

학습자료 (빅데이터 쉽게 이해하기)

1차시. 빅데이터의 개요

*빅데이터 4가지 구성요소

- 규모(Volume)의 증가
- 다양성(Variety)의 증가
- 속도(Velocity)의 증가
- 복잡성(Complexity)

* 빅데이터의 개념

- 많은 데이터로 양적인 의미를 벗어나 데이터 분석과 활용을 포괄하는 개념이다.
- 대량의 정형 또는 비정형 데이터 세트 및 이러한 데이터로부터 가치를 추출하고 결과를 분석하는 기술이다.
- 기존의 관리 및 분석 체계로는 감당할 수 없을 정도의 거대한 데이터의 집합으로 대규모 데이터와 관계된 기술 및 도구를 모두 포함하는 개념이다.

*빅데이터 플랫폼의 애플리케이션 레벨

- 실제 빅데이터와 사용자가 만나게 되는 커뮤니케이션 레벨
- 인프라 및 데이터 처리기술 분야보다 상대적으로 낮은 진입 장벽
- 빅데이터 의미파악, 이해능력, 분석결과를 서비스화 할 수 있는 애플리케이션

*빅데이터의 의의

- 분석결과를 갖고 국가기관, 민간 등의 분석기획 담당자들이 자유롭게 이용할 수 있어야 한다.
- 빅데이터는 잠재적 가치와 잠재적 위험이 공존하며 사회 · 경제적으로 성패를 좌우하는 핵심 원천이 될 것으로 평가되고 있다.
- 빅데이터 분석을 통해서 나타난 복잡한 정보를 한눈에 볼 수 있도록 도표나 3D 형태의 정보의 시각화가 이루어져야 한다.

2차시. 빅데이터의 가치와 영향

*부문별 빅데이터의 영향

- 기업부문 : 원가의 절감
- 기업부문 : 제품의 차별화
- 공공부문 : 재난정보의 도출
- 기업부문 : 소비자행동 패턴의 분석

*데이터

- 자본이나 노동력과 거의 동등한 레벨의 경제적 투입자본으로 비즈니스의 새로운 원자재 역

할을 수행한다.

*빅데이터의 사회 · 경제적 의미(산업적 도구)

- 데이터 효율적인 관리와 분석을 통해 기업의 경쟁 우위 확보
- 데이터를 신속하게 처리해 실시간 의사결정에 지원
- 데이터 분석 역량이 기업 경쟁력을 좌우
- 데이터 산업 혁명(Industrial Revolution)

*데이터의 다양한 형태(Variety) 측면의 적응성 확보

- 새로운 부가가치 창출을 위해 데이터 분석을 통해 적응성을 확보해야한다.

*빅데이터를 통한 새로운 부가가치 창출

- 데이터의 다양성에 초점(focus)을 맞추는 것 보다는 다양한 데이터를 처리하고 활용할 수 있는 적응성을 확보하는 것이 더 중요하다
- 데이터의 대용량성이 적응성을 높이는 상황에서 부가가치 효율성이 높아진다.

3차시. 빅데이터 비즈니스 모델

*비즈니스 모델의 성공적 개발을 위한 고려사항

- 제품이나 서비스를 제공할 고객
- 제품과 서비스의 유형
- 거래형태

*빅데이터 판매 비즈니스 모델

- 원시 판매 비즈니스
- 가공 판매 비즈니스
- 분석판매 비즈니스

4차시. 빅데이터 활용 전략

*가트너의 데이터 경제 시대의 4단계

- 사일로(저장)
- 교환(공유)
- 데이터풀(통합)
- 공유데이터(공동창출)

