

PART 2. 데이터 분석 기획

데이터 준전문가

ADSP, Advanced Data Analytics semi-Professional

류영표 강사

ryp1662@gmail.com



류영표

Youngpyo Ryu

동국대학교 수학과/응용수학 석사수료

現 Upstage AI X 네이버 부스트 캠프 AI tech 1~4기 멘토

前 Innovation on Quantum & CT(IQCT) 이사

前 한국파스퇴르연구소 Image Mining 인턴(Deep learning)

前 (주)셈웨어(수학컨텐츠, 데이터 분석 개발 및 연구인턴)

강의 경력

- 현대자동차 연구원 강의 (인공지능/머신러닝/딥러닝/강화학습)
- (주)모두의연구소 Aiffel 1기 퍼실리테이터(인공지능 교육)
- 인공지능 자연어처리(NLP) 기업데이터 분석 전문가 양성과정 멘토
- 공공데이터 청년 인턴 / SW공개개발자대회 멘토
- 고려대학교 선도대학 소속 30명 딥러닝 집중 강의
- 이젠 종로 아카데미(파이썬, ADSP 강사) / 강남 : ADSP
- 최적화된 도구(R/파이썬)을 활용한 애널리스트 양성과정(국비과정) 강사
- 한화, 하나금융사 교육
- 인공지능 신뢰성 확보를 위한 실무 전문가 자문위원
- 보건 · 바이오 AI활용 S/W개발 및 응용전문가 양성과정 강사
- Upstage AI X KT 융합기술원 기업교육 모델최적화 담당 조교

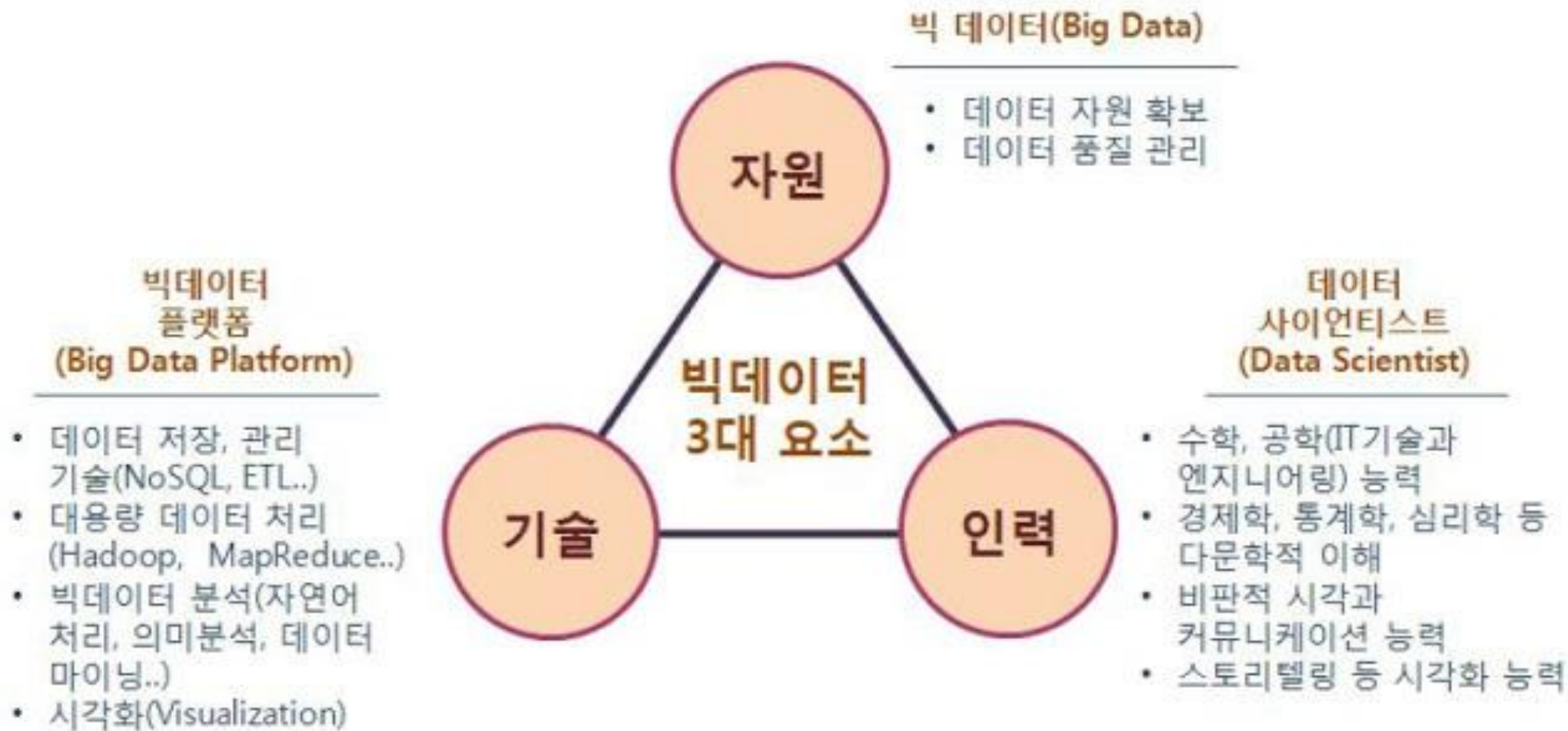
주요 프로젝트 및 기타사항

- 개인 맞춤형 당뇨병 예방·관리 인공지능 시스템 개발 및 고도화(안정화)
- 폐플라스틱 이미지 객체 검출 경진대회 3위
- 인공지능(AI)기반 데이터 사이언티스트 전문가 양성과정 1기 수료
- 제 1회 산업 수학 스터디 그룹 (질병에 영향을 미치는 유전자 정보 분석)
- 제 4,5회 산업 수학 스터디 그룹 (피부암, 유방암 분류)
- 빅데이터 여름학교 참석 (혼잡도를 최소화하는 새로운 노선 건설 위치의 최적화 문제)

분석 기획

- 실제 분석을 수행하기 앞서 분석을 수행할 **과제의 정의** 및 **의도했던 결과를 도출**할 수 있도록 이를 적절하게 **관리 할 수 있는 방안을 사전에 계획**하는 일련의 작업
- 분석 과제 및 프로젝트를 직접 수행하는 것은 아니지만, 어떠한 목표를 달성하기 위하여 어떤 데이터를 가지고 어떤 방식으로 수행할지에 대한 **일련의 계획을 수립하는 작업**이기 때문에 성공적인 분석 결과를 도출하기 위한 중요한 사전 작업.
- 분석을 기획한다는 것은 해당 문제 영역에 대한 전문성 역량 및 수학 / 통계학적 지식을 활용한 분석 역량과 분석의 도구인 데이터 및 프로그래밍 기술 역량에 대한 **균형 잡힌 시각**을 가지고 방향성 및 계획을 수립해야 한다는 것을 의미.

빅데이터의 주요 요소 3가지



데이터 사이언티스트의 역량



분석의 대상 (What)

분석의
방법
(How)

