

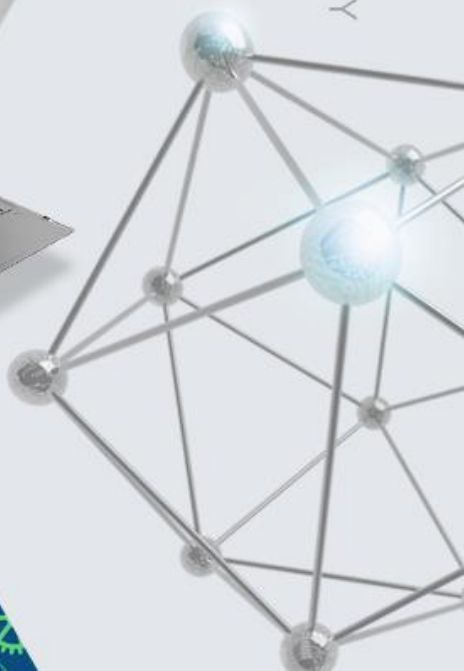


한국기술교육대학교
온라인평생교육원

The 4th Industrial Revolution is characterized by super connectivity and super intelligence, where various products and services are connected to the network, and artificial intelligence and information communication technologies are used in 3D printing, unmanned transportation, robotics. Of the world's most advanced technologies.

빅데이터 입문

묘사적 빅데이터 기술분석 수행하기



The 4th Industrial Revolution is characterized by super connectivity and super intelligence, where various products and services are connected to the network, and artificial intelligence and information communication technologies are used in 3D printing, unmanned transportation, robotics. Of the world's most advanced technologies.





묘사적 빅데이터 기술분석 수행하기



학/습/목/표

1. 묘사적 빅데이터 분석을 정의하고 분류할 수 있다.
2. 기술통계, 클러스터링, 연관규칙분석 기법을 수행할 수 있다.



