** ADsP 시험에 출제되는 핵심 내용 요약 정리

AIEDU 대표 : 김 계 철

001 데이터의 정의 ★

① 데이터의 존재적 특성

- 데이터는 객관적 사실로서 존재하는 모든 형태의 정보를 의미
- 개별 데이터는 단순히 객관적 사실을 나타내는 정보일 뿐, 의사결정이나 인사이트 도출에 부족함

  예 "서울의 오늘 기온은 25도입니다."

② 데이터의 당위적 특성

- 데이터는 추론, 예측, 전망, 추정 등 의사결정을 위한 근거로 기능

  예 "서울의 평균 기온이 5년간 상승했으므로 기후 변화 대응이 필요하다."

002 데이터의 유형 ★★

- 정성적 데이터와 정량적 데이터는 **데이터의 형태, 분석 방식, 결과 해석** 등에서 차이가 있다.

① 정성적 데이터(Qualitative Data)

- 데이터의 형태와 구조가 정해져 있지 않음
- 정성적 데이터는 **수치로 직접 표현할 수 없고**, 맥락과 주관적 해석에 의존
- 텍스트, 음성, 이미지, 동영상 등의 다양한 형식을 포함

  예 트위터, 페이스북에서 사용자의 게시글과 댓글

② 정량적 데이터(Quantitative Data)

- 정량적 데이터는 **숫자, 기호, 또는 도형**으로 표현되며, 정밀한 계산이 가능
- 계산과 분석 용이
- 데이터양이 증가하더라도 저장, 검색, 분석이 효율적

  예 풍속(10m/s), 강수량(50mm)

003 암묵지 vs 형식지★

- 암묵지와 형식지는 지식의 표현 방식, 전달 방법, 저장 가능성 등 여러 기준에 따라 구분

① 암묵지(Tacit Knowledge)

- 학습과 체험을 통해 개인에게 습득된 지식으로, 겉으로 명확히 드러나지 않는 무형의 지식을 의미함.
- 시행착오와 오랜 경험을 통해 체득된 지식으로, 개인의 행동과 직관에 내재되어 있음.
- 개인에게 체화 되어 외부로 표출하거나 전달하기 어려움.
- 문서화 또는 구조화가 어렵고, 주로 관찰, 시연, 경험적 학습을 통해 전수됨

② 형식지(Explicit Knowledge)

- 교과서, 매뉴얼, 비디오, 데이터베이스(DB)와 같이 형상화되고 구조화된 지식을 의미함.
- 유형의 자료로 표현되어 있어 누구나 쉽게 이해하고 사용할 수 있음.
- 지식의 전달과 공유가 쉬움.
- 문서화, 데이터베이스화, 시스템화를 통해 조직 내 활용 가능

③ 콘텐츠와 지식의 관계

- 콘텐츠는 다양한 매체(동영상, 웹사이트, SNS 등)를 통해 광범위한 대상에게 전달 가능
- 콘텐츠는 지식이 특정 상황이나 맥락에 맞게 변형되고 활용된 형태라고 할 수 있다.
- 콘텐츠는 디지털화된 지식의 형태로, 암묵지와 형식지를 모두 표현할 수 있다.

004 암묵지와 형식지의 상호 작용 ★★

- 암묵지와 형식지의 상호작용은 지식을 창출하고 공유하며, 이를 조직의 학습 및 혁신으로 연결하는 핵심 과정이다. 이 과정은 다음 네 가지 단계로 구분

① 공통화(Socialization)

- 암묵지(개인의 경험과 직감)를 다른 사람과 공유하여 새로운 암묵지를 생성하는 과정

② 표출화(Externalization)

- 암묵지를 형식지로 전환하는 과정. 개인의 경험, 직감을 문서화하거나 시각적 자료로 표현하여 공유 가능하게 만듦

③ 연결화(Combination)

- 기존의 형식지를 조합하거나 새로운 지식을 추가하여 더 발전된 형식지로 확장하는 과정.
- 데이터를 통합하고, 자료를 재구성하여 새로운 관점이나 정보를 생성함

④ 내면화(Internalization)

- 형식지를 학습하여 개인의 암묵지로 습득하는 과정
- 문서, 매뉴얼, 교육 자료 등을 기반으로 학습하여 실무에 적용하고 경험을 통해 내재화함

005 데이터와 정보의 관계 ★★

- 데이터의 당위적 특성인 근거의 기능에 주목하여, 데이터와 정보의 관계를 이해하기 위해 DIKW 피라미드를 정의할 필요가 있다.
- DIKW 피라미드란 Data → Information → Knowledge → Wisdom의 과정을 통해 데이터를 시작으로 최종적으로 지혜에 도달하는 과정을 나타냄

① Data(데이터)

- 가공되지 않은 순수한 수치나 기호로, 특정 맥락 없이 단순히 존재하는 상태

예 연필 가격 : A 마트 100원, B 마트는 200원

② Information(정보)

- 데이터를 가공하고 상관관계를 분석하여 의미를 부여한 상태
- 패턴을 인식하고, 데이터를 특정 맥락에서 해석 가능

예 "A 마트의 연필 가격이 B 마트보다 더 싸다."

③ Knowledge(지식)

- 서로 연결된 정보 간의 패턴을 이해하고, 이를 바탕으로 의사결정이나 예측을 도출한 상태
- 정보의 활용을 통해 실질적인 행동을 가능하게 함

예 "A 마트가 상대적으로 저렴하니, A 마트에서 연필을 사야겠다."

④ Wisdom(지혜)

- 근본 원리와 맥락에 대한 깊은 이해를 바탕으로 통찰적이고 창의적인 아이디어를 도출한 상태

예 "A 마트의 연필이 저렴하니, 다른 상품도 B 마트보다 저렴할 가능성이 높다.
따라서 앞으로 A 마트를 우선적으로 고려해야겠다."

006 데이터베이스 정의 ★

① 데이터베이스의 정의

- 데이터베이스는 동시에 복수의 적용 업무를 지원하며, 다수 이용자의 요구를 수용하고 저장 및 공급하기 위해 일정한 구조에 따라 **조직된 데이터의 집합이다**.
- **(독립성)** 응용프로그램 독립성은 데이터베이스가 **응용프로그램으로부터 독립되어** 설계되고 관리된다는 것을 의미

② 데이터베이스 관리시스템(DBMS) 정의

- DBMS는 데이터베이스를 효율적으로 관리하고 활용할 수 있는 **소프트웨어 도구로**, 데이터 검색, 저장, 보안, 장애 복구 등의 기능을 제공한다.
- **(중개 역할)** 데이터베이스와 사용자(또는 응용 프로그램) 사이에서 데이터를 요청하고 처리하는 역할을 수행
- 데이터 검색 및 수정
- 사용자 권한 관리 및 보안
- 데이터 무결성 유지 및 장애 복구

007 DBMS의 발전 과정 ★

- DBMS(Database Management System)는
  데이터베이스 기술의 발전과 함께 여러 세대를 거쳐 발전

① 1세대 : 네트워크 DBMS, 계층 DBMS

- 데이터 구조를 네트워크형 또는 계층형으로 구성
- 구조가 복잡하며, 데이터의 변경이나 확장이 어려움

② 2세대 : 관계형(Relational) DBMS

- 데이터를 테이블 형태로 구성하여 직관적이고 간결한 데이터 관리가 가능
- SQL(Structured Query Language)을 통해 데이터 관리와 검색이 용이함

③ 3세대 : 객체 지향 DBMS 및 객체 관계형 DBMS

- **객체 지향(Object-Oriented) DBMS :**
  - 멀티미디어 데이터와 같은 비정형 데이터를 처리하기 위해 등장
  - 같은 속성과 행위를 갖는 데이터를 클래스로 그룹화
  - 클래스 연산을 수행하기 위해 메소드(함수)를 정의함
  - 관계형 데이터 모델로 표현하기 어려운 복잡한 데이터를 효과적으로 관리
- **객체 관계형 모델(ORDBMS) :**
  - 관계형 DBMS에 객체 지향 DBMS의 장점을 통합한 모델
  - 기존의 관계형 모델을 확장하여, 다양한 데이터 구조를 지원

④ 4세대 : NoSQL DBMS

- 데이터 구조를 미리 정의하지 않으므로 비정형 데이터를 저장하고 처리하는 데 적합.
- 데이터의 분산 저장 및 대규모 데이터 처리에 유리함

008 데이터 유형 분류 ★★★

- 데이터는 구조화 정도와 저장 및 처리 방식에 따라 정형, 반정형, 비정형 데이터로 구분

① 정형 데이터(Structured Data)

- 정해진 형식과 구조를 가지며, RDBMS(Relational Database Management System)에 저장되는 데이터
- 고정된 필드와 스키마에 의해 데이터 구조가 정의
- SQL을 사용하여 데이터 검색과 관리가 용이함

② 반정형 데이터(Semi-Structured Data)

- 데이터의 속성(메타데이터)를 포함하며, 정형 데이터와 비정형 데이터의 중간 형태
- 고정된 스키마는 없지만, 태그나 키-값 구조를 통해 데이터의 구성이 정의됨
- 일반적으로 파일 형태로 저장됨

  예 HTML, JSON, XML 파일, 로그 데이터

③ 비정형 데이터(Unstructured Data)

- **정해진 형식이나 구조가 없으며**, 일반적으로 대용량 데이터로 취급됨
- 이미지, 동영상, 오디오 등 복잡한 형태의 데이터
- 분석 및 처리가 까다로우며, 특수한 도구나 기술을 필요로 함

예 이미지 파일, 동영상 파일, 소셜 미디어 데이터

009 데이터베이스 설계 ★

① 요구 조건 분석

- 데이터베이스의 목적과 사용자의 요구 사항을 분석

② 개념적 설계(E-R 모델)

- 엔티티와 관계를 정의하여 데이터베이스의 전체적인 구조를 설계

③ 논리적 설계(테이블 설계)

- E-R 모델을 기반으로 테이블 구조를 설계하고 스키마를 정의

④ 물리적 설계(데이터 구조화)

- 데이터베이스를 실제 환경에서 구현하며, 저장소 및 성능을 최적화

010 데이터베이스의 특징 ★★

① 통합된 데이터(Integrated Data)

- 여러 데이터가 중복 없이 논리적으로 통합

② 저장된 데이터(Stored Data)

- 컴퓨터 시스템에 저장된 상태로 유지되는 데이터

③ 공용 데이터(Shared Data)

- 여러 사용자와 응용 프로그램이 공동으로 사용할 수 있음

④ 변화되는 데이터(Changed Data)

- 데이터는 시간의 흐름에 따라 추가, 삭제, 갱신이 발생함

011 데이터웨어하우스(Data Warehouse)와 데이터 마트(Data Mart)★★★

- 데이터웨어하우스와 데이터 마트는
  데이터 관리와 분석을 위한 저장소라는 공통점이 있지만, 규모와 용도에서 차이가 있다.

① 데이터웨어하우스(Data Warehouse)

- 기업의 모든 데이터를 중앙 집중화하여 통합 저장하고, 의사결정을 지원하기 위한 대규모 데이터 저장소
- 통합된 데이터 : 여러 출처에서 수집한 데이터를 하나의 시스템에 통합
- 주제 중심적 : 조직 전체에서 사용할 수 있도록 데이터가 설계됨
- 시간 가변적 : 과거 데이터와 현재 데이터를 모두 포함하여 분석 가능
- 비휘발성 : 데이터를 저장한 후에는 변경하지 않으며, 분석을 목적으로만 사용
- 규모 : 대규모의 데이터를 다룸
- 사용자 : 기업의 전사적 의사결정을 담당하는 사용자

② 데이터 마트(Data Mart)

- 데이터웨어하우스의 데이터를 특정 부서나 비즈니스 영역에 최적화하여 제공하는 소규모 데이터 저장소
- 단일 주제 : 특정 부서나 비즈니스 영역(예 마케팅, 판매)에 초점
- 독립적 운영 가능 : 데이터웨어하우스 없이 독립적으로 운영되기도 함
- 규모 : 데이터웨어하우스보다 작고 간단함
- 사용자 : 특정 부서나 팀의 실무자

012 OLTP vs OLAP ★★

① OLTP(On-Line Transaction Processing)

- 네트워크 상에서 여러 사용자가 실시간으로 데이터 베이스의 데이터를 갱신하거나 조회하는 단위 작업을 처리하는 시스템
- 실시간 처리 : 초 단위 또는 실시간으로 데이터를 입력하고 수정
- 데이터 무결성 유지 : 정확성과 일관성이 중요한 시스템

예 은행에서 고객이 ATM을 통해 입출금을 요청하면 OLTP 시스템이 실시간으로 처리

② OLAP(On-Line Analytic Processing)

- 데이터베이스에서 수집된 데이터를 분석하여 의사결정에 활용할 수 있는 정보를 도출하는 시스템
- 비실시간 처리 : 실시간이 아닌 주기적으로 데이터 분석
- 대규모 데이터 : 과거 데이터와 현재 데이터를 통합하여 처리

예 회사의 연간 판매 데이터를 분석하여 지역별 매출 트렌드를 파악

013 CRM와 SCM ★★

- CRM과 SCM은 2000년대부터 기업 DB 구축 및 운영의 핵심 화두로 등장한 두 가지 주요 솔루션
- 두 시스템은 기업의 내부 운영과 외부 고객 관리에서 각각 중요한 역할을 담당하며, 서로 밀접하게 연계되어 활용

① CRM(Customer Relationship Management)

- 고객과의 관계를 관리하고 강화하여 장기적인 고객 유지 및 수익 창출을 목표로 하는 솔루션
- 고객 데이터를 기반으로 고객의 요구를 파악하고 맞춤형 서비스를 제공

② SCM(Supply Chain Management)

- 제조, 물류, 유통 등 공급망의 모든 과정을 최적화하고, 정보기술을 활용해 재고 관리와 비용 절감을 목표로 하는 솔루션

014 ERP와 BI와 BA★★

- ERP, BI, BA는 기업의 데이터 관리와 의사결정을 지원하는 핵심 기술
- 각각의 목적과 기능은 다르지만, 상호 보완적으로 사용되어 조직의 효율성과 경쟁력을 높이는 데 기여

① ERP(Enterprise Resource Planning)