Known

Un-Known

Known

Un-Known

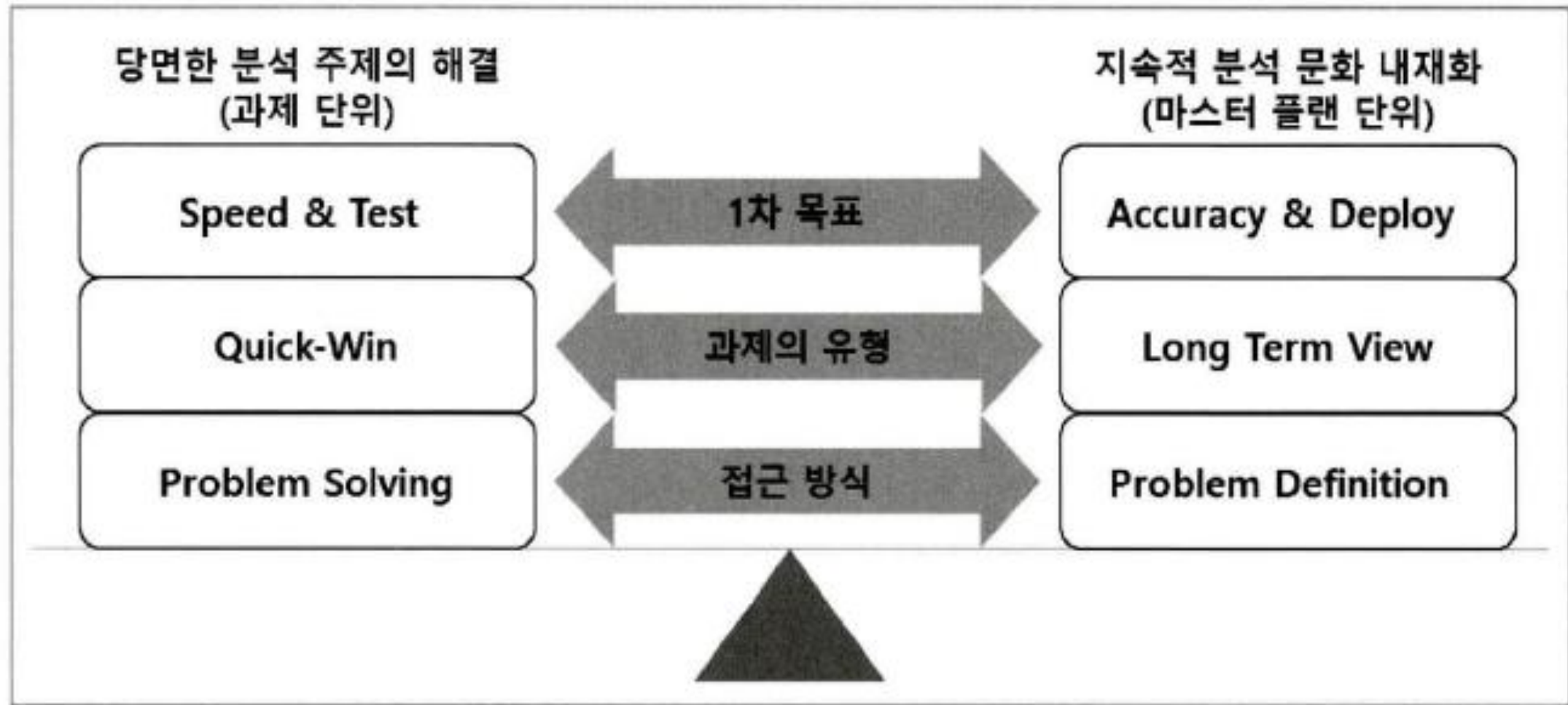
① Optimization

③ Insight

② Solution

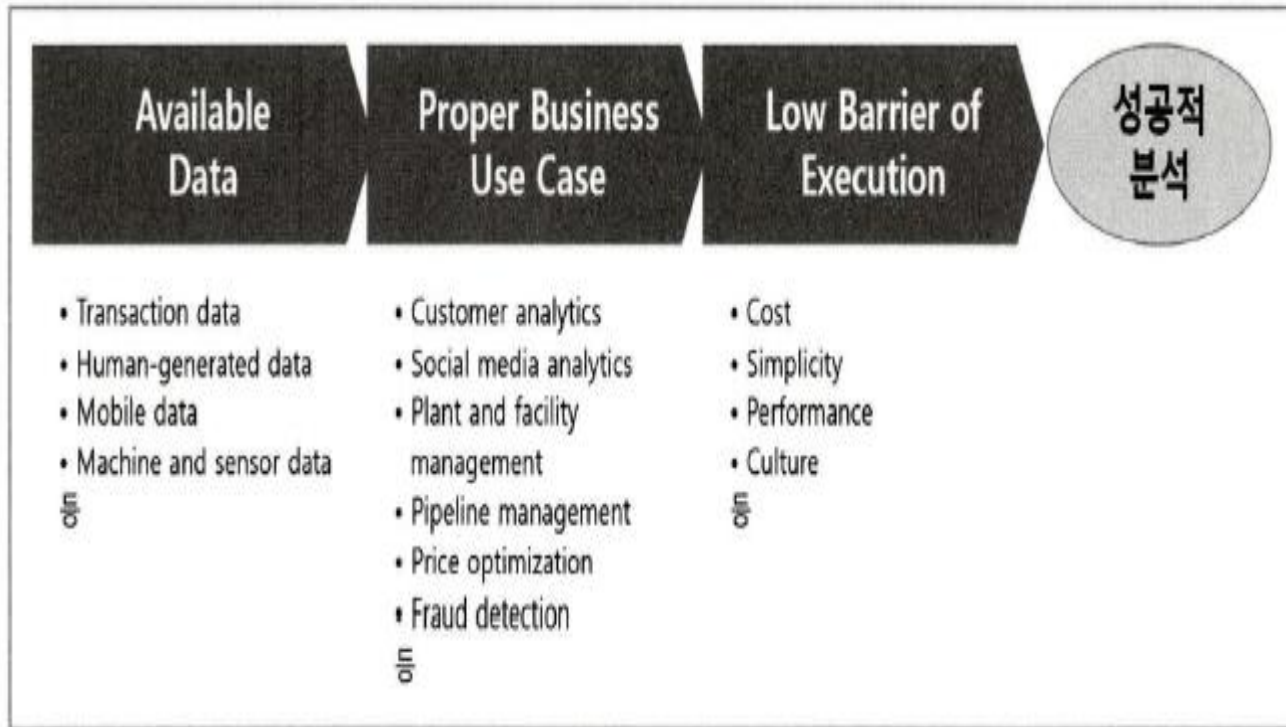
④ Discovery

목표시점 별 분석 기획 방안

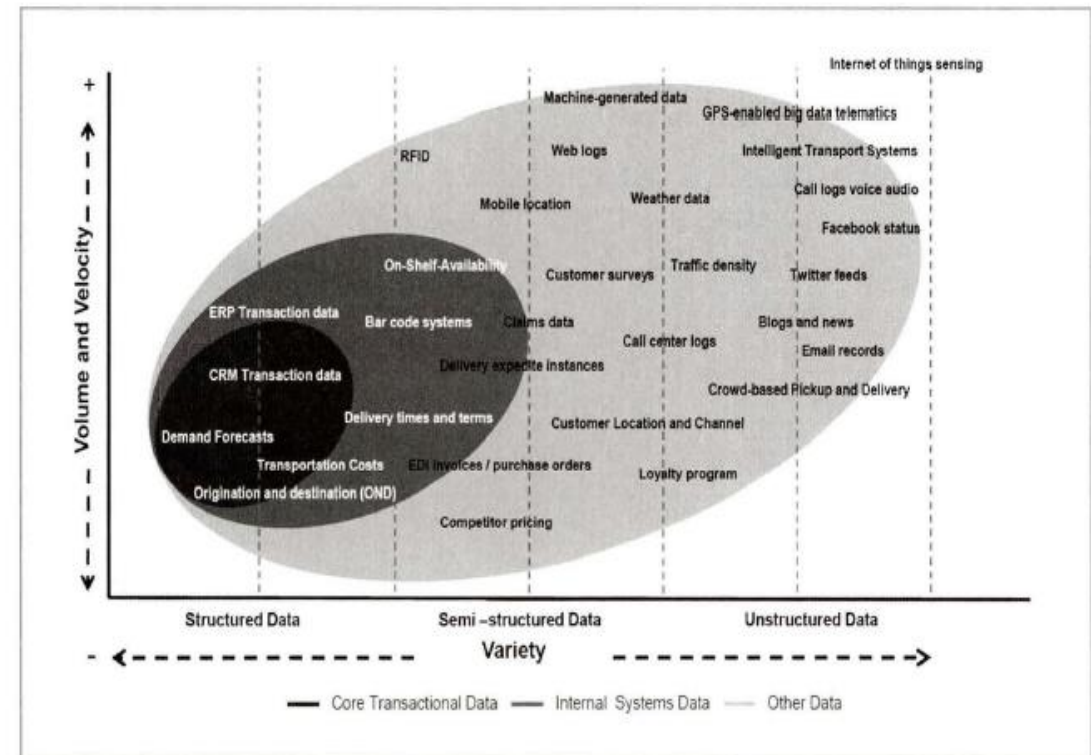


[그림 III-1-3] 목표 시점 별 분석 기획 방안

분석 기획시 고려사항



[그림 III-1-6] 분석 기획 시 고려사항



[그림 III-1-5] 다양한 데이터 유형

- 분석가는 3가지의 기본 역량에 더해 프로젝트 관리 역량 및 리더십 역량이 필요
- IT + Business + Communication + **Project Management + Leadership**

분석 기획시 고려사항

- 데이터에 대한 고려
 - 분석을 위한 데이터의 확보가 우선이며, 데이터의 유형에 따라 적용 가능한 솔루션 및 분석방법이 다르기 때문에 유형에 따라 분석이 선행적으로 이루어져야 한다.
- Use case의 탐색
 - 기존에 유사한 케이스 또는 시나리오를 최대한 활용

