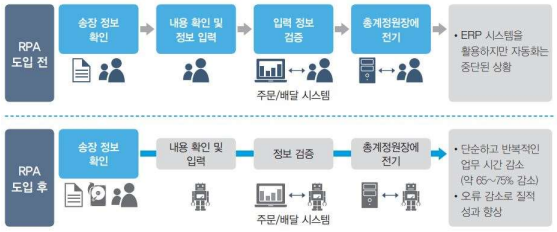


그림 1-4 임금 지급 업무에 RPA를 도입한 예

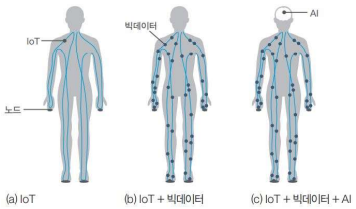
02. 4차 산업혁명을 실현하는 데이터 과학

■ 데이터 과학과 {IoT + 빅데이터 + AI}

- IoT를 구성하는 센서와 기기의 노드: 감각 및 행동 기관
- 인터넷/4G/5G: 인지된 자극과 명령을 전달하는 신경계
- AI: 인지된 자극을 처리하고 분석하여 명령을 내리는 두뇌
- IoT와 빅데이터, AI가 함께 선순환하며 발전하고 진화해야 함

■ 데이터 과학과 21세기 핵심 역량

- 세계 경제 포럼
  - 2020년에는 모든 산업 분야 가운데 3분의 1 이상이 복합 문제 해결 능력을 필요로 할 것
  - 기초 문해 6가지 (문해, 수해, 과학 문해, ICT 문해, 재정 문해, 문화 및 시민 문해)
  - 복잡한 문제를 대하는 방법에 관한 역량 4가지(비판적 사고/문제 해결, 창의성, 의사소통, 협력)
  - 변화하는 환경에 대한 대처 능력으로 인성 자질 6가지(호기심, 주도성, 일관성/끈기, 적응력, 리더십, 사회 및 문화 의식)



(a) IoT (b) IoT + 빅데이터 (c) IoT + 빅데이터 + AI

그림 1-5 IoT와 빅데이터, AI 결합의 의미

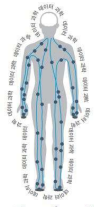


그림 1-6 {IoT + 빅데이터 + AI}와 데이터 과학의 완전체 개념도

03. 4차 산업혁명 서비스 사례

■ 자율주행차와 커넥티드 카

- 자율주행차
  - 자율주행차는 인지-판단-제어라는 3가지 단계로 동작
  - 도로 환경에서 빅데이터를 수집하여 상황을 인지하고 판단한 뒤 신속하게 제어
  - 자율주행차는 운전자와 차량의 역할 비중에 따라 5가지 레벨로 구분

- 자율주행차와 커넥티드 카에 필요한 핵심 기반 기술
  - 주변 환경의 빅데이터 수집 및 분석(AI), 차량 부품의 빅데이터 수집 및 진단(AI), 차량 제어, 고속 무선통신

표 1-3 자율주행자의 5가지 레벨

레벨	설명
레벨 0	자동화 기능이 미적용된 상태
레벨 1	운전자 보조주행: 운전자가 속도 또는 방향을 통제
레벨 2(현재)	부분적 자율주행: 차간 거리 및 속도 유지 등이 가능하지만 운전자가 주행에 적극 개입해야 하는 상태
레벨 3	조건부 자율주행: 자율주행 시스템을 운행하지만 비상시 몇 초 안에 운전자가 개입해야 하는 상태
레벨 4	고수준 자율주행: 비상시 차량이 일정 시간은 자체 대응하는 상태로 운전자가 차량 내에서 책을 읽어도 되는 수준
레벨 5	완전 자율주행: 어떠한 도로 환경에서도 무인 자율주행이 가능한 상태

03. 4차 산업혁명 서비스 사례

■ 자율주행차와 커넥티드 카

- 커넥티드 카
  - 정보통신기술과 자동차를 연결시킨 것으로 양방향 인터넷 및 모바일 서비스가 가능한 차량
  - 차량과 도시의 모든 곳이 연결되어 스스로 위험을 감지하고 다른 자동차와의 거리나 속도를 제어하며 운전할 수 있음
  - 스스로 고장을 진단하여 필요한 조치를 취함
  - 영화 스트리밍 서비스나 실시간 날씨 및 뉴스 검색, 소셜 네트워크 서비스 등 다양한 운전자 맞춤형 서비스를 제공

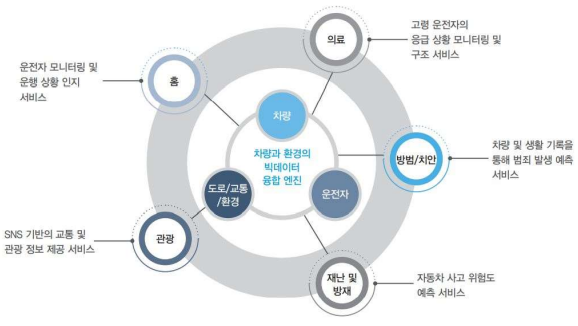


그림 1-7 커넥티드 카의 운전자 맞춤 서비스 개념도(출처: ETRI IoT 추진 계획)

03. 4차 산업혁명 서비스 사례

■ 스마트 시티

- 스마트 시티
  - 도시 구성원과 시설 기관들이 네트워킹이 가능하도록 인터넷과 IoT 등의 통신 인프라가 갖춰진 것
  - 빅데이터 및 AI와 융합하여 보다 편리하고 안전한 생활 및 업무 환경을 구현하는 4차 산업혁명 시대의 진화된 도시

- 스마트 시티에 필요한 핵심 기반 기술
  - IoT와 AI, 빅데이터 분석, AR/VR/MR, 건강/교통/교육/기기제어 등의 요소 기술

표 1-4 스마트 시티의 구성 요소

구성 요소	기능
스마트 홈/사무실	• 주거, 사무실, 학교, 편의시설 등이 상호 유기적으로 연결되어 사용자 요구를 예측해서 해결 • 개인별 편의성 극대화 • 재택근무 등 업무 환경의 제한 완화
스마트 시설 관리	• 발전, 교량, 환경 등 사회 기간 시설의 실시간 관제를 통해 에너지 절약 및 운영 효율화
스마트 교통	• 교통 시설이나 도로 상황의 실시간 지능형 관제를 통해 시간 단축 및 운영 효율화 • 개인별 이동 상황에 따른 맞춤형 교통 편의 제공
스마트 교육	• 학생별 학습 수준에 따라 맞춤형 교육을 제공하는 AI 기반의 튜터링 시스템 보급
스마트 치안	• 빅데이터 분석을 기반으로 범죄, 테러, 사고 등의 징후 예측 및 예방 • 유사시 효과적인 구조 조치를 통해 안전한 생활 환경 구축
스마트 환경	• 신재생 및 청정 에너지 기술, 생활 환경의 위생 상태 측정 및 관리, 자원 재활용, 환경오염의 측정/예방/처리가 융합되어 쾌적하고 청결하며 안전한 생활 환경 구축
스마트 문화/여가	• 문화, 콘텐츠, 스포츠와 VR, AR, MR 기술이 융합하여 개인 맞춤형의 건강, 재미, 지식을 제공하는 복합적 오락, 운동, 문화 체험 환경 제공