학/습/내/용

1. 묘사적 빅데이터 분석 살펴보기
2. 묘사적 빅데이터 기술분석의 대표적 기법

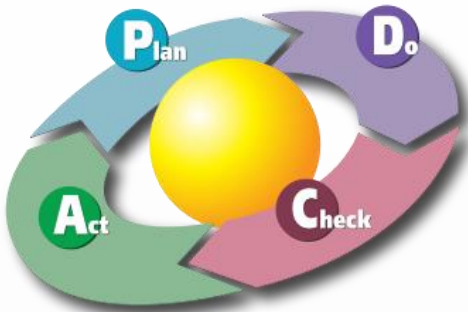
1. 묘사적 빅데이터 분석 살펴보기

1) 빅데이터 기술 분석의 시작

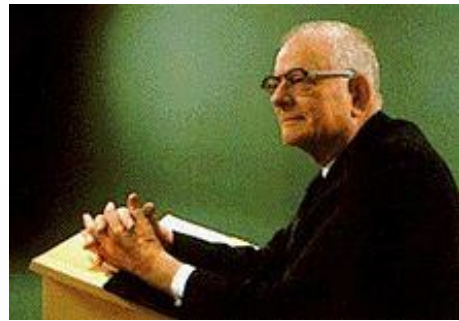
(1) 기업들이 빅데이터 분석을 하는 가장 큰 목적

- 수집/저장된 데이터 분석
- 기업경영에 도움이 되는 즉, 가치 있는 지식을 추출하여 활용

(2) 4단계로 업무적 프로세스 추진



Deming's PDCA Cycle



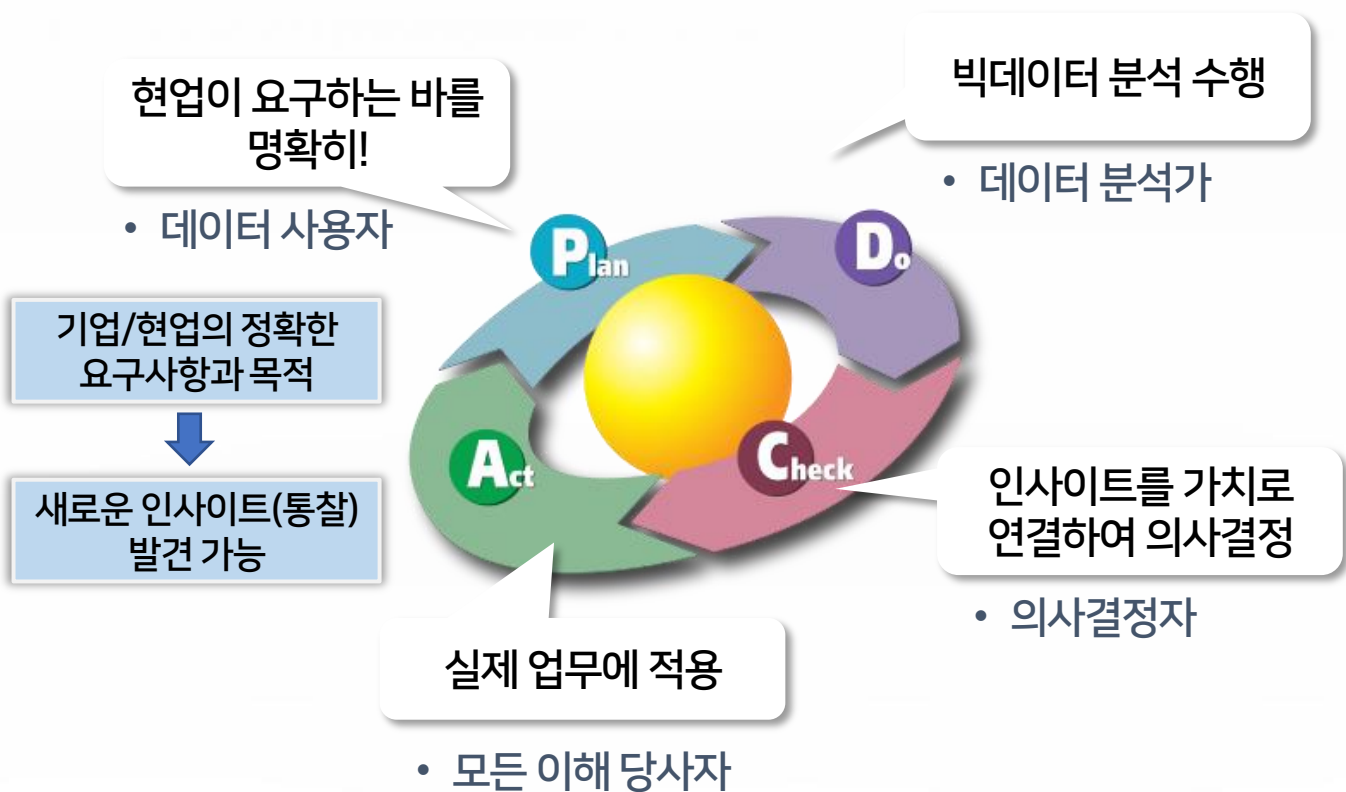
W. Edwards Deming

1. 묘사적 빅데이터 분석 살펴보기

1) 빅데이터 기술 분석의 시작

(1) PLAN

- 빅데이터의 분석을 통해 얻고자 하는 것이 무엇인가?



데이터 사용자: 기업·현업의 정확한 요구사항과 목적이 분명해야
인사이트 발견 가능

1. 묘사적 빅데이터 분석 살펴보기

1) 빅데이터 기술 분석의 시작

(1) PLAN

- 구글 검색, 구글 포토, 인스타그램, 유튜브, 등 사용료가 왜 무료일까?



검색 데이터



사진 데이터



Instagram

사진 데이터



동영상 데이터



동영상 데이터



NAVER

검색 데이터

서비스를 제공하고 데이터를 수집



새로운 인사이트 창출,
비즈니스 활용 가치로 연결

1. 묘사적 빅데이터 분석 살펴보기

1) 빅데이터 기술 분석의 시작

(2) 공공 분야에서의 다양한 데이터

- 공공데이터포털 : <https://www.data.go.kr>
- 아고다 : 항공권 예약 → 숙박 예약
 - 웹 사이트의 평점과 이용 후기는 특정 사용자의 중요한 도구로 사용됨
 - 사용자가 아고다에 자발적으로 빅데이터를 제공함
 - 제공의 대가로 다른 사용자의 이용후기 열람 가능
- FOURSQUARE : 사용자들의 후기와 음식점의 평점을 볼 수 있음
 - 사용자 이용후기와 사진 데이터 수집
 - 포스퀘어 사용자들에 또 다른 가치 제공
 - 특정 지역에 대한 정보와 예측으로 가치 창출, 인사이트 연결
- ✓ 즉 제 3의 기업에 제공하여 수익창출 가능