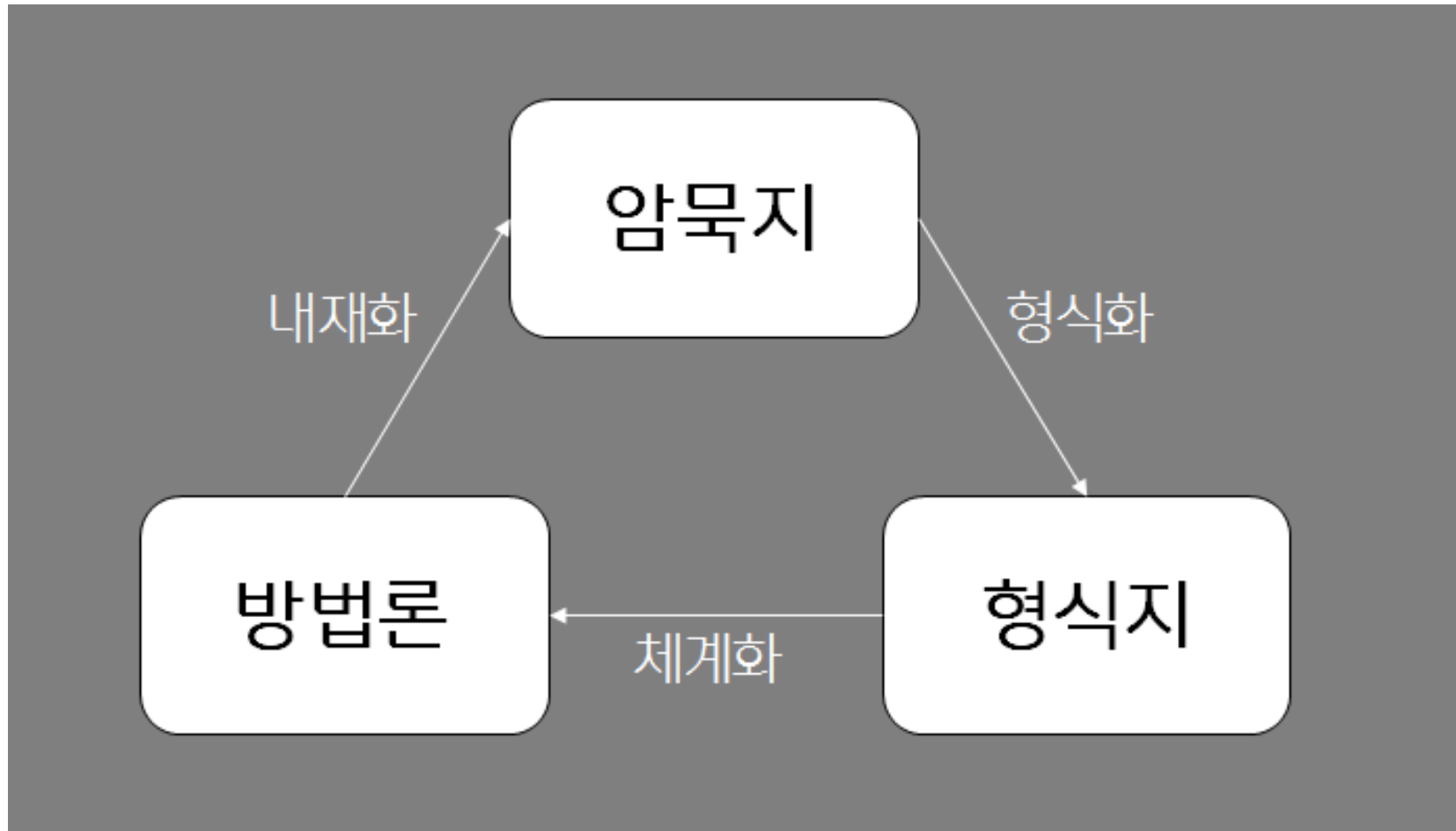
종류	정형 데이터	반정형 데이터	비정형 데이터
특징	<ul style="list-style-type: none">• 데이터 자체로 분석가능• RDB구조의 데이터• 데이터베이스로 관리	<ul style="list-style-type: none">• 데이터로 분석이 가능하지만 해석이 불가능하며 데이터 정보를 활용해야 해석이 가능	<ul style="list-style-type: none">• 데이터 자체로 분석이 불가능,• 특정한 처리 프로세스를 거쳐 분석데이터로 분경 후 분석
유형	<ul style="list-style-type: none">• ERP, CRM, SCM등• 정보시스템	<ul style="list-style-type: none">• 로그데이터, 모바일 데이터, 센싱데이터	<ul style="list-style-type: none">• 영상, 음성, 문자 등

분석 방법론

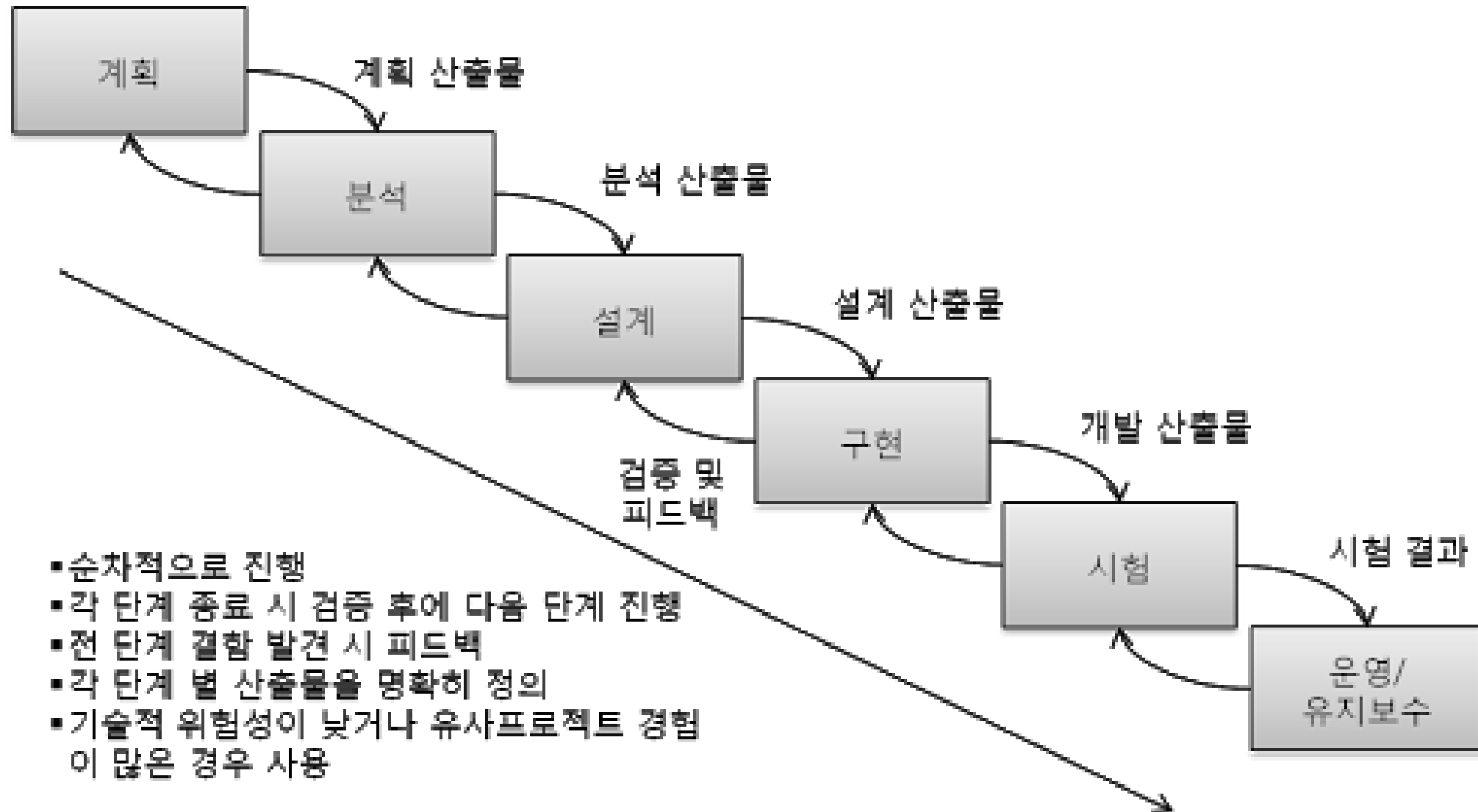
- 데이터 분석이 효과적으로 기업 내에 정착하기 위해서는 이를 체계화한 절차와 방법이 정리된 데이터 **분석 방법론의 수립**이 필수적.
- 방법론은 절차, 방법, 도구와 기법, 템플릿과 산출물로 구성되어 어느 정도 지식만 있으면 활용 가능.
- **데이터 기반 의사 결정 필요성**
 1. 경험과 감에 따른 의사결정 -> 데이터 기반의 의사결정
 2. 기업의 합리적 의사결정을 가로막는 장애요소:

고정관념, 편향된 생각, 프레이밍 효과(문제의 표현 방식에 따라 동일한 사건이나 상황임에도 불구하고 개인의 판단이나 선택이 달라질 수 있는 현상 등)