03. 4차 산업혁명 서비스 사례

■ 스마트 헬스 케어

- 스마트 헬스 케어
  - 개인의 건강에 대한 의료 정보, 기기, 시스템, 플랫폼을 다루는 산업 분야
  - 건강 관련 서비스와 의료 IT가 융합된 종합 의료 서비스
  - 고령화와 의료비 지출 증가라는 사회적 요 인과 AI, 빅데이터, IoT, 5G 등의 기술 발전에 따라 지속적으로 성장하는 추세

- 스마트 헬스 케어에 필요한 핵심 기반 기술
  - 종합 건강 정보 빅데이터 구축, 분야별 지식베이스 구축
  - 건강 빅데이터에 대한 실시간 수집 및 분석, 진단 및 처방용 AI스마트시티의 개념

표 1-5 헬스 케어 서비스와 ICT 융합의 발전 과정

구분	Tele-헬스 케어	e-헬스 케어	u-헬스 케어	스마트 헬스 케어
시기	1999년대 중반	2000년대 초반	2000년대 후반	2010년 이후
핵심 서비스	병원 내 치료	치료, 의료 정보 제공	e-헬스 케어 + 원격 의료, 만성 질환자 관리로 질병 예방	u-헬스 케어 + 운동 및 식사량 등의 건강 생활 관리, 복지, 안전
공급자	병원	병원	병원, ICT 기업	병원, ICT 기업, 보험사, 헬스 케어 서비스 기업
주요 이용자	의료인	의료인, 환자	의료인, 환자, 일반인	의료인, 환자, 일반인
핵심 ICT 기술		초고속 인터넷	무선 인터넷	스마트 기기, 앱, AI, 빅데이터

01. 빅데이터의 이해

■ 빅데이터의 개념

- 빅데이터의 정의
  - 디지털 환경에서 발생하는 대량의 모든 데이터
  - 대규모의 데이터를 저장·관리·분석할 수 있는 하드웨어 및 소프트웨어 기술, 데이터를 유통·활용하는 모든 프로세스를 포함
  - 빅데이터 플랫폼을 구성하는 하드웨어, 소프트웨어, 애플리케이션 간의 유기적 순환에 의해 가치를 창출

표 2-1 빅데이터의 정의

기관	정의
맥킨지	일반적인 데이터베이스 소프트웨어가 수집, 저장, 관리, 분석할 수 있는 범위를 초과하는 대규모의 데이터다.
가트너	향상된 시사점과 더 나은 의사결정을 위해 사용되는 것으로 비용 효율이 높고 혁신적이며 대용량 고속 및 다양성을 가지는 정보 자산이다.
위키피디아	기존 데이터베이스 관리 도구의 수집, 저장, 관리, 분석 역량을 넘어서는 대량의 정형 또는 비정형 데이터셋 및 이러한 데이터로부터 가치를 추출하고 결과를 분석하는 기술이다.
국가전략위원회	대용량 데이터를 활용 및 분석하여 가치 있는 정보를 추출하고 생성된 지식을 바탕으로 능동적으로 대응하거나 변화를 예측하기 위한 정보화 기술이다.
삼성경제연구소	기존의 관리 및 분석 체계로는 감당할 수 없을 정도의 기대한 데이터 집합으로 대규모 데이터와 관계된 기술 및 도구(수집, 저장, 검색, 공유, 분석, 시각화 등)를 모두 포함한다.
한국정보화진흥원	저장, 관리, 분석할 수 있는 범위를 초과하는 규모의 데이터와 이것을 저장, 관리, 분석할 수 있는 하드웨어 및 소프트웨어 기술, 데이터를 유통 및 활용하는 과정을 통틀어 나타낸다. 즉, 빅데이터를 구성하는 하드웨어, 소프트웨어 그리고 이를 포괄하는 모든 프로세스를 의미하는 거대 플랫폼이다.

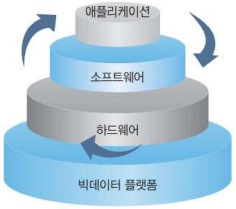


그림 2-1 빅데이터 플랫폼

01. 빅데이터의 이해

■ 빅데이터의 개념

- 빅데이터의 출현
  - 기술의 발달과 비용 저하, 소셜 네트워크 서비스 발달, 그림자 정보와 사물 정보 증가 등의 ICT 패러다임의 변화
  - 빅데이터에 전문 역량과 기술을 더하여 전략적으로 활용할 방법이 주목됨
  - 경제적 가치 창출, 사회 문제 해결, 새로운 ICT 패러다임 견인이라는 신가치 창출

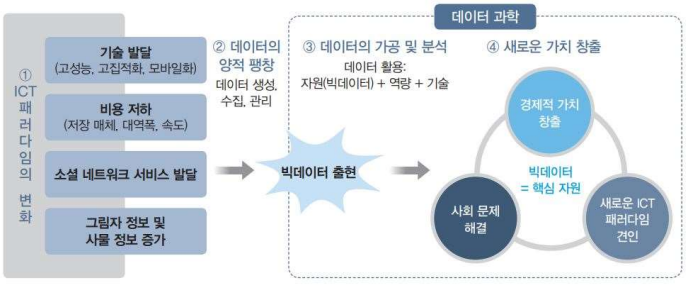


그림 2-2 빅데이터의 출현과 신가치 창출의 흐름

01. 빅데이터의 이해

■ 빅데이터의 분류

- 정형 데이터
  - 일정한 규칙으로 체계적으로 정리된 것으로 그 자체로 해석이 가능하여 바로 활용할 수 있음
- 반정형 데이터
  - 고정된 필드에 저장되어 있지는 않지만 XML, HTML 등의 메타데이터와 스키마를 포함하는 것으로 파일 형태로 저장
- 비정형 데이터
  - 고정된 필드나 스키마가 없는 것
  - 스마트 기기에서 페이스북, 트위터, 유튜브 등으로 생성되는 소셜 데이터
  - IoT 환경에서 생성되는 위치 정보나 센서 데이터와 같은 사물 데이터 등