1. 묘사적 빅데이터 분석 살펴보기

1) 빅데이터 기술 분석의 시작

(3) 가트너그룹의 빅데이터 분석 방법

- 묘사 분석: 지금 일어난 현상에 대한 설명
→ 실제 업무에서 많이 활용

기술통계

클러스터링
(군집화)

연관규칙분석

- 진단 분석: 그러한 일이 발생한 원인을 제시
- 예측 분석: 어떤 일이 발생할지를 나타냄
→ 축적된 데이터를 활용하여 미래 값을 추적, 예측
- 처방 분석: 우리가 해야 될 일을 제시

1. 묘사적 빅데이터 분석 살펴보기

2) 묘사적 기술분석의 의미

(1) 묘사적 분석(Descriptive Approach)

- 특정 기업이나 기관에서 발생한 일 혹은 사건을 정리해서 깔끔하게 제공



자동차 계기판과 같이 기업의 현상을 데이터 형태로 요약, 분류, 수치화된 형태로 사용자에게 제공



지표의 지정, 숫자를 표나 도식으로 시각화하여 표현하는 것이 중요

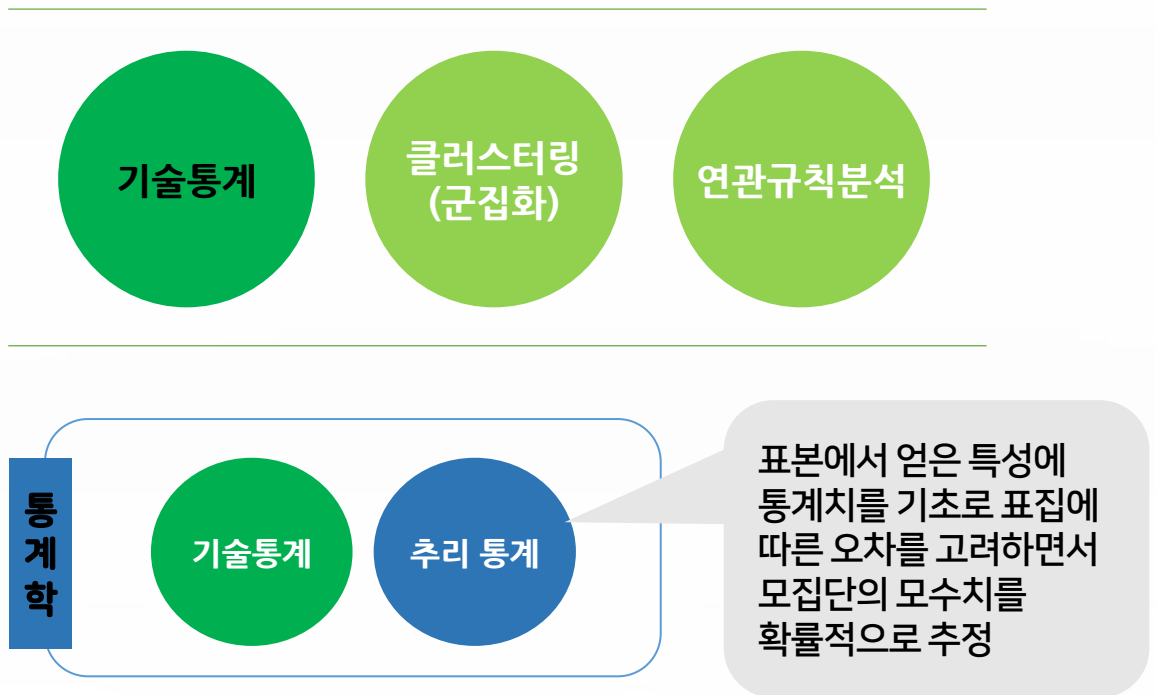
- 기업의 발생하는 일을 일목요연하게 요약, 시각화하여 제시
→ 이해관계자들에게 인사이트(통찰)를 제공