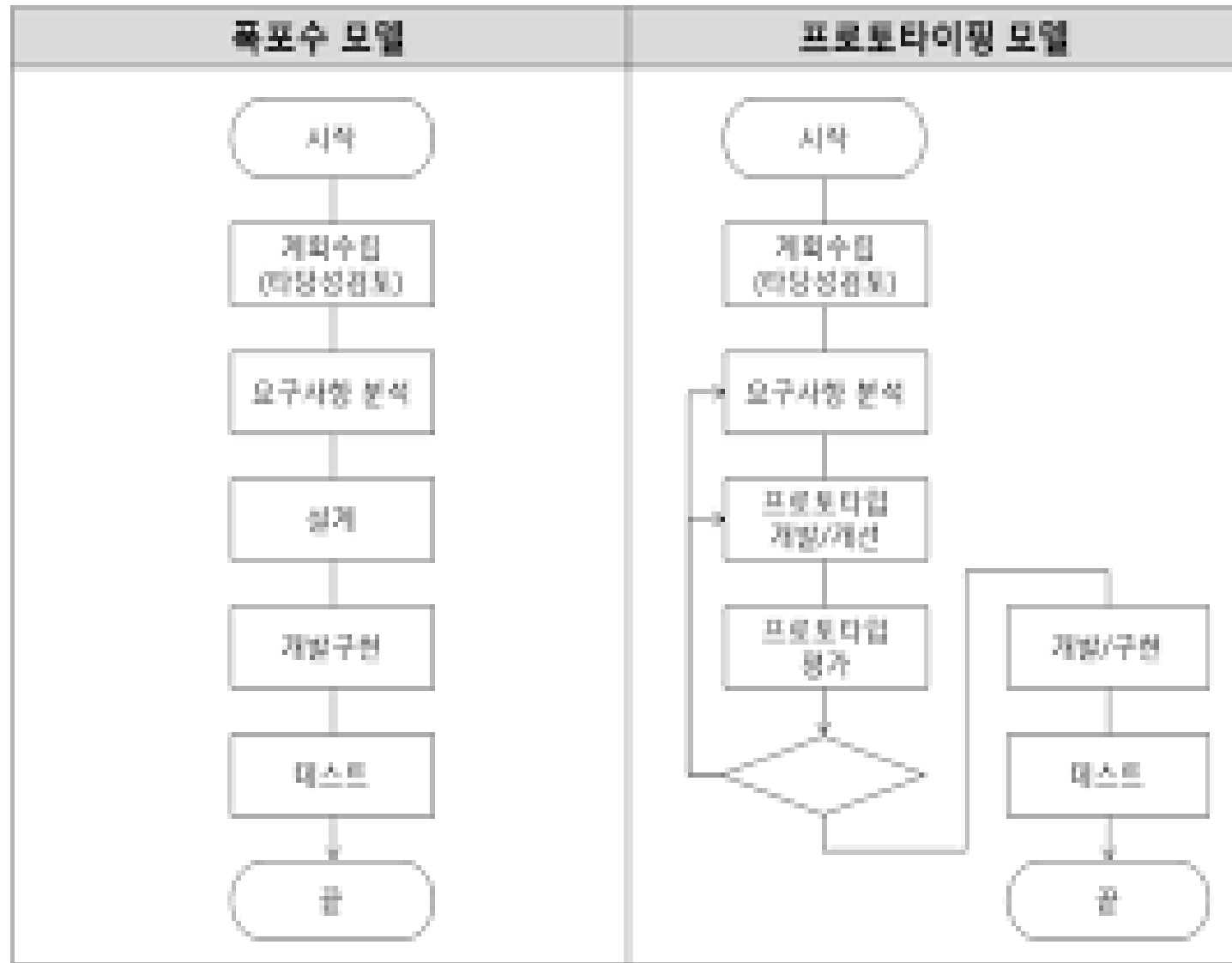
분석 방법론



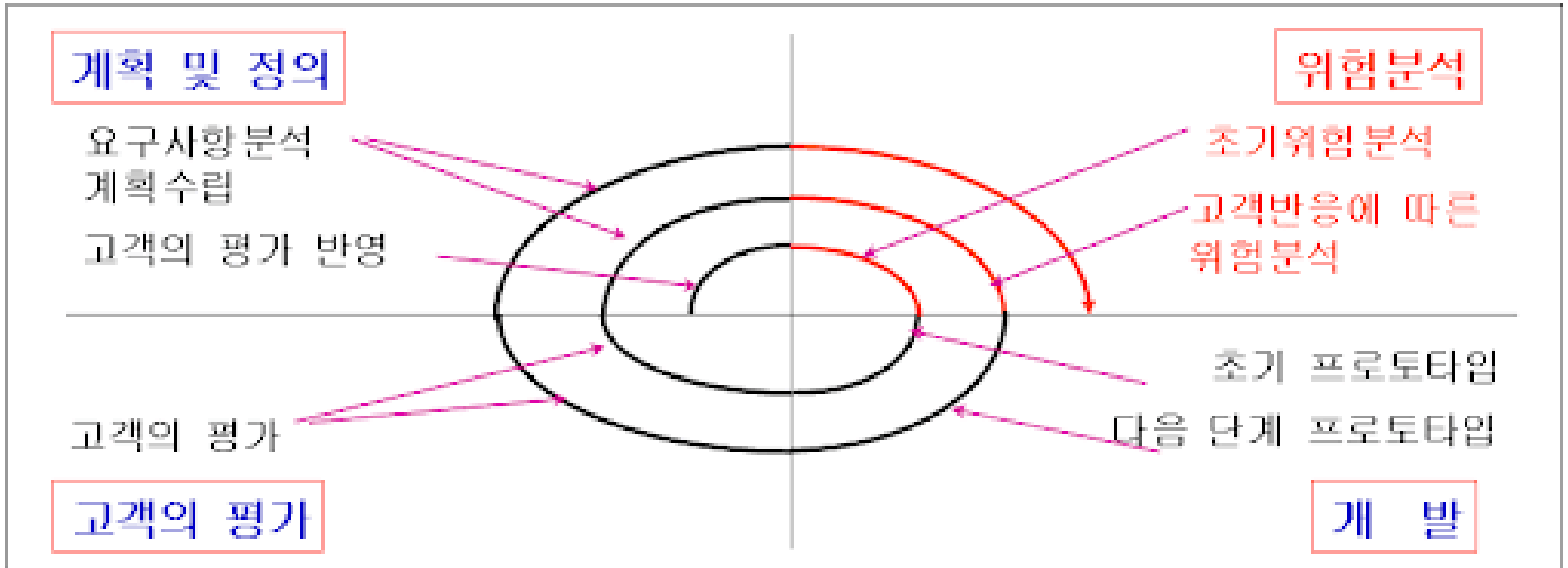
분석 방법론-폭포수 모델(Waterfall Model)



분석 방법론-프로토타입 모델(Prototype model)



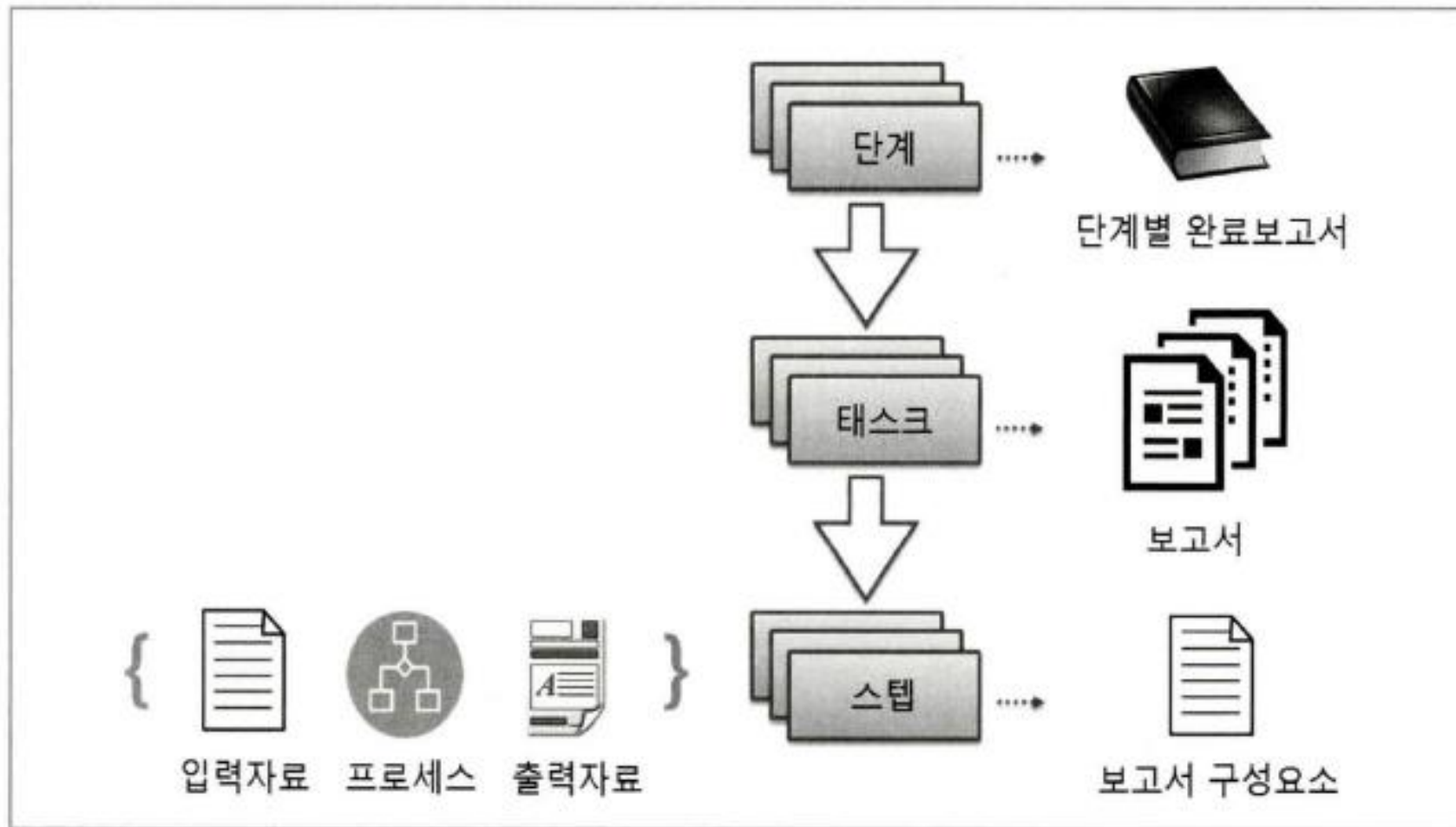
분석 방법론-나선형 모델(Spiral model)



