

빅데이터 쉽게 이해하기

1차시 빅데이터의 개요

1. 빅데이터 4가지 구성요소

- 빅데이터는 3V로 대표되는 규모(Volume), 다양성(Variety), 속도(Velocity)의 증가뿐 아니라 복잡성(Complexity) 등 4가지 구성요소를 갖추어야 한다.

2. 빅데이터란?

- 기존의 관리 및 분석체계로는 감당할 수 없을 정도의 거대한 데이터의 집합을 지칭하며 대량의 정형 또는 비정형 데이터세트 및 이러한 데이터로부터 가치를 추출하고 결과를 분석하는 기술을 말한다.

3. 빅데이터의 개념

- 고객정보와 같은 정형화된 자산정보(내부)뿐만 아니라 외부 데이터 및 비정형, 소셜, 실시간 데이터 등이 복합적으로 구성되어 있다.
- Google, Naver, Apple, Daum 등 대기업의 프로젝트에 분석하는 대용량의 정형화된 개념은 빅데이터의 개념과 다르다.

4. 빅데이터 플랫폼의 애플리케이션 레벨

- 빅데이터를 저장할 수 있는 레벨은 하드웨어 레벨이며 분석하는 레벨은 소프트웨어 레벨이다.
- 실제 빅데이터와 사용자가 만나게 되는 커뮤니케이션 레벨
- 인프라 및 데이터 처리기술분야 보다 상대적으로 낮은 진입장벽
- 빅데이터 의미파악, 이해능력, 분석결과를 서비스화 할 수 있는 애플리케이션

5. 빅데이터의 의의

- 빅데이터는 범람하는 정보 속에서 새로운 가치창출을 하고자 데이터를 저장하고 분석하는 과정으로 단순히 범람하는 정보를 처리·삭제하는 과정이 아니다. 또한 대용량 데이터를 보관하기 위한 시스템을 갖추는 것이지 용량을 줄이는 것이 아니다.
- 분석결과를 갖고 국가기관 민간 등의 분석기획 담당자들이 자유롭게 이용할수있어야함
- 잠재적 가치와 잠재적 위험이 공존하며 사회/경제적으로 성패를 좌우하는 핵심 원천이 될것으로 평가됨
- 분석을 통해서 나타난 복잡한 정보를 한눈에 볼수있도록 도표나 3D형태의 정보의 시각화가 이루어져야함

6. 반정형데이터란?

- 빅데이터의 유형 중 하나로 반정형데이터로는 웹로그, 소셜미디어 피드가 있다.

7. 빅데이터 플랫폼의 3가지 구성요소

- 하드웨어, 소프트웨어, 애플리케이션

2차시 빅데이터의 가치와 영향

1. 맥킨지가 제시한 빅데이터의 사회경제적 가치

산업의 투명성 증대	<ul style="list-style-type: none"> ● 빅데이터를 시기적절하게 관련 부문에 제공하는 것만으로 검색 및 처리 시간 절감
소비자 니즈 발견 트렌드 예측성과 향상을 위한 실험	<ul style="list-style-type: none"> ● 기업들이 더 많은 거래 데이터를 디지털형태로 축적하면서 보다 정확하고 상세한 성과 데이터 수집 가능 ● 자연적으로 일어나는 통제된 실험에 의해 일어나는 성과의 변동성과 근본적 원인 및 결과 분석에 데이터 이용 가능
소비자 맞춤형 비즈니스를 위한 고객 세분화	<ul style="list-style-type: none"> ● 기업들이 매우 구체적인 고객 분류를 통해 고객 니즈에 맞춘 맞춤형 서비스 제공
자동 알고리즘을 통한 의사결정지원과 대행	<ul style="list-style-type: none"> ● 정교한 분석에 의해 의사결정 향상, 위험 최소화, 가치 있는 인사이트 발굴 가능
비즈니스 모델, 상품, 서비스 혁신	<ul style="list-style-type: none"> ● 새로운 상품·서비스 개발, 기존 상품·서비스 향상, 새로운 비즈니스 모델 설계 가능

2. 빅데이터의 사회·경제적 측면에서의 의미

천연자원적 (Natural Resources)	<ul style="list-style-type: none"> ● 데이터에 내포된 가치와 가능성에 대해 주목 ● 사회적으로 현안과 위험을 해결할 수 있는 잠재력에 기대 ● 새로운 경제적 가치의 원천으로 활용 ● 새로운 원유, 데이터 골드러쉬, 데이터 금맥 찾기 (Data Mining)
새로운 재난적 (Natural Disasters)	<ul style="list-style-type: none"> ● 정보의 범람으로 기회 파악과 규정준수 어려움 ● 늘어나는 데이터로 인해 현 상태를 유지하는데 ICT예산이 사용되어 혁신을 위한 새로운 동력에 투자가 어려움 ● 데이터 처리의 낮은 응답 속도가 기업의 생산성 저하로까지 이어질 우려 ● 데이터 토네이도 (Data Tornado), 데이터 홍수(Data Deluge)
산업적 도구적	<ul style="list-style-type: none"> ● 데이터의 효율적 관리와 분석을 통해 기업의 경쟁 우위 확보 ● 데이터를 신속하게 처리해 실시간 의사결정에 지원 ● 데이터 분석 역량이 기업 경쟁력을 좌우 ● 데이터 산업 혁명(Industrial Revolution)이라 불리기도 한다.