표 2-2 정형화 정도에 따른 빅데이터의 분류

구분	설명	수집 및 처리 난이도
정형 데이터	• 고정된 필드에 저장 • 관계형 데이터베이스처럼 스키마 형식에 맞게 저장 • 예: RDB, 스프레드시트	• 내부 시스템에 의한 데이터라 수집하기 쉬움 • 파일 형태의 스프레드시트는 형식을 가지고 있어 처리하기 쉬움 • 처리 난이도: 하
반정형 데이터	• 고정된 필드에 저장되어 있지는 않지만 메타데이터나 스키마 등을 포함 • 예: XML, HTML, JSON, 웹 문서, 웹 로그	• API 형태로 제공되므로 데이터 처리 기술이 필요함 • 처리 난이도: 중
비정형 데이터	• 데이터 구조가 일정하지 않음 • 규격화된 데이터 필드에 저장되지 않음 • 예: 소셜 데이터, 텍스트 문서, 이미지/동영상/음성 데이터, 문서 파일(PDF)	• 파일을 데이터 형태로 파싱해야 하므로 처리하기 어려움 • 처리 난이도: 상

01. 빅데이터의 이해

■ 빅데이터의 특징

- 데이터 측면
  - 초기에는 빅데이터의 특징을 3V로 일컬어지는 규모, 다양성, 속도로 나타냄
  - 빅데이터를 통한 가치 창출이 중요해지면서 정확성과 가치를 추가한 5V로 나타냄

표 2-3 빅데이터의 특징 – 데이터 측면

구분	특징	설명
1차 특징	규모	• ICT 기술의 발전으로 디지털 정보량이 기하급수적으로 폭증하여 제타바이트 시대로 진입
	다양성	• 데이터의 종류 증가: 로그 기록, 소셜/위치/소비/현실 데이터 등 • 데이터의 유형 다양화: 텍스트 외에 멀티미디어 등의 비정형 데이터 증가
	속도	• 센서, 모니터링 등의 사물 정보와 스트리밍 등의 실시간 정보가 증가하면서 데이터의 생성 및 이동(유통) 속도 증가 • 대규모 데이터를 처리하고 가치 있는 정보를 활용하기 위한 데이터 처리 및 분석 속도 증가
추가 특징	정확성	• 방대한 데이터를 기반으로 분석을 수행하므로 정확성 향상
	가치	• 빅데이터 분석으로 도출된 최종 결과물이 문제 해결을 위한 통찰력을 제공하므로 새로운 가치 창출 가능

01. 빅데이터의 이해

■ 빅데이터의 특징

- 분석 환경 측면
  - 데이터 분석 시스템의 구성 요소인 데이터, 하드웨어, 소프트웨어 분석 방법은 분석 환경에 따라 다른 특징을 나타냄

표 2-4 빅데이터의 특징 - 분석 환경 측면

요소	과거의 데이터 분석 환경	현재의 빅데이터 분석 환경
데이터	• 정형화된 수치 중심의 자료	• 비정형의 다양한 데이터 • 예: 문자 데이터(SMS, 검색어), 영상 데이터(CCTV, 동영상), 위치 데이터 등
하드웨어	• 고가의 저장 장치 • 데이터베이스 • 대규모 데이터웨어하우스	• 클라우드 컴퓨팅: 비용 대비 효율성 증대
소프트웨어 분석 방법	• 관계형 데이터베이스: RDBMS • 통계 패키지: SAS, SPSS • 데이터 마이닝 • 머신러닝 • 지식 발견	• 오픈 소스 형태의 무료 소프트웨어 • 오픈 소스 통계 솔루션: R • 텍스트 마이닝 • 오피니언 마이닝 • 감성 분석

01. 빅데이터의 이해

■ 빅데이터의 특징

- 처리 방식 측면
  - 빅데이터는 기존 데이터베이스 관리 시스템(DBMS)으로 처리하던 것에 비해 100배 이상 많은 정형, 비정형 데이터를 처리

표 2-5 빅데이터의 특징 - 처리 방식 측면

구분	이전의 데이터 처리 방식	빅데이터 처리 방식
데이터 트래픽	• 테라바이트 수준	• 페타바이트 수준: 최소 100테라바이트 이상 • 정보의 장기간 수집 및 분석 • 방대한 처리량
데이터 유형	• 정형 데이터 중심	• 비정형 데이터 비중이 높음: SNS 데이터, 로그 파일, 클릭스트림 데이터, 콜센터 로그 통신, CDR 로그 등 • 처리 복잡성 증대
프로세스 및 기술	• 단순한 프로세스 및 기술 • 정형화된 처리 및 분석 결과 • 원인 및 결과 규명 중심	• 다양한 데이터 소스와 복잡한 로직 처리 • 처리 복잡도가 높아 분산 처리 기술 필요 • 새롭고 다양한 처리 방법 필요: 정의된 데이터 모델/상관관계/절차 등이 없음 • 상관관계 규명 중심 • 하둡, NoSQL 등 개방형 소프트웨어 사용

01. 빅데이터의 이해

■ 빅데이터의 가치

- 혁신과 창조의 도구
  - 빅데이터 분석이 제공하는 스마트 서비스는 기존 비즈니스에 효율화, 개인화, 그리고 미래 예측력을 통한 혁신을 제공
  - 단순히 새로운 기술이나 비즈니스 모델이 아니라 새로운 패러다임으로의 변화를 의미
  - 빅데이터 자체부터 이를 활용한 사용자 애플리케이션까지 광범위하여 빅데이터 플랫폼과 에코시스템으로 확장

표 2-6 빅데이터 분석을 통한 비즈니스 혁신 방향

방향	설명
효율화	• 빅데이터를 이용해 과거 및 현재의 현상을 파악할 수 있다. • 물류, 재무, 기획, 마케팅 등 경영 전반의 데이터를 실시간으로 분석한 후 최선의 의사결정을 할 수 있다.
개인화	• 온라인 이용자의 활동 정보와 SNS 등으로 축적된 개인 정보를 결합하여 사용자에게 특화된 서비스를 제공할 수 있다. • 현재 개인 정보는 광고 분야에 활용 중인데 이를 넘어 의료, 교육 등 모든 분야로 확대가 가능하다.
미래 예측력	• 과거 및 실시간 데이터를 분석하여 축적한 개인 정보로 개인 또는 조직 전체의 행동 및 의사결정 패턴을 도출할 수 있다. • 미래에 적용 가능한 시나리오를 제시하고 예측 가능한 행동 및 발생 가능한 문제점을 사전에 방지하는 서비스가 가능하다.

