

빅데이터의 이해	
성공회대학교	김종율 교수

2주차 “빅데이터의 개념 및 특징”

I. 데이터는 무엇일까?

여러분 모두 ‘데이터’란 말을 많이 들어 보았을 겁니다.

그리고 우리가 듣고 있는 강의 주제인 ‘빅데이터’는 무엇일까요?

데이터란 무엇인지 생각해 봅시다.

1. 데이터의 정의

‘데이터(Data)’라는 단어는 라틴어인 ‘Datum’의 복수형인 ‘Data’에서 그 유래를 찾을 수 있다.

‘datum’은 선물, 주어진 것, 차변(회계용어) 등의 의미를 담고 있다.

옥스포드 대사전에서는 데이터를 추론과 추정의 근거를 이루는 사실이라고 정의하고 있다.

위키피디아에서는 문자, 숫자, 소리, 그림, 형상, 단어 등의 형태로 된 의미 단어이며, 보통 연구나 조사들의 바탕이 되는 자료를 데이터라고 정의한다.

일반적으로 데이터란 이론을 세우는 데 기초가 되는 사실 또는 바탕이 되는 자료, 혹은 컴퓨터가 처리할 수 있는 문자, 숫자, 소리, 그림 따위의 형태로 된 자료 등을 의미한다.

여기서 중요한 것은 데이터는 정보가 아니라 자료라는 점이며 의미를 담고 있느냐 없느냐의 차이이다.

데이터가 의미를 담고 있으면 정보가 되지만, 의미가 없다면 그냥 자료로 남는다.

데이터는 마치 일기예보를 위한 기초자료와 같다.

온도, 습도, 풍향, 기압 등의 자료가 모여 내일 비가 온다든지 혹은 맑다든지 하는 일기예보를 가능하게 한다.

여기서 온도, 습도, 풍향, 기압 등은 데이터이고, 이를 근거로 일기예보라는 추론과 추정이 가능하다.

이처럼 데이터는 단순한 정보 객체로서의 가치뿐만 아니라 다른 객체(온도, 습도, 풍향, 기압 등)와의 상호관계 속에서 가치(일기예보)를 갖는다.

데이터는 객관적 사실이라는 존재적 특성을 갖는 동시에 추론•예측•전망•추정을 위한 근거로서의 당위적 특성을 함께 가진다.

2. 데이터의 유형

- **정량적 데이터(정형 데이터):** 수치로 표현할 수 있는 숫자•도형•기호 등의 데이터를 말하며, 저장•검색•분석 활용에 용이하다. 앞에서 온도, 습도, 풍향, 기압 등은 정량적 데이터이다

- **정성적 데이터(비정형 데이터):** 언어•문자 등의 정형화되지 않는 데이터를 말한다. '오늘 무릎이 무척 시리니 비가 올 것이다'라는 예측을 한다면 '무릎이 시린 정도'는 수치로 정형화 할 수 없는 정성적 데이터다,

'영화 감상평', 'SNS 실시간 검색어' 등의 정성적 데이터 수집과 분석에는 상대적으로 많은 비용과 기술적 투자가 필요하다.

영화 감상평, SNS 실시간 검색어는 어떻게 분석할까요?

3. 암묵지와 형식지

- 암묵지: 학습과 체험을 통해 개인에게 습득되어 있지만, 겉으로 드러나지 않는 상태의 지식을 말한다. 머리속에 존재하는 지식으로, 언어나 문자를 통해 나타나지 않는 지식이다.

암묵지는 대개 시행착오와 같은 경험을 통해 체득하는 경우가 많다.

- 형식지: '문서나 매뉴얼처럼 외부로 표출돼 여러 사람이 공유할 수 있는 지식'을 말한다.

교과서, 데이터베이스, 신문, 비디오와 같이 어떤 형태로든 형상화된 지식은 형식지라고 할 수 있다.

- 암묵지와 형식지의 상호작용: 공유화되기 어려운 암묵지가 형식지로 표출되고 연결되면 그 상호작용으로 지식이 형성된다. 암묵지가 고도화되거나 형식지화해 공유되는 등의 변환 과정을 거쳐 더 높은 가치를 창조하게 된다고 말한다.