- 기업의 핵심 비즈니스 프로세스를 통합적으로 관리하는 시스템
- 제조, 구매, 재고, 주문, 공급망, 고객 서비스 등 다양한 프로세스를 지원

② BI(Business Intelligence)

- 데이터를 기반으로 의사결정을 지원하기 위해 설계된 리포트 중심 도구
- 과거 데이터를 분석하여 현재의 상태를 시각적으로 이해하고 판단을 돕는 데 초점

③ BA(Business Analytics)

- BI보다 고급 분석 기술을 적용하여 미래를 예측하거나, 특정 전략의 결과를 예측하는 도구
- 통계적 및 수학적 모델링과 머신러닝을 활용한 데이터 분석

015 기타 기업 내부 데이터베이스 솔루션 ★

① EAI(Enterprise Application Integration)

- 기업 애플리케이션 통합 솔루션으로, 기업 내 다양한 시스템(ERP, CRM, SCM 등)을 상호 연동하여 통합적으로 운영할 수 있도록 지원

② 블록체인(Blockchain)

- 데이터 분산 처리 기술로, 네트워크에 참여하는 모든 사용자가 거래 데이터를 분산 저장하여 데이터의 위·변조를 방지
- 블록과 체인 구조로 이루어져, 데이터의 투명성과 보안을 보장
- 중앙 서버가 아닌 네트워크 참여자들이 데이터를 공유 및 저장
- 중앙 서버가 없어 특정 서버 해킹으로 인한 데이터 손실이 방지됨

016 사회기반 데이터 베이스 솔루션

① 종합물류정보망

- 운행 차량의 위치 및 상태를 실시간으로 추적하여 운송회사와 서비스 가입자의 합리적 의사결정을 지원

② 부가가치통신망(VAN, Value Added Network)

- 통신회선에 부가가치를 첨가한 정보 전송 네트워크로, 정보를 축적, 가공, 변환하여 제공

③ 국가지리정보체계(NGIS, National Geographic Information System)

- 국가 및 지역별 지리 데이터를 관리하고 활용하는 정보 체계

④ 지능형교통시스템(ITS, Intelligent Transportation System)

- 교통 상황 데이터를 실시간으로 수집하고, 이를 활용해 교통 흐름을 최적화하는 시스템

⑤ 의료 EDI(Electronic Data Interchange)

- 병원, 약국, 보험사 간 전자적으로 의료 정보를 교환하는 시스템

⑥ 교육행정정보시스템(NEIS, National Education Information System)

- 전국 초. 중. 고등학교의 교육 행정 데이터를 통합 관리하는 시스템

017 빅데이터(Big Data)의 5V 특징 ★

- 빅데이터는 데이터 관리 및 분석에서 중요한 개념으로, Volume(규모), Variety(다양성), Velocity(속도), Value (가치)와 Veracity 5가지 핵심 특징으로 정의

① Volume(데이터의 크기)

- 방대한 양의 데이터 규모를 의미하며, 생성되는 모든 데이터를 수집하고 저장하는 것을 목표

② Variety(데이터의 다양성)

- 데이터의 형식과 구조가 다양하며, 정형, 반정형, 비정형 데이터를 포함

③ Velocity(데이터의 속도)

- 데이터가 생성되고, 처리되고, 분석되는 속도를 의미

④ Value(데이터의 가치)

- 방대한 데이터를 분석하여 유용한 정보와 통찰을 도출하는 것

⑤ Veracity(데이터의 정확성)

- 데이터의 신뢰성과 정확성을 의미하며, 데이터 품질 관리가 중요

018 빅데이터가 등장하게 된 결정적 요인★

- 빅데이터의 등장은 기술, 조직, 사회적 변화 등 여러 요인이 복합적으로 작용하여 가능

① 기술적 변화

- 대량의 데이터를 저장하고 처리하기 위해 고가의 서버와 데이터센터를 소유하지 않고도, **클라우드 기반 플랫폼**을 통해 저비용으로 데이터 관리 가능
- 하둡(Hadoop)과 같은 분산 데이터 저장 기술 등장으로 대규모 데이터를 효율적으로 저장하고 처리 가능

② 인재 및 조직적 변화

- 데이터의 폭발적 증가로 인해 이를 분석하고 통찰 을 도출할 수 있는 데이터 사이언티스트의 중요성이 대두

019 빅데이터 출현 배경 ★★

- 빅데이터는 완전히 새로운 현상이 아니라, 기존 데이터와 기술, 그리고 이를 다루는 사람과 조직의 변화가 축적되어 나타난 결과

① 산업계 : 양질 전환 법칙

- 데이터를 양적으로 많이 축적하다 보면, 어느 순간 이를 활용해 질적으로 새로운 가치를 창출할 수 있는 전환점이 도래한다는 개념

② 학계 : 빅데이터를 다루는 현상의 증가

- 빅데이터는 학문적으로도 새로운 분석 도구와 방법론의 필요성을 가져왔고, 이를 연구하는 사례가 증가

③ 관련 기술 발전

- 디지털화와 기술 발전, 특히 클라우드 컴퓨팅, 저장 기술, 인터넷 및 모바일 기술이 빅데이터 시대를 열었다.
- 이미지, 동영상, 소셜미디어 텍스트 등 기존의 정형 데이터와는 다른 데이터 유형의 급증

020 빅데이터의 기능 ★

① 빅데이터는 렌즈에 비유

- 빅데이터는 세상을 분석하고 이해하는 새로운 시각을 제공하며, 데이터를 통해 보이지 않던 패턴과 통찰을 발견할 수 있음
- "렌즈"는 데이터 분석 도구나 시각화를 통해 세상을 이해하는 방식을 상징

② 빅데이터는 플랫폼에 비유

- 빅데이터는 다양한 차원에서 활용 가능하며, 이를 기반으로 새로운 비즈니스와 생태계를 구축할 수 있는 기반 인프라 역할을 함
- "플랫폼"은 공동 활용을 목적으로 구축된 구조물로, 이를 통해 데이터의 유통, 공유, 분석이 이루어짐

예 페이스북 : 사용자 데이터를 기반으로 광고 플랫폼 구축
운영체제(OS) : 다양한 앱과 기능을 지원하는 기본 구조물

021 빅데이터가 만들어 내는 본질적 변화★★★

- 빅데이터의 등장과 활용은 데이터 처리 방식, 분석 방법, 그리고 의사결정 패러다임에 근본적인 변화를 가져옴

① 정보의 사전처리에서 사후처리 시대로

- 과거 : 데이터를 분석하기 전에 필요한 정보를 선별 하고 처리하는 방식이 일반적이었음
- 현재 : 데이터를 모두 수집한 후 분석하며, 저장 비용이 감소하면서 데이터의 선별 없이 대규모 데이터 분석이 가능해짐

② 표본조사에서 전수조사로

- 과거 : 표본조사를 통해 제한된 데이터만 분석하여 전체 경향성을 추정
- 현재 : 빅데이터 기술로 인해 전수조사가 가능해져 정확성과 신뢰도가 향상됨

③ 질보다 양으로

- 과거 : 데이터의 품질에 초점을 맞추어 제한된 데이터로 분석
- 현재 : 데이터의 양이 많아질수록 더 정밀한 결과를 도출할 수 있다는 패러다임 변화

④ 인과관계에서 상관관계로

- 과거 : 데이터를 분석하여 인과관계를 도출하려는 시도가 중심이었음
- 현재 : 빅데이터 분석에서는 상관관계를 찾아내는데 초점을 맞춤

022 빅데이터의 가치 산정이 어려운 이유★

- 빅데이터는 방대한 양, 다양한 형식, 높은 처리 속도 등의 특성으로 인해 가치 산정이 쉽지 않다.

① 데이터 활용 방식

- 재사용 및 다목적 활용 : 데이터는 한 번 사용되고 끝나는 것이 아니라, 재사용되고 다양한 목적으로 활용

② 데이터의 기존에 없던 가치 창출

- 데이터는 기존에 정의되지 않은 방식으로 새로운 가치를 창출

③ 분석 기술 발달의 영향

- 새로운 기술로 데이터 가치 창출

023 빅데이터 활용 대표 사례 ★

- 빅데이터는 기업, 정부, 개인 등 다양한 주체가 각자의 목적에 맞게 데이터를 활용하여 새로운 가치를 창출

① 기업 활용 사례

- 구글 검색 : 검색 로그 데이터를 활용하여 기존 페이지랭크 알고리즘을 개선
- 월마트 구매 패턴 분석 : 연관 규칙 분석을 통해 고객의 구매 패턴을 파악
- IBM 왓슨 : 인공지능(AI)을 활용해 의료 데이터를 분석하여 진단 및 치료 방안을 제시

② 정부 활용 사례

- 환경 탐색 : 실시간 교통정보 수집,기후 정보 분석
- 상황 분석 : 소셜미디어 분석,CCTV 및 통화 기록

③ 개인 활용 사례

- 정치인의 SNA(Social Network Analysis) 활용
- 가수 팬들의 청취 분석

024 빅데이터 활용 기법 ★★★

- 빅데이터 분석에서는 다양한 기법이 활용되며, 각각의 기법은 특정한 문제를 해결하거나 패턴을 발견하는데 사용된다.
- 아래는 대표적인 빅데이터 활용 기법과 그 적용 사례이다.

① 연관규칙학습(Association Rule Learning)

- 데이터에서 변수 간 상관관계를 찾아내는 방법으로, 어떤 항목이 함께 발생할 가능성이 높은지를 분석

예 마트에서 상관관계가 높은 상품을 함께 진열, 온라인 쇼핑몰에서 고객의 구매 패턴 분석

② 유형 분석(Classification Tree Analysis)

- 데이터를 특정 기준에 따라 그룹화하거나 분류하는데 사용
- "사용자가 어떤 특성을 가진 집단에 속하는가?"와 같은 질문에 답을 제공

예 온라인 수강생들의 특성에 따라 학습 그룹 분류

③ 유전 알고리즘(Genetic Algorithms)

- 자연선택과 유전학 원리를 활용하여 최적화 문제를 해결하는 방법

예 최대의 시청률을 얻기 위한 방송 시간대 최적화
연료 효율적인 자동차 개발
응급실에서 의사 배치 최적화

④ 기계 학습(Machine Learning)

- 데이터에서 학습하고 패턴을 발견하여 미래를 예측하거나 의사결정을 지원하는 알고리즘
- 훈련 데이터를 기반으로 모델을 생성
- 넷플릭스의 영화 추천 시스템, 전자상거래에서 고객 맞춤형 상품 추천

⑤ 회귀분석(Regression Analysis)

- 독립변수와 종속변수 간의 관계를 파악하여 변수 간 상관성을 분석

  예 구매자의 나이가 구매 차량 타입에 미치는 영향 분석

⑥ 감성분석(Sentiment Analysis)

- 텍스트 데이터에서 긍정적, 부정적 감정을 파악하여 의견과 평가를 분석

  예 새로운 환불 정책에 대한 고객의 평가는 어떤가?

⑦ 소셜 네트워크 분석(Social Network Analysis)

- 소셜 네트워크 데이터를 분석하여 연결성과 영향력 있는 인물을 식별

  예 소셜커머스에서 고객 간의 관계와 영향력 분석

025 빅데이터 시대의 위기 요인과 통제 방안 ★★★

- 빅데이터 시대는 정보와 기술의 발전으로 많은 기회를 제공하지만, 동시에 다양한 윤리적, 사회적 위기를 초래할 수 있다.

① 사생활 침해(Privacy Violation)

- 위기 요인
  - 정보 수집 센서(M2M)의 증가
  - 데이터의 2차, 3차 활용 가능성 증가
  - 사생활 침해와 사회적 위협
- 통제 방안
  - 동의에서 책임으로 전환 : 개인이 매번 동의해야 하는 현재의 시스템은 비효율적이며, 데이터 사용 자가 더 큰 책임을 지도록 하는 체계가 필요

② 책임 원칙의 훼손(Erosion of Accountability)

- 위기 요인
  - 빅데이터 분석으로 개인의 행동, 성향, 위험 요인을 예측하는 알고리즘이 발전- 예측 결과만으로 개인을 단죄하거나 제한을 가하는 사회로 변질될 가능성
  - 이는 개인의 기본권 침해와 민주주의 원칙 훼손으로 이어질 수 있음
- 통제 방안
  - 책임 원칙 보강 : 예측 결과만으로 처벌이나 불이 익을 주는 것을 금지하고, 투명성과 정당성을 확보

③ 데이터의 오용(Misuse of Data)

- 위기 요인
  - 빅데이터는 과거에 일어난 사건과 데이터를 기반으로 분석하고 예측
  - 그러나 과거 데이터만으로 미래를 예측하는 데는 한계가 있으며, 새로운 상황이나 변수를 반영하지 못할 수 있음이는 개인의 기본권 침해와 민주주의 원칙 훼손으로 이어질 수 있음
- 통제 방안
  - 알고리즘에 대한 접근권 허용 : 데이터 알고리즘의 설계와 작동 방식에 대한 투명성 확보
  - 객관적 인증 방안 도입 : 알고리즘이 편향 없이 설계되었는지, 신뢰할 수 있는 결과를 도출하는지 검증하는 공인 인증 시스템 구축
  - 알고리즘 전문가 역할(알고리즘니스트) : 책임과 투명성 강화

026 데이터 3법 ★

- 데이터 3법은 데이터의 이용을 활성화하면서도 개인 정보를 보호하기 위해 마련된 법률로, 다음의 세 가지 법률을 포함

① 개인정보보호법

② 정보통신망 이용촉진 및 정보보호 등에 관한 법률(정보통신망법)

③ 신용정보의 이용 및 보호에 관한 법률(신용정보법)

027 개인정보 비식별화(De-identification of Personal Information) ★★

- 개인정보 비식별화는 개인정보를 특정 개인과 연관 지을 수 없도록 처리하는 방법으로, 데이터 활용과 개인정보 보호를 동시에 달성하기 위해 중요하게 다뤄진다.