3. 각 부문에 미치는 빅데이터의 영향

경제, 기업 부문	<ul style="list-style-type: none"> ● 소비자 행동 패턴 분석 및 시장 동향 예측을 통해 기업의 기존 비즈니스 프로세스의 문제점 파악 후, 개선 및 수정에 활용 ● 기업의 기존 ERP, SCM 등 생산 전반의 데이터를 활용하여 <u>원가 절감</u> 및 <u>제품 차별화</u>에 활용 ● <u>소비자 행동패턴의 분석</u>
공공 부문	<ul style="list-style-type: none"> ● 기상 데이터, 인구 데이터, 각종 통계지표 등 공공 데이터를 수집하고 이를 토대로 전반적인 재난 및 재해에 대한 <u>정보 도출</u>에 활용 ● 시스템 다이내믹스, 복잡계이론 등과 같은 분석을 통해 미래전략 수집에 활용

사회 부문	<ul style="list-style-type: none"> ● 빅데이터 분석을 통해 사회적 약자를 위한 활동 수행에 활용 ● 빅데이터 분석을 통해 다양한 사회적 기회 창출에 기여
기타 부문	<ul style="list-style-type: none"> ● 다양한 통찰을 통해 사회의 여러 가지 문제를 해결하는데 새로운 돌파구 발견 가능성 ● 유명인사들이 빅데이터를 활용하는 사례가 점차 증가하고 있음.

4. 기업에 있어 데이터란?

- 자본이나 노동력과 거의 동등한 레벨의 경제적 투입자본으로 비즈니스의 새로운 원자재 역할을 수행한다.

5. 빅데이터의 가치측정 방식

데이터 활용 방식	<ul style="list-style-type: none"> ● 데이터는 본연의 목적 외에도 활용되기도 하며, 이를 통해 새로운 가치를 창출할 수 있다. * <i>전기자동차 같은 경우는 다른 자동차들에 비해 배터리 정보(충전 및 충전시간 등)에 관한 정보가 중요하기에 해당 정보를 통해 기업들에서는 새로운 가치 창출의 방편이 될 수 있다.</i>
가치 창출 방식	<ul style="list-style-type: none"> ● 데이터를 창의적으로 조합하게 되면서 기존 방식으로는 절대 풀 수 없었던 문제를 해결하는데 큰 영향을 줄 수 있다.
분석 기술 발전	<ul style="list-style-type: none"> ● 과거 가치 없다고 판단한 데이터들도 새로운 데이터분석기법들의 등장으로 거대한 가치가 부여될 수 있다. ● 최근 클라우드 분산 컴퓨팅으로인해 저렴한 비용으로 데이터 분석이 가능해지면서 점점 그 활용도도 증가하고 있다.

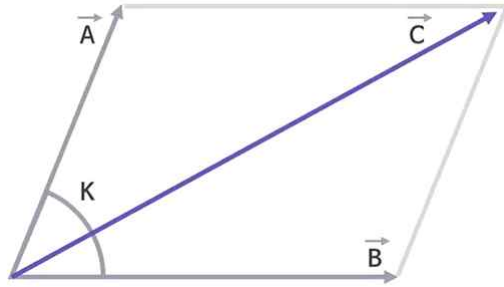
6. 빅데이터 가치산점을 위해 확보해야 하는 조건(3V)

데이터 량(Volume)	<ul style="list-style-type: none"> ● 수집된 대용량의 데이터를 수용할 수 있는 대용량성 확보 필요
다양한 형태(Variety)	<ul style="list-style-type: none"> ● 정형, 비정형, 반정형 등 다양한 형태의 데이터를 처리하고 활용할 수 있는 적응성 필요
빠른 생성 속도 (Velocity)	<ul style="list-style-type: none"> ● 데이터를 빠르게 처리 및 확보하여 활용할 수 있는 실시간성 필요

7. 새로운 부가가치 창출을 위해 데이터 분석을 통해 확보해야 하는 것

- 데이터의 다양한 형태(Variety) 측면의 적응성을 확보해야함.

8. 빅데이터의 3V(특징요소)를 통한 새로운 부가가치 창출의 효율적인 모델



- 대용량성(A)이 커질수록 적응성(B)도 같이 높아지는 상황에서 빅데이터의 부가가치의 효율성(C)이 높아짐(벡터각 $K < 90^\circ$)
- 빅데이터를 통해 새로운 부가가치 창출을 하기 위해서는 데이터의 다양성에 초점(focus)을 맞추는 것 보다는 다양한 데이터를 처리하고 활용할 수 있는 적응성을 확보하는 것이 더 중요하다고 할 수 있다.
- 데이터의 대용량성이 적응성을 높이는 상황

3차시 빅데이터 비즈니스 모델

1. 비즈니스 모델

- 빅데이터 비즈니스 모델을 판별하는 주요 기준은 수익성이다.