4. 데이터와 정보



자료:Debra M. Amidon, 1997

- 데이터(Data): 개별 데이터 자체는 의미가 중요하지 않은 객관적인 사실을 말한다,.
- 정보(Information): 데이터는 가공•처리와 데이터 간 연관 관계 속에서 의미가 도출된 것을 말한다. 하지만 정보가 내포하는 의미는 유용하지 않을 수 있다.
- 지식(Knowledge): 데이터를 통해 도출된 다양한 정보를 구조화하여 유의미한 정보를 분류하고 개인적인 경험을 결합해 고유의 지식으로 내재화된 것을 말한다.
- 지혜(Wisdom): 지식의 축적과 아이디어가 결합된 창의적 산물이다.

이번 코로나 사태에서의 데이터, 정보, 지식, 지혜, 비전의 형성 과정에 대해서 살펴볼까요?

II. 데이터베이스

여기는 IT 전공 수업이 아니라 교양학부 수업입니다. 때문에 데이터베이스를 설명하는 것이 옳은 것인가에 대해 생각해 보았습니다.

하지만, 사용할 줄 몰라도, 깊이 알고 있지는 않아도 '빅데이터'를 이해하는데 있어 데이터베이스가 무엇인가에 대해서는 기본적으로 알아야 합니다.

왜냐하면, 데이터를 저장해야 하니까요. 그러한 데이터를 저장하는 플랫폼/소프트웨어/시스템이 바로 데이터베이스입니다.

아주 깊게 이해할 필요는 없지만, 이게 뭔지는 알아야겠죠.

1. 데이터베이스의 연혁

1950년대 미군에서 군수물자를 관리하기 위해 수집된 자료를 일컫는 '데이터(Data)', '기지(base)'의 합성어로 데이터베이스(Database)라는 용어가 처음 등장했다.

이후 1963년 미국 SDC(System Development Corporation)가 개최한 심포지엄에서 '대량의 데이터를 축적하는 기지'라는 개념으로 데이터베이스라는 용어가 공식적으로는 처음 사용됐다.

같은 해 GE의 C. 바크만은 최초로 현대적 의미의 데이터베이스 관리 시스템인 IDS를 개발했다. 이후 저장장치에 직접 접근하기 위한 다양한 데이터 모델과 데이터베이스 관리 시스템이 개발됐다.

2. 데이터베이스의 다양한 정의

- EU의 '데이터베이스의 법적 보호에 관한 지침': 데이터베이스란 체계적이거나 조직적으로 정리되고 전자식 또는 기타 수단으로 개별적으로 접근할 수 있는 독립된 저작물, 데이터 또는 기타 소재의 수집물이다.

- 우리나라의 '저작권법': 데이터베이스란 소재를 체계적으로 배열 또는 구성한 편집물로서 개별적으로 그 소재에 접근하거나 그 소재를 검색할 수 있게 한 것이다(법률적으로 기술 기반 저작물로 인정).

- '컴퓨터용어사전', '정보통신용어사전(TTA)': 데이터베이스란 동시에 복수의 적용 업무를 지원할 수 있게 복수 이용자의 요구에 대응해서 데이터를 받아들이고 저장·공급하기 위해 일정한 구조에 따라 편성된 데이터의 집합이다.

- '위키피디아': 데이터베이스란 체계적으로 정렬된 데이터 집합을 의미한다. 데이터의 양과 이용이 늘어나면서 데이터베이스는 대용량의 데이터를 저장·관리·검색·이용할 수 있는 컴퓨터 기반의 데이터베이스로 진화했다.

- '한국데이터산업진흥원': 데이터베이스란 문자, 기호, 음성, 화상, 영상 등 상호 관련된 다수의 콘텐츠를 정보처리 및 정보통신기기에 의하여 체계적으로 수집·축적하여 다양한 용도와 방법으로 이용할 수 있게 정리한 정보의 집합체를 의미한다.

여기서 콘텐츠란 다양한 의미 전달 매체에 의해 표현된 데이터, 정보, 지식, 저작물 등의 인식 가능한 모든 자료를 의미한다.