① 개인정보란

- 살아 있는 개인에 관한 정보로, 특정 개인을 직접적 또는 간접적으로 식별할 수 있는 정보를 의미

② 비식별화란

- 정보의 일부 또는 전부를 삭제하거나 대체하여 특정 개인을 식별할 수 없도록 하는 조치
- 데이터를 다른 정보와 결합하더라도 개인 식별이 어렵도록 처리

③ 개인정보 비식별화 기술

비식별 기술	제거방법	예시
가명처리	식별요소를 다른 값으로 대체	홍길동, 35세, 서울 거주, 한국대 재학 → 임꺽정, 30대 서울 거주, 국제대 재학
총계처리 또는 평균값 대체	데이터를 총합으로 표시하여 개별 데이터값을 보이지 않도록 함	임꺽정 180cm, 홍길동 170cm → 1-5반 학생 키 합 350cm, 평균 키 175cm
데이터 값 삭제	개인 식별을 인식할 수 있는 값 삭제	홍길동, 35세, 서울 거주, 한국대 졸업 → 35세, 서울 거주
범주화	범주의 값으로 변환	홍길동, 35세 → 홍씨, 30-40세
데이터 마스킹	개인 식별자가 보이지 않도록 처리	홍길동, 35세 → 홍**, 35세

028 빅데이터의 활용에 필요한 3요소 ★

- 빅데이터의 성공적인 활용을 위해서는 데이터, 기술, 인력이라는 세 가지 핵심 요소가 필요

① 데이터(Data)

- 빅데이터 활용의 핵심 자원으로, 다양한 형태와 방대한 양의 데이터를 의미
- 데이터를 수집, 저장, 가공하는 과정에서 모든 현상을 데이터화(Datafication) 하는 것이 중요

② 기술(Technology)

- 데이터를 처리하고 분석하기 위한 도구와 기술로, 빅데이터의 양적, 질적 분석을 가능하게 함

③ 인력(People)

- 데이터를 분석하고, 기술을 활용하여 가치를 창출할 수 있는 전문 인력이 필요
- 데이터 사이언티스트 : 데이터를 분석하고 모델을 설계하여 비즈니스 문제를 해결
- 알고리즈미스트(Algorithmist) : 알고리즘 설계와 검토, 공정성 보장을 담당. 데이터 분석 피해자가 제기한 문제를 조사하고, 잘못된 결과를 조정하거나 수정

029 빅데이터 열풍과 회의론★★

- 빅데이터의 중요성과 활용이 강조되는 가운데, 이에 대한 성과와 기대가 항상 일치하지 않으며, 빅데이터 회의론이 대두되고 있다.
- 빅데이터 분석이 성공하려면 단순히 데이터의 양과 기술적 최적화에 의존하기보다는, 의미 있는 통찰과 전략적 가치를 도출하는 데 중점을 둬야 한다.

① 빅데이터 분석의 성과 창출 관건

- 빅데이터 분석의 목적은 데이터에서 통찰(Insight)을 도출하여 의사결정과 비즈니스 성과를 창출하는 것
- 하지만, 복잡하고 방대한 데이터를 최적화하는 기술만으로는 최고의 가치를 창출할 수 없음

② 데이터 크기보다 통찰력에 집중

- 빅데이터는 단순히 크기(Volume)가 중요한 것이 아니라, 데이터를 통해 의미 있는 시각과 통찰을 도출하는 것이 관건
- 빅데이터 활용의 장애물은 비용 문제보다 분석 방법과 성과에 대한 이해 부족에 있음

③ 성과와 데이터 기반 의사결정의 한계

- 성과가 높은 기업일수록 데이터 기반 의사결정을 채택하지만, 전략적 통찰력과 가치 분석 능력에서 한계를 보임
- 이는 빅데이터 분석의 결과를 기업 전략에 효과적으로 내재화 하는 것이 어렵다는 것을 의미

030 일차적인 분석 vs 전략 도출을 위한 가치 기반 분석 ★

- 데이터 분석은 일차적인 분석에서 시작하지만, 기업의 성과와 지속 가능성을 높이기 위해서는 전략적 인사이트를 제공하는 가치 기반 분석으로 나아가야 한다.

① 일차적인 분석

- 일차적인 분석은 주로 특정 부서나 팀의 업무 개선을 목표로 하는 분석으로, 비교적 좁은 범위와 내부 문제에 초점을 둠

② 전략 도출을 위한 가치 기반 분석

- 가치 기반 분석은 조직의 경쟁력을 강화하기 위해 전략적 인사이트를 도출하는 데 중점을 둔 분석
- 업계 내부 데이터 뿐만 아니라, 인구통계학적 변화, 경제사회 트렌드, 고객 니즈의 변화 등 외부 요인까지 고려
- 전략적 가치기반 분석은 기존 성과를 유지하는 데 초점이 아니라, 비즈니스 성과를 개선하고 혁신하는 것이 주된 목적

031 데이터 사이언스 vs 데이터 마이닝 vs 통계학 ★★

- 데이터를 다루는 세 가지 주요 학문인 데이터 사이언스, 데이터 마이닝, 통계학은 서로 밀접하게 관련되어 있지만, 목적, 접근 방식, 분석 대상에서 차이가 있다.

① 데이터 사이언스(Data Science)

- 데이터를 활용하여 의미 있는 정보를 추출하고 이를 효과적으로 구현하고 전달하는 학문
- 데이터 분석 뿐만 아니라 데이터 관리, 모델 개발, 시각화, 전달까지 모든 과정을 포함

② 데이터 마이닝(Data Mining)

- 데이터에서 패턴과 관계를 탐색하고, 이를 통해 유용한 정보를 추출하는 기술과 과정
- 주로 분석에 초점을 둠

③ 통계학(Statistics)

- 데이터를 수집, 분석, 해석, 표현하는 방법론을 연구하는 학문
- 주로 정형화된 데이터와 실험 데이터를 대상으로 함

032 데이터 사이언스의 핵심 구성 요소 ★

① IT(Data Management)

- 데이터를 수집, 저장, 관리, 처리하는 기술과 인프라를 의미

② Analytics(분석적 영역)

- 데이터를 분석하여 패턴을 발견하고 예측 모델을 구축하는 과정

③ 비즈니스 분석

- 데이터 분석 결과를 실제 비즈니스 문제 해결과 의사결정 지원에 활용

033 데이터 사이언티스트가 갖춰야 할 역량(가트너 기준) ★

- 데이터 사이언티스트는 데이터 분석과 관련된 기술적 능력 뿐만 아니라, 비즈니스 이해와 소프트 스킬을 겸비해야 한다.
- 가트너(Gartner)가 제시한 데이터 사이언티스트의 주요 역량

① 데이터 관리

- 데이터를 수집, 정리, 저장, 가공하는 능력으로, 데이터의 구조와 성격을 깊이 이해해야 함

② 분석 모델링

- 데이터를 분석하고 통계적. 머신러닝 모델을 설계하고 구현하는 능력

③ 비즈니스 분석

- 분석 결과를 활용해 비즈니스 문제를 해결하고, 조직의 전략적 목표를 지원하는 능력

④ 소프트 기능

- 데이터 분석 결과를 조직에 효과적으로 전달하고, 협업을 통해 프로젝트를 성공적으로 이끄는 능력

034 데이터 사이언티스트 요구 역량 (Hard Skill & Soft Skill) ★★★

- 데이터 사이언티스트는 하드 스킬(기술적 역량)과 소프트 스킬(비기술적 역량)을 모두 갖춰야 한다.
- 하드 스킬은 데이터 분석 기술의 전문성을, 소프트 스킬은 분석 결과를 활용해 조직 내 가치를 창출하는 능력을 나타낸다.

① Hard Skill

- 빅데이터에 대한 이론적 지식
- 분석 기술에 대한 숙련
- 최적화 설계 및 노하우 축적

② Soft Skill

- **통찰력 있는 분석** : 창의적 사고,논리적 비판, 호기심
- **설득력 있는 전달** : 분석 결과를 효과적으로 전달하고, 의사결정을 지원할 수 있는 스토리텔링과 시각화 능력
- **다분야 간 협력** : 분석 과정에서 다양한 팀과 협력하고, 커뮤니케이션을 통해 조화를 이루는 능력

035 가치 패러다임의 변화 ★★

- 신기술, 신상품, 서비스의 성공 여부는 이러한 가치 패러다임의 작동 원리에 부합하는지에 달려 있다.
- 급변하는 환경 속에서 가치 패러다임의 변화 흐름을 읽고 대응하는 것이 중요하다

① 디지털화(Digitalization)

- 아날로그 정보를 디지털로 전환하여 저장, 처리, 활용 가능하게 만드는 과정

② 연결(Connection)

- 디지털화된 정보와 대상이 서로 연결됨으로써 가치가 증폭되는 단계

③ 에이전시(Agency)

- 연결된 환경에서 복잡한 관계를 관리하고, 시스템이 자율적으로 작동하도록 지원

036 분석 주제 유형 ★★

- 분석의 대상(what)과 방법(how)에 따라 4가지 유형으로 구분

① Optimization(최적화)

- 정의 : 분석 대상과 방법이 명확하며, 문제를 최적화 하여 해결하는 방식
- 목표 : 효율성 극대화, 비용 최소화, 성능 개선

② Solution(솔루션)

- 정의 : 분석 대상은 명확하지만, 구체적인 분석 방법이 불분명할 때 해결책을 찾는 방식
- 목표 : 문제 해결 및 실질적 성과 도출

③ Insight(통찰)

- 정의 : 분석 방법은 알고 있으나 분석 대상이 불명확 한 경우 새로운 통찰을 얻는 방식
- 목표 : 데이터에서 의미 있는 패턴과 관계 도출

④ Discovery(발견)

- 정의 : 분석 대상과 방법 모두 불명확한 상황에서 새 로운 데이터와 관계를 탐구
- 목표 : 분석 대상 자체를 새롭게 정의하고, 데이터 기반 문제를 새롭게 발견

037 목표 시점별 기획 방안 ★

- 분석 및 기획 방안은 **목표 시점에 따라 단기적인 과제 중심 접근과 장기적인 마스터플랜** 접근으로 나뉜다.
- 각각의 방식은 해결해야 할 과제의 긴급성과 지속 가능성에 따라 선택되며, **당면한 문제 해결과 분석 문화 내재화라는** 상반된 목표를 균형 있게 달성해야 한다.

구분	과제 중심적 접근 방식	장기적인 마스터플랜 방식
목표	당면한 분석 주제를 빠르게 해결	지속적이고 체계적인 분석 문화 내재화
분석단위	과제 단위	마스터플랜 단위
핵심가치	Speed & Test	Accuracy & Deploy
과제유형	Quick-Win	Long-Term View
접근방식	Problem Solving	Problem Definition

038 분석기획 시 고려사항 ★★

▪ 효율적이고 성공적인 분석기획을 위해 데이터, 유스케이스, 장애요소를 포함한 다양한 요소를 사전에 고려해야 한다.

① 가용한 데이터(Available Data)

- 데이터 유형(정형, 반정형, 비정형)에 따라 적용 가능한 분석 방법과 솔루션이 달라지므로, 데이터를 이해하는 것이 첫 번째 단계

② 적절한 유스케이스(Proper Use-Case) 탐색

- 비즈니스 목표와 분석 목표를 일치시키기 위해 유사한 분석 사례(Use-Case)를 탐색
- 유스케이스는 분석의 가이드라인 역할을 하며, 기존솔루션과 방법론 활용 가능

③ 장애요소(Barriers) 사전 계획

- 분석 프로젝트에서 발생할 수 있는 잠재적 장애요소를 미리 파악
- 장애요소에 대한 사전 대응책을 마련하여 프로젝트 진행에 차질이 없도록 함

039 데이터 저장 방식 ★★

- 데이터 저장 방식은 데이터의 구조와 사용 목적에 따라 크게 RDB(Relational Database)와 NoSQL로 나뉜다.

① RDB(Relational Database)

- 정의 : 관계형 데이터베이스로, 데이터를 테이블 형식으로 저장하며 테이블 간 관계를 정의
- SQL 사용으로 표준화된 데이터 관리 가능
- 수직적 확장(서버의 성능 향상으로 확장)
- 주요 도구(Oracle, MySQL, MSSQL)

② NoSQL

- 정의 : 비정형 데이터, 반정형 데이터 등 구조화되지 않은 데이터를 저장하고 처리하기 위한 비관계형 데이터베이스
- 대규모 데이터 처리와 분산 저장에 최적화 데이터 모델링이 자유로워 유연성 제공
- 수평적 확장(여러 대의 서버를 추가)
- 주요 유형 및 도구(MongoDB, Redis, Cassandra, HBase)

040 기업의 합리적 의사결정 장애 요소 ★

① 고정관념(Stereotype)

- 특정 사물, 상황, 사람에 대한 고정된 신념이나 이미지로 인해 새로운 정보를 왜곡하거나 거부

② 편향된 생각(Bias)

- 개인적 경험, 감정, 선호로 인해 객관적 데이터나 사실을 왜곡하는 경향

③ 프레이밍 효과(Framing Effect)

- 문제의 표현 방식에 따라 동일한 사건이나 상황에 대한 판단이나 선택이 달라지는 현상

041 분석 방법론의 구성 요소 ★

① 상세한 절차(Procedure)

- 분석 수행을 위한 단계적 가이드라인으로, 작업의 체계성과 일관성을 유지

예 데이터 수집 → 전처리 → 이탈 예측 모델링 → 결과 해석

② 방법(Methods)

- 분석 문제 해결을 위한 구체적인 접근 방식

예 통계적 분석, 기계 학습

③ 도구와 기법(Tools & Techniques)

- 분석을 수행하기 위한 소프트웨어 및 알고리즘

예 Python,R,Tableau,Power BI

④ 템플릿과 산출물(Templates & Outputs)

- 분석 과정의 표준화를 돕는 템플릿과 결과물을 문서화한 자료

예 분석 계획서

042 폭포수(Waterfall), 나선형(Spiral), 프로토타입(Prototype) 모델 비교 ★★

▪ 소프트웨어 개발 및 프로젝트 관리에서 사용되는 폭포수 모델, 나선형 모델, 프로토타입 모델은 각기 다른 특징과 적합한 상황을 가진다.