2. 묘사적 빅데이터 기술분석의 대표적 기법

1) 기술통계

(1) 기술통계

- 데이터를 조직하고 요약하는 하는 한 방법으로 데이터를 이해하기 쉽게 수치로 표현

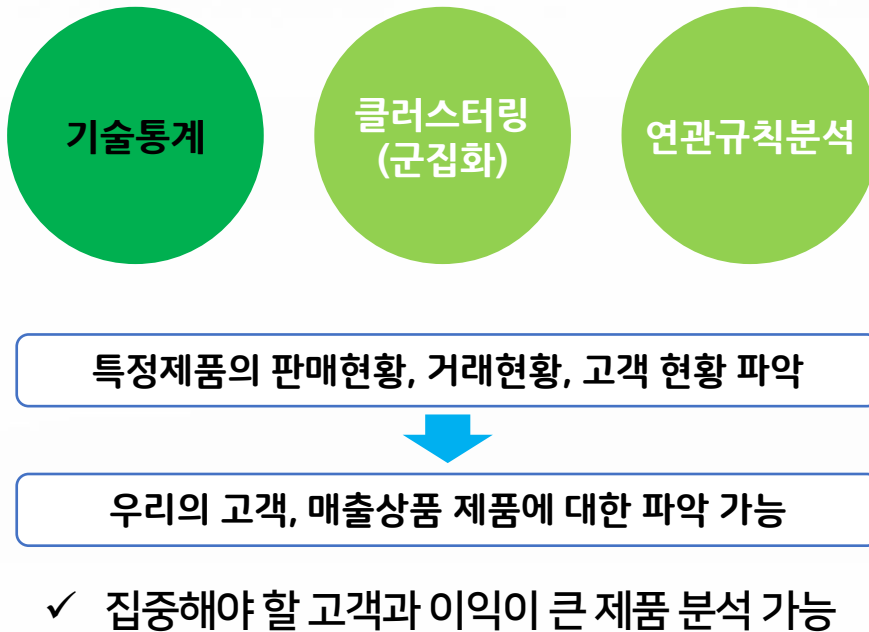


2. 묘사적 빅데이터 기술분석의 대표적 기법

1) 기술통계

(1) 기술통계

- 기업에서의 기술 통계 분석 활용



2. 묘사적 빅데이터 기술분석의 대표적 기법

1) 기술통계

(1) 기술통계

- 기업에서의 기술 통계 분석 활용 - 편의점



나이 : 20살
주소 : 서울 중구
직업 : 대학생
한 달에 5번 방문



멤버십 카드를
제시하면 할인 가능

- 멤버십 카드 제시 → 나의 정보를 제공 → 축적된 데이터 기반으로 기술통계 실시
 - 편의점은 축적된 데이터를 기반으로 기술 통계 실시
 - 30대 남성이 저녁 8시에 가장 많이 구입한 제품은?
 - 동시에 구입하는 제품은?
 - 방문 빈도 및 가장 최근 방문 시점은?
- ✓ 연령별, 성별, 지역별 분류와 요약 후 관계자에게 제시 할 수 있음

2. 묘사적 빅데이터 기술분석의 대표적 기법

1) 기술통계

(1) 기술통계

- 기술통계의 또 다른 예 → 검색엔진
 - 데이터를 제공, 서비스 무료 사용: 특정 개인 정보 제외한 데이터 바탕으로 자주 입력된 검색어를 파악, 제시
 - 편의점의 경우 제품별, 일자별, 지역별, 지점별, 주 연령별, 직영점별, 가맹점별로 현재 매출 체계적으로 파악
- 자동차 계기판 형태와 유사한 대시보드로 관찰



2. 묘사적 빅데이터 기술분석의 대표적 기법

2) 클러스터링(군집화)

(1) 클러스터링(군집화)

- 유사 정도에 따라 다수 객체를 군집/클러스터로 나누는 작업
 - 동일 군집의 개체간 유사도가 높도록 군집을 구성
 - 대표적 비지도 학습방법, 데이터 분할/요약에 널리 이용
- 데이터에 유용한 지식을 추출하여 이해관계자/사용자에게 제시

기술통계

클러스터링
(군집화)