01. 빅데이터의 이해

■ 빅데이터의 가치

- 사회·경제적 가치
  - 빅데이터는 정치, 사회, 경제, 문화, 과학 기술 등 사회 전반에 걸쳐 가치 있는 정보를 제공
  - 데이터의 도입과 활용은 산업 경쟁력 제고, 생산성 향상, 혁신을 위한 새로운 가치 창출을 할 것으로 기대

표 2-7 맥킨지가 제시한 빅데이터를 이용한 사회 · 경제적 가치 창출 방법

방법	설명
정보의 투명성	• 이해 관계자가 적시에, 보다 쉽게 빅데이터에 접근할 수 있게 하는 것만으로도 가치 창출이 가능하다. • 예: 공공 부문에서 분리된 부서가 관련 데이터에 보다 쉽게 접근할 수 있으면 데이터 검색과 처리 시간이 절감된다.
실험을 통한 소비자의 요구 발견, 트랜드 예측, 성과 관리	• 더 많은 거래 데이터를 디지털 형태로 축적함에 따라 더욱 정확하고 상세하게 소비자 요구를 발견하거나 트랜드 예측을 할 수 있다. • 예: 관리자가 빅데이터를 사용하여 자연스럽게 발생하거나 통제된 실험으로 일어나는 성과의 변동성을 분석하고 나아가서 근본적인 원인과 결과를 분석하면 더 높은 수준으로 성과를 관리할 수 있다.
소비자 맞춤 비즈니스를 위한 고객 세분화	• 빅데이터를 통해 더 구체적으로 고객을 세분화하여 고객의 요구에 맞는 더 정확한 맞춤형 서비스를 제공할 수 있다. • 예: 공공 부문에서 시민을 세분화하여 필요한 서비스를 제공할 수 있다.
자동화된 알고리즘을 통한 의사결정	• 빅데이터 기술을 사용하여 전체 데이터셋을 정교하게 분석함으로써 의사결정을 개선하고 위험을 최소화할 수 있으며 가치 있는 인사이트를 발굴할 수 있다. • 예: 판매 정보에 실시간 대응하여 재고 및 가격을 자동으로 조정하는 자동화 알고리즘은 소매업체의 의사결정을 최적화할 수 있다.
새로운 비즈니스 모델, 상품, 서비스의 혁신	• 빅데이터를 통해 새로운 상품 및 서비스를 개발하거나 기존 상품 및 서비스를 강화하여 완전히 새로운 비즈니스 모델을 개발할 수 있다. • 예: 실시간 위치 데이터를 이용하여 자동차를 운전하는 장소와 방법에 따라 내비게이션을 제공하고 상해보험 가격도 책정하는 완전히 새로운 위치 기반 서비스가 가능하다.

02. 빅데이터의 활용

■ 빅데이터의 역할

- 미래 사회의 특성은 불확실성, 리스크, 스마트, 융합으로 대변됨
- 빅데이터를 활용해 여러 가지 가능성에 대한 시나리오 시뮬레이션을 하면 불확실한 상황 변화에 유연하게 대처 가능
- 빅데이터에 기반한 정보 패턴 분석으로 리스크에 대응할 수 있음
- 개인화 및 지능화된 서비스를 제공하여 삶의 질을 향상시킴

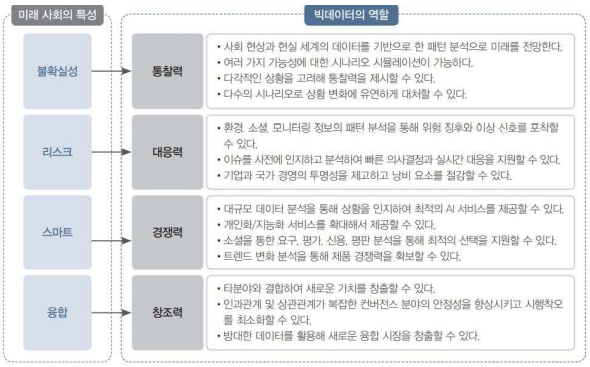


그림 2-3 미래 사회의 특성과 대응되는 빅데이터의 역할

02. 빅데이터의 활용

■ 빅데이터 활용 전략

- 기업의 성공적인 빅데이터 활용
  - 리더십, 역량 관리, 기술 도입, 의사결정, 기업 문화가 필요 (맥아이,브린욜프슨)

표 2-8 성공적인 빅데이터 활용 조건

조건	내용
리더십	목표 설정을 위해 빅데이터를 활용한 성공이 무엇인지를 명확히 정의하고 이를 강력하게 추진할 수 있는 리더십이 필요하다.
역량 관리	데이터 과학자, 시스템 개발자 등과 같은 전문 인력의 역량을 관리해야 한다.
기술 도입	빅데이터 관련 시스템에 최적화된 기술을 도입하고 조직 내·외부의 데이터를 통합 및 가시화하는 기술을 도입해야 한다.
의사결정	빅데이터 분석에 기반한 의사결정으로 조직의 유연성을 보장해야 한다.
기업 문화	빅데이터를 활용할 수 있는 조직 문화가 필요하다.

  - 자원, 기술, 인력의 3가지 요소에 대한 전략을 수립

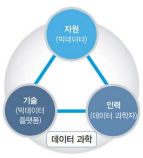


그림 2-4 성공적인 빅데이터 활용을 위한 3요소

02. 빅데이터의 활용

■ 빅데이터 활용 전략

- 활용 가능한 빅데이터 발견하기
  - 가트너: 미래 사회에는 '데이터 경제 시대'가 도래할 것으로 전망
  - 상호 연결과 협력으로 데이터 활용 영역이 확장되면 데이터 자원이 단계적으로 무한해질 것
  - 그에 따라 자원을 확보하는 방안도 단계적으로 확장

표 2-9 데이터 자원 확보를 위한 단계적 방법

단계	내용과 과제	방법
저장	• 조직의 독자적인 데이터를 생성 및 저장하는 단계 • 인터넷을 통해 외부 데이터 수집(검색) 가능 • 데이터의 신뢰성과 품질 제고를 위한 노력이 필요	생성, 저장, 수집(검색)
공유	• 기업 데이터를 외부 기관과 상호 교환하는 단계 • 1:1 또는 1:n의 공유 및 연계 가능	연계, 공유
통합	• 특정 활동이나 목적을 위하여 연합, 그룹, 클럽이 상호 협력하는 공동의 장(집단)을 형성하는 단계 • 표준 데이터 풀과 연계하여 국경을 초월한 정보 교환과 상호 이용이 가능	참여, 협력
공동 창출	• 오픈 플랫폼으로 데이터를 공유하는 단계 • 상호 협력과 참여로 공동의 자원을 창조	오픈, 창조