① 폭포수 모델(Waterfall Model)

- 프로젝트를 단계별로 진행하며, 각 단계가 완료되고 승인된 후에 다음 단계로 넘어가는 방식

- 장점
  - 각 단계가 명확히 정의되어 있어 체계적이고 이해하기 쉬움

- 단점
  - 단계별 진행이므로 변경사항 적용이 어렵고 비용이 증가
  - 반복과 피드백이 제한적

② 나선형 모델(Spiral Model)

- 여러 번의 개발 반복을 통해 점진적으로 프로젝트를 완성하는 방식

- 장점
  - 위험 관리가 용이하며, 반복을 통해 점진적으로 품 질 향상
  - 초기에 요구사항이 명확하지 않아도 점진적으로 구체화 가능
  - 대규모 프로젝트에서 유용

- 단점
  - 관리가 복잡하며, 반복을 제대로 관리하지 못하면 비용과 일정 초과

③ 프로토타입 모델(Prototype Model)

- 초기 단계에서 사용 가능한 프로토타입(시제품)을 개발하여 사용자로부터 피드백을 받고, 이를 바탕으로 반복적으로 개선하는 방식

- 장점
  - 사용자의 요구사항을 명확히 파악할 수 있음
  - 초기 오류를 신속히 발견하고 개선 가능
  - 반복적으로 개선해 나가는 개발 방식으로 개발 시간을 단축하는 데 유리

- 단점
  - 지속적인 피드백 과정으로 시간과 비용 증가

043 KDD 분석 절차 (Knowledge Discovery in Databases) ★★★

- KDD는 데이터베이스에서 유용한 지식을 발견하기 위한 체계적인 분석 절차
- 데이터 준비부터 분석 결과의 해석과 활용까지 포괄적인 단계를 포함하며, 데이터 마이닝 기법의 효과적인 적용을 지원

① 비즈니스 이해와 목표 설정

- 무엇을 분석하고, 어떤 결과를 원하는지 정의

② 데이터 선택

- 목적에 맞는 데이터 세트를 선정

③ 데이터 전처리(Preprocessing)

- 분석의 정확성과 신뢰성을 높이기 위해 데이터를 정리하고 준비하는 과정
- 잡음 제거 : 데이터에 포함된 불필요한 정보 삭제
- 결측값 처리 : 삭제, 평균 대체, 예측값 대체
- 이상값(Outlier) 처리 : 데이터 입력 오류.측정 오류등을 대체

④ 데이터 변환(Transformation)

- 분석 목적에 맞게 데이터셋을 조정하거나 변경하는 단계
- 변수 선택(Feature Selection) : 분석 목적과 관련성이 높은 변수를 선택
- 차원 축소(Dimensionality Reduction) : 데이터의 차원을 줄여 분석 속도를 높이고 과대적합(Overfitting)을 방지
- 데이터 스케일링(Scaling) : 데이터 값의 범위를 조정하여 모델 학습 과정에서 특정 변수의 영향을 과도하게 받지 않도록 함

⑤ 데이터 마이닝(Data Mining)

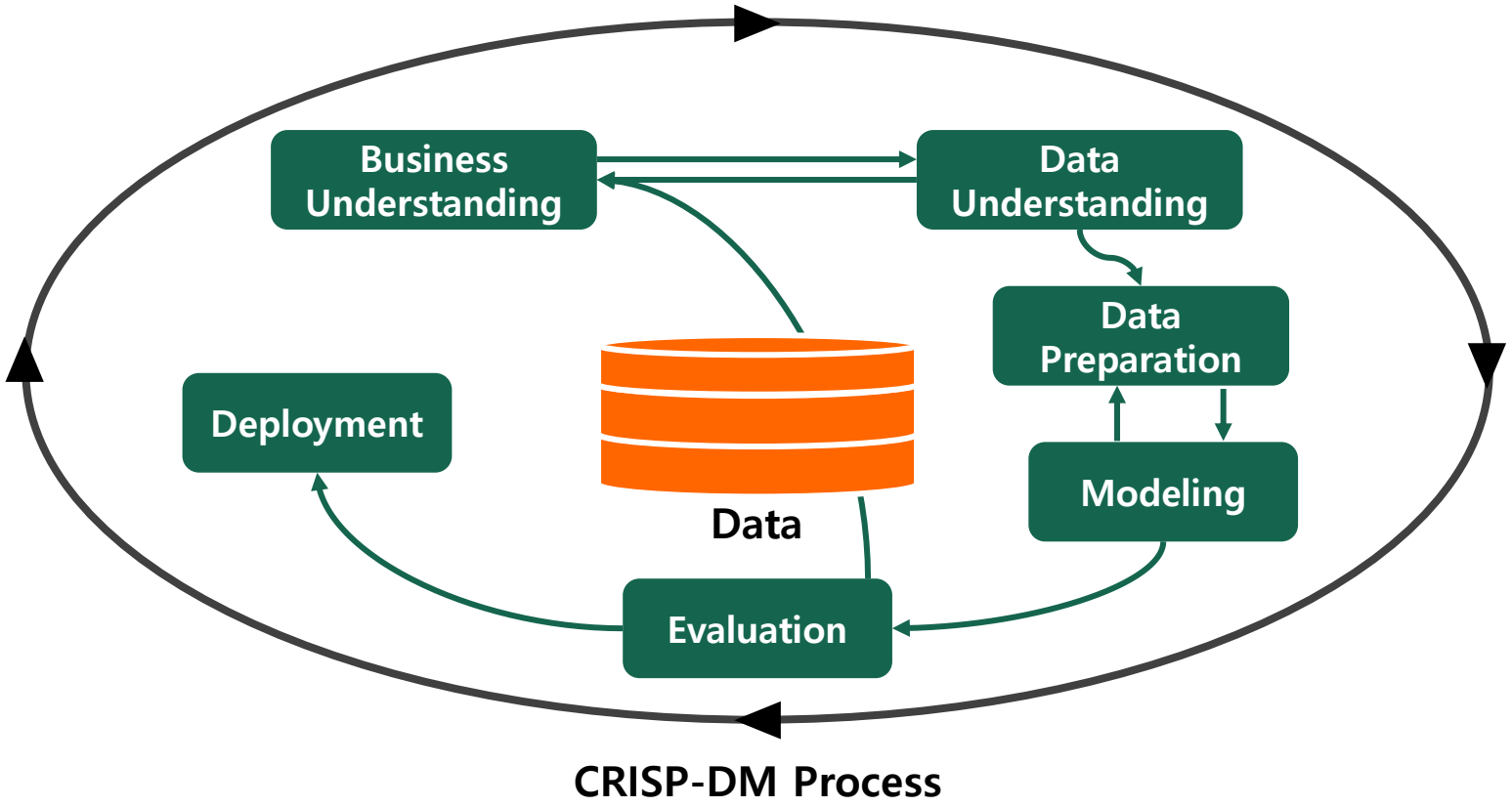
- 변환된 데이터셋을 활용하여 패턴을 발견, 데이터를 분류, 예측 작업을 수행하는 과정
- 데이터 마이닝은 다양한 기법과 알고리즘을 사용하여 비즈니스 목표에 맞는 분석 결과를 도출하는데 중점

⑥ 데이터 마이닝 결과 평가(Interpretation / Evaluation)

- 데이터 마이닝 결과 평가 단계는 분석 결과를 해석, 평가, 활용하는 과정으로, 데이터 마이닝의 성패를 결정하는 중요한 단계

044 CRISP-DM 분석 방법론 ★

- CRISP-DM(Cross Industry Standard Process for Data Mining)은 다양한 산업 분야에서 데이터 마이닝 프로젝트를 체계적으로 수행하기 위해 설계된 표준 프로세스
- 6단계로 구성되어 있으며, 각 단계는 피드백을 통해 완성도를 높이고 유연성을 제공



CRISP-DM Process

045 CRISP-DM 6단계 ★★★

① 업무 이해(Business Understanding),단계(Phase) 계층

- 업무 이해 단계는 데이터 마이닝 프로젝트의 목적과 요구사항을 비즈니스 관점에서 이해하고, 이를 데이터 분석 문제로 정의하는 기초 단계

- **업무 목적 파악** : 프로젝트의 비즈니스 목표를 명확히 정의 , **태스크(Task) 계층**
- **상황 파악** : 기존 프로세스 및 시스템 검토, **태스크(Task) 계층**
- **데이터 마이닝 목표 설정** : 데이터 마이닝의 구체적 목표 정의(분류, 예측, 군집화 등) **태스크(Task) 계층**

② 데이터 이해(Data Understanding)

- 데이터 이해는 프로젝트에서 사용할 데이터를 수집, 탐색, 분석하여 데이터의 특성과 품질을 파악하고, 숨겨진 인사이트를 발견하는 단계

- **초기 데이터 수집** : 분석을 수행하기 위해 데이터를 수집하고 정리
- **데이터 기술 분석** : 데이터를 요약하고 기본적인 특성을 확인
- **데이터 탐색** : 데이터의 분포와 변수 간 관계를 시각적으로 탐색
- **데이터 품질 확인** : 데이터셋의 품질 문제를 식별하고 기록

③ 데이터 준비(Data Preparation)

- 데이터 준비 단계는 분석 기법에 적합한 데이터 세트를 구성하기 위해 수집된 데이터를 정리하고 변환하는 과정

- **분석용 데이터 세트 선택** : 분석에 필요한 데이터만 선택하여 세트를 구성
- **데이터 정제(Data Cleaning)** : 데이터의 결함을 수정하여 품질을 개선
- **데이터 통합(Data Integration)** : 서로 다른 소스에서 데이터를 결합하여 통합 데이터 세트를 생성
- **데이터 포맷팅(Data Formatting)** : 데이터를 분석 기법에 적합한 형식으로 변환

④ 모델링(Modeling)

- 모델링 단계는 다양한 알고리즘과 기법을 선택하여 데이터를 분석하고, 모델을 구축, 평가, 최적화하는 과정

- **모델링 기법 선택** : 분석 목적과 데이터 특성에 따라 적합한 알고리즘과 기법을 선택
- **모델 테스트 계획 설계** : 모델의 성능을 검증하기 위한 테스트 데이터와 평가 기준을 설정
- **모델 작성** : 데이터를 기반으로 모델을 학습시키고, 패턴을 학습하여 예측값 생성
- **모델 평가** : 모델의 성능과 신뢰성을 검토하고, 과적합 문제를 진단

⑤ 평가(Evaluation)

- 평가 단계는 데이터 마이닝 과정에서 도출된 모델의 성능과 결과가 프로젝트의 목적에 부합하는지를 판단하는 단계

- **분석 결과 평가** : 모델이 생성한 결과가 비즈니스 목표를 얼마나 충족하는지 확인
- **모델링 과정 평가** : 모델링 과정에서의 문제점이나 개선 가능성을 분석
- **모델 적용성 평가** : 모델이 실제 비즈니스 환경에서 실행 가능하고, 실질적인 가치를 제공하는지 확인

⑥ 전개(Deployment)

- 전개 단계는 모델링과 평가를 통해 완성된 모델을 실제 업무에 적용하고, 이를 운영 환경에서 안정적으로 활용하기 위해 계획을 수립하고 실행하는 단계

- **전개 계획 수립** : 모델을 실제 운영 환경에 적용하기 위한 실행 계획 수립
- **모니터링 및 유지보수 계획 수립** : 운영 중인 모델의 성능을 모니터링하고, 필요 시 업데이트
- **프로젝트 종료 보고서 작성** : 프로젝트의 전 과정과 성과를 요약한 보고서를 작성
- **프로젝트 검토** : 프로젝트 전체를 리뷰하고 성공 요인 및 개선점을 도출

046 빅데이터 분석 방법론의 계층적 프로세스 모델 ★★

▪ 빅데이터 분석을 체계적으로 수행하기 위해 설계된 계층적 프로세스 모델 (Stepwised Process Model)은 3계층(단계, 태스크, 스텝)으로 구성

① 단계(Phase)

- 분석 프로젝트의 최상위 계층으로, 프로세스 그룹(Process Group)을 기반으로 주요 목표를 달성하기 위한 큰 흐름을 정의
- 각 단계는 기준선(Baseline)으로 설정되어 관리되며, 변경 사항은 버전 관리(Configuration Management)로 통제

② 태스크(Task)

- 단계를 구성하는 세부 활동 단위로, 각 태스크는 물리적 또는 논리적 작업으로 정의
- 여러 태스크로 단계가 구성됨

③ 스텝(Step)

- 태스크를 세분화한 세부 프로세스 단위로, WBS(Work Breakdown Structure)의 워크 패키지에 해당
- 각 스텝은 입력자료(Input), 처리 및 도구(Process & Tool), 출력자료(Output)로 구성
- 구체적이고 실행 가능한 작업

047 빅데이터 분석 방법론 ★★★

분석기획 (Planning)	데이터 준비 (Preparing)	데이터 분석 (Analyzing)	시스템 구현 (Developing)	평가 및 전개 (Deploying)
비즈니스 이해 및 범위설정	필요 데이터 정의	분석용 데이터 준비	설계 및 구현	모델 발전 계획
프로젝트 정의 및 계획 수립	데이터 스토어 설계	텍스트 분석	시스템 테스트 및 운영	프로젝트 평가 보고
프로젝트 위험 계획 수립	데이터 수집 및 적합성 점검	탐색적 분석		평가 및 전개
		모델링		
		모델 평가 및 검증		

048 빅데이터 단계별 프로세스 ★★★

프로세스 순서	단계 (Phase)	태스크 (Task)	내용
1	분석 기획	비즈니스 이해 범위 및 범위 설정	• 프로젝트에 참여하는 관계자들(Stakeholders)의 이해를 일치시키기 위하여 구조화된 프로젝트 범위 정의서인 SOW (Statement of Work)를 작성
		프로젝트 정의 및 계획 수립	• 프로젝트의 목표 및 KPI(핵심성과지표), 목표 수준 등을 구체화하여 상세 프로젝트 정의서를 작성 • 프로젝트 수행 계획서를 작성하는 단계로서 프로젝트의 목적 및 배경, 기대효과, 수행방법, 일정 및 추진조직, WBS를 작성한다.
		프로젝트 위험계획 수립	• 데이터 분석 위험 식별, 계획 수립 단계에서 빅데이터 분석 프로젝트를 진행하면서 발생 가능한 모든 위험을 식별한다. • 예상되는 위험에 대한 대응은 회피(Avoid), 전이(Transfer), 완화(Mitigate), 수용(Accept)으로 구분하여 위험 관리 계획서를 작성한다.