연관규칙분석

2. 묘사적 빅데이터 기술분석의 대표적 기법



2) 클러스터링(군집화)

(2) 대표적 비지도 학습방법

- 지도학습: 데이터의 특정 값이 존재하고 이에 따라 값을 예측하거나 분류하는 학습방법
- 비지도학습: 조력자의 도움 없이 처음부터 스스로 학습하는 형태
- 반지도학습: 지도학습과 비지도학습을 혼합한 학습 방식
- 강화학습: 컴퓨터의 동작이 적절한지에 대한 피드백이 있는 학습방법

2. 묘사적 빅데이터 기술분석의 대표적 기법

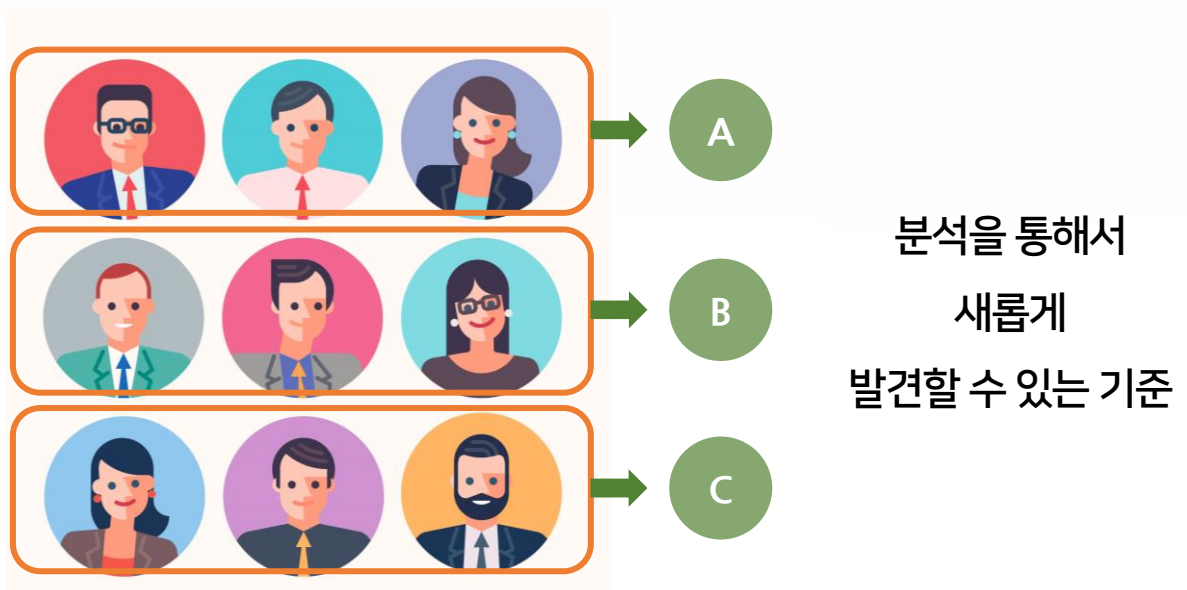
2) 클러스터링(군집화)

(3) 세분화(Segmentation) VS 클러스터링(군집화)

세분화(Segmentation)	클러스터링(군집화)
<ul style="list-style-type: none">인구 통계학적 분류대표적인 비통계학적 방법	<ul style="list-style-type: none">처음부터 유사성을 기반으로 하는 것이 아님유사성을 찾아 내는 자동화되고 통계적인 방법

- 클러스터링(군집화): 고객에 대한 여러 요소를 사용하여 고객 기반 속에 존재하는 세분화(Segmentation)를 자동으로 발견하는 방법

(4) 군집화 예시 - 특정 행위에 따른 분류



2. 묘사적 빅데이터 기술분석의 대표적 기법

2) 클러스터링(군집화)

(4) 군집화 예시 - 인터넷 서점

- 유사한 성향을 가진 집단이 샀던 책을 추천하여 구입 가능성을 높임
- 군집분석을 사용하여 특정집단에 속한 고객에게 추천하는 기능 제시 가능

(4) 군집화 예시 - OTT(Over-The-Top) 방송사업자

- 특정 영화를 본 사용자 → 해당 사용자가 속한 집단이 다시 선택한 영화

2. 묘사적 빅데이터 기술분석의 대표적 기법

3) 연관규칙분석

(1) 연관규칙분석

- 데이터의 숨어있는 항목 간의 관계를 탐색하는 분석방법
 - 얼마나·자주·동시에 발생하는지를 표현하는 규칙 또는 조건을 발견함
- 비지도학습의 일종, 인과관계와 구분해야 함