02. 빅데이터의 활용

■ 빅데이터 활용 전략

- 빅데이터 처리 단계와 신기술 이해하기
  - 빅데이터는 데이터의 생성 → 수집 → 저장 → 분석 → 표현의 단계를 거치며 세부 영역과 관련 기술이 개발
  - 조직과 기업의 혁신 전략으로 적용할 수 있게 빅데이터 플랫폼, 빅데이터 분석 기법 및 기술에 대한 이해가 필요
  - 분석 기술
    - » 통계, 데이터 마이닝, 머신러닝, 딥러닝, 자연어 처리, 패턴 인식, 소셜 네트워크 분석, 비디오·오디오·이미지 프로세싱 등
  - 빅데이터의 활용·분석·처리를 포함하는 인프라
    - » BI, DW, 클라우드 컴퓨팅, 분산 데이터베이스 (NoSQL), 분산 병렬 처리, 분산 파일 시스템 등
  - 빅데이터 관련 신기술
    - » 대용량 데이터 처리를 위한 분산 처리 기술인 하둡과 인메모리, 의미 분석 기술인 데이터 마이닝, 자연어 처리, 머신러닝, 딥 러닝, 그리고 비정형 데이터 처리를 위한 NoSQL 기술

02. 빅데이터의 활용

■ 빅데이터 활용 전략

- 빅데이터 처리 단계와 신기술 이해하기

표 2-10 빅데이터의 처리 단계별 기술 영역

단계	기술 영역	내용
데이터 소스	내부 데이터	데이터베이스, 파일 관리 시스템
	외부 데이터	파일, 멀티미디어, 스트리밍
수집	크롤링crawling	검색 엔진 로봇을 이용한 데이터 수집
	ETL: 추출Extraction, 변환Transformation, 적재Loading	소스 데이터의 추출, 전송, 변환, 적재
저장	데이터 관리: NoSQL	비정형 데이터 관리
	저장소	빅데이터 저장
	서버	초경량 서버
처리	맵리듀스mapReduce	데이터 추출
	작업 처리	다중 작업 처리
분석	신경 언어 프로그래밍NLP, Neuro Linguistic Programming	자연어 처리
	머신러닝	데이터 패턴 발견
	직렬화serialization	데이터 간 순서화
표현	시각화visualization	데이터를 도표나 그래픽으로 표현
	획득acquisition	데이터의 획득 및 재해석

02. 빅데이터의 활용

■ 빅데이터 활용 전략

- 데이터 과학자 역량 강화하기
  - 빅데이터 시대에는 데이터를 분석하고 관리할 수 있는 인력에 대한 중요성이 커짐
  - 대규모 데이터 속에서 숨겨진 정보를 찾아내는 데이터 과학자는 '빅데이터 시대의 연금술사'
  - 존 라우치: 데이터 과학자에게 필요한 6가지 기본 자질

① 수학 역량
② 공학 역량
③ 데이터를 분석할 때 필수적인 가설을 세우거나 검증할 때 필요한 비판적 시각
④ 이를 잘 작성할 수 있는 글쓰기 역량
⑤ 다른 사람에게 잘 전달할 수 있는 대화 능력
⑥ 호기심과 개인의 행복

  - 데이터 과학자는 외부보다는 내부 인력으로 내재화하여 활용

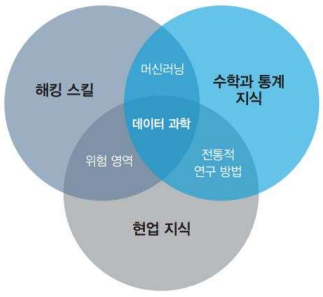


그림 2-5 데이터 과학자가 지녀야 할 역량 벤다이어그램

01. 빅데이터 산업의 이해

■ 빅데이터 산업을 설명하는 용어

- 빅데이터 산업은 관련된 여러 분야가 유기적으로 결합된 시스템
- 빅데이터 플랫폼
  - 데이터 관점에서 빅데이터를 수집·저장·분석하는 프로세스와 그에 필요한 자원의 유기적 결합을 나타냄
- 빅데이터 에코시스템
  - 빅데이터 플랫폼에 서비스 산업을 결합하여 고객에게 가치를 전달 하는 유기적 공동체를 나타냄
- 빅데이터 서비스 프레임워크
  - 빅데이터 에코시스템에서 서비스 공급자를 분류하고 서비스 유형과 수준을 파악한 것을 나타냄

01. 빅데이터 산업의 이해

■ 빅데이터 플랫폼

- 데이터 플랫폼의 발전
  - 데이터 플랫폼은 정형화된 형태로 데이터를 저장하는 파일 시스템으로 시작
  - 이후에 다수가 동시에 사용할 수 있는 데이터베이스와 데이터 웨어하우스(DW)로 발전
  - 폭발적으로 증가하는 데이터를 저장 및 유통하기 위한 빅데이터 플랫폼으로 진화
- 빅데이터 플랫폼의 개념
  - 빅데이터를 처리하는 것
  - 대량의 데이터를 저장 및 분석, 처리할 수 있는 대용량의 고속 저장 공간과 고성능 계산 능력의 컴퓨팅 인프라를 보유
  - 실시간으로 발생하는 빅데이터를 처리 및 분석하여 일관성을 유지하는 데이터 분석도 필요
  - 빅 데이터에서 발생하는 개인 정보를 위한 정보 보안 관리체계 지원도 필요
  - 빅데이터 플랫폼은 오픈 소스인 하둡을 근간으로 많이 사용