2과목 데이터의 분석 기획

프로세스 순서	단계 (Phase)	태스크 (Task)	내용
2	데이터 준비	필요 데이터 정의	•정형·비정형·반정형 등의 모든 내·외부 데이터를 포함하고 데이터의 속성, 데이터 오너, 데이터 관련 시스템 담당자 등을 포함하는 데이터 정의서를 작성한다.
		데이터 스토어 설계	•일반적으로 관계형 데이터베이스(RDBMS)를 사용하고 데이터의 효율적인 저장과 활용을 위하여 데이터 스토어의 논리적, 물리적 설계를 구분하여 설계한다. •하둡, NoSQL 등을 이용하여 비정형 또는 반정형 데이터를 저장하기 위한 논리적, 물리적 데이터 스토어를 설계한다.
		데이터 수집 및 정합성 점검	•크롤링 등의 데이터 수집을 위한 ETL 등의 다양한 도구와 API, 스크립트 프로그램 등을 이용하여 데이터를 수집하고, 수집된 데이터를 설계된 데이터를 스토어에 저장한다.

2과목 데이터의 분석 기획

프로세스 순서	단계 (Phase)	태스크 (Task)	내용
3	데이터 분석	분석용 데이터 준비	• 분석 계획 단계에서 비즈니스 이해, 도메인 문제점 인식, 프로젝트 정의 등을 이용하여 프로젝트의 목표를 정확하게 인식한다. • 데이터 스토어로 분석에 필요한 정형·비정형데이터를 추출한다.
		텍스트분석	• 감성 분석(Sentimental Analysis), 토픽 분석(Topic Analysis), 오피니언 분석(Opinion Analysis), 소셜 네트워크 분석(SNA) 등을 실시 하여 텍스트로부터 분석 목적에 맞는 적절한 모델을 구축한다.
		탐색적분석	• 다양한 관점별로 기초 통계량을 산출하고 데이터의 분포와 변수 간의 관계 등 데이터특성 및 데이터의 통계적 특성을 이해하고 모델링을 위한 기초 자료로 활용한다.
		모델링	• 모델링이란 분석용 데이터를 이용한 가설 설정을 통하여 통계 모델을 만들거나 기계학습을 이용한 데이터의 분류, 예측, 군집 등의 기능을 수행하는 모델을 만드는 과정이다.

2과목 데이터의 분석 기획

프로세스 순서	단계 (Phase)	태스크 (Task)	내용
3	데이터 분석	모델링	**1) 데이터 분할** • 모델의 과적합과 일반화를 위하여 분석용 데이터셋을 모델 개발을 위한 훈련용 데이터와 모델의 검증력을 테스트하기 위한 데이터로 분할한다. **2) 데이터 모델링** • 기계학습 등을 이용한 모델링은 훈련용 데이터를 활용하여 분류, 예측, 군집 등의 모델을 만들어 가동중인 운영 시스템에 적용한다. **3) 모델적용 및 운영 방안** • 모델을 가동중인 운영시스템에 적용하기 위해서는 모델에 대한 상세한 알고리즘 설명서 작성이 필요하다. 알고리즘 설명서는 시스템 구현 단계에서 중요한 입력자료로 활용되므로 필요 시 의사코드 수준의 상세한 작성이 필요할 수도 있다.
		모델평가 및 검증	• 프로젝트 정의서의 모델 평가 기준에 따라 모델을 객관적으로 평가하고 품질관리 차원에서 모델 평가 프로세스를 진행한다.

2과목 데이터의 분석 기획

프로세스 순서	단계 (Phase)	태스크 (Task)	내용
4	시스템 구현	설계 및 구현	• 모델링 태스크에서 정의된 알고리즘 설명서와 데이터 시각화 보고서를 이용하여 시스템 및 아키텍처 설계, 사용자 인터페이스 설계를 진행한다. • 설계서를 바탕으로 패키지를 활용하여 프로그램을 구축한다.
		시스템 테스트 및 운영	• 구축된 시스템의 검증을 위하여 단위테스트, 통합테스트, 시스템 테스트 등을 실시한다.
5	평가 및 전개	모델발전 계획수립	• 개발된 모델의 지속적인 운영과 기능 향상을 위한 발전 계획을 상세하게 수립한다.
		프로젝트 평가 보고	• 프로젝트의 성과를 정량적, 정성적으로 평가하고 프로젝트 진행 과정에서 산출된 지식, 프로세스, 출력 자료를 지식 자산화하고 프로젝트 최종보고서를 작성한다.

049 분석 과제 도출 ★★

- 빅데이터 분석 과제를 정의하고 수행하기 위해서는 문제를 **데이터 분석 문제로 변환**하는 과정이 필요하다.
- 분석 과제 도출은 하향식(Top Down)과 상향식(Bottom Up) 접근 방법을 통해 이루어질 수 있으며, 이 두 가지 접근 방식은 상황에 따라 상호 보완적으로 활용된다.

① 하향식 접근 방법(Top Down Approach)

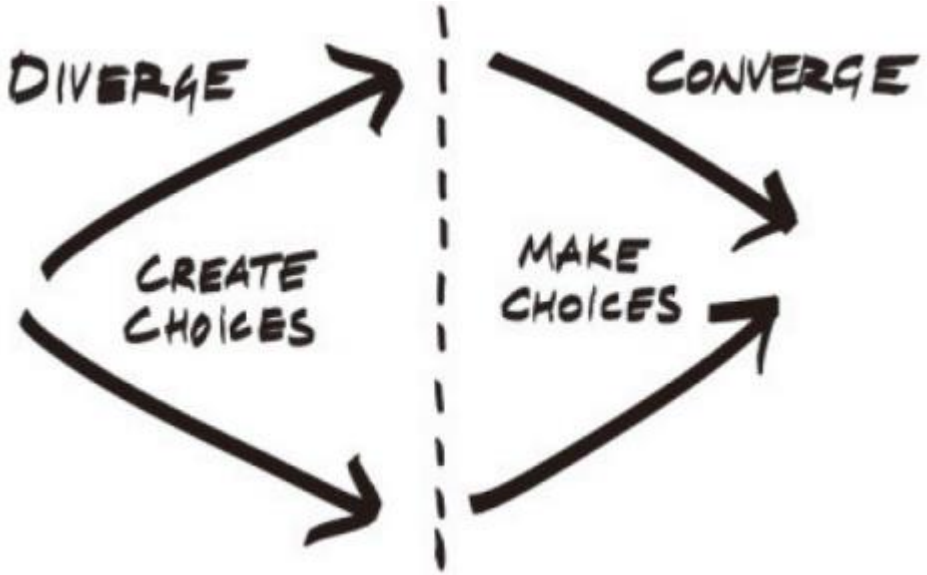
- **문제가 명확히** 주어진 경우, 이를 체계적으로 분석하여 해결방안을 도출하는 방식.
- 각 과정이 체계적으로 단계화 되어 수행되며, 프로젝트 목표와 요구 사항이 **명확할 때 적합**

② 상향식 접근 방법(Bottom Up Approach)

- **문제정의 자체가 어려운 경우, 데이터를 기반으로 문제를 재정의**하고 해결방안을 탐색하는 방식
- 시행착오를 통해 문제와 해결 방안을 점진적으로 개선하며, 새로운 인사이트 도출에 유용

050 상향식 및 하향식 접근 방식의 혼용 : 디자인 씽킹 ★

- 디자인 씽킹(Design Thinking)은 상향식 접근 방식(Bottom-Up)과 하향식 접근 방식(Top-Down)을 혼합하여 활용하는 혁신적 문제 해결 방법론
- 사용자와 공감하고, 다양한 아이디어를 **발산한 후 이를 수렴** 및 검증하는 반복적이고 체계적인 방식으로, 다양한 도전 과제와 복잡한 문제를 효과적으로 해결



상향식 접근 방식(diverse) 하향식 접근 방식(converse)

051 디자인 씽킹 프로세스 : 5단계 ★

- 디자인 씽킹(Design Thinking)은 사용자 중심으로 문제를 정의하고 해결방안을 찾아가는 반복적이고 유연한 프로세스이다.
- 이 과정은 공감 → 정의 → 아이디어 발산 → 프로토타이핑 → 테스트로 구성되며, 각 단계는 고객의 문제를 깊이 이해하고 효과적인 해결책을 도출하는데 초점이 맞춰져 있다.

① **Empathize(공감)**
- 사용자 인터뷰와 관찰을 통해 고객의 문제를 이해하고 공감

② **Define(문제정의)**
- Empathize 단계에서 얻은 인사이트를 바탕으로 고객의 진짜 문제를 명확히 정의

③ **Ideate(아이디어 발산)**
- 문제정의를 기반으로 자유롭게 다양한 해결 방안을 제시

④ **Prototype(프로토타입 제작)**
- Ideate 단계에서 도출된 아이디어를 구체화하여 시제품(프로토타입)이나 서비스 시나리오 제작

⑤ **Test(테스트)**
- 프로토타입을 사용자에게 테스트하여 피드백 수집

052 하향식 접근 방식 (Top Down Approach) ★★★

- 하향식 접근 방식은 명확히 주어진 문제나 전략적 목표를 기반으로 분석을 체계적으로 설계하고 실행하는 접근법이다.
- 이 방식은 현황 분석에서 시작해, 문제를 데이터 문제로 변환하여 구체적이고 실행 가능한 해결방안을 도출하는 데 초점을 맞춘다.
- 하향식 접근법(Top-Down Approach)은 문제가 명확히 정의된 상황에서 문제를 해결하는 데 적합하지만, 새로운 문제를 탐색하거나 발견하는 데는 몇 가지 한계를 가진다.

① 문제탐색(Problem Discovery)

- 현황 분석을 통해 확인된 문제나 기회를 심층적으로 탐색
- 문제의 원인과 연관성을 파악하여 데이터를 기반으로 해결 가능성을 검토
- 세부적인 구현 및 솔루션에 중점을 두는 것이 아니라, 문제를 해결했을 때 발생하는 가치에 집중

② 문제정의(Problem Definition)

- 문제 정의 단계는 비즈니스 문제를 데이터 분석 문제로 변환하여, 필요한 데이터와 분석 기법을 명확히 정의하는 단계
- 데이터 분석의 효율성을 높이고, 해결 과정을 체계적으로 설계하는 데 중요한 역할을 한다.

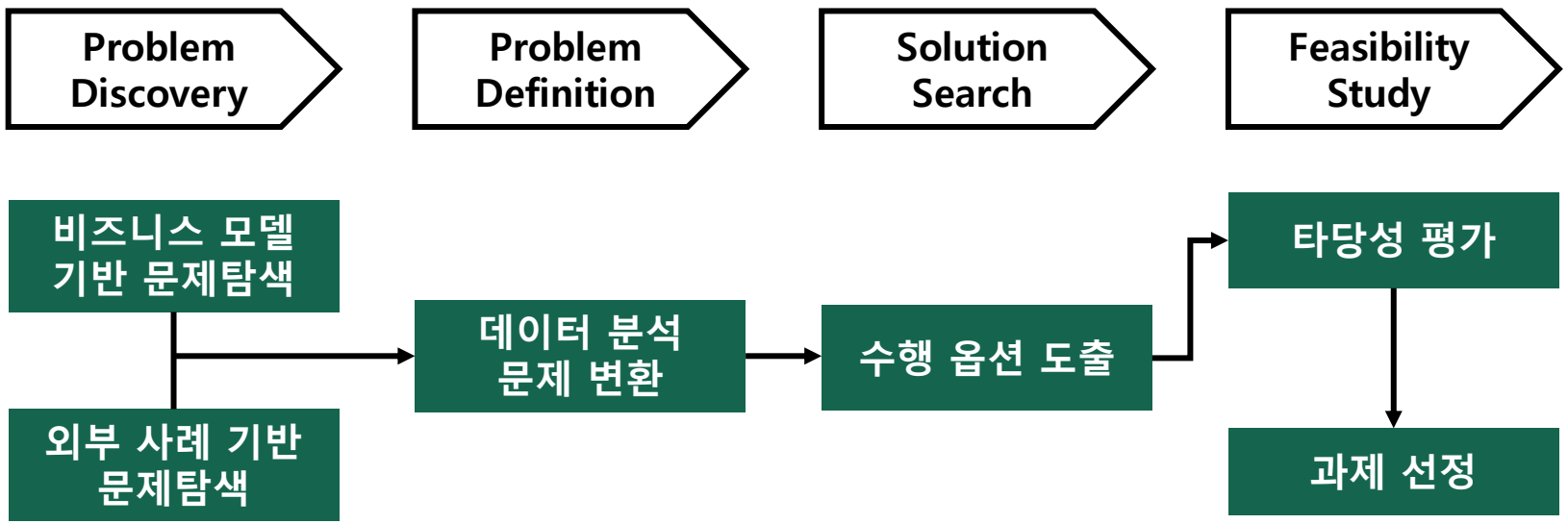
③ 해결방안 탐색(Solution Search)

- 해결방안 탐색 단계는 이전 단계에서 정의된 데이터 분석 문제를 해결하기 위해 다양한 방안을 탐색하고 비교하는 과정
- 분석의 효율성과 효과를 극대화하기 위해 사용 가능한 데이터, 도구, 예산, 기술적 요구 사항 등을 다각적으로 검토하는 데 중점을 둠

④ 타당성 검토(Feasibility Study)

- 타당성 검토 단계는 해결 방안 탐색에서 도출된 대안 을 실제로 실행 가능한지 평가하는 과정이다.
- 이 단계는 경제적 타당도와 데이터 및 기술적 타당도 의 두 가지 관점에서 분석하여, 프로젝트의 성공 가능성을 사전에 확인하고 리스크를 최소화한다.

2과목 데이터의 분석 기획



하향식 접근 방식(converse)

053 문제 탐색(Problem Discovery) 단계 : 비즈니스 모델 기반 문제 탐색 ★

- 비즈니스 모델 기반 문제 탐색은 비즈니스 모델 캔버스를 활용하여 기업 내 가치 창출과 관련된 문제를 구조적으로 도출하는 단계이다.

- 이 과정은 **업무, 제품, 고객, 규제와 감사, 지원 인프라 등 5가지 핵심 영역**에서 문제를 발굴하여 데이터 분석 기회를 식별하는 데 초점이 맞춰져 있다.

054 혁신 관점에서 중장기적 분석 기회 발굴 및 확장 ★

▪ 혁신의 관점에서 중장기적 분석 기회 발굴은 기업이 지속 가능한 경쟁 우위를 확보하고 미래 시장에서 성공적으로 자리 잡기 위해 필수적이다.