2. 수익성 모델

- 수익을 창출하는 모델로 돈이나 기타 재산을 벌어들일 수 있도록 하는 경영방식이다.
- 빅데이터 비즈니스 모델을 판별하는 주요 기준

3. 비즈니스 모델의 성공적 개발을 위한 고려사항

제품·서비스 제공 고객	● 제품과 서비스를 제공할 고객에 대한 세밀한 조사
제품과 서비스의 유형	● 고객에 제공한 제품과 서비스에 대한 검토
거래형태	● 방문판매에서 인터넷 등을 활용한 판매를 포함한 거래형태를 결정
차별화된 서비스 전략	● 서비스를 제공하는 차별화 전략에 대한 문제 고려
적절한 기반기술 검토	● 기반으로 하는 정보통신기술(ICT)에 대한 검토

4. 빅데이터 판매 비즈니스 모델 유형

원시 판매 비즈니스	● 가공되기 전의 데이터 판매
가공 판매 비즈니스	● 원시 데이터를 분석·가공하여 필요한 부분만 추출한 데이터 판매
분석 판매 비즈니스	● 원시 데이터를 분석한 데이터 판매

4차시 빅데이터 활용 전략

1. 가트너의 데이터 경제 시대의 4단계

단계	데이터 종류	방법
1단계 : 사일로(저장)	<ul style="list-style-type: none"> 조직의 독자적인 데이터 생성, 저장 중심 단계 외부 데이터는 인터넷을 통해 수집(검색) 가능 데이터의 신뢰성과 품질제고 노력 필요 	생성, 저장, 수집(검색)
2단계 : 교환(공유)	<ul style="list-style-type: none"> 기업의 데이터를 외부 기관들과 상호 교환하는 단계 (1:1 또는 1:N의 공유, 연계) 	연계, 공유
3단계 : 데이터풀(통합)	<ul style="list-style-type: none"> 특정한 활동이나 목적을 위해 모인 연합, 그룹, 클럽들이 상호 협력과 공유의 장 형성(집단) 표준화된 데이터풀(Data Pool)의 연계를 통해 국경을 초월한 정보 교환과 상호이용 가능 	참여, 협력
4단계 : 공유데이터 (공동창출)	<ul style="list-style-type: none"> 오픈 방식 플랫폼을 통한 데이터 공유 상호 협력과 참여를 통해 공동의 자원 창조 	오픈, 창조

2. 각 나라별 빅데이터 전략

미국	빅데이터 연구개발 이니셔티브 (빅데이터의 가능성을 파악하고 정부 주도의 전략 추진)
영국	오픈 데이터 전략 (정보공유·활용을 위한 데이터 개발 중심의 정책 추진)
일본	Active Japan ICT 전략 내 일부 포함 (사회현안 해결을 위한 데이터에 대한 가치 재평가)
한국	빅데이터를 활용한 스마트 정부 구현 (데이터의 창조적 활용을 통한 스마트 강국 실현)

3. 맥아피와 브린올프손의 기업이 빅데이터를 성공적으로 활용하기 위한 다섯 가지

- ① 기업문화(Company Culture)
- ② 의사결정(Decision Making)
- ③ 역량관리(Talent Management)
- ④ 기술도입(Technology)
- ⑤ 리더십(Leadership)

4. 빅데이터 수용 4단계

- 교육 > 탐색 > 시험 > 실행

5. 성공적인 빅데이터 활용을 위한 3대 요소

지원	활용할 수 있는 빅데이터를 발견
기술	빅데이터 프로세스와 신기술의 이해
인력	데이터 사이언티스트(Data Scientist) 역량 향상

6. 존 라우저의 데이터 사이언티스트(Big Data Scientist)가 갖춰야할 6가지 자질

- ① 수학(통계) 능력
- ② 공학(IT기술과 엔지니어링) 능력
- ③ 스토리텔링 등 시각화 능력
- ④ 비판적 시각
- ⑤ 커뮤니케이션 능력
- ⑥ 호기심과 개인의 행복

5차시 비즈니스 프로세스와 의사결정

1. 의사결정 단계

- 탐색 > 설계 > 선택 > 수행

탐색 (Intelligence)	● 필요한 데이터를 수집, 문제를 발견 및 새로운 기회를 파악하는 단계
설계 (Design)	● 탐색된 문제를 구조적으로 연구, 의사결정의 대안을 만들고 예상결과를 평가하는 단계
선택 (Choice)	● 비교, 평가된 대안을 분석하여 최적의 대안을 선택하는 단계 ● 미래의 불확실성과 위험을 감소시킬 수 있는 방안을 선택하는 단계
수행 (implementation)	● 대안의 실행과 피드백 단계

2. 비즈니스 인텔리전스(BI : Business Intelligence)란?

- 기업이 보유하고 있는 수많은 데이터를 정리하고 분석하여 전략적 의사결정에 활용하는 일련의 프로세스
- 관점에 따라 고객, 제품, 서비스, 운영, 공급자, 파트너에 대한 개별 정보 및 모든 관련 거래 데이터를 모으고 관리하며 분석하는 것으로 정의
- 의사결정에 대한 지원에 사용되는 기술, 프로세스, 스킬, 응용프로그램 등을 모두 포함하는 포괄적인 개념

3. 프로세스란?

- 한 개 이상의 입력을 통해서 가치 있는 산출물로 전환하는 활동
- 커다란 프로세스는 한 개 이상의 부프로세스를 가지며, 이 부프로세스는 여러 개의 활동들을 가진다.
- 활동은 더 작고 간단한 직무로 나누어진다.
- 직무는 단순 단계로 구성되어 있다.