*나라별 빅데이터 전략

- 영국-오픈 데이터 전략
- 미국-빅데이터 연구개발 이니셔티브
- 일본-Active Japan ICT 전략 내 일부 포함

- 한국-빅데이터를 활용한 스마트 정부 구현

*맥아피와 브린올프슨이 제시한 기업이 빅데이터를 성공적으로 활용하기 위해 필요한 다섯 가지 조건

- 리더십
- 역량관리
- 기술도입
- 의사결정
- 기업문화

*빅데이터 활용의 3대 전략요소

- 기술-자원-인력

*빅데이터 사이언티스트(Big Data Scientist)

- 수학·공학(IT기술과 엔지니어링) 능력
- 스토리텔링 등 시각화 능력
- 비판적 시각과 커뮤니케이션 능력

*가트너에서 제시한 데이터 경제 시대의 4단계중 사일로 단계

- 데이터의 신뢰성과 품질제고 노력 필요
- 조직의 독자적인 데이터 생성, 저장 중심의 단계
- 외부 데이터는 인터넷을 통해 수집 가능

5차시. 비즈니스 프로세스와 의사결정

*의사결정단계

- 의사결정단계는 탐색→설계→선택→수행의 순이다.

*비즈니스 프로세스 구성요소 중 자원

- 역할이 지정된 사람, 시설

*비즈니스 인텔리전스(BI)

- BI는 고객, 제품, 서비스, 운영 등의 개별 정보 및 모든 거래 데이터를 모으고 관리하며 분석하는 것을 말한다.

*프로세스

- 한 개 이상의 입력을 통해서 가치 있는 산출물로 전환하는 활동을 말한다.
- 활동은 더 작고 간단한 직무로 나누어진다.
- 직무는 아주 단순 단계로 구성되어 있다.

*비즈니스 프로세스의 구성요소

- 활동
- 의사결정
- 역할

6차시. 마케팅 애널리틱스

*고객정보 데이터

- 업무상 데이터 : 웹사이트나 이메일 리스트, 소셜미디어 가입 시에 필요한 기초 데이터
- 보안 데이터 : 소득수준
- 보안 데이터 : 휴대폰 번호, 신용카드번호

*애널리틱스(Analytics)의 데이터 분석 역량

- 데이터 프로그래밍 역량
- 수학적 역량
- 전문적 지식 역량

*마케팅 애널리틱스의 가격/프로모션의 결정

- B2C기업에서 있어 가장 중요한 의사결정은 가격이다.
- 고객들의 가격민감도는 상품과 상황에 따라 천차만별로 달라질 수 있다.
- 애널리틱스 성공 기업은 공통적으로 가격 의사결정에 많은 노력을 투입하고 있다.

*마케팅 애널리틱스의 분석 대상

- 마케팅 실적 평가
- 시장 트렌드
- 고객 니즈

7차시. 생산운영 애널리틱스

*제조실행시스템(MES)

- 빅데이터를 활용할 수 있는 대표적인 정보시스템이다.
- 생산관리의 종류에는 공정위치, 생산량 및 작업시간이 있다.
- 원가관리의 종류에는 재료사용량, 에너지사용량 및 재고가 있다.

*생산운영 애널리틱스

- 최근 제조, 금융, 유통 및 다양한 산업분야에서 이러한 대용량 데이터분석을 통해 새로운 분석능력을 가지고 운영의 효율성을 증대시키고 기업의 경쟁력을 발전시킨 사례가 증가하면서 새로운 혁신적 패러다임으로 부각되고 있다.
- 비즈니스 환경이나 특히 제조환경에서는 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.