분석 방법론-나선형 모델(Spiral model)

구분	폭포수 모델	프로토타입 모델	나선형 모델
Concept			
특징	순차적 접근	프로로 타입 개발	위험분석~개발 반복수행
장점	이해가 쉽고 관리가 편함	요구분석 용이 개발 타당성 검증가능	위험성 감소와 변경에 유연한 대처
단점	초기 요구분석이 어려워 후반 문제발생 가능성 有	프로토 타입 평가 후 취소 시 폐기비용발생	단계 반복에 따른 공정관리가 어려움

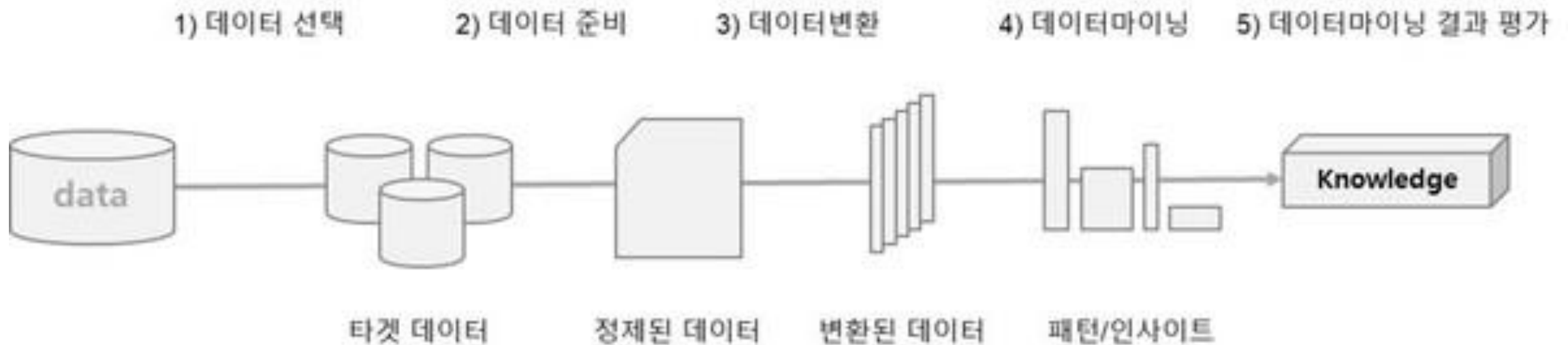
방법론의 구성



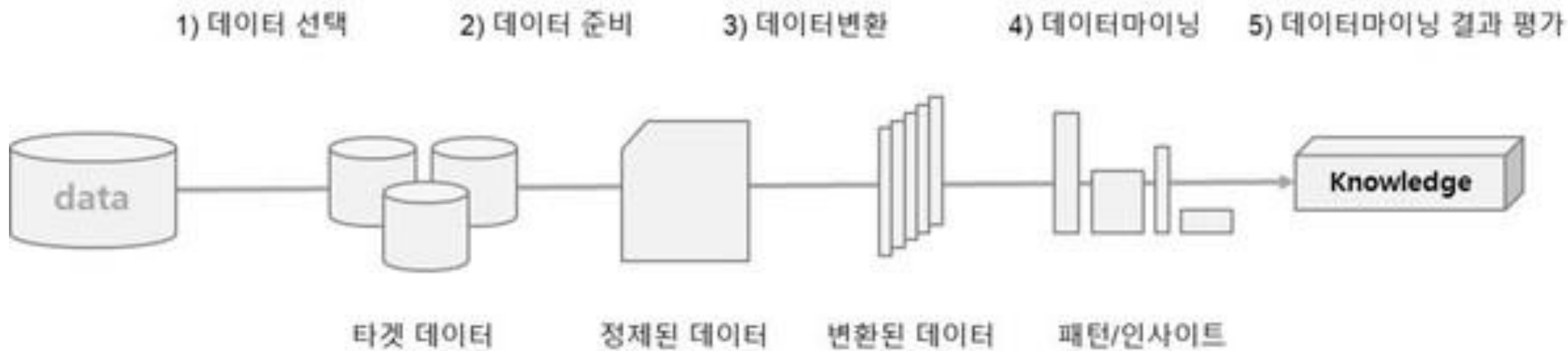
[그림 III-1-8] 방법론의 구성

KDD 분석 방법론

- KDD(Knowledge Discovery in Database) 분석 방법론은 1996년 Fayyad가 소개한 방법론으로 데이터를 통해 통계적 **패턴**이나 **지식을 찾을 수 있도록 정리한 데이터마이닝** 프로세스이다. 데이터마이닝, 기계학습, 인공지능, 패턴인식, 데이터 시각화에 응용 될 수 있는 구조를 갖고 있다. KDD 분석 방법론은 데이터셋 선택, 데이터 전처리, 데이터 변환, 데이터마이닝, 결과 평가로 이루어져 있다.



KDD 분석 방법론



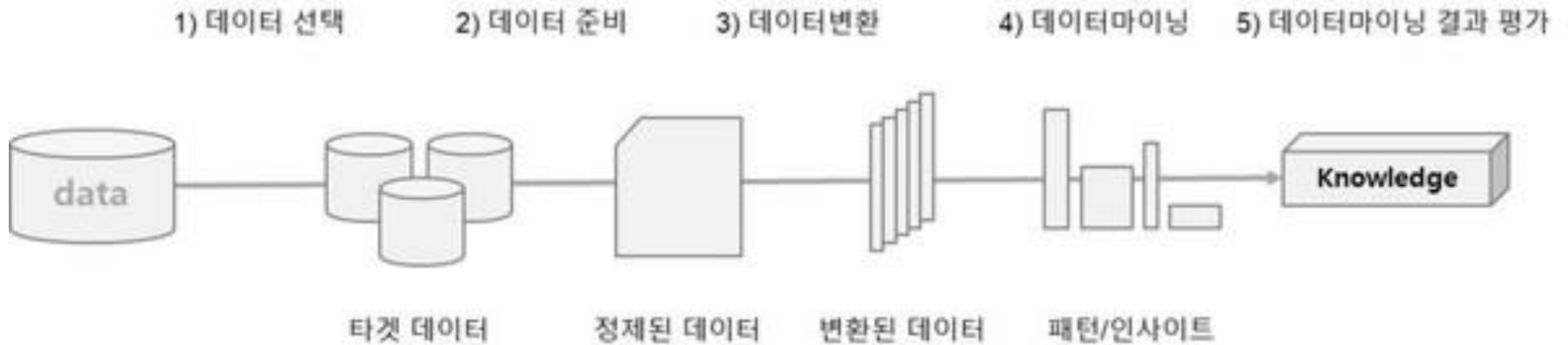
1) 데이터 선택(Selection)

- 데이터셋 선택에 앞서 분석 대상의 **비즈니스 도메인**에 대한 이해와 **프로젝트 목표 설정이 필수**이며 데이터베이스 또는 원시 데이터에서 **분석에 필요한 데이터를 선택**하는 단계.

2) 데이터 전처리(Preprocessing)

- 추출된 분석 대상용 데이터 셋에 포함되어 있는 잡음(Noise)과 이상치(Outlier), 결측치(Missing value)를 식별하고 필요시 제거하거나 **의미있는 데이터로 재처리하여 데이터 셋을 정제**하는 단계

KDD 분석 방법론



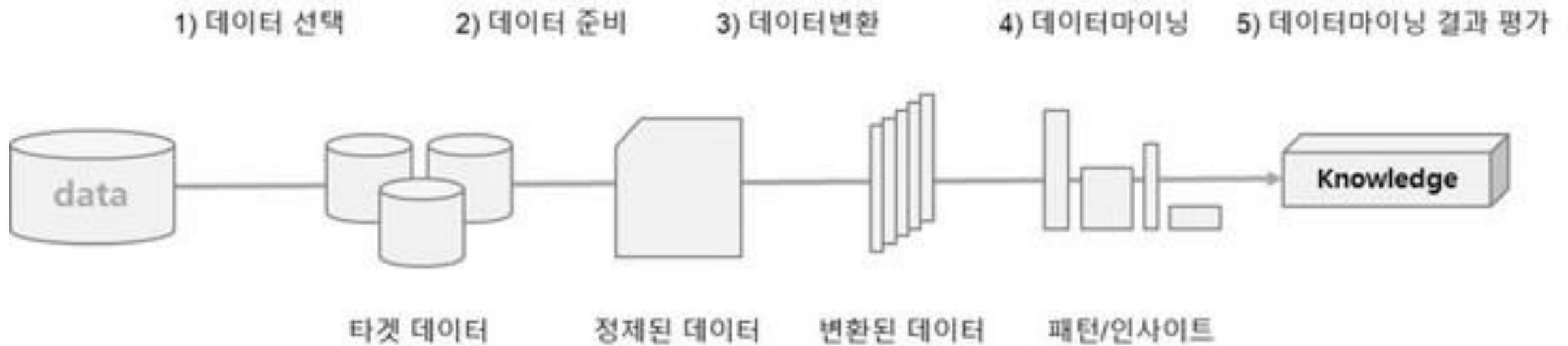
3) 데이터 변환(Transformation)