2. 데이터베이스와 데이터베이스 관리시스템

DB (Data Base)	체계적으로 수집·축적하여 다양한 용도와 방법으로 이용할 수 있게 정리한 정보의 집합체
DBMS (Data Base Management System)	이용자가 쉽게 데이터베이스를 구축·유지할 수 있게 하는 관리 소프트웨어

3. 데이터베이스의 일반적 특징

- 통합된 데이터(Integrated data): 동일한 내용의 데이터가 중복되지 않게 통합되어 있다.
- 저장된 데이터(Stored data): 컴퓨터 기술을 바탕으로 컴퓨터가 접근할 수 있는 저장 매체에 저장되어 있다.
- 공용 데이터(shared data): 여러 사용자가 서로 다른 목적으로 데이터베이스의 데이터를 공동으로 이용할 수 있다.
- 변화되는 데이터(Operational data): 새로운 데이터의 삽입, 기존 데이터의 삭제, 갱신으로 변화하므로 항상 현재의 정확한 데이터 상태를 유지한다

- 데이터베이스의 특징

정보의 축적 및 전달 측면	<p>- 기계 가독성: 대량의 정보를 일정한 형식에 따라 컴퓨터 등의 정보처리기가 읽고 쓸 수 있다.</p> <p>- 검색 가능성: 다양한 방법으로 필요한 정보를 검색할 수 있다.</p> <p>- 원격조작성: 정보통신망을 통해 원거리에서도 즉시 온라인으로 이용이 가능하다.</p>
정보이용 측면	이용자의 정보요구에 따라 다양한 정보를 신속하게 획득할 수 있고 원하는 정보를 정확하고 경제적으로 찾아낼 수 있다,
정보관리 측면	정보를 일정한 질서와 구조에 따라 정리·저장하고 검색·관리할 수 있게 하여 방대한 양의 정보를 체계적으로 축적하고, 새로운 내용 추가나 갱신이 용이하다.
정보기술발전 측면	데이터베이스는 정보처리, 검색·관리 소프트웨어, 관련 하드웨어, 정보 전송을 위한 네트워크 기술 등의 발전을 견인할 수 있다.

경제·산업적 측면	데이터베이스는 다양한 정보를 필요에 따라 신속하게 제공·이용할 수 있는 인프라의 특성을 가지고 있어 경제, 산업, 사회 활동의 효율성을 제고하고 국민의 편의를 증진하는 수단의 의미를 가진다.
-----------	--

4. 데이터베이스의 활용

4.1 기업 내부의 데이터베이스

- 인하우스 DB : 1990년대 정보통신망 구축이 가속화되면서 기업 경영 전반에 관한 모든 자료를 연계하여 일관된 체계로 구축 및 운영하는데 중점을 두었다. 이후 경영 활동의 기반이 되는 전사자원관리시스템(ERP)로 확대 되었다.
- OLTP(online Transaction processing) 시스템: 1990년대 중반 이전, 정보의 수집과 이를 조직 내에서 공유하기 위한 경영정보시스템(MIS)과 생산자동화, 통합 자동화 등 기업 활동에서 영역별로 구축되던 단순 자동화 중심의 시스템을 말한다.
- OLAP(Online Analytical Processing) 시스템 : 데이터마이닝 등의 기술이 등장하면서 단순한 정보의 수집과 공유에서 탈피하여 분석이 중심이되는 시스템 구축으로 변화하게 되었는데, 이를 OLAP이라고 부른다.
- CRM과 SCM: 2000년대에 들어서면서 기업 DB 구축의 화두는 CRM과 SCM으로 변화했다. 유통·판매 및 고객 데이터가 CRM과 연동되어 CRM과 SCM은 상호 밀접한 관련을 가지게 됐다.

4.2 사회 기반 구조의 데이터베이스

1990년대 사회 각 부문의 정보화가 본격화되면서 DB 구축이 활발하게 추진됐다. 이후 무역, 통관, 물류, 조세, 국세, 조달 등 사회간접자본(SOC) 차원에서 EDI활용이 본격화되면서 부가 가치통신망(VAN)을 통한 정보망이 구축되었다.

지리·교통 부문의 데이터베이스는 고도화되고, 의료·교육·행정 등 사회 각 부문으로 공공 DB의 구축·이용이 확대됐다.