① 거시적 관점(STEEP)

- 사회, 기술, 경제, 환경, 정치적 요소를 포함한 거시적 관점에서 분석 기회를 발굴

② 경쟁자 확대 관점

- 기존 및 잠재적 경쟁자를 포함하여 시장에서의 위협 요소를 식별하고 분석 기회 발굴
- **대체재 영역** : 현재 제품 및 서비스를 대체할 수 있는 신기술, 신제품의 등장 가능성 분석
- **경쟁자 영역** : 주요 경쟁자의 동향 및 전략 분석
- **신규 진입자 영역** : 신규 진입자가 시장에 미칠 잠재 적 위협 분석

③ 시장의 니즈 관점

- 고객, 채널, 이해관계자(영향자) 관점에서 제품과 서비스의 개선 및 확장 기회를 탐색
- **고객 영역** : 고객 구매 패턴, 행동 분석을 통해 숨겨 진 니즈 파악
- **채널 영역** : 제품 및 서비스 전달 경로의 효율성 및 확 장 가능성 분석
- **영향자 영역** : 주요 이해관계자(투자자, 협회, 주주)의 관심사를 분석하여 비즈니스 방향 설정

④ 역량의 재해석 관점

- 기업의 내부 역량과 파트너 네트워크를 재해석하여 경쟁력을 강화할 기회를 발굴
- **내부 역량 영역** : 내부 노하우(기술, 지식) 및 유형 자 산의 효율적 활용
- **파트너와 네트워크 영역** : 협력 관계를 활용하여 부족한 역량 보완

055 외부 참조 모델 기반 문제 탐색 ★

- 외부 참조 모델 기반 문제 탐색은 **유사 • 동종 업종과 업무 서비스 사례를 벤치마킹**하여, 새로운 문제와 분석 기 회를 발굴하는 데 초점을 둔 접근법

- 산업별, 업무 서비스별 분석 테마 후보 그룹을 활용해 빠르고 효율적으로 분석 과제를 도출하고, 이를 기업 환경 에 적용할 수 있도록 지원

056 분석 유즈 케이스 (Analytics Use Case) ★

- 분석 유즈 케이스는 도출된 분석 기회를 구체적인 데이터 분석 문제로 전환하기 전, 문제의 상세 설명과 문제 해결 시 기대 효과를 명시하는 단계이다.

- 이를 통해 분석 과제를 체계적으로 정리하고, 적합성 평가 및 데이터 분석 설계의 기초로 활용한다.

057 상향식 접근 방식 (Bottom-Up Approach) ★★

- 상향식 접근 방식은 문제정의가 명확하지 않거나 초기 가설 설정이 어려운 상황에서 데이터 기반으로 문제를 탐색하고 해결방안을 개선하는 방식

① 문제 정의 전 데이터 기반 탐색

② 비지도 학습 활용

③ 기존의 인과관계 분석 대신, 상관관계와 연관분석을 통해 문제 해결에 도움

058 프로토타이핑 프로세스 ★

- 프로토타이핑 접근법은 사용자의 요구사항이나 **데이터의 명확성이 부족한 상황**에서, 빠르게 해결책의 초안을 개발하고 이를 **반복적으로 개선하는 방식**
- **상향식 접근 방식의 특징을 가지며**, 특히 빅데이터 환경 에서 유용하게 사용

① 장점

- **빠른 결과 도출** : 초기 단계에서 신속히 결과물을 제공하여 문제를 명확히 정의
- **유연성** : 데이터와 요구사항이 부족한 상황에서도 시작 가능
- **피드백 기반 개선** : 사용자 피드백을 통해 점진적으로 개선 가능
- **상향식 접근** : 초기 데이터를 통해 문제를 도출하고 구체화

② 단점

- **완성도 부족** : 초기 프로토타입은 완전하지 않을 수 있음
- **리소스 소모** : 반복적인 개선 과정에서 시간과 비용 이 증가
- **추가 데이터 의존성** : 결과 개선을 위해 추가 데이터가 필요할 수 있음

059 빅데이터 환경에서 프로토타이핑의 역할 ★

① 문제 정의의 불명확성 해소

- 프로토타입을 통해 문제를 시각화하고, 데이터를 분석해 초기 결과물을 제시함으로써 문제의 본질을 이 해하고 구체화하는데 도움을 제공

② 필요 데이터 존재 여부 확인

- 프로토타입은 필요한 데이터를 빠르게 탐색하고, 데이터 부족 또는 부적합 문제를 조기에 발견할 수 있도록 돕는다.

③ 데이터 사용 목적의 가변성 지원

- 프로토타입은 초기 목적에 얽매이지 않고, 새로운 목적이나 목표로 전환할 수 있는 유연성을 제공

060 분석 과제 정의서 ★

- 분석 과제 정의서는 **프로젝트의 목표, 데이터 요구사항, 분석 방법, 수행 주기** 등을 체계적으로 정리한 문서

- 이를 통해 **이해관계자 간의 협업을 강화**하고, 분석 결과의 신뢰성과 활용성을 극대화할 수 있다.

061 분석 프로젝트 관리 방안 ★★★

- 분석 프로젝트는 도출된 분석 기회를 통해 구체적인 가치를 증명하고 목표를 달성하기 위한 체계적인 관리가 필요하다.
- 일반적인 프로젝트 관리 요소(범위, 일정, 품질, 리스크 등) 외에도, **분석 프로젝트는 데이터와 분석 모델의 특 수성을 반영한 5가지 주요 속성을 중심**으로 관리가 이루어져야 한다.

① Data Size(데이터 크기)

- 데이터 양이 방대할수록 처리 속도와 자원 관리가 중요

② Data Complexity(데이터 복잡성)

- 정형 데이터 뿐만 아니라 비정형. 반정형 데이터(텍스트, 이미지 등)를 포함할 경우, 데이터 확보와 모델 선정이 더욱 까다로움

③ Speed(속도)

- 분석 결과를 도출하는 시간적 요구사항

④ Analytic Complexity(분석 복잡성)

- 분석 모델의 복잡도가 증가하면 정확도(Accuracy)는 높아지지만, 모델 해석과 적용이 어려워짐
- 복잡성과 정확도 간의 균형을 고려한 기준점 설정

⑤ Accuracy & Precision(정확도와 정밀도)

- Accuracy와 Precision 속성은 트레이드 오프(trade-off) 관계를 가지며, 분석 모델의 해석 및 적용 단계에서 사전 고려가 필요
- 정확도(Accuracy) : 모델의 예측 값이 실제 값과 얼마나 가까운지
- 정밀도(Precision) : 모델을 반복 적용했을 때 결과가 얼마나 일관되게 나오는지
- Accuracy와 Precision을 고려한 모델 설계에서 활용이 중요하다면 Accuracy를 우선, 안정성이 중요하다면 Precision을 우선

062 분석 프로젝트 영역별 주요 관리 항목 ★

▪ 데이터 분석 프로젝트는 다양한 관리 영역에서 세부적인 관리를 통해 성공적인 결과를 도출할 수 있다.

① 범위(Scope)

- 프로젝트의 목표와 범위를 명확히 정의

② 시간(Time)

- 타임박싱(Timeboxing) 기법을 활용하여 일정 관리를 유연하게 진행
- 품질을 보장하면서 일정 초과를 방지하는 관리 전략 적용

③ 원가(Cost)

- 프로젝트 수행에 필요한 데이터 처리 비용, 분석 도구, 인력 비용 등을 관리

④ 품질(Quality)

- 데이터 처리 및 분석 프로세스 준수 여부 확인

⑤ 통합(Integration)

- 다양한 데이터 소스와 분석 기법의 통합관리

⑥ 조달(Procurement)

- 조달은 프로젝트를 수행하는데 필요한 자원(인프라, 데이터, 소프트웨어, 기술 등)을 확보하고, 이를 효과적으로 관리하여 프로젝트 관리 프로세스가 통합적으로 운영될 수 있도록 지원하는 중요한 활동
- PoC 프로젝트는 인프라를 구매하지 않고도 클라우드 서비스와 외부 소싱을 통해 신속하고 유연하게 진 행할 수 있다.
- 이를 통해 초기 투자 비용을 절감하고, 프로젝트 성공 가능성을 높이는 동시에 내부 역량을 보완할 수 있다.

⑦ 자원(Resource)

- 프로젝트 수행에 필요한 인력, 기술, 시스템 등의 리소스를 관리

⑧ 리스크(Risk)

- 데이터 품질, 분석 오류 등 프로젝트와 관련된 위험을 식별하고 대응 방안을 수립

⑨ 의사소통(Communication)

- 이해관계자 및 팀 간 명확하고 원활한 의사소통 체계를 구축

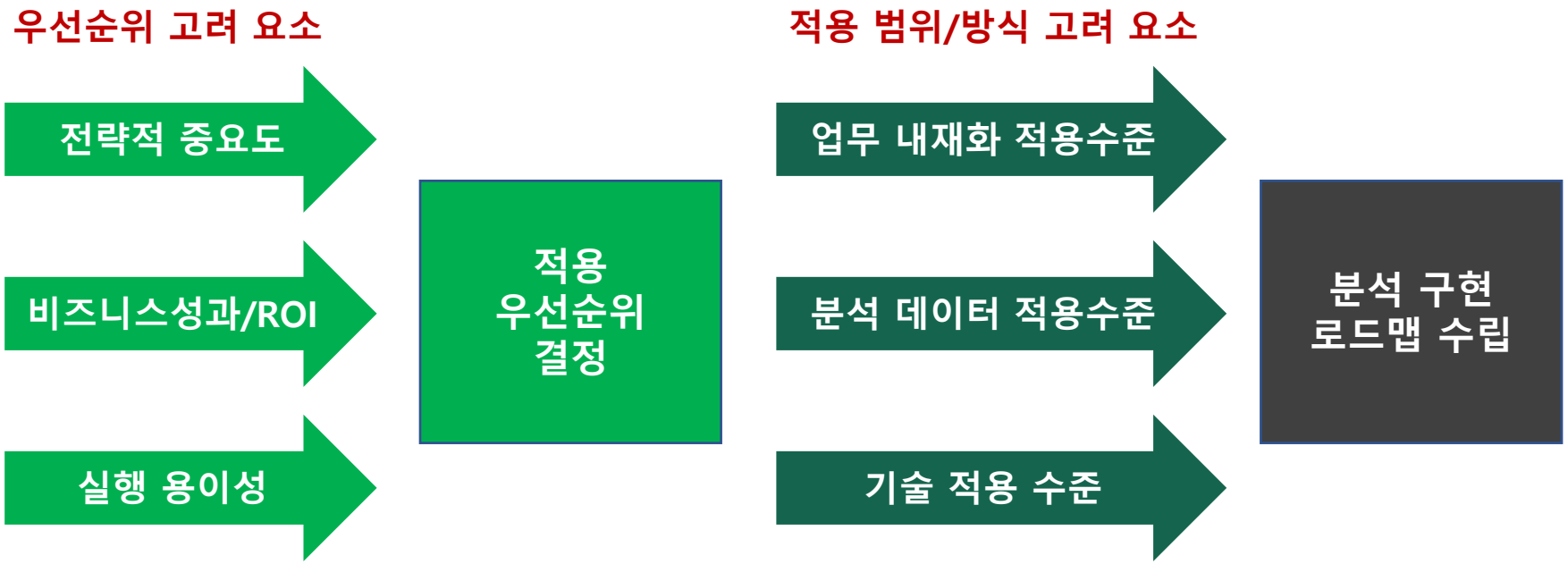
⑩ 이해관계자(Stakeholder)

- 프로젝트에 영향을 미치는 모든 이해관계자를 식별하고 관리

063 분석 마스터 플랜 ★

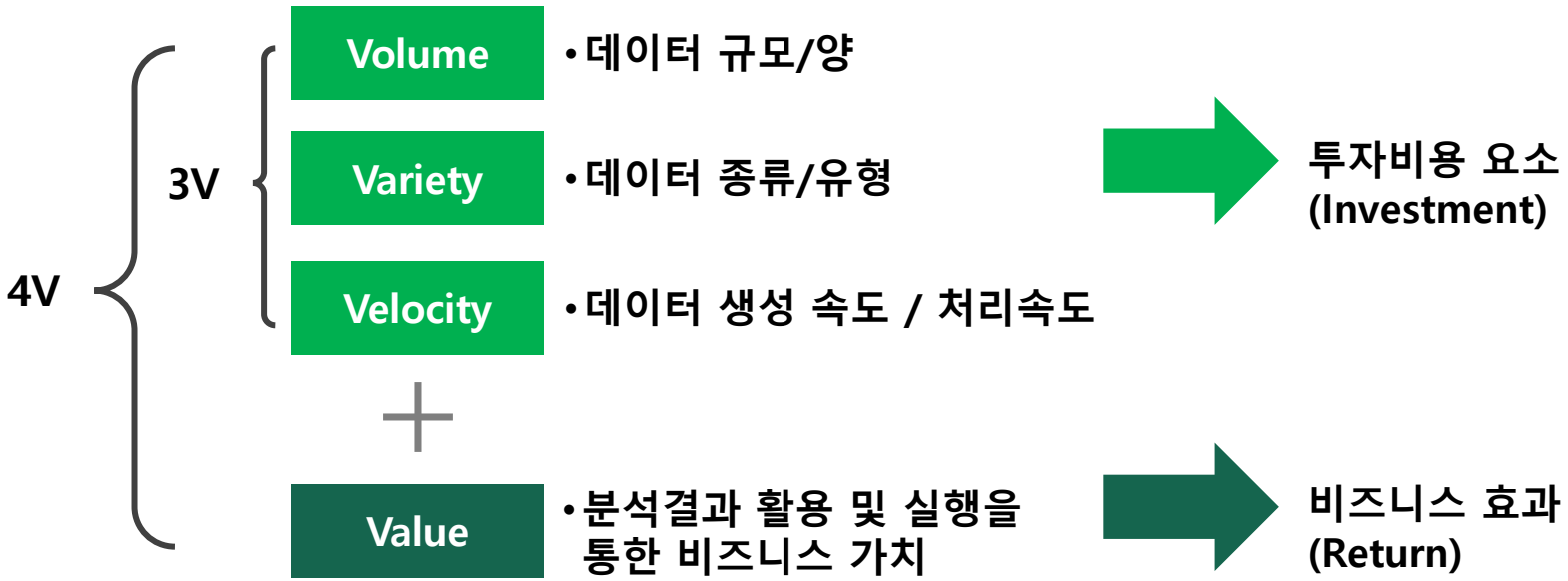
- 분석 마스터 플랜은 조직이 데이터 분석을 통해 목표를 달성하기 위해 **분석 과제 도출, 우선순위 평가, 세부 실행 계획 수립, 중. 장기 로드맵 설계**를 체계적으로 작성하는 계획

064 마스터 플랜 수립 프레임 워크 ★★



065 ROI 관점에서 보는 빅데이터 4V ★

- ROI 관점에서 빅데이터 4V는 Volume, Variety, Velocity가 투자 비용 요소로 작용하며, Value는 비즈니스 효과를 나타낸다.