- 데이터 전처리 과정을 통해 정제된 데이터에 분석 목적에 맞게 변수를 생성, 선택하고 **데이터의 차원을 축소하여 효율적으로 데이터 마이닝을 할 수 있도록 변경하는 단계** / 프로세스를 진행하기 위해 학습용 데이터와 검증용 데이터로 분리하는 단계

4) 데이터 마이닝(Data Mining)

- 학습용 데이터를 이용하여 분석목적에 맞는 데이터 마이닝 기법을 선택하고, 적절한 알고리즘을 적용하여 데이터 마이닝 작업을 수행 / **필요에 따라 전처리와 데이터 변환 프로세스를 추가로 실행하여 최적의 결과를 산출.**

KDD 분석 방법론

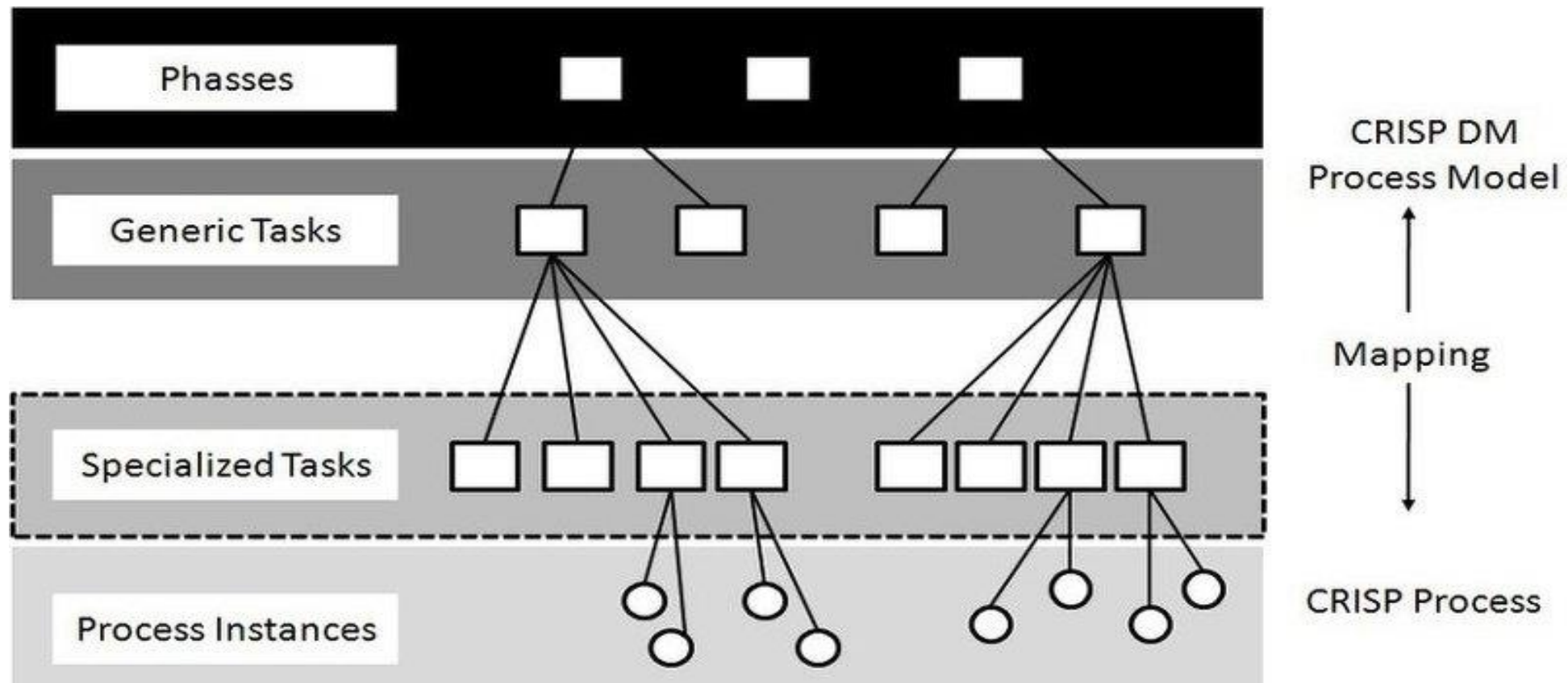


5) 데이터 마이닝 결과 평가(Interpretation / Evaluation)

- 데이터 마이닝 **결과에 대한 해석과 평가**, 그리고 분석목적과의 일치성을 확인.
- 데이터 마이닝을 통해 발견한 **지식을 업무에 활용**하기 위한 방안 마련의 단계
- 필요에 따라 데이터 선택 프로세스에서 데이터 마이닝 프로세스를 반복 수행.

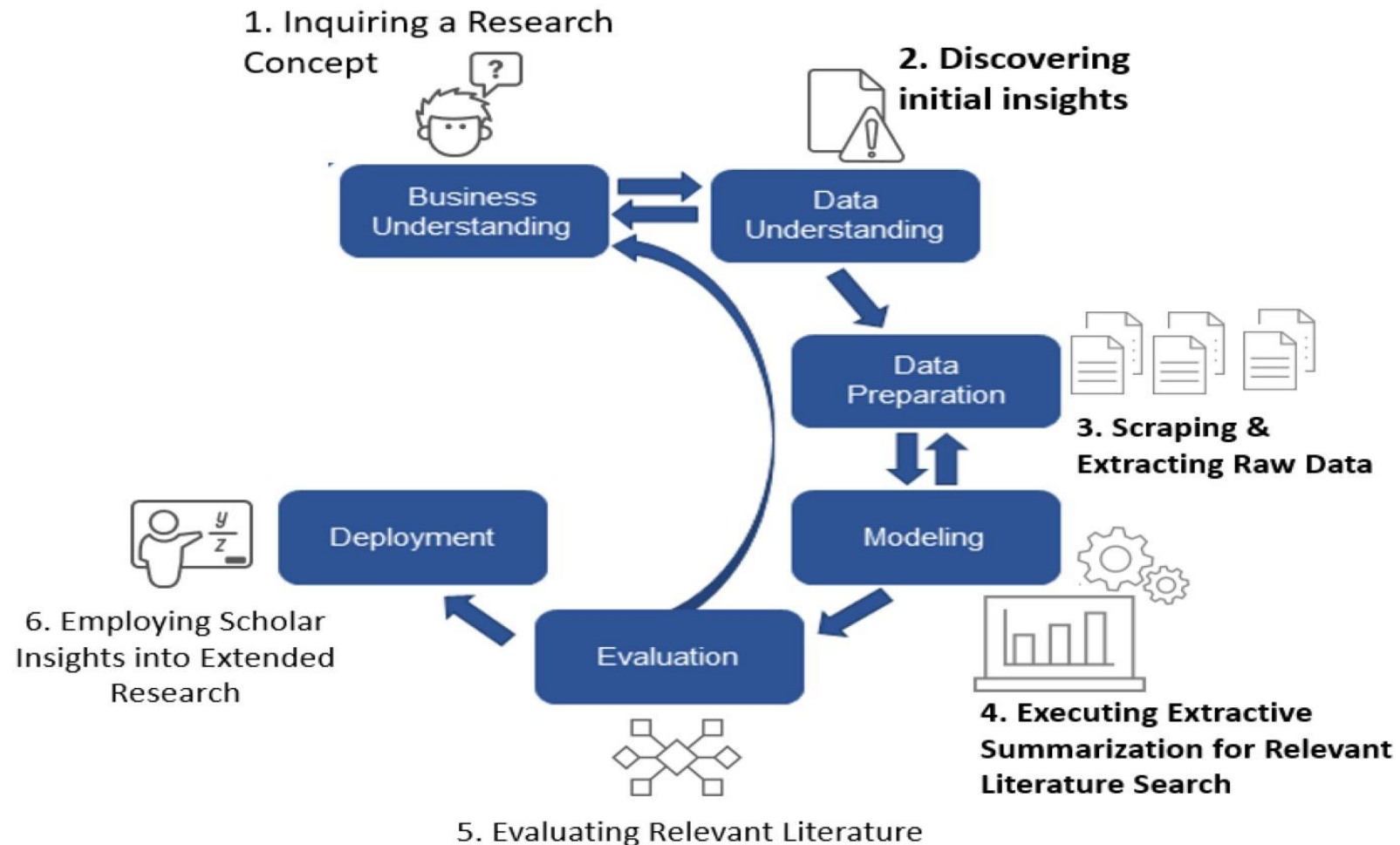
CRISP-DM 분석 방법론

- CRISP-DM(Cross Industry Standard Process for Data Mining) 방법론은 전 세계에서 가장 많이 사용되는 데이터마이닝 표준 방법론으로 단계, 일반 과제, 세부과제, 프로세스 실행 등의 4가지 레벨로 구성된 계층적 프로세스 모델이기도 하다.