II. 빅데이터의 이해

1. 빅데이터의 정의

- 일반적 정의: 빅데이터의 '빅(Big)'에는 단순히 양적인 개념뿐만 아니라 복잡하고 다양한 질적인 개념도 포함되어 있다. 일반적으로 빅데이터란 큰 용량과 복잡성으로 기존 애플리케이션이나 툴로는 다루기 어려운 데이터셋의 집합을 의미한다.
- 가트너의 정의: 빅데이터란 향상된 시사점과 더 나은 의사결정을 위해 사용되는 비용 효율이 높고 혁신적이며 대용량, 고속 및 다양성의 특성을 가진 정보 자산을 말한다(2012)
- 메킨지 정의 : 빅데이터란 일반적으로 데이터베이스 소프트웨어가 저장, 관리, 분석할 수 있는 범위를 초과하는 규모의 데이터를 말한다(2011)
- IDC 정의: 빅데이터란 다양한 종류의 대규모 데이터에서 낮은 비용으로 가치를 추출하고, 데이터의 초고속 수집과 발굴을 지원하도록 고안된 차세대 기술 및 아키텍처를 말한다(2011)
- 일본 노무라연구소 정의: 데이터와 데이터 처리, 저장 및 분석 기술에 의미 있는 정보 도출은 물론이고, 그에 필요한 인재나 조직까지도 넓은의미의 빅데이터에 포함시킬 것을 제안했다.
- 더그 레이니의 정의(3V): 빅데이터는 데이터의 양(Volume), 데이터의 유형과 소스의 다양성(Variety), 데이터 수집과 처리 측면에서의 속도(Velocity)가 급격히 증가하면서 나타나는 현상이다.
- 마이어쉴베르크와 쿠키어의 정의: 빅데이터란 대용량 데이터를 활용해 작은 용량으로 얻을 수 없었던 새로운 통찰이나 가치를 추출해내는 일이다. 나아가 이를 활용해 시장,

기업 및 시민과 정부의 관계 등 많은 분야에 변화를 가져오는 일이다.

- 한국데이터산업진흥원 정의: 빅데이터란 데이터에 대한 기존의 접근 방식으로는 얻을 수 없었던 통찰과 가치를 창출하는 모든 것을 말한다.

각자의 빅데이터에 대한 정의를 기준으로 빅데이터는 무엇일까요?

빅데이터라는 용어를 처음 만들어 낸 것은 2000년도 정보 폭발을 가장 먼저 일으킨 천문학 및 계측 연구 분야였다. 천문학에서 나오는 우주로부터 들어오는 대량의 정보와 계측 연구에서 나오는 DNS 염기서열 정보를 가지고 과학자들은 놀라지 않을 수 없었다. 과거와는 비교할 수 없을 정도로 데이터의 규모가 커지고 종류도 다양해졌기 때문이다.

2. 빅데이터의 특징

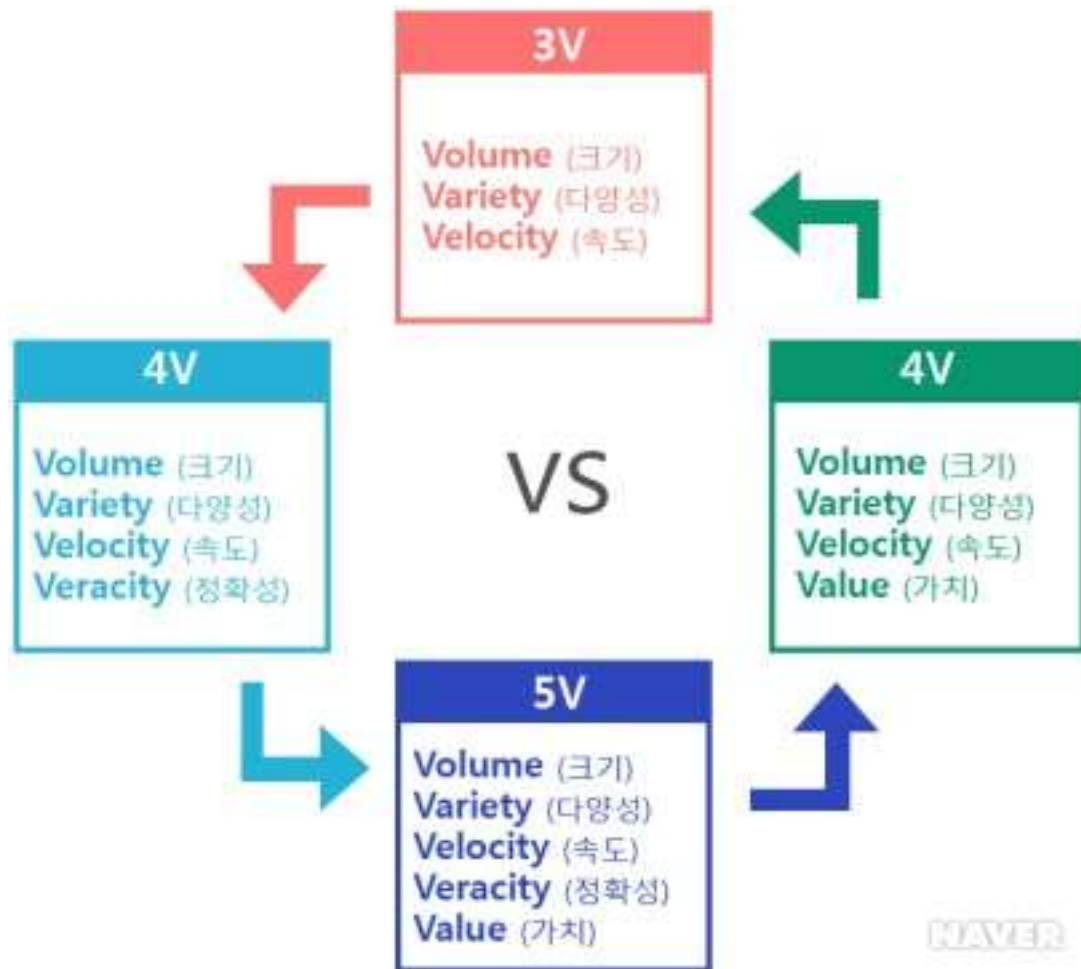
- 일반적으로 빅데이터의 가장 기본적인 특징을 3V, 즉 Volume(크기), Variety(다양성), Velocity(속도)로 정의한다.