01. 빅데이터 산업의 이해

■ 빅데이터 플랫폼

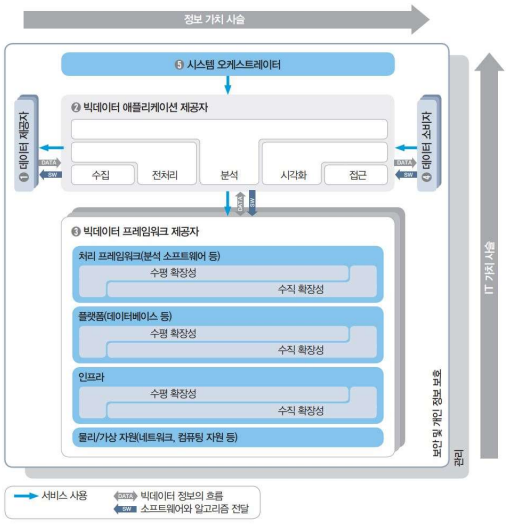


그림 3-1 대표적인 표준화 빅데이터 플랫폼인 NIST의 빅데이터 참조 아키텍처

01. 빅데이터 산업의 이해

■ 빅데이터 서비스 프레임워크

- 빅데이터 서비스 프레임워크는 빅데이터 시장을 효율적으로 이해하기 위한 것
- 에코시스템 안에서 서비스 공급자를 분류하고 서비스 유형과 수준을 파악하는 것이 필요
- 공급하는 서비스의 유형과 수준에 따라 빅데이터 서비스 공급자와 애플리케이션 공급자로 분류

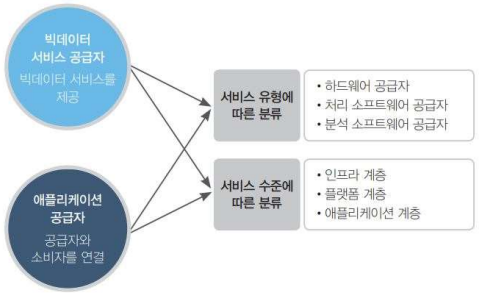


그림 3-3 빅데이터 시장의 공급자 분류

01. 빅데이터 산업의 이해

■ 빅데이터 서비스 프레임워크

- 공급 서비스 유형에 따른 분류
  - 하드웨어 공급자
    - 자체 데이터센터 및 클라우드 시스템을 통해 빅데이터 서비스를 위한 인프라를 공급
  - 처리 소프트웨어 공급자
    - 서비스 소비자가 저장한 빅데이터를 효과적으로 저장 및 처리할 수 있는 소프트웨어를 제공한다.
  - 분석 소프트웨어 공급자
    - 서비스 소비자의 빅데이터를 분석할 소프트웨어를 제공
- 공급 서비스 수준에 따른 분류
  - 인프라 계층
    - 빅데이터를 위한 기초 작업을 담당하는 하드웨어나 운영체제를 제공
    - 자체 인프라를 구축하거나 가상화를 위한 클라우드 컴퓨팅 서비스가 여기에 속함
  - 플랫폼 계층
    - 클라우드 컴퓨팅 서비스나 하드웨어에 종속되지 않는 처리 및 분석 소프트웨어 등을 제공
  - 애플리케이션 계층
    - 소비자가 빅데이터와 소통하는 매커니즘을 제공한다. 빅데이터 처리 결과를 바탕으로 소비자가 원하는 분석 결과를 제공하거나 시장에 유통

01. 빅데이터 산업의 이해

■ 빅데이터 서비스 프레임워크

- 빅데이터 서비스 공급자 분류를 위한 빅데이터 서비스 프레임워크

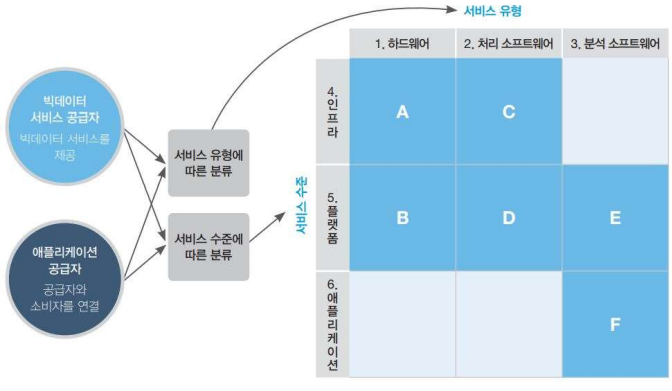


그림 3-4 빅데이터 서비스 공급자 분류를 기반으로 나타낸 빅데이터 서비스 프레임워크

01. 빅데이터 산업의 이해

■ 빅데이터 서비스 프레임워크

- A: 하드웨어-인프라 유형
  - 기업 등에서 자체 데이터센터를 구축할 수 있게 해주는 서비스 유형
  - 이 유형은 사적 데이터를 중심으로 하는 기업형 솔루션과 공적 데이터를 중심으로 하는 플랫폼 서비스로 구분할 수 있음
  - IBM, HP, 오라클 등의 기업용 하드웨어 솔루션 제품이 여기에 해당
- B: 하드웨어-플랫폼 유형
  - 클라우드를 기반으로 서비스를 제공하는 유형
  - 기존의 클라우드 컴퓨팅 시스템을 사용해 빅데이터 서비스를 제공
- C: 처리 소프트웨어-인프라 유형
  - 하드웨어와 소프트웨어를 함께 제공하는 서비스 유형
  - 대용량 데이터를 다루기 위해 필요한 분산 저장 및 병렬 처리 인프라에 처리 솔루션까지 제공
  - 기업용 솔루션 사업을 하는 오라클, IBM, HP, EMC 등의 기업에서 자사의 하드웨어와 특화된 소프트웨어를 통합해서 제공

01. 빅데이터 산업의 이해

■ 빅데이터 서비스 프레임워크

- D: 처리 소프트웨어-플랫폼 유형
  - 오픈 소스 기반의 소프트웨어 플랫폼을 제공하는 서비스 유형
  - 공급자는 오픈 소스를 기반으로 하는 빅데이터 처리 프로그램을 공급
  - 소비자는 공급자가 제공하는 클라우드 서비스를 통해 빅데이터 처리 서비스를 이용할 수 있음
- E: 분석 소프트웨어-플랫폼 유형
  - 일반 소비자를 위한 분석 소프트웨어를 제공하는 서비스 유형
  - 빅데이터를 솔루션으로 상품화하고 클라우드 컴퓨팅과 결합하여 제공
  - 소비자는 자체 서버와 솔루션을 구축하는 대신에 클라우드 컴퓨팅 인프라에서 데이터를 저장 및 분석하는 프로그램을 이용할 수 있음
- F: 분석 소프트웨어-애플리케이션 유형
  - 고객 맞춤형 솔루션 서비스 유형으로 데이터의 의미를 파악하고 이를 분석해서 활용하는 서비스를 제공
  - 축적된 데이터를 바탕으로 분석 후 결과의 의미를 파악하여 제공
  - 소비자의 검색 패턴을 이용해 독감 확산을 예측했던 구글 분석이 대표적 사례

02. 빅데이터 분석 방법과 접근법

■ 빅데이터 분석 방법

- 분석 목적에 따른 구분

① 통계 분석
- 통계 기법에 의한 분석 방법으로 가장 대표적인 유형

② 예측 분석
- 과거의 데이터와 변수 간의 관계를 이용하여 새로운 변수를 추정

③ 데이터 마이닝 분석
- 많은 데이터 속에 숨겨진 유용한 패턴을 추출하여 분류, 군집, 연관, 이상 탐지 분석 등을 수행

④ 최적화 분석
- 주어진 제한 조건을 만족하면서 목적 함수를 최대화 또는 최소화하는 방법 을 찾는다.