4. 비즈니스 프로세스 구성요소

활동 (Activity)	한 가지 형태의 자원과 정보를 다른 형태의 자원과 정보로 변환
의사결정 (Decision Making)	여러 대안 중에서 하나의 행동을 고르는 일
역할 (Role)	처리의 집합
자원 (Resource)	역할이 지정된 사람, 시설, 혹은 컴퓨터 프로그램
저장소 (Storage)	비즈니스 기록들의 모음, 저장공간

6차시 마케팅 애널리틱스

1. 고객데이터 유형

고객 데이터	업무상 데이터, 신체 데이터, 보안 데이터, 개인상세 데이터 등
업무상 데이터	웹사이트나 이메일 리스트, 소셜미디어 가입 시에 필요한 소비자들이 통상 공유하는 기초 데이터로 생일, 주소, 이메일 주소 등이 있다.
신체 데이터	키와 몸무게 등
보안 데이터	소득수준, 휴대폰 번호, 신용카드번호 등
개인상세 데이터	종교, 정치색, 신조, 성적기호 등

2. 애널리틱스(Analytics)란?

- 비즈니스의 당면 이슈를 기업 내·외부 데이터의 통계적·수학적인 분석을 이용하여 해결하는 의사결정 방법론
- 전략적, 전술적 운영적 비즈니스 의사결정문제를 통계적·수학적, 자료프로그래밍, 전문 분야 지식 기반 데이터분석을 이용하여 해결하려는 강력한 방법

3. 애널리틱스에게 필요한 데이터 분석 역량 3가지

- 수학적/통계적 역량
- 데이터 프로그래밍 역량
- 전문적 지식 역량

4. 마케팅 애널리틱스의 가격/프로모션의 결정에 대한 주요 고려사항

가격	<ul style="list-style-type: none"> ● B2C기업에 있어 가장 중요한 의사결정은 가격이며 애널리틱스 성공 기업은 공통적으로 가격 의사결정에 많은 노력을 투입하고 있다. ● 상품과 상황에 따라 고객들이 느끼는 가격민감도는 천차만별로 달라질 수 있다. ● 빅데이터를 이용하면 가격과 프로모션의 의사결정에 도움을 줄 수 있다.
----	---

5. 마케팅 애널리틱스 분석 대상

- 빅데이터와 결합해 사용되는 마케팅 분석은 마케팅 실적을 적절히 평가하고, 고객들의 구매 습관, 시장 트렌드와 고객니즈에 대한 통찰을 얻고, 증거 기반의 마케팅 결정을 내리는 데 도움을 받기 위해 진행된다.

7차시 생산운영 애널리틱스

1. 생산운영 애널리틱스란?

- 생산운영과 관련되어 있는 비즈니스 이슈를 기업 내·외부 생산운영관련 데이터의 통계적·수학적 분석을 이용하여 해결하는 해결책의 총칭

2. 생산운영 애널리틱스의 필요성

- 최근 제조, 금융, 유통 및 다양한 산업분야에서 이러한 대용량 데이터분석을 통해 새로운 분석능력을 가지고 운영의 효율성을 증대시키고 기업의 경쟁력을 발전시킨 사례가 증가하

면서 새로운 혁신적 패러다임으로 부각되고 있다.

- 비즈니스 환경과 제조회업에서는 빅데이터 개념을 새롭게 확장할 필요가 있다.
- 데이터 자체의 크기나 형태보다는 다양한 데이터를 복합적으로 분석하여 기존에 파악하지 못한 새로운 분석력을 가지고 이를 바탕으로 생산성을 높이는 일련의 작업과 인력 그리고 비즈니스 프로세스를 통합한다는 의미로 해석할 필요가 있다.

3. 제조실행시스템(MES)란?

- 빅데이터를 활용할 수 있는 대표적인 정보시스템
- 생산관리, 원가관리, 설비관리, 품질관리, 생산계획, 자재조달 등의 기능들이 관련 데이터에 의해 처리되는 데이터 중심적 시스템

4. 제조실행시스템(MES) 기능별 수집 및 활용 데이터 종류

생산관리	생산능력 및 설비 수, 생산정보, 공정위치, 생산량 및 작업시간 등
원가관리	가동시간, 재료사용량, 에너지사용량 및 재고 등
설비관리	가동상태, 고장시간 및 비가동 내역 등
품질관리	운전조건, 복구시간, 불량원인 및 불량 수 불량 내용 등

5. 맥켄지의 글로벌 제조회업체들의 가치사슬 단계별 빅데이터 활용 분야

R&D 및 설계 단계	정보시스템이 수집한 데이터 및 정보는 하나의 집합체로 통합 관리하여 통찰력 있는 제품생산주기관리가 가능해짐
공급 사슬 단계	수요 예측, 실시간 재고 현황 관리가 가능하여 운영 품질을 높일 수 있음
생산 단계	사물 인터넷, RFID 센서기술 등을 통한 데이터를 분석하여 제조 품질 향상 및 설비에 대한 예측 정비 실현 등에 기여 가능

6. 제조회업에서 사용되는 대표적인 빅데이터 유형과 활용

구분	주요유형	예시
제조 장비 데이터	장비로그 데이터(비정형)	● 반도체 장비에서 생성되는 이벤트 로그 데이터
운영 통합 데이터	세일즈-마케팅-물류 통합 데이터(정형)	● 고객 구매(POS) 데이터, 주문 데이터, 물류 및 생산 통합 데이터
고객 경험 데이터	제품 사용 후기(비정형), 실시간 유입되는 제품 사용정보(정형)	● 고객 댓글, 사용 후기, 애프터서비스 데이터