빅데이터 특징을 고려한 분석 ROI 요소

066 ROI 요소를 고려한 우선순위 평가 기준 ★★

① 시급성(Urgency)
- 전략적 중요도 : 과제가 조직의 전략적 목표와 얼마나 밀접하게 연결되어 있는가
- 시점의 중요성 : 현재 해결이 필요한 문제인지, 미래에 진행해도 되는 문제인지

② 난이도(Difficulty)
- 비용 : 과제를 수행하는 데 소요되는 자금 및 자원
- 범위 : 과제가 미치는 영향과 이를 구현하는 데 필요한 노력
- 기술적 준비도 : 현재 기술과 데이터 환경에서 과제를 실행하기에 적합한 상태인지 여부



분석 우선순위 평가 기준

067 포트폴리오 사분면 분석을 통한 과제 우선순위 선정 ★

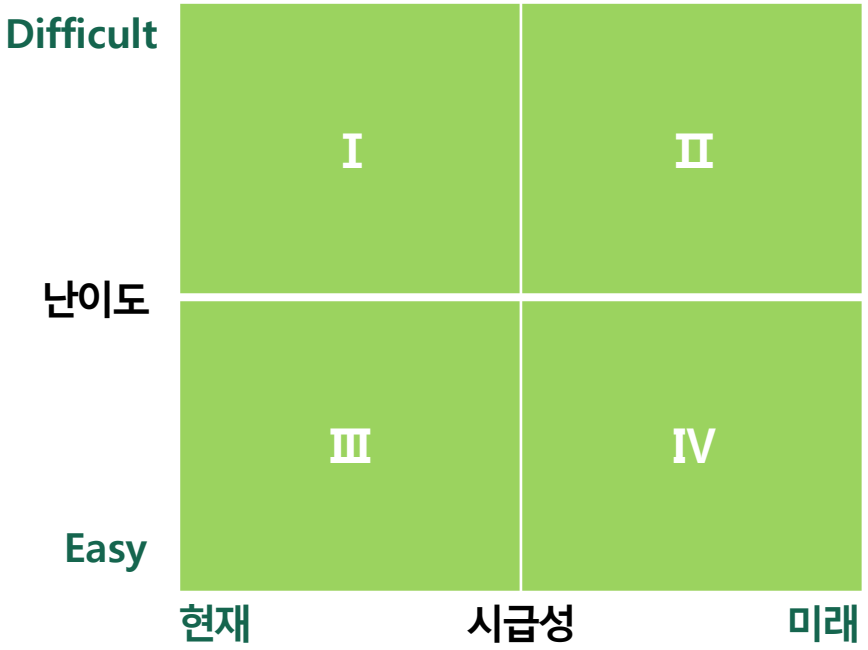
- 포트폴리오 사분면 분석은 과제를 **시급성과 난이도**라는 두 가지 기준으로 구분하여, 자원의 효율적 배분과 우선순위 결정을 돕는 기법
- 시급성이 높으면서도 난이도가 낮은 업무는 신속히 처리할 수 있어 전체 효율성을 높일 수 있기 때문에 우선적으로 수행해야 한다.

① 시급성을 기준으로 한 우선순위

- 시급성을 기준으로 과제를 수행하는 우선순위는
  Ⅲ → Ⅳ → Ⅱ 영역 순서

② 난이도를 기준으로 한 우선순위

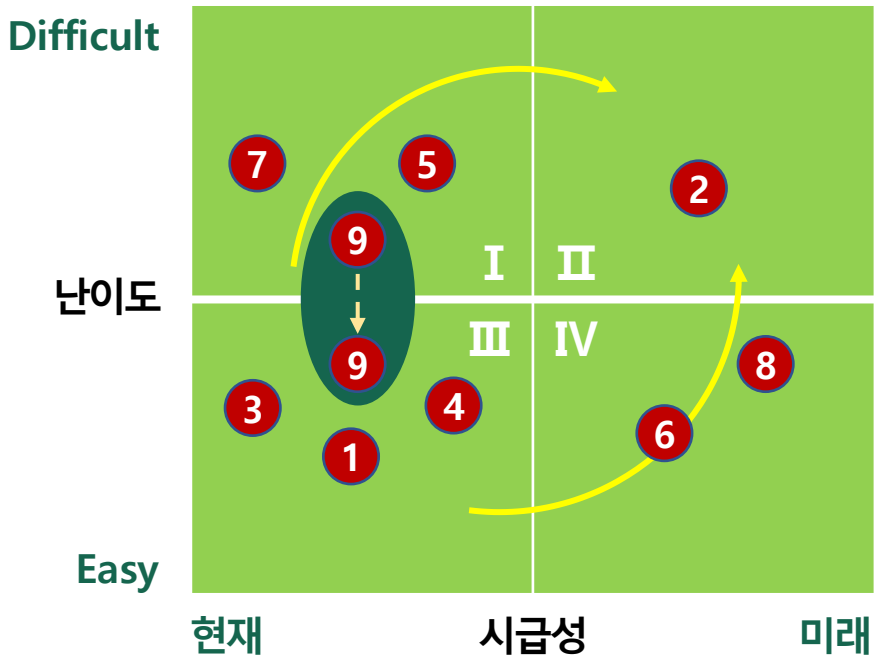
- 난이도를 기준으로 과제를 수행하는 우선순위는
  Ⅲ→ Ⅰ → Ⅱ 영역 순서



분석 과제 우선순위 선정 매트릭스

068 분석 과제 우선순위 조정 ★

- 분석 과제의 우선순위는 다양한 요소에 따라 조정할 수 있으며, **데이터 특성, 기술적 요소, 분석 범위**가 주요 기 준으로 작용



분석 과제 우선순위 선정 조정

069 분석 거버넌스 (Analytics Governance) ★

- **분석 거버넌스**는 조직 내 데이터와 분석 활동이 어떻게 관리되고, 유지되며, 규제되는지를 체계적으로 정의한 내부 관리 체계를 말한다.
- **분석 거버넌스**는 데이터를 활용하여 조직의 목표를 달 성하기 위해 **데이터 관리와 분석 과정의 표준화, 책임 소재, 규제 준수** 등을 지원한다.

070 분석 거버넌스 체계 구성 요소 ★★

▪ 분석 거버넌스 체계는 조직이 데이터를 수집하고 축적 하는 데 그치지 않고, 이를 효과적으로 분석하고 활용할 수 있도록 관리 체계를 수립하는 데 초점을 맞춘다.

① 조직(Organization)

- 분석 활동과 데이터 관리에 관련된 조직 구조 및 역할을 정의

② 과제 기획 및 운영 프로세스(Task Planning and Operation Process)

- 분석 과제를 발굴하고 기획하며, 실행 및 평가하는 일련의 프로세스

③ 분석 관련 시스템(IT 시스템 및 프로그램)

- 데이터를 수집, 처리, 분석, 저장 및 시각화하는 데 필요한 기술 인프라와 도구

④ 데이터(Data Governance)

- 데이터의 수집, 저장, 처리, 배포 및 삭제까지의 전체 라이프사이클을 관리

⑤ 분석 관련 교육 및 마인드 육성 체계(Education and Culture)

- 데이터와 분석 활용에 대한 직원들의 기술 역량 강화와 데이터 중심 마인드를 육성

071 분석 준비도(Readiness) ★★★

- 분석 준비도(Readiness)는 기업이 데이터 분석을 도입 하고 성공적으로 실행할 수 있는 환경과 역량을 갖추었 는지 파악하기 위한 진단 방법이다.
- 주요 진단 영역은 분석 업무, 인력 및 조직, 분석 기법, 분석 데이터, 분석 문화, IT 인프라 등 6가지로 구성

[분석준비도 6개 영역]

① 분석 업무 파악

- 발생한 사실 여부 파악
- 예측 분석업무
- 시뮬레이션 분석업무
- 최적화 분석업무
- 분석업무 정기적 개선

② 인력 및 조직

- 분석전문가 직무 존재
- 분석전문가 교육 훈련 프로그램
- 관리자들의 기본적 분석 능력
- 전사 분석업무 총괄 조직 존재
- 경영진 분석업무 이해 능력

2과목 데이터의 분석 기획

③ 분석기법

- 업무별 적합한 분석기법 사용
- 분석업무 도입 방법론
- 분석기법 라이브러리
- 분석기법 효과성 평가
- 분석기법 정기적 개선

④ 분석 데이터

- 분석업무를 위한 데이터 충분성
- 분석업무를 위한 데이터 신뢰성
- 분석업무를 위한 데이터 적시성
- 비구조적 데이터 관리
- 외부 데이터 활용 체계
- 기준데이터 관리

⑤ 분석문화

- 사실에 근거한 의사결정
- 관리자의 데이터 중시
- 회의 등에서 데이터 활용
- 경영진의 직관보다 데이터
- 데이터 공유 및 협업문화

⑥ IT 인프라

- 운영시스템 데이터 통합
- EAI, ETL 등 데이터 유통체계
- 분석 전용 서버 및 스토리지
- 빅데이터 분석 환경
- 통계 분석 환경
- 비주얼 분석 환경

072 분석 성숙도 ★★★

- 분석 성숙도 진단은 조직이 데이터 분석을 얼마나 효과적으로 도입하고 활용하고 있는지를 평가하기 위한 방법론이다.
- CMMI(Capability Maturity Model Integration) 모델 기반으로 개발되었으며, 조직의 데이터 분석 역량과 활용 수준을 평가하기 위해 비즈니스 부문, 조직 • 역량 부문, IT 부문을 기준으로 도입단계, 활용단계, 확산단계, 최적화 단계로 구분

2과목 데이터의 분석 기획

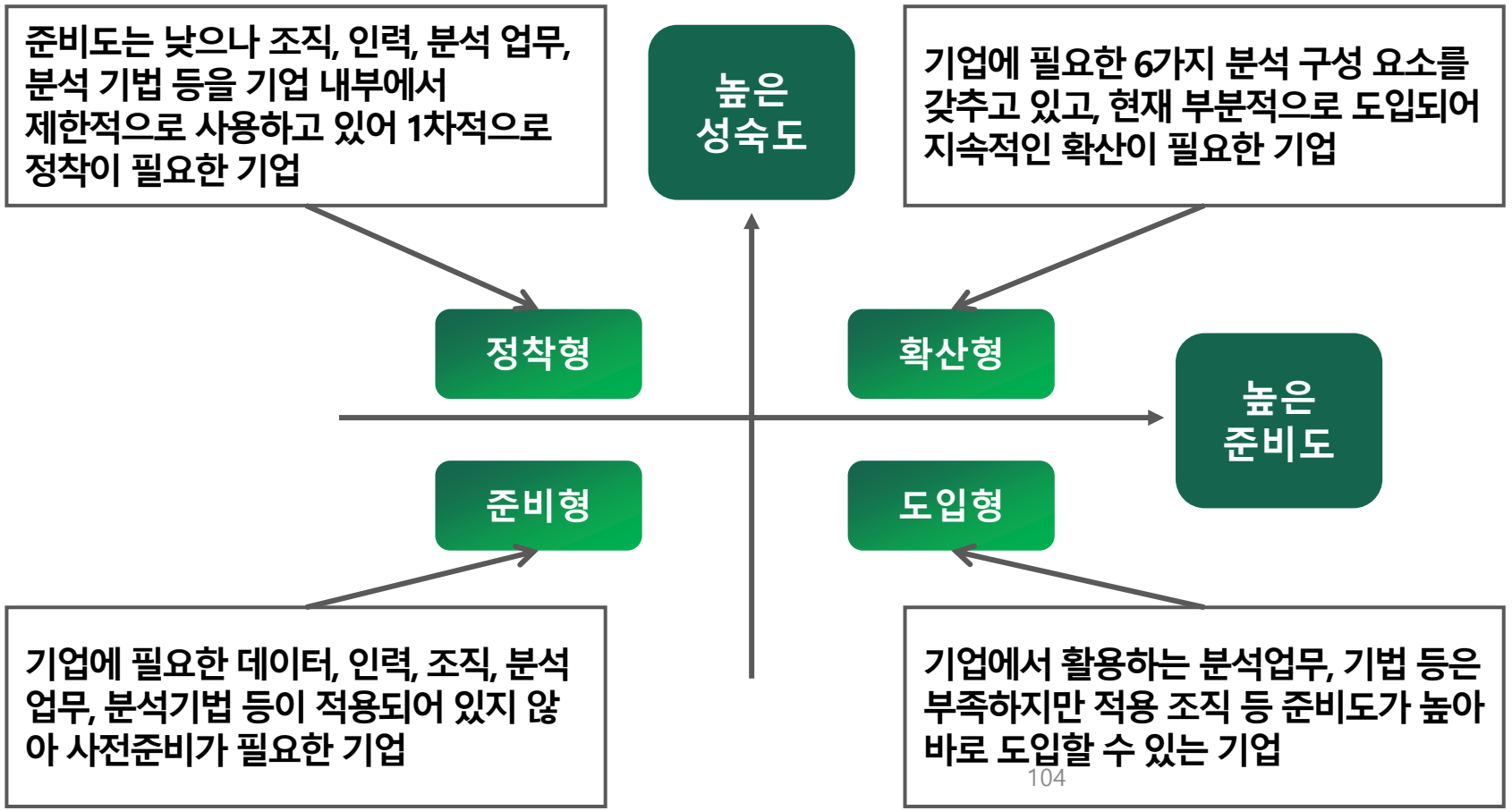
단계	도입	활용	확산	최적화
설명	분석을 시작하여 환경과 시스템을 구축	분석 결과를 실제 업무에 적용	전사 차원에서 분석을 관리하고 공유	분석을 진화 시켜서 혁신 및 성과 향상에 기여
비즈니스 부문	• 실적분석 및 통계 • 정기보고 수행 • 운영 데이터 기반	• 미래 결과 예측 • 시뮬레이션 • 운영 데이터기반	• 전사 성과 실시간 분석 • 프로세스혁신 3.0 • 분석규칙 관리 • 이벤트 관리	• 외부환경 분석 활용 • 최적화 업무 적용 • 실시간 분석 • 비즈니스 모델 진화
조직역량 부문	• 일부 부서에서 수행 • 담당자 역량에 의존	• 전문 담당부서 수행 • 분석 기법 도입 • 관리자가 분석 수행	• 전사 모든 부서 수행 • 분석 CoE 조직 운영 • 데이터 사이언티스트 확보	• 데이터 사이언스 그룹 • 경영진 분석 활용 • 전략 연계
IT부문	• 데이터웨어 하우스 • 데이터 마트 • ETL/EAI • OLAP	• 실시간 대시보드 • 통계분석 환경	• 빅데이터 관리 환경 • 시뮬레이션 최적화 • 비주얼 분석 • 분석 전용 서버	• 분석 협업 환경 • 분석 Sandbox • 프로세스 내재화 • 빅데이터 분석