02. 빅데이터 분석 방법과 접근법

■ 빅데이터 분석 접근법

• 하향식 접근법
- 문제 해결 방법을 찾기 위해 필요한 데이터를 수집 및 분석하는 방식
- 문제 해결을 위해 근본 원인을 파악하고 분석 과제를 도출한 뒤 해결 방안을 도출
- 도출된 해결 방안에 대한 실현 가능성과 우선순위를 결정하기 위해 데이터를 수집, 가공, 분석하는 접근법
- 분석 과제를 도출하기 위해 '수요 기반 분석 과제 도출 방식'을 사용
- 데이터 분석은 문제 해결을 가능하게 하는 실행 동인 역할

• 상향식 접근법
- 현재 보유하고 있는 데이터를 분석하여 의미 있는 관계나 패턴을 찾아 지식을 발견하고 문제를 해결하는 방식
- 정형 데이터는 물론이고 다양한 원천의 비정형 데이터를 조합 하고 시각화를 통해 의미 있는 패턴을 파악한 뒤 이를 적용하여 문제를 해결하는 데이터 기반의 접근
- 분석 과제를 도출하기 위해 '데이터 주도 분석 과제 도출 방식'을 사용
- 데이터는 추진 동인 역할

• 프로토타이핑 접근법
- 빅데이터 환경의 불확실성을 고려한 방식
- 소비자의 요구 사항이나 데이터를 규정하기가 어렵고 데이터 원천도 명확히 파악하기 어려운 경우 사용
- 일단 프로토타입을 만들어 분석을 시도한 뒤 결과를 확인하고 개선하고 이를 반복

03. 빅데이터 분석을 위한 데이터 과학 방법론

■ 데이터 과학 방법론

- 여섯 단계로 구성되며 필요에 따라 특정 단계를 반복해서 수행 가능

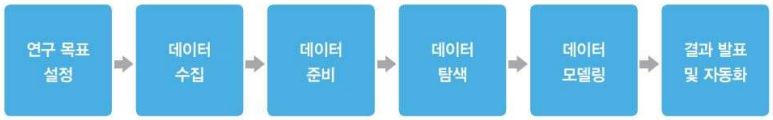


그림 3-5 데이터 과학 방법론의 6단계 구성

03. 빅데이터 분석을 위한 데이터 과학 방법론

■ [1단계] 연구 목표 설정

- 프로젝트와 관련된 모든 참여자가 연구 목표를 함께 정의하고 산출물과 일정 등의 계획에 합의한 뒤 프로젝트 헌장 작성

표 3-1 프로젝트 헌장 예시

프로젝트 헌장(Project Charter)			
프로젝트 명 (Project Name)			
프로젝트 설명 (Project Description)			
프로젝트 매니저 (Project Manager, PM)		승인 날짜 (Date Approved)	
프로젝트 스폰서 (Project Sponsor)		사명 (Signature)	
비즈니스 케이스(Business Case)		목표(Goals) / 산출물(Deliverables)	
팀 구성원(Team Member)			
이름(Name)	역할(Role)		
위험과 제약사항(Risk and Constraints)		주요 일정(Milestones)	

03. 빅데이터 분석을 위한 데이터 과학 방법론

■ [2단계] 데이터 수집

- 프로젝트에 필요한 데이터의 위치와 형태를 확인하고 원시 데이터를 수집
  - 필요한 데이터를 수집할 때는 이미 가지고 있는 내부 데이터베이스나 데이터 저장소를 이용
  - 외부에서 수집하는 경우 다양한 수집 기술을 활용할 수 있음
  - 수집할 데이터의 유형과 종류를 파악한 뒤 그에 맞는 수집 기술을 선택해서 사용

표 3-2 개방 데이터를 제공하는 사이트

사이트	설명
http://data.go.kr	한국 정부에서 제공하는 공공데이터
http://kostat.go.kr	한국 통계청에서 공개하는 데이터
http://opendata.hira.or.kr	한국 보건 의료 빅데이터 개방 시스템
http://www.localdata.kr	한국 지방행정 인허가 데이터
https://www.mcst.go.kr	한국 문화체육관광부 문화 데이터
http://data.seoul.go.kr	서울시 열린데이터 광장
https://data.gg.go.kr	경기도 공공데이터 개방 포털
http://data.gov	미국 정부의 공공데이터
http://data.worldbank.org	세계 은행에서 제공하는 개방 데이터
http://open.fda.gov	미국 식약청의 개방 데이터

표 3-3 다양한 데이터 수집 기술

수집 기술	설명	수집 데이터
크롤링	• SNS, 뉴스, 웹 정보처럼 인터넷에서 제공하는 데이터를 수집할 수 있다.	웹 추출 데이터
FTP	• TCP/IP 프로토콜을 활용하는 인터넷 서버에서 각종 파일을 송수신할 수 있다. • 보안을 강화하려면 SFTP 사용을 고려해야 한다. • 서버 간 연동시에는 전용 네트워크 구축을 고려해야 한다.	파일
Open API	• 서비스, 데이터 등을 어디서나 쉽게 이용하도록 개방된 API로 데이터 수집 방식을 제공한다. • 다양한 애플리케이션을 개발할 수 있도록 개발자와 소비자에게 공개되어 있다.	실시간 수집 데이터
RSS	• 웹 기반의 최신 정보를 공유하기 위한 XML 기반의 콘텐츠 배급 프로토콜이다.	XML 기반 웹 콘텐츠
스트리밍	• 인터넷에서 실시간으로 음성/오디오/비디오 데이터를 수집하는 기술이다.	음성/오디오/비디오의 실시간 수집 데이터
로그 수집기	• 웹 서버 로그, 웹 로그, 트랜잭션 로그, 클릭 로그, DB 로그 등 각종 로그 데이터를 수집하는 오픈 소스 기술이다. • Chukwa, Flume, Scribe 등이 있다.	로그
RDB 수집기	• 관계형 데이터베이스에서 정형 데이터를 수집한 뒤 HDFS(하둡 분산 파일 시스템)나 HBase와 같은 NoSQL에 저장하는 오픈 소스 기술이다. • Sqoop, Direct JDBC/ODBC 등이 있다.	RDB 기반 데이터