7. 로그분석 방식

프로세스 마이닝 (Process Mining)	로그로 이뤄진 작업이 일어난 시간과 형태를 분석하여 작업의 상관관계를 분석하는 로그 데이터 분석 방식
텍스트 마이닝 (Text Mining)	텍스트로 이뤄진 로그 데이터에서 의미 있는 데이터 포인트를 파악하여 정형화하여 필요한 정보를 추출하는 데이터 분석 방식 * 자동차의 센서 정보를 취합해 리콜 여부를 결정하거나 제품 개발에 활용되는 등의 사례도 있다.

8. 공급 사슬 단계

제조업체들이 판매 촉진 관련 데이터, 제품 출시 데이터, 재고 데이터 등을 통합하여 분석함으로써 수요를 예측하거나 실시간으로 재고를 관리할 수 있어 운영품질을 높이는 효과를 볼 수 있는 가치사슬 단계

8차시 회계 애널리틱스

1. 전통적 회계절차 6단계

①거래발생 > ②분개장 > ③원장 > ④시산표 > ⑤수정분개 > ⑥회계보고서

2. 회계데이터 발생 절차

거래상태 > 분개 > 전기 > 시산표 정리 > 수정분개 > 회계보고서

3. 재무상태표 데이터란?

- 시점까지의 기업의 자산, 부채, 자본 데이터를 통합 정리하여 정보이용자에게 제시하는 데이터
- 회계 절차를 통해 도출된 데이터는 각 구성 요소별로 재무제표에 기재되어 활용할 수 있는 형태로 제공됨.
- 회계보고서 데이터

4. 매출상태표 데이터 분류

- 자산 = 부채 + 자본

자산	매출채권, 미수금, 미수수익, 선급금, 선급비용, 대여금, 선대금, 가지급금
부채	매입채무, 미지급금, 미지급비용, 선수금, 선수수익, 차입금, 예수금, 가수금
자본	자본금, 자본잉여금, 이익잉여금, 자본조정

5. 재무정보

- 자금 조달과 운용의 전반적인 활동을 의미한다.

6. 재무정보의 유형 4가지

- ① 안정성: 기업의 재무상태가 얼마나 양호한지를 측정
- ② 활동성: 기업자산이 얼마나 효율적으로 사용되고 있는가를 측정
- ③ 수익성: 기업의 이익 여부를 판단
- ④ 성장성: 기업의 미래 가치를 측정

9차시 재무/인적자원 애널리틱스

1. 재무관리란?

- 기업의 활동 중 자금과 관련한 활동을 운영하고 관리하는 것
- 자금 조달과 운용에 관한 전반적인 활동에 관한 것
- 재무계획과 재무통제로 구성

2. 재무계획 분류

- 자본의 흐름과 관련한 자금관리와 함께 이익계획과 이익통제의 실행을 의미

이익계획	설비투자계획, 제품개발 계획, 판로개척 계획
자금계획	설비자금계획, 제품개발, 자금계획

3. 재무정보란?

- 자금의 조달과 운용에 관한 전반적인 활동정보
- 도출된 회계정보를 바탕으로 경영 의사결정 활동을 위한 정량적 시사점을 도출할 수 있다.

4. 재무정보의 4가지 유형

안정성	기업의 재무상태가 얼마나 양호한지 측정하는 유형
활동성	기업자산이 얼마나 효율적으로 사용되고 있는가를 측정하는 유형
수익성	기업의 이익여부를 판단하는 유형
성장성	기업의 미래 가치를 측정하는 유형

5. 재무 애널리틱스의 절차

- 회계데이터 발생(저장)→회계보고서 작성(처리)→재무(비율)정보 발생(분석)→재무계획 재무통제(의사결정)이다.

6. 인적자원관리 업무와 데이터

업무	세부 업무	데이터
조서관리	분업, 부분화, 권한위양	직급, 호봉테이블, 부문별 이력
인사관리	경력관리, 이직관리, 승급관리, 상벌관리	인사기록카드, 인사발령이력, 조직원정보, 조직원이력
급여관리	기본급/성과급 관리, 임금경정, 보험관리	급여, 상여금, 연장근무, 4대 보험
복리후생관리	연금, 휴가, 상병, 체육시설, 장학금, 건강	조직원정보, 조직원이력, 시설관리, 건강관리표, 4대 보험
채용관리	직무관리, 모집관리, 선발관리, 배치관리	이력서, 연봉/직급 테이블, 조직원정보
교육훈련관리	교육, 훈련	교육참가이력, 시험결과, 성과평가, 승진이력
노사관리	교섭, 협약, 양자간협약	노동조합정보, 계약서, 법률문서, 탄원서, 교섭이력, 파업이력

7. 인적자원 애널리틱스란?

- 인적자원 데이터는 기업전략에 따른 인적자원계획 수립을 위한 애널리틱스 활동으로 연결
- 인적자원계획, 인적자원 모집 및 선발, 퇴직 및 이직관리 등이 있다.