***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

***제조장비에서 발생하는 빅데이터**

- 장비에서 생성되는 이벤트 로그데이터

***텍스트 마이닝**

- 고객이 제품에 느끼는 일종의 느낌이나 감정(sentiment) 등을 파악해 좀 더 구체적으로 고객의 니즈를 파악할 수 있는 기술
- 자동차의 센서 정보를 취합해 리콜 여부를 결정하거나 제품 개발에 활용하는 방식을 설명하는 기술

8차시. 회계 애널리틱스

***매주상태표의 데이터 분류의 자산**

- 매출채권
- 미수금
- 선급금

***재무정보**

- 자금의 조달과 운용에 관한 전반적인 활동을 의미

***재무정보의 유형중 활동성**

- 기업자산이 얼마나 효율적으로 사용되고 있는가를 측정

***전통적 회계절차의 6단계**

- 거래발생-> 분개장 -> 원장 -> 시산표 -> 수정분개

9차시. 재무/인적자원 애널리틱스

***재무 애널리틱스의 절차**

- 회계데이터 발생→회계보고서 작성→재무(비율)정보 발생→재무계획 재무통제

***유동비율**

- 유동자산/유동부채로 산출

*자금계획의 장기계획

- 설비자금계획
- 제품개발
- 자금계획

*이익계획의 장기계획

- 설비투자계획
- 제품개발 계획
- 판로개척 계획

*활동성

- 기업자산이 얼마나 효율적으로 사용되고 있는가를 측정하는 재무정보 유형

*인적자원관리의 업무

- 조직관리
- 인사관리
- 채용관리

*인적자원 애널리틱스

- 인적자원 계획
- 인적자원 모집 및 선발
- 퇴직관리

10차시. 빅데이터 분석 기획의 이해

*기획과 계획

- 기획은 왜 그 일을 해야 하는가를 정의하는 것
- 기획은 무엇을 추진할 것인가를 결정하는 것
- 계획은 목표달성을 위해 구체적인 절차를 정하는 것

*빅데이터 분석 기획의 정의

- 목표달성 최적화를 위해 의사결정과 실행 과정에 필요한 정보와 인사이트를 과학적 분석을 통해 제공하는 분석체계를 수립

*분석단계에서 타당성 분석의 평가단계

- 경제적 타당성
- 데이터 및 기술적 타당성
- 운영적 타당성

*빅데이터 분석 예산 수립

- 하드웨어, 소프트웨어 등 시스템 예산

- 외부 컨설팅 비용
- 데이터 확보 비용

11차시. 빅데이터 분석 및 기획 접근법

*데이터분석 과제

- 과제란 해결해야할 문제나 이슈를 의미한다.
- 분석 과제는 분석활동의 대상이 된다.
- 문제는 기대상태와 현재상태의 차이라고 정의할 수 있다.

*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.

*수요기반 분석 과제 도출

- 문제에 대한 해결이 가능하도록 해주는 하나의 “실행동인” 역할을 한다.
- 내/외부 데이터의 수집/가공/분석을 통하여 수행하는 일련의 접근방법이다.
- 조직의 핵심성과지표(KPI)의 수준을 바람직한 방향으로 강화시키거나 또는 부정적으로 악화시키는 요인이 된다.

*데이터 주도 분석과제 도출

- 선행적인 정의 없이 분석기법을 활용하여 과제 도출

12차시. 통계 기반 데이터 분석

*추측통계학

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

*추론통계

- 추정
- 가설검정
- 회귀분석
- 상관분석

*기술통계

- 그래프

*등간척도에 해당하는 자료

- 상대적 크기를 나타내는 것으로 절대 영점이 없다.
- 온도 등

***가설**

- 귀무가설
- 대립가설
- 연구가설
- 결론가설

13차시. 마이닝과 머신러닝 기반 데이터 분석

***데이터마이닝 기법**

- 사례기반추론(Case Base Reasoning)
- 군집분석(Cluster Analysis)
- 인공신경망(Neural Networks)

***데이터마이닝의 특징**

- 기업들의 데이터베이스 필요성의 증가
- 기업업무와 생상의 효율성 증대
- 활용분야는 카드사의 사기 발견, 금융권의 대출 승인, 투자분석

***정형데이터 분석을 다루는 기법**

- 인공신경망 기법
- 의사결정나무 기법
- 사례기반 추론

14차시. 비정형(텍스트) 마이닝

***소셜데이터 분석**

- 텍스트마이닝
- 소셜 네트워크 분석
- 웹마이닝

***비정형 데이터**

- 정형화 되지 않는 데이터이다.
- 미리 정의된 데이터 모델(구조)을 가지고 있지 않는 데이터이다.
- 전자 문서 속에서도 발견 된다.