분석 성숙도 모델

073 분석 수준 진단 결과 유형별 특성 구분 ★

- 기업의 데이터 분석 준비도와 성숙도에 따라 기업을 **준비형, 정착형, 도입형, 확산형**으로 구분
- 각 유형은 **데이터 분석 도입 및 활용 정도**를 나타내며, 조직의 현황과 전략적 방향성을 설정하는 데 유용

2과목 데이터의 분석 기획



사분면 분석

074 데이터 거버넌스 (Data Governance) ★★

- 데이터 거버넌스는 조직의 모든 데이터를 효과적으로 관리하기 위해 **정책, 지침, 표준, 운영조직 및 책임 체계를 수립**하고 운영하는 **표준화된 관리 체계**이다.
- 이를 통해 데이터의 가용성, 유용성, 통합성, 보안성, 안정성을 확보하여 조직 전반의 **데이터 활용 역량을 강화하는데 필수적이다**.
- 데이터 거버넌스는 조직의 데이터를 효과적으로 관리하기 위해 **마스터 데이터, 메타데이터, 데이터 사전**을 중요한 **관리 대상**으로 설정

075 데이터 거버넌스의 구성 요소 ★★

- 데이터 거버넌스는 조직 내 데이터를 효율적으로 관리 하고 활용하기 위해 **원칙, 조직, 프로세스**의 세 가지 주 요 요소로 구성

① 원칙(Principles)

- 데이터 관리와 유지보수를 위한 지침과 가이드라인을 제시

② 조직(Organization)

- 데이터 관리와 관련된 역할과 책임을 명확히 정의하고, 이를 수행할 조직과 인력을 구성

③ 프로세스(Processes)

- 데이터 관리와 운영을 위한 일련의 체계적 활동과 작업 절차

076 데이터 거버넌스 체계 요소 ★★★

▪ 데이터 거버넌스 체계는 조직이 데이터를 효율적으로 관리하고 활용하기 위한 **표준화, 관리 체계, 저장소 관 리, 표준화 활동**으로 구성

① 데이터 표준화

- **데이터 표준 용어 설명** : 데이터의 명확한 정의와 용어 표준화
- **명명 규칙(Naming Rule)** : 데이터 필드, 테이블, 파일 등에 대해 명명 규칙 수립
- **메타데이터 구축** : 데이터 속성, 출처, 관계 등을 정의한 메타데이터 체계 수립
- **데이터 사전 구축** : 데이터 구조와 속성을 체계적으로 정리한 데이터 사전 마련

② 데이터 관리 체계

- **관리 원칙 수립** : 메타데이터 관리 지침 및 데이터 사전 관리 기준 정의
- **데이터 품질 관리** : 데이터의 정확성, 완전성, 일관성을 유지하기 위한 프로세스 수립
- **접근 권한 관리** : 데이터 접근 권한 및 보안 정책 설정

③ 데이터 저장소 관리

- **저장소 구성** : 전사 차원의 데이터 저장소를 설계 및 운영
- **워크플로우 및 관리 소프트웨어 지원** : 데이터 저장, 변경, 접근 등에 대한 워크플로우와 관리 도구 제공
- **시스템 간 인터페이스** : 데이터 관리 대상 시스템과 의 연계 및 통제
- **사전영향평가** : 데이터 구조 변경 시 발생할 수 있는 영향을 사전에 평가

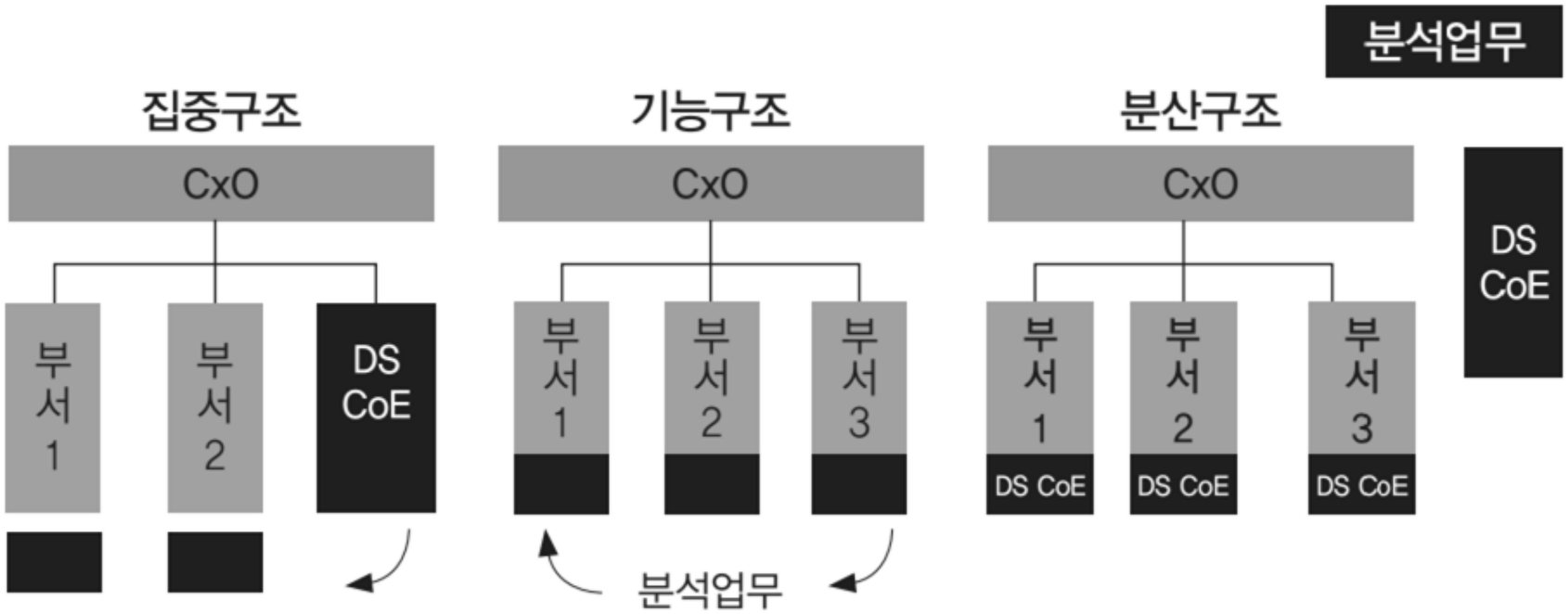
④ 표준화 활동

- **표준 준수 여부 점검** : 데이터 표준화 및 관리 체계가 잘 준수되고 있는지 정기적으로 평가
- **모니터링** : 데이터 품질 및 보안 상태를 지속적으로 점검
- **지속적 개선** : 점검 결과를 바탕으로 데이터 표준 및 관리 체계 개선

077 분석 업무 수행 주체에 따른 3가지 조직구조 유형 ★★

- 데이터 분석 조직의 구조는 기업이 처한 환경과 특성에 따라 다르며, 분석 업무 수행 주체에 따라 크게 **집중형, 기능 중심형, 분산형**으로 나눌 수 있다.

2과목 데이터의 분석 기획



- 전사 분석 업무를 별도의 분석 전담 조직에서 담당
- 전략적 중요도에 따라 분석 조직이 우선순위를 정해서 진행 가능
- 현업 업무부서의 분석업무와 이중화/이원화 가능성 높음

- 일반적인 분석 수행 구조
- 별도 분석조직이 없고 해당 업무 부서에서 분석수행
- 전사적 핵심 분석이 어려우며, 부서 현황 및 실적통계 등 과거 실적에 국한된 분석 수행 가능성 높음

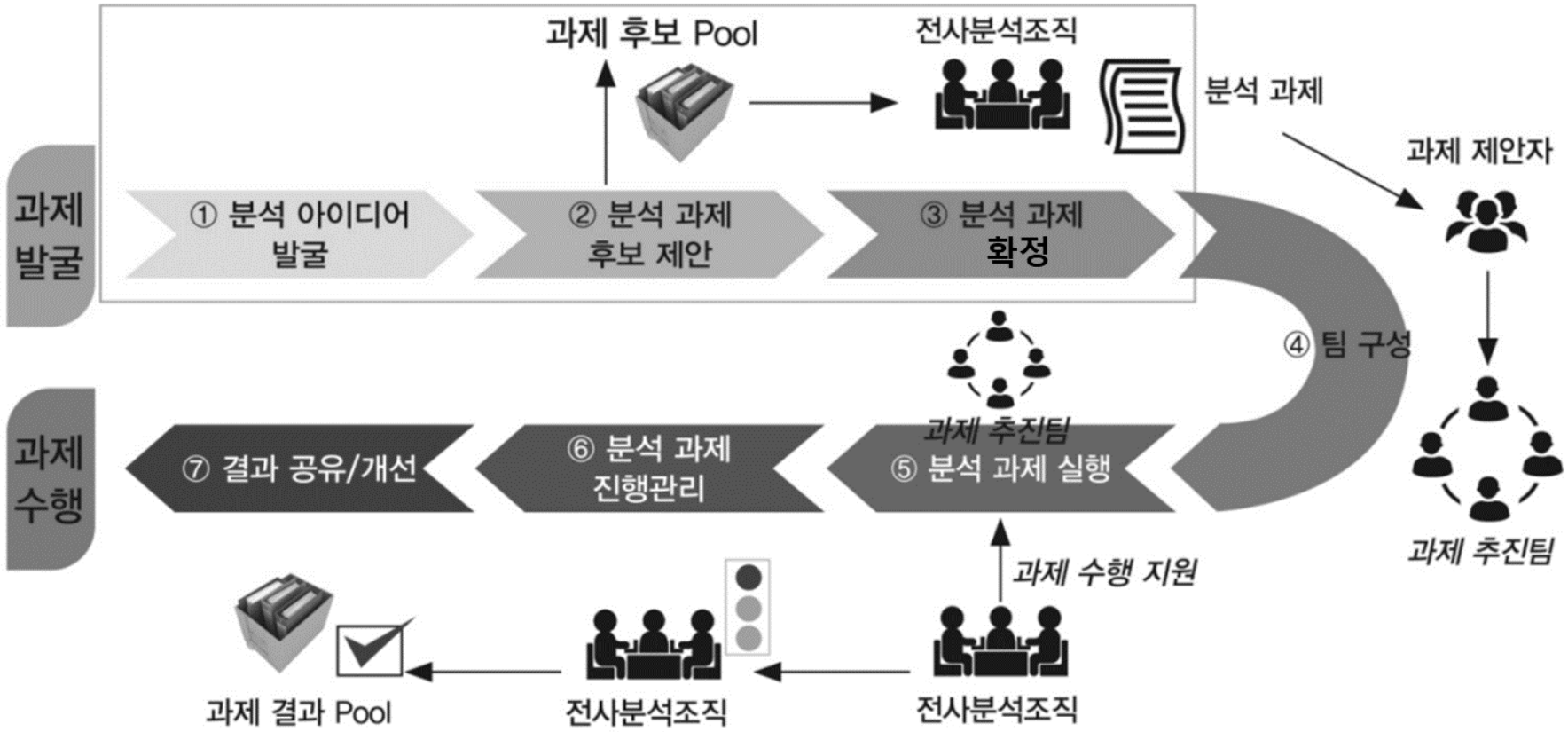
- 분석조직 인력들을 현업부서로 직접 배치하여 분석 업무 수행
- 전사 차원의 우선순위 수행
- 분석 결과에 따른 신속한 Action가능
- 베스트 프랙티스 공유가능
- 부서 분석 업무와 역할 분담 명확히 해야 함

분석업무 수행 주체에 따른 3가지 유형

078 분석 과제 관리 프로세스 수립 ★

- 분석 마스터 플랜의 성공적인 초기 실행 이후, 조직 내에 서 지속적으로 발생하는 분석 니즈와 기회를 체계적으 로 관리하기 위해 분석 과제 관리 프로세스를 수립하는 것이 중요하다.

- 분석 과제 관리 프로세스는 **과제 발굴과 과제 수행 및 모 니터링의 두 단계**로 나누어진다.


과제 후보 Pool
전사분석조직
분석 과제
과제 제안자
과제 발굴
① 분석 아이디어 발굴
② 분석 과제 후보 제안
③ 분석 과제 확정
④ 팀 구성
과제 추진팀
과제 수행
⑦ 결과 공유/개선
⑥ 분석 과제 진행관리
⑤ 분석 과제 실행
과제 추진팀
과제 수행 지원
과제 결과 Pool
전사분석조직
전사분석조직

분석 과제 관리 프로세스

079 분석 교육 및 변화관리 ★

- **분석 교육의 목표는** 단순히 데이터 분석 툴의 사용법을 교육하는 것을 넘어, 조직 구성원의 **분석 역량을 확보하고 강화**하여 업무에 효과적으로 적용할 수 있도록 하는데 중점을 둔다.

- 이를 통해 조직 전체가 분석 기반의 **업무 문화를 내재화** 할 수 있도록 해야 한다.