CRISP-DM의 프로세스

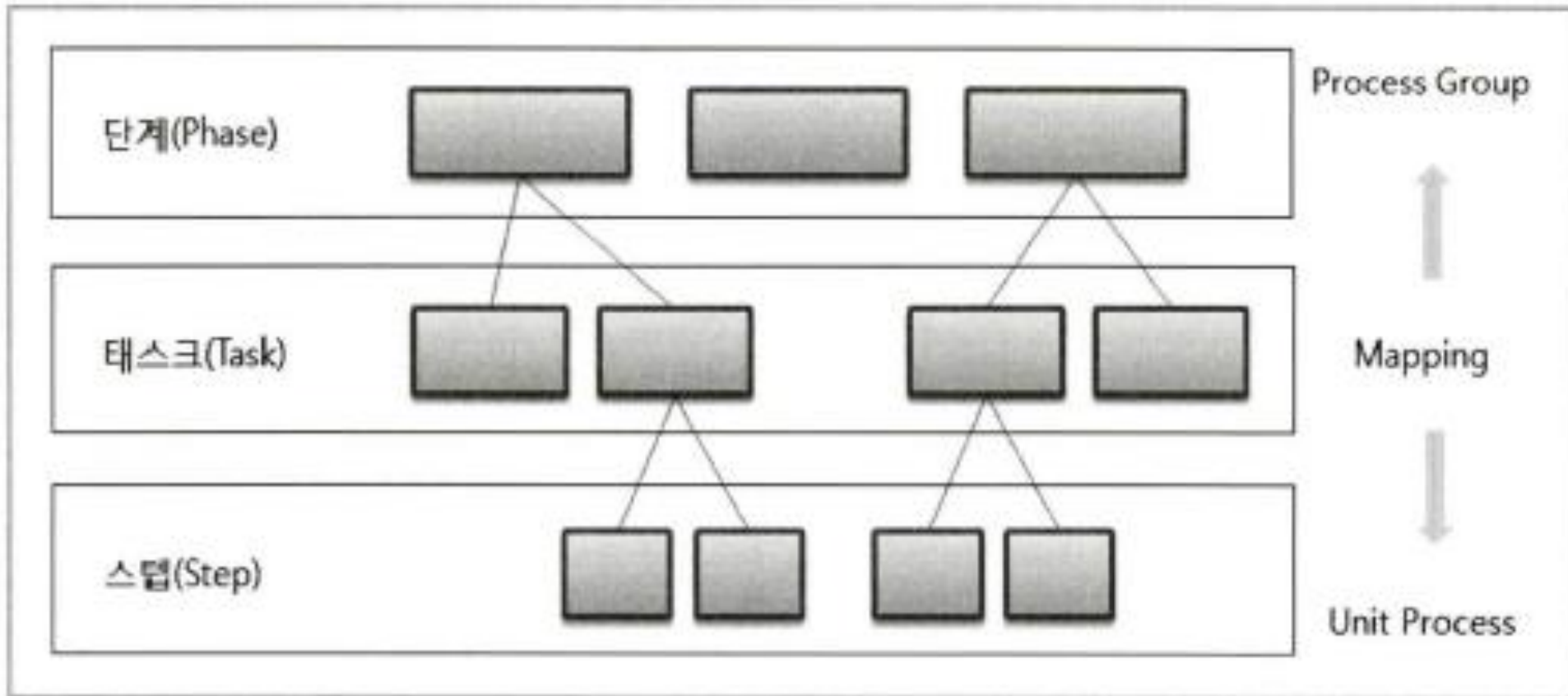
- CRISP-DM의 절차는 6단계로 구성되어 있는데 각 단계들은 순차적으로 진행되는 것이 아니라, 필요에 따라 단계 간의 반복 수행을 통해 분석의 품질을 향상시킨다.



KDD와 CRISP-DM의 비교

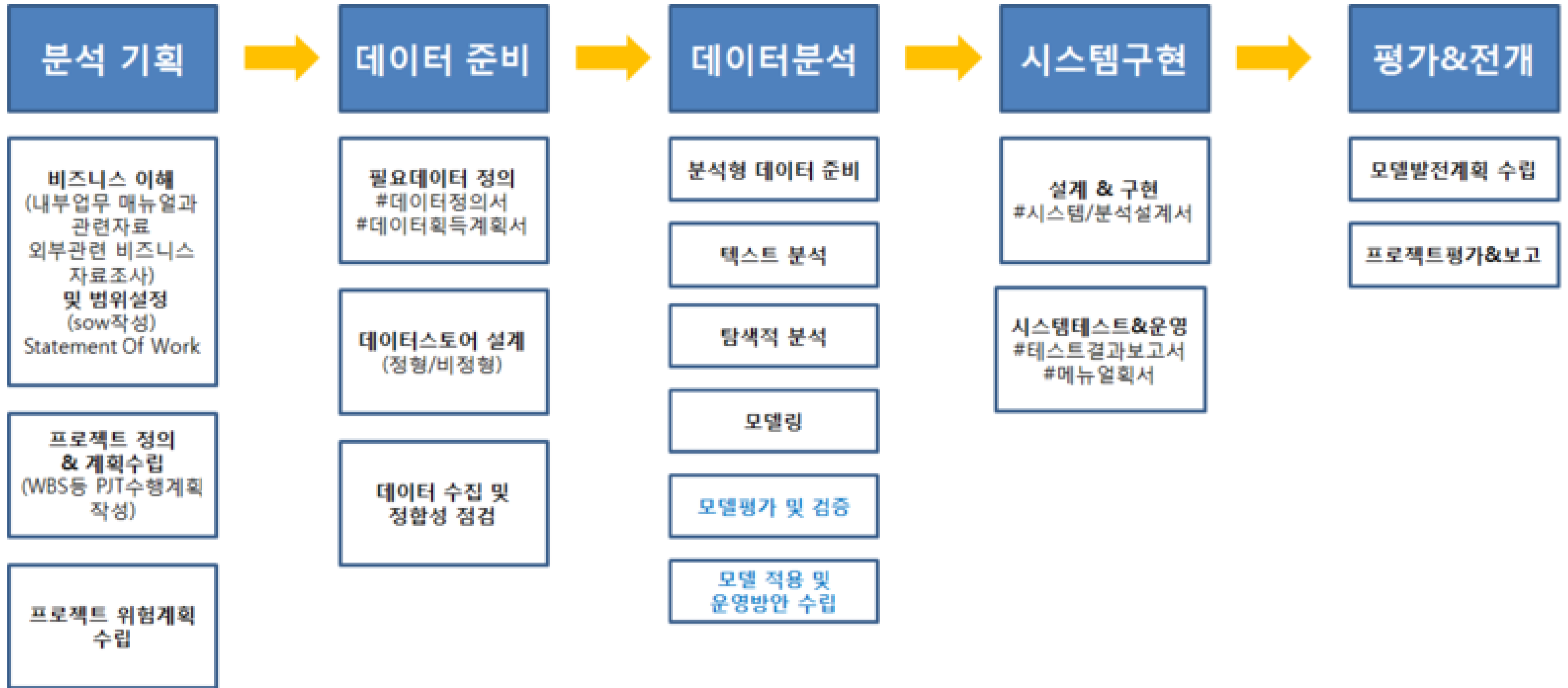
KDD	CRISP-DM
분석대상 비즈니스 이해	업무 이해
데이터셋 선택	데이터의 이해
데이터 전처리	
데이터 변환	데이터 준비
데이터 마이닝	모델링
데이터 마이닝 결과 평가	평가
데이터 마이닝 활용	전개