8. 유동비율

- 유동비율은 유동자산/유동부채로 산출된다.

10차시 빅데이터 분석 기획의 이해

1. 기획과 계획의 차이

① 기획

- 왜 그 일을 해야 하는 가를 정의하는 것
- 무엇을 추진할 것인가를 결정하는 것

② 계획

- 어떻게 그 일을 할 것인가를 결정하는 것
- 목표달성을 위해 구체적인 절차를 정하는 것

2. 빅데이터 분석 기획이란?

- 목표달성 최적화를 위해 의사결정과 실행 과정에 필요한 정보와 인사이트를 과학적 분석을 통해 제공하는 분석체계를 수립하는 것을 말한다.

3. 빅데이터 분석 기획 단계 중 분석단계에서 타당성 분석의 평가단계

①경제적 타당성

②데이터 타당성

③기술적 타당성

④운영적 타당성과 같은 다각적인 타당성 분석이 수행

4. 빅데이터 분석 예산 수립에 고려할 요소

①하드웨어 시스템예산

②소프트웨어 시스템 예산

③외부 컨설팅 비용

④데이터 확보 비용 등

11차시 빅데이터 분석 및 기획 접근법

1. 데이터분석 과제란?

①과제란 해결해야할 문제나 이슈

②문제란 기대상태와 현재상태의 차이

③분석 과제는 분석활동의 대상이 됨.

2. 수요기반 분석 과제

- 선행적으로 문제해결에 대한 시나리오를 먼저 정의한다.
- 활용되는 데이터는 문제 해결이 가능하도록 해주는 실행동인으로서의 역할을 한다.
- 비즈니스 상의 이슈나 문제들의 해결방안에 초점을 맞춘다.

3. 수요기반 분석과제도출

- 선행적으로 문제해결 시나리오 먼저 정의 후 적합한 데이터 및 분석기법을 찾아 활용해야 함
- ‘실행동인’ 역할 수행
- 내·외부 데이터의 수집, 가공, 분석을 통하여 수행하는 일련의 접근방법
- 조직의 핵심성과지표(KPI)의 수준을 바람직한 방향으로 강화시키거나 또는 부정적으로 악화시키는 요인이 되기도 한다.

4. 데이터주도 분석과제 도출

- 특정 이슈에 대한 선행적 정의 없이 분석기법을 활용함
- ‘추진동인’ 역할 수행

12차시 통계 기반 데이터 분석

1. 통계학이란?

- 자연 및 사회현상에 나타나는 다양한 상황이나 측정값들을 요약하여 표현하는 것
- 기술통계(Descriptive Statistics)와 추론통계(Inferential Statistics)가 있다.

기술통계	그래프
추론통계	추정, 가설검정, 회귀분석, 상관분석 등

2. 추측통계학이란?

- 모집단의 일부인 표본의 특성을 이용하여 모집단의 특성을 추정하거나 가설을 검정하는 방법을 다룬 통계분야

* 모집단(population))이란? 조사대상이 되는 모든 개체들의 전체 집합

3. 측정과 척도

- 측정 : 특정하게 명시된 규정에 의해 수치나 혹은 다른 기호들을 통해 조사한 대상물의 특성을 기록한 것
- 척도 : 측정대상이 가지는 고유한 특성을 기록하는 것은 측정 대상물의 유형에 따라 적절하게 기록해야 하며, 이때 적절하게 기록하는 기준을 의미

척도	기본특성	일상적인 활용사례
명목	대상을 확인, 분류	주민등록번호, 운동선수 유니폼 번호 등
서열	대상의 상대적 순서위치	품질순위, 팀별 순위, 결승선 통과순위 등
등간	비교대상들간 차이, 크기 등	온도계의 온도 등
비율	절대영점이 존재, 척도값 비율을 계산하여 이용	길이, 무게 등

4. 가설의 종류

- ①귀무가설
- ②대립가설
- ③연구가설

5. 가설검증이란?

- 모집단의 모수에 대한 가설을 설정하고 표본으로부터 조사한 결과에 따라 그 가설을 선택할 것인지를 통계적으로 결정하는 분석방법
- 귀무가설을 기각함으로써 대립가설을 채택하기 위한 의도록 실시함
- 가설검증에는 귀무가설, 대립가설이 있다.

귀무가설	영가설(Null Hypothesis)
대립가설	연구가설(Alternative Hypothesis)

13차시 마이닝과 머신러닝 기반 데이터 분석

1. 데이터마이닝이란?

- 대용량의 데이터로부터 자동 또는 반자동적인 방법을 통하여 의미 있는 패턴, 규칙, 관계를 찾는 방법
- 데이터마이닝 기법은 크게 정형 데이터 분석과 비정형 데이터 분석으로 나뉜다.

정형 데이터 분석	연관관계분석 기법, 의사결정나무 기법, 인공신경망 기법, 사례기반추론, 군집분석 기법
비정형 데이터 분석	웹 문서, 소셜 데이터를 주로하는 텍스트 마이닝, 웹 마이닝, 오피니언 마이닝, 소셜 네트워크 분석

2. 데이터마이닝의 특징

- 기업들의 데이터베이스 필요성의 증가
- 사용자의 경험이나 판단을 배제하고, 전적으로 데이터에 기반하여 지식과 패턴을 추출
- 기업 업무와 생상의 효율성 증대
- 활용분야는 카드사의 사기 발견, 금융권의 대출 승인, 투자분석 등이 있다.