표 3-4 데이터의 유형과 종류에 따라 사용할 수 있는 수집 기술의 예

유형	종류	수집 기술
정형 데이터	RDB, 스프레드시트	ETL, FTP, Open API
반정형 데이터	HTML, XML, JSON, 웹 문서, 웹 로그, 센서 데이터	크롤링, RSS, Open API, FTP
비정형 데이터	소셜 데이터, 문서(워드, 한글), 이미지, 오디오, 비디오, IoT	크롤링, RSS, Open API, 스트리밍, FTP

03. 빅데이터 분석을 위한 데이터 과학 방법론

■ [3단계] 데이터 준비

- 수집한 원시 데이터의 품질을 높이기 위해 정제 후 사용 가능한 형태로 가공하는 단계
- 수집한 데이터를 다음 단계에서 사용할 수 있게 오류를 여과 하거나 수정하여 정제
- 필요에 따라서는 데이터를 통합하거나 형태를 변환

■ [4단계] 데이터 탐색

- 데이터와 변수 간의 관계나 상호 작용을 이해하기 위한 단계
- 변수 간의 관련성, 데이터의 분포, 편차, 패턴 존재 여부를 확인하는 탐색적 데이터 분석(EDA)이라고도 함
- 데이터를 쉽게 이해하기 위해 꺾은선 그래프나 히스토그램, 분포도 등과 같은 그래픽 기법을 많이 사용

표 3-5 데이터 준비에 필요한 작업

종류	설명
데이터 여과	• 오류 발견, 보정, 삭제, 중복성 확인 등의 과정을 통해 데이터 품질을 향상시킨다.
데이터 정제	• 결측치는 채워 넣고 이상치는 식별 또는 제거하고 잡음이 섞인 데이터는 평활화하여 데이터 불일치성을 교정한다.
데이터 통합	• 데이터 분석이 용이하도록 유사 데이터 및 연계가 필요한 데이터(또는 데이터베이스)를 통합한다.
데이터 축소	• 분석 시간을 단축하기 위해 분석에 사용하지 않는 항목은 제거한다.
데이터 변환	• 데이터 분석에 용이한 형태로 데이터 유형을 변환한다. • 정규화normalization, 집합화aggregation, 요약summarization, 계층 생성 등의 방법을 활용한다. • ETLExtraction, Transformation, Loading 도구를 제공한다.

03. 빅데이터 분석을 위한 데이터 과학 방법론

■ [5단계] 데이터 모델링

- 이전 단계에서 얻은 데이터 탐색 결과로 프로젝트에 대한 답을 찾는 단계
- 변수를 선택하여 모델을 구성하고 실행 및 평가하는 과정을 반복 수행하 여 문제 해결 모델을 완성
- 이때 분석하려는 데이터의 특성과 목적에 따라 모델 유형을 선택할 수 있음

표 3-6 데이터 분석 모델의 종류

유형	종류 및 설명	
통계 분석 모델	전통적인 분석 기법이다. 주로 수치형 데이터에 사용하며 확률을 기반으로 현상을 추정 및 예측한다.	
	기술 통계	대표적인 것으로 평균(산술평균, 중앙값, 최빈값), 분산, 표준편차가 있다.
	상관 분석	두 변수가 어떤 선형적 관계를 가지는지 분석하는 기법이다. 두 변수는 서로 독립적 관계일 수도 있고 상관된 관계일 수 있는데 이러한 관계의 강도를 상관관계라고 한다.
	회귀 분석	연속형 변수에 대해 독립 변수와 종속 변수 사이의 상관관계에 따른 수학적 모델인 선형적 관계식을 구하여 어떤 독립 변수가 주어졌을 때 이에 따른 종속 변수를 예측하거나 수학적 모델이 얼마나 잘 설명하고 있는지를 판별하기 위한 적합도를 측정하는 분석 기법이다.
	분산 분석	두 개 이상 다수의 집단을 비교할 때 집단 내의 분산, 총평균과 각 집단의 평균의 차이로 생긴 집단 간 분산의 비교를 통해 만들어진 F분포로 가설을 검증하는 기법이다.
	주성분 분석	다양한 변수를 분석하는 다변량 분석으로 많은 변수로부터 몇 개의 주성분을 추출하는 기법이다. 이때 주성분 분석은 차원 축소를 위한 것이다.
데이터 마이닝 모델	패턴 인식, AI, 머신러닝, 딥러닝 등을 이용하여 대용량 데이터에 숨겨진 데이터 간의 상호 관련성 및 유용한 정보를 추출하는 기법이다.	
	예측	대용량 데이터 집합 내의 패턴을 기반으로 미래를 예측한다(예: 수요 예측).
	분류	일정한 집단에 대해 특정한 정의로 분류 및 구분을 추론한다.
	군집화	구체적인 특성을 공유하는 자료를 분류한다. 미리 정의된 특성에 대한 정보를 가지지 않는다는 점에서 분류와 다르다(예: 유사 행동 집단의 구분).
	패턴 분석	동시에 발생한 사건 간의 상호연관성을 탐색한다(예: 장바구니 속 상품의 관계).
	순차 패턴 분석	연관 규칙에 시간 개념을 반영하여 시계열에 따른 패턴의 상호연관성을 탐색한다(예: 금융 상품 사용을 위한 반복 방문).
텍스트 마이닝 모델	텍스트 기반의 데이터로부터 새로운 정보를 발견할 수 있도록 정보 검색, 추출, 체계화, 분석을 모두 포함하는 텍스트 처리 과정 및 기법이다.	
소셜 네트워크 분석 모델	언어 분석 기반의 정보 추출을 통해 대용량의 소셜 미디어 데이터에서 이슈를 탐지하고 시간 경과에 따라 이슈가 유통되는 전체 과정을 모니터링하고 향후 추이를 분석하는 기법이다.	

03. 빅데이터 분석을 위한 데이터 과학 방법론

■ [6단계] 결과 발표 및 분석 자동화

- 프로젝트 수행 결과가 연구 목표를 달성했는지를 이해 당사자, 특히 의사 결정자에게 이해시키고 가능하다면 이후의 유사 프로젝트 수행을 위해 분석 과정을 자동화하는 단계
- [1단계]에서 작성한 프로젝트 헌장에 명시된 목표를 달성했는지 산출물이 제대로 작성되었는지, 일정과 예산은 계획대로 진행되었는지 여부를 확인
- 모든 참여자를 대상으로 분석 결과를 발표
- 분석 과정을 재사용할 수 있도록 자동화