기술통계

클러스터링
(군집화)

연관규칙분석

2. 묘사적 빅데이터 기술분석의 대표적 기법

3) 연관규칙분석

(1) 연관규칙분석

- 데이터에서 여러 번 발생하는 속성을 찾아내고 이 중에서 서로 연관 있는 규칙을 발견하는 기법
- 대용량 데이터베이스에서 특정 형태의 지식을 찾아내는 것

데이터 웨어하우스



VS

데이터



- ✓ 주제지향적
- ✓ 통합성
- ✓ 시간변이적
- ✓ 비휘발성

2. 묘사적 빅데이터 기술분석의 대표적 기법

3) 연관규칙분석

(1) 연관규칙분석

- 상품이나 서비스의 거래기록 데이터로부터 연관성이 많은 상품들을 그룹화하는 방법
- 장바구니 분석
 - 상품의 근접 배치, 교차 판매 전략에 적용
 - 돼지고기 구매 코너와 상추 구매 코너를 근접 배치하여 고객만족과 매출 증대 효과

2. 묘사적 빅데이터 기술분석의 대표적 기법

3) 연관규칙분석

(1) 연관규칙분석

- ① 교차 판매: 제품을 같이 판매하게 하는 상품 전략
- ② 상향 판매: 제품을 바꿀 시점에 업그레이드 제품을 제시하고 추천함
- ③ 재판매: 마일리지와 포인트 적립을 통해 고객이 다시 와서 구매하도록 유도함

거래 발생 시의 구매 성향에 관한 물품을
그룹핑하여 연관관계를 규칙으로 표시

If A than B

새로운 생물학적 관계 규명, 시스템
모니터링 등 다양한 분야에 적용

지지도, 신뢰도, 상승도에 대한
이해가 필요함

2. 묘사적 빅데이터 기술분석의 대표적 기법

3) 연관규칙분석

(1) 연관규칙분석

- 지지도: Support
 - 전체 자료에서 A를 구매한 후 B를 구매하는 거래의 리브
- 신뢰도: Confidence
 - A가 포함된 거래 중에서 B를 포함한 거래비율 이른바 조건부확률
- 지지도: Lift
 - 하위 항목들이 독립해서 얼마나 벗어났는지의 정도를 측정하는 값

장점	단점
<ul style="list-style-type: none">• 이해가 쉬움• 계산과정이 간단함• 데이터 변환 없이 활용 가능함	<ul style="list-style-type: none">• 품목 수 증가 시 계산 과정이 복잡함• 지나친 세분화로 인한 의미 손실



1. 묘사적 빅데이터 분석 살펴보기

1) 기업들이 빅데이터 분석을 하는 목적

- 수 많은 데이터를 수집하고 저장하는 것도 중요하지만 수집되고 저장된 데이터를 분석을 해서 가치 있는 지식을 추출하고 이를 업무에 활용하기 위함

2) 예측적 분석

- 예측적 분석은 과거의 데이터를 기반으로 관측되지 않았던 변수의 미래의 값을 평가하는 것을 의미





2. 묘사적 빅데이터 기술분석의 대표적 기법

1) 분석기술

- 빅데이터 분석 실무에서 가장 많이 활용되는 두 가지 분석기술은 묘사적 분석과 예측적 분석이 있음

2) 군집분석

- 묘사적 데이터 분석 중 군집분석은 서로 유사한 정도에 따라 다수의 객체를 군집으로 나누는 작업으로 동일한 군집에 속하는 객체 간의 유사도가 그렇지 않은 객체 간의 유사도보다 평균적으로 높도록 군집을 구성

3) 연관분석

- 묘사적 데이터 분석 중 연관분석은 데이터에 숨어있는 항목 간의 관계를 탐색하는 것으로 데이터에서 여러 번 발생하는 속성을 찾아내고, 이 중에서 서로 연관이 있는 규칙을 발견하는 기법