3. 데이터마이닝 기법의 종류

- ①사례기반추론
- ②군집분석
- ③인공신경망

14차시 비정형(텍스트) 마이닝

1. 비정형 데이터란?

- 정형화 되지 않은 데이터
- 미리 정의된 데일 모델을 가지고 있지 않은 데이터
- 아주 많은 양의 데이터를 가지고 있으면서 구조와 형태가 다르고 정형화 되지 않는 전자 문서, 영상, 음성

2. 비정형 데이터 분석 종류

비정형 데이터 분석	웹 문서, 소셜 데이터를 주로하는 텍스트 마이닝, 웹 마이닝, 오피니언 마이닝, 소셜 네트워크 분석
-----------------------	--

3. 웹 마이닝이란?

- 데이터 마이닝 기술의 응용분야로서 인터넷을 통해 웹 서비스를 이용하면서 웹에서 패턴을 발견 하는 것
- 웹 사용 마이닝, 웹 콘텐츠 마이닝, 웹 구조 마이닝 등이 있다.

4. XML이란?

- 인터넷에 연결된 시스템 끼리 데이터를 쉽게 주고받을 수 있도록 지원하는 웹 언어 체계를 말한다.

5. 텍스트 마이닝이란?

- 대규모 문서에서 정보를 추출하거나, 연계성을 파악 분류, 군집화 하는 것으로 대규모의 비구조적인 문서 데이터에서 의미 있는 정보를 추출하는 것을 뜻한다.

6. 소셜데이터분석 기법

- ①텍스트마이닝
- ②소셜네트워크분석
- ③웹마이닝

15차시 데이터 시각화 기법과 이해

1. 데이터 시각화란?

- 추상적인 데이터를 컴퓨터의 쌍방향적 시각 표현을 통해서 이해와 인식의 정도를 증폭하는 것으로 방대한 양의 데이터를 탐색하거나 이해할 때 가장 좋은 방법이다.

2. 데이터 시각화의 절차

1단계	데이터의 획득
2단계	데이터 구조화 및 분류
3단계	관심 데이터 추출
4단계	통계적인 방법 또는 데이터마이닝 기법 적용

5단계	바 그래프, 리스트 또는 트리 등의 기본적 시각모델 선택
6단계	보다 명확하게 재정의, 매력적 표현으로 개선
7단계	데이터 변경 또는 보여지는 특질을 조작하는 방법 추가, 상호작용

3. 프로그래밍 방법

-데이터 시각화 방법 중 프로그래밍 언어를 이용하여 새로운 데이터 조작하는 방법이다.

4. 시계열 데이터의 시각화 사례

- 스캐터 플롯, 라인 차트, 계단식 차트

16차시 빅데이터 플랫폼과 관리

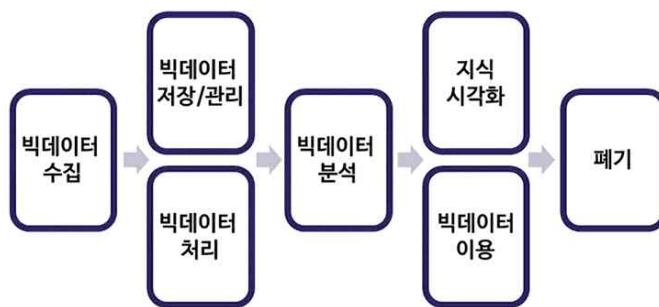
1. 빅데이터 플랫폼이란?

- 다양한 데이터 소스에서 수집한 데이터를 분석하여 지식을 발굴하고, 이를 기반으로 지능화된 서비스를 제공하는데 필요한 ICT 환경을 말한다.
- 빅데이터로부터 지식을 얻어 활용하기까지는 여러 단계와 단계별 수많은 기술이 활용된다.

2. 빅데이터 처리 플랫폼이란?

- 다양한 데이터 소스로부터 데이터 수집, 저장/관리 기술, 처리·분석 및 지식 시각화를 통한 지식 이용까지 각 단계를 지원하는데 필요한 공통 소프트웨어
- 빅데이터 수집, 빅데이터 저장관리, 빅데이터 처리가 이뤄진다.

3. 빅데이터 분석 프로세스



빅데이터 수집	<ul style="list-style-type: none"> ● 데이터를 검색하여 수집하고 변환 과정을 통해 정제된 데이터를 확보하는 과정 ● 세부 절차로는 대상 데이터 선정 > 세부계획 수립 > 수집 실행 순으로 수집됨
빅데이터 저장/관리	<ul style="list-style-type: none"> ● 검색 수집한 데이터를 분석에 사용하기에 적합 ● 안전하게 영구적인 방법으로 보관하는 것 ● 대용량의 다양한 형식의 데이터를 고성능으로 저장하고, 필요 시 데이터 검색, 수정, 삭제 또는 원하는 내용을 읽어오는 방법 제공까지 포함 ● 빅데이터 전/후처리와 빅데이터 저장으로 분류