***XML**

- 인터넷에 연결된 시스템 끼리 데이터를 쉽게 주고받을 수 있도록 지원하는 웹 언어 체계

***웹 마이닝**

- 웹 사용 마이닝은 분석 대상에 따른 웹 마이닝 분류 중 하나이다.
- 웹 콘텐츠 마이닝 분석 대상에 따른 웹 마이닝 분류 중 하나이다.
- 웹 구조 마이닝 분석 대상에 따른 웹 마이닝 분류 중 하나이다.

15차시. 데이터 시각화 기법과 이해

***데이터 시각화의 개념**

- 방대한 양의 데이터를 탐색하거나 이해할 때 가장 좋은 방법

***데이터 시각화의 단계**

- 획득→구조화→추출→마이닝→시각화→재정의→상호작용

***프로그래밍 방법**

- 데이터 시각화 방법 중 자신의 목적에 맞는 프로그래밍 언어를 이용하여 새로운 데이터 조작 방법을 적용하는 방법

16차시. 빅데이터 플랫폼과 관리

***빅데이터 전처리**

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

***빅데이터 분석의 순환 과정**

- 수집-저장-처리-분석-시각화-이용-폐기의 과정으로 이루어진다.

*** ETL의 기능**

- 도메인 검증
- 데이터의 요약
- 데이터기 값의 재구성

***내부 데이터 수집 방법**

- ETL(Extraction, Transformation, Loading)
- 다양한 소스 시스템으로부터 필요한 데이터를 추출(extract)하여 변환(transformation) 작업을 거쳐 저장하거나 분석을 담당하는 시스템으로 전송 및 적재(loading)하는 모든 과정을 포함

***빅데이터 폐기**

- 빅데이터 폐기단계란 데이터 분석을 위해 사용되었던 데이터를 사용 후 삭제하는 단계이다.

- 정보의 가치가 없는 데이터들은 이용목적 달성 후 지체 없이 폐기해야 한다.
- 소프트웨어적으로는 데이터를 저장하는 장소에 다른 데이터를 여러 번 덮어쓰기(Over writing)를 해서 데이터를 삭제하는 방법도 있다.

*빅데이터 플랫폼

- 빅데이터 처리 플랫폼은 빅데이터 수집, 저장/관리 기술, 시각화 기술 등을 적용하여 구현한다.
- 빅데이터로부터 지식을 얻어 활용하기까지는 여러 단계가 필요하고, 그 단계마다 수많은 기술이 활용된다.
- 빅데이터 플랫폼은 데이터를 수집해서 지식을 발굴하는 데 필요한 처리 플랫폼 기술, 컴퓨팅 기반을 제공하는 빅데이터 컴퓨팅 인프라 기술을 구성된다.

17차시. 빅데이터 분석 도구

*빅데이터 시각화 도구

- SAS
- SPSS
- R

*엑셀의 데이터 입력과 분석

- 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.

*R

- 통계분석이 가능하다.
- 오픈소스 프로그램이다.
- 다양한 통계기법이 빠르게 적용된다.

* SPSS

- 데이터 획득에서부터 리포팅(reporting)까지 전 과정을 메뉴와 대화 상자를 통해 수행 가능하다.
- 변수보기 창에서 변수의 측도를 지정해 주어야 한다.
- 데이터 입력법으로는 스프레드시 입력법과 불러오기 방법이 있다.

*텍스트마이닝

- 대규모의 비구조적인 문서 데이터에서 의미있는 정보를 추출하는 것을 뜻한다.

*SAS의 화면 구성 중 확장편집기

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