구분	세부 내용
Volume(크기)	대량의 데이터 증가 발생으로 기존 데이터 수집, 관리 한계
Variety(다양성)	비정형 데이터(영상, SNS등)의 발생으로 다양한 데이터 형식 증가
Velocity(속도)	실시간 정보 발생으로 데이터의 유입, 처리 속도 요구

- 3V에 추가로 Value(가치) 혹은 Veracity(정확성)를 포함해 4V로 빅데이터의 특징을 설명하기도 한다. 일부 학자들은, Visualization(시각화), Variability(가변성) 등을 추가하기도 한다.

구분	세부 내용
----	-------

Value(가치)	데이터 전체를 파악하고 패턴을 발견하기가 어렵게 되면서 가치(Value)의 중요성 강조
Veracity(정확성)	빅데이터 기반의 예측 분석 결과에 대한 신뢰성이 중요하게 됨



3. 빅데이터 출현 배경

과학기술의 발달에 따라 빅데이터는 왜 늘어나게 되었을까요?

- **산업계의 변화:** 산업계에서 일어난 빅데이터 현상을 '양질 전환의 법칙'으로 설명하기도 한다(한국데이터산업진흥원). '양질 전환의 법칙'이란 헤겔의 변증법에 기초를 둔 개념으로, 양적인 변화가 축적되면 질적인 변화도 이루어진다는 개념이다. 정보가 지속적으로

축적되면서 거대한 데이터는 새로운 기술을 만나 새로운 가치를 창출할 수 있는 변화의 상태가 된다는 것이다. 이제 기업들이 직면한 도전은 '보유 데이터 속에 숨어있는 가치를 발굴해 새로운 성장동력원으로 만들어 낼 수 있는 빅데이터 기술의 확보'라고도 할 수 있다.

- **학계의 변화:** 하계에서도 빅데이터를 다루는 현상이 증가하고 있다. 거대한 데이터를 다루는 학문 분야가 늘어나면서 필요한 기술 아키텍처 및 통계 도구도 지속해서 발전하고 있다(예: 게놈 프로젝트, 대형 강입자 충돌기, NASA의 기후 시뮬레이션 등)
- **관련 기술의 발전:** 빅데이터가 출현한 배경을 기술 발전에서 찾을 수 있다. 디지털화의 급진전, 저장 기술의 발전과 가격 하락, 인터넷의 발전과 모바일 시대 돌입, 클라우드 컴퓨팅 보편화 등은 빅데이터의 출현과 매우 밀접하다. 특히 클라우드 컴퓨팅은 많은 정보가 클라우드에 수집하는 것도 의미가 있지만, 무엇보다 빅데이터의 처리 비용을 획기적으로 낮추었다는 점에서 그 의의가 있다. 대용량의 데이터를 클라우드 분산 병렬처리 시스템으로 처리할 경우 비용이 혁신적으로 줄어든다는 점도 빅데이터를 분석하고 새로운 가치를 창출하는 데 기여했다고 볼 수 있다.