빅데이터 처리	<ul style="list-style-type: none"> ● 빅데이터에서 유용한 정보와 의미 있는 지식을 찾아내기 위한 데이터 가공이나 데이터 분석 과정을 지원하는 과정 ● 지속적으로 발생하는 스트림 데이터나 기존 저장소에 저장된 대규모 저장 데이터의 적시 처리를 지원
빅데이터 분석	<ul style="list-style-type: none"> ● 빅데이터로부터 의미 있는 지식을 얻고 효율적인 의사결정에 활용하기 위한 분석 과정 ● 효과적인 분석방법과 다양한 인프라가 필요함.
빅데이터 분석 시각화	<ul style="list-style-type: none"> ● 빅데이터 속에서 정보와 가치를 찾을 수 있게 직관적인 방법으로 표현한 기술 ● 분석한 결과를 어떻게 표현하느냐에 따라 직관이 달라지기에 중요함.
폐기	<ul style="list-style-type: none"> ● 데이터 분석을 위해 이용된 데이터를 삭제하는 단계

4. 빅데이터 수집

- 수집 대상 데이터 선정
- 수집 세부 계획 수립
- 데이터 수집 실행 : 데이터소스 위치에 따라 내부 데이터 수집과 외부 데이터 수집으로 구분

내부 데이터 수집	<ul style="list-style-type: none"> ● 주로 자체적으로 보유한 내부 파일 시스템, DBMS, 센서 등에 접근하여 수집 ● 대표적인 수집 방법은 ETL(Extraction Transformation Loading)이 있다.
외부 데이터 수집	<ul style="list-style-type: none"> ● 인터넷으로 연결된 외부에서 수집 ● 대표적인 수집 방법은 크롤링 엔진(Crawling Engine)이 있다.

- 빅데이터 변환/통합

5. 빅데이터 폐기

- ① 빅데이터 폐기단계란 데이터 분석을 위해 사용되었던 데이터를 사용 후 삭제하는 단계이다.
- ②정보의 가치가 없는 데이터들은 이용목적을 달성 후 지체 없이 폐기해야 한다.
- ③소프트웨어적으로는 데이터를 저장하는 장소에 다른 데이터를 여러 번 덮어쓰기(Overwriting)를 해서 데이터를 삭제하는 방법도 있다.
- ④데이터 분석을 위해 이용된 데이터를 삭제하는 단계이며 특히 개인정보와 같은 데이터이거나 또는 정보의 가치가 없는 데이터들은 이용목적을 달성 후 지체 없이 폐기해야 한다.

5. ETL(Extraction Transformation Loading)의 기능

- ①도메인 검증
- ②데이터 요약
- ③레코드 통합
- ④데이터키 값의 재구성

6. 빅데이터 전처리(Pre-Processing)란?

- 수집과 변환 과정에서 빅데이터 저장소에 적재하기 위하여 수집한 데이터를 필터링(filtering)하거나 유형 변환(transformation), 정제(cleansing) 등을 거치는 과정

7. Sqoop란?

- RDBMS와 HDFS간의 데이터를 연결해 주는 기능으로 SQL데이터를 Hadoop으로 로드하는 도구로 사용된다.

17차시 빅데이터 분석 도구

1. 빅데이터 분석 및 시각화 도구

- Excel, SAS, SPSS, R, Python

Excel	<ul style="list-style-type: none">● 마이크로소프트사에서 개발한 윈도 환경의 스프레드시트 프로그램● 다른 분석툴에 비해 사용이 비교적 쉽다.● '데이터' 리본메뉴에서 제공되는 다양한 방법을 마우스 클릭을 통해 사용할 수 있다.● 기존에 존재하는 데이터를 불러와 수정, 사용하는 방법이 있어 사용자의 상황에 따라 선택적 사용이 가능하다.
SAS	<ul style="list-style-type: none">● 고가의 라이선스가 필요한 프로그램● 공인되어 있는 대부분의 통계분석을 포괄하여 수행 가능● 매우 정밀한 결과 제공● 보고서 작성과 그래픽 작업도 가능● 데이터 입력 및 편집을 위한 DATA STEP 단계와 본격적인 데이터 분석을 위한 PROC STEP 단계가 있다.
SPSS	<ul style="list-style-type: none">● IBM에서 상용화한 사회과학 자료 분석을 위해 고안된 통계분석 전용 프로그램으로 라이선스가 필요하다.● 사용법이 간편하여 데이터 획득에서부터 리포팅(reporting)까지 전 과정을 메뉴와 대화 상자를 통해 수행 가능하다● 데이터 입력법으로는 스프레드시 입력법과 불러오기 방법이 있다.● 변수보기 창에서 변수의 측도를 지정해 주어야 한다.
R	<ul style="list-style-type: none">● 데이터 분석을 위한 통계분석 기법과 알고리즘, 데이터 시각화 기능을 지원하는 오픈 소스 프로그램● 빅데이터와 관련된 분석을 위한 툴(Tools)로 주목● 다양한 통계기법에 빠르게 적용
Python	<ul style="list-style-type: none">● 동적 타이핑(Dynamic typing) 범용 프로그래밍 언어로 다양한 플랫폼에서 사용 가능● 풍부한 라이브러리(모듈) 제공● 순수한 프로그램 언어 기능 외의 다른 언어로 쓰여진 모듈들을 연결하는 풀언어(Glue language)로 자주 이용 됨.

2. SAS의 확장편집기

- 사용자가 data step과 proc step과 같은 프로그램 명령문을 직접 입력하는 공간이다.