- 빅데이터에 거는 기대

빅데이터는 "산업혁명의 석탄•철"	제조업뿐만 아니라 서비스 분야의 생산성을 획기적으로 끌어올려 사회•경제•문화•생활 전반에 혁명적 변화를 가져올 것으로 기대
빅데이터는 "21세기 원유"	빅데이터도 원유처럼 각종 비즈니스, 공공기관 대 국민 서비스, 그리고 경제 성장에 필요한 정보를 제공하여 산업 전반의 생산성을 향상시키고 새로운 범주의 산업을 만들어낼 것으로 기대
빅데이터는 "렌즈"	렌즈는 통해 현미경이 생물학 발전에 끼쳤던 영향만큼, 빅데이

	터도 렌즈처럼 산업 발전에 큰 영향을 줄 것으로 기대
빅데이터는 “플랫폼”	플랫폼은 공동 활용의 목적으로 구축된 유무형의 구조물을 말한다. 빅데이터는 플랫폼으로서 다양한 서드파티 비즈니스에 활용될 것으로 기대된다.

- 빅데이터가 만들어내는 변화

- 사전처리 → 사후처리: 데이터를 사전 처리하지 않고, 가능한 많은 데이터를 모으고 데이터를 다양한 방식으로 조합하여 숨은 인사이트를 발굴한다.
- 표본조사 → 전수조사 : IOT•클라우드 기술의 발전으로 데이터 처리 비용이 감소하게 되면서, 데이터 활용 방법이 표본조사에서 전수조사로 변화되었다.
- 질 → 양: 수집 데이터의 양이 증가할수록 분석의 정확도가 높아져 양질의 분석 결과 산출에 긍정적인 영향을 주었다.
- 인과관계→상관관계: 특정한 인과관계가 중요시되던 과거와 달리, 데이터 양이 급격하게 늘어나면서 상관관계를 통해 특정 현상의 발생 가능성이 포착되고 그에 상응하는 행동을 추천하는 등 상관 관계를 통한 인사이트 도출이 점점 확산되고 있다.

4. 빅데이터의 가치와 영향

- 빅데이터의 가치에 대해 생각해 봅니다.
- 빅데이터 가치 산정의 어려움: 빅데이터 가치 산정은 데이터 활용 방식, 가치 창출 방식, 분석 기술의 발전이라는 세 가지의 이유로 쉽지 않다.

데이터 활용 방식	빅데이터의 재사용이나 재조합, 다목적용 데이터 개발 등이 일반화되면서 특정 데이터를 누가, 언제, 어떻게, 어디서 활용하는지 알
-----------	---

	수 없게 되었기 때문에 가치 산정이 어렵다.
가치 창출 방식	빅데이터는 기존에 없던 새로운 가치를 창출함에 따라 그 가치를 산정하기 어렵다
분석 기술의 발전	데이터 분석 기술의 발전으로 가치 있는 데이터와 가치 없는 데이터의 경계를 나누기 어려워졌다. 오늘의 가치 없는 데이터가 내일은 가치 있는 데이터가 될 수도 있기 때문에 빅데이터의 가치 산정은 어렵다.

- 빅데이터의 영향

- 빅데이터가 가치를 만들어내는 5가지 방식 (맥킨지의 빅데이터 보고서, 2011)
- 투명성 제고로 연구개발 및 관리 효율성 제고
- 시뮬레이션을 통한 수요포착 및 주요 변수 탐색으로 경쟁력 강화
- 고객 세분화 및 맞춤 서비스 제공
- 알고리즘을 활용한 의사결정 보조 혹은 대체
- 비즈니스 모델과 제품, 서비스의 혁신 등

빅데이터의 영향

기업	빅데이터를 활용해 소비자의 행동을 분석하고, 시장 변동을 예측해 비즈니스 모델을 혁신하거나 신사업을 발굴
정부	빅데이터 활용 부문은 크게 환경 탐색, 상황 분석, 미래 대응으로 나눌 수 있음 <ul style="list-style-type: none"> ● 미래 대응: 법제도 및 거버넌스 시스템 정비, 미래성장 전략, 국가 안보 대응 등
개인	개인의 목적에 따라 빅데이터의 활용이 확산되면서 스마트라이프로 변화

	됨
--	---

5. 빅데이터와 비즈니스 모델

- 빅데이터 활용 사례

- 기업혁신 사례: 구글 검색 기능, 월마트 매출 향상, 질병 예후 진단 등 의료 분야에 적용
- 정부 활용 사례: 실시간 교통정보수집, 기후정보, 각종 지질 활동 등에 활용, 국가 안전 확보 활동 및 의료와 교육 개선에의 활용 방안 모색
- 개인 활용 사례 : 정치인과 연예인의 SNS 활용

6. 빅데이터 활용 기본 테크닉 7가지

1) 연관 규칙 학습

- 어떤 변인들 간에 주목할 만한 상관관계가 있는지를 찾아내는 방법이다.
- '커피를 구매하는 사람이 탄산음료를 더 많이 사는가?'
- '우유구매자가 기저귀를 더 많이 구매하는가?' 등을 해결하고자 할 때 사용한다.

2) 유형 분석

- 통계적 분류를 활용하여 새로운 사건이 속하게 될 범주를 찾아내는 방법이다.
- "이 사용자는 어떤 특성을 가진 집단에 속하는가?" 와 같은 문제를 해결하고자 할 때 사용한다.
- 기존 자료를 바탕으로 만들어진 훈련용 분류틀이 미리 갖춰져 있어야 한다.
- 이 기법은 문서를 분류하거나 조직을 그룹으로 나눌 때, 혹은 온라인 수강생들을 특성에 따라 분류할 때 사용할 수 있다.

3) 유전 알고리즘

- 최적화가 필요한 문제의 해결책을 자연선택, 돌연변이 등과 같은 메커니즘을 통해 점진적으로 진화(evolve)시켜 나가는 방법이다.
- '최대 시청률을 얻으려면 어떤 프로그램을 어떤 시간대에 방송해야하는가?'
- '응급실에서 의사를 어떻게 배치하는 것이 효율적인가?'

4) 기계 학습

- 훈련 데이터로부터 학습한 알려진 특성을 활용해 '예측' 하는 일에 초점을 맞춘다.

'기존의 시청 기록을 바탕으로 시청자가 현재 보유한 영화 중에서 어떤 것을 가장 보고 싶어할까?'와 같은 문제를 해결할 때 사용한다.

이를 위해 기계 학습은 데이터를 통해 학습할 수 있는 소프트웨어를 포함하고 있다. 사용자의 기호를 학습해 추천서비스를 제공할 때도 사용된다. (유튜브 관련 동영상 뜨는 것도 비슷한 원리!)

5) 회귀 분석

분석가는 독립변수를 조작하며, 종속변수가 어떻게 변하는지를 보면 두 변인의 관계를 파악한다.

'사용자의 만족도가 충성도에 어떤 영향을 미치는가?'

'이웃들과 그 규모가 집값이 어떤 영향을 미치는가?'

'구매자의 나이가 구매 차량의 타입에 어떤 영향을 미치는가?'

등을 해결하고자 할 때 쓰인다.

6) 감정 분석

특정 주제에 대해 말하거나 글을 쓴 사람의 감정을 분석한다.

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

소셜 미디어에서 나타난 의견을 바탕으로 고객이 원하는 것을 찾아낼 때 이 기법이 활용된다.

7) 소셜 네트워크 분석 (사회관계망 분석)

오피니언 리더, 즉 영향력이 있는 사람을 찾아낼 수 있으며, 고객들 간 소셜 관계를 파악할 수 있다.

'특정인과 다른 사람이 몇 촌 정도의 관계인가?' 를 파악할 때 사용된다.

7. 빅데이터의 위기 요인과 통제 방안

위기 요인	통제 방안
사생활 침해	동의에서 책임으로
책임 원칙 훼손	결과 기반 책임 원칙 고수
데이터 오용	알고리즘 접근 허용

참고 서적 : 2021 빅데이터 분석기사 필기(위키북스)