빅데이터 분석의 계층적 프로세스

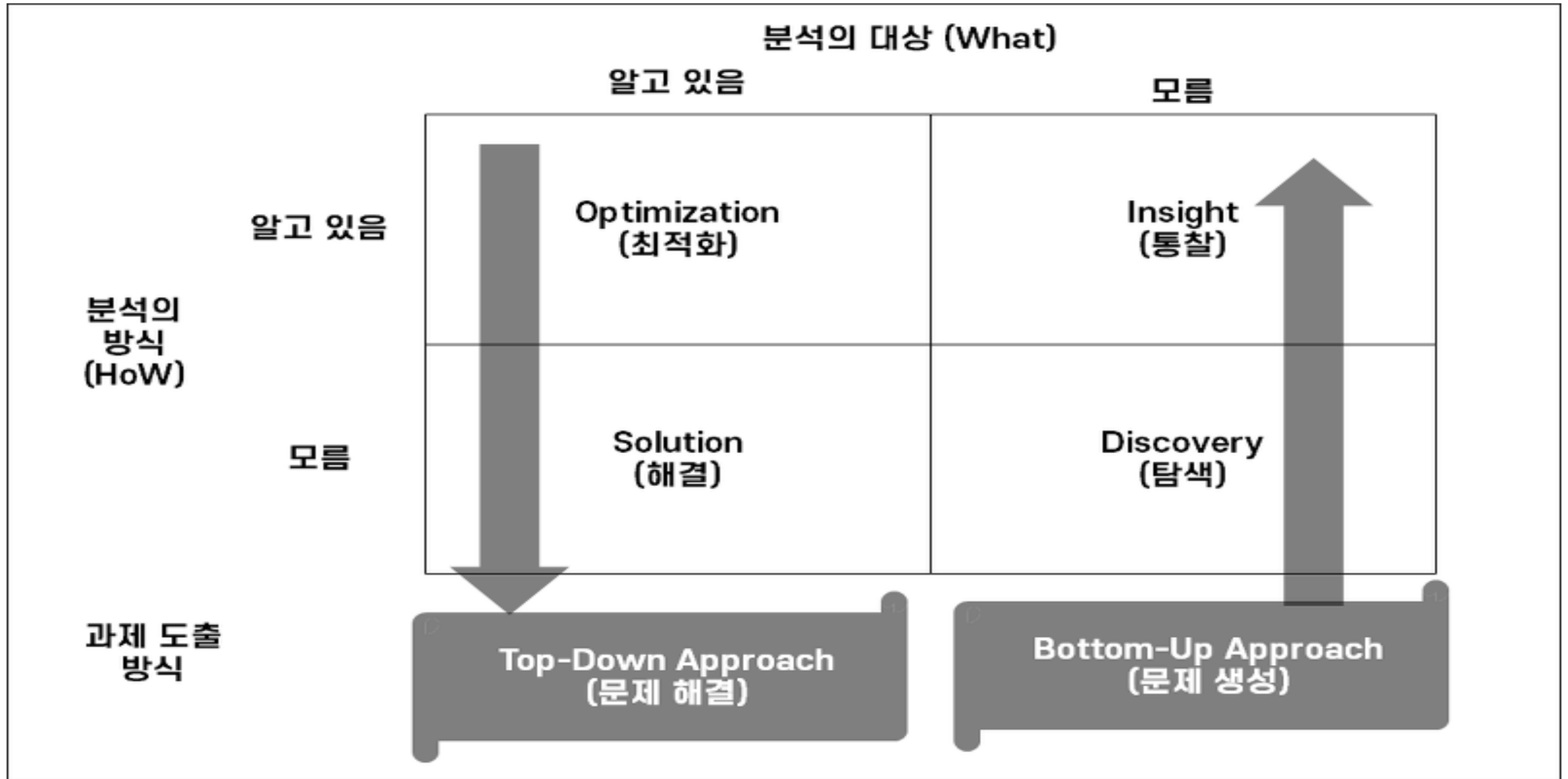


[그림 III-1-12] 빅데이터 분석 방법론 3개층 구조

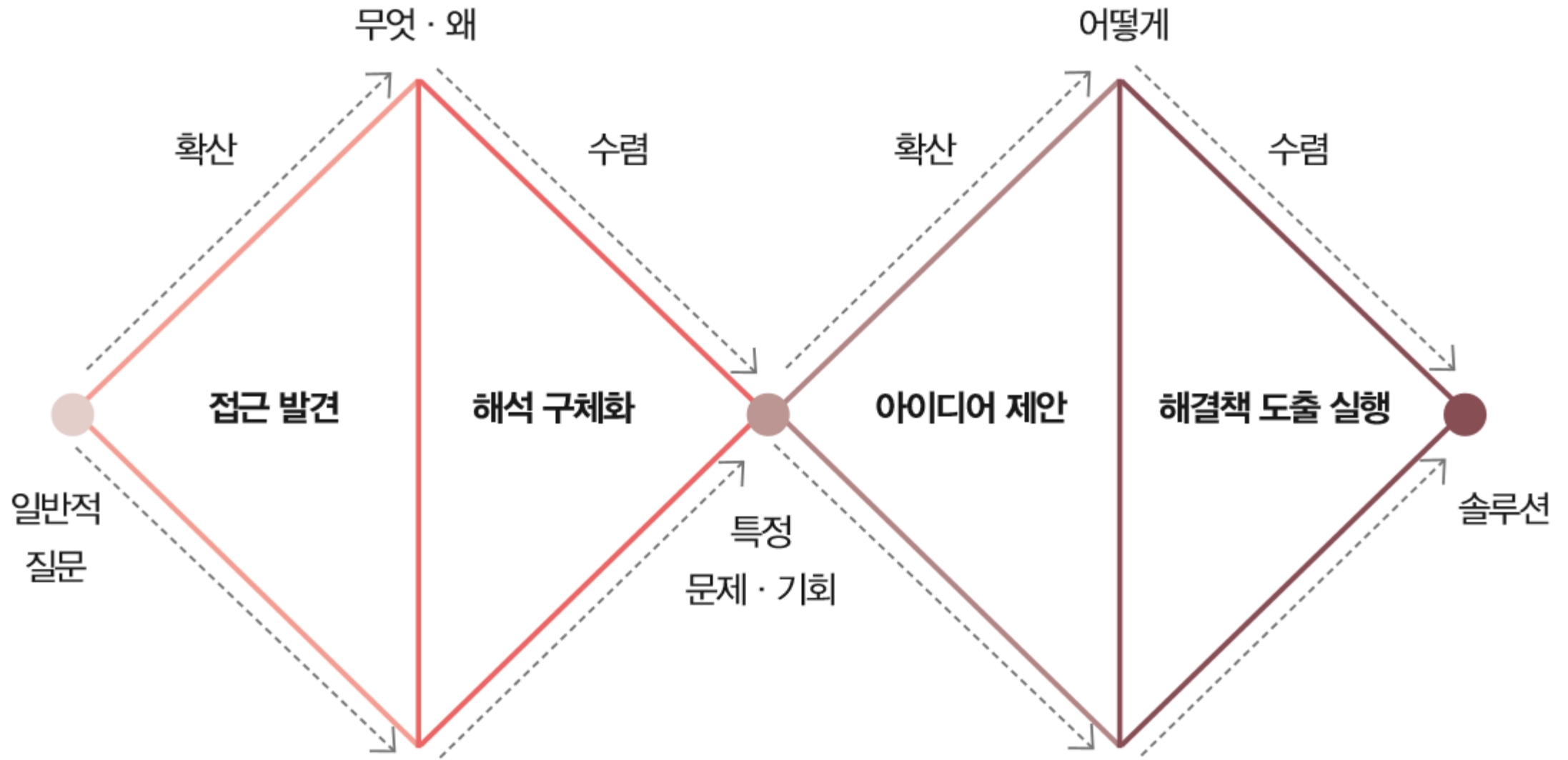
빅데이터 분석 5단계



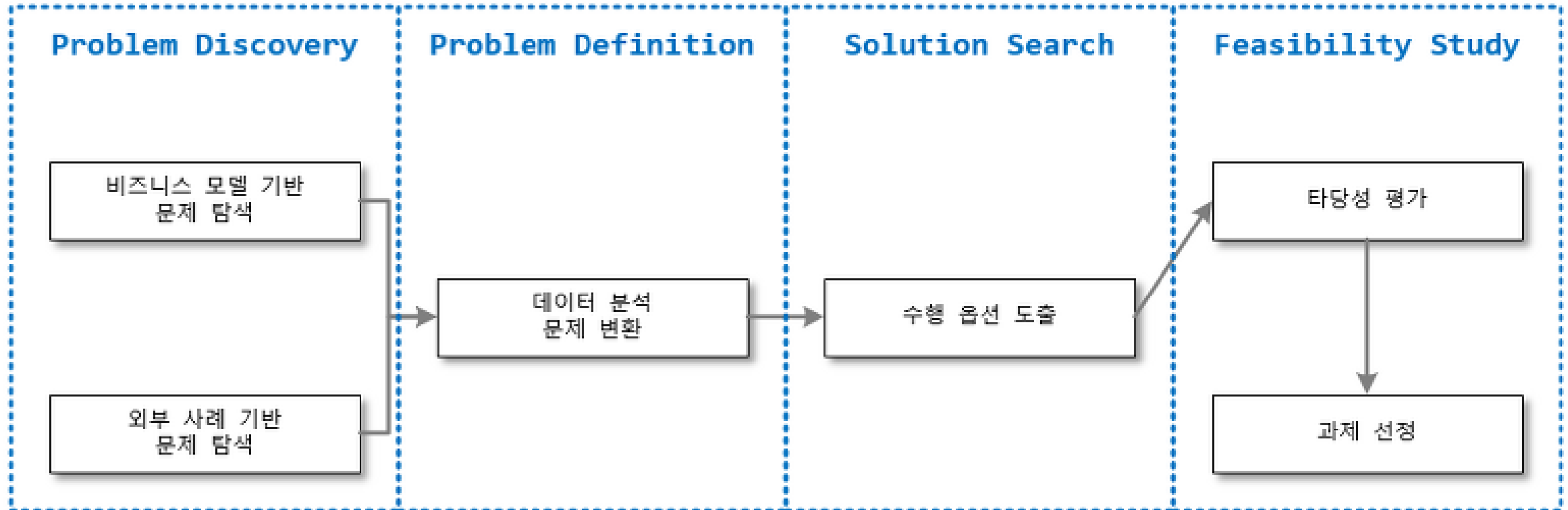
분석 과제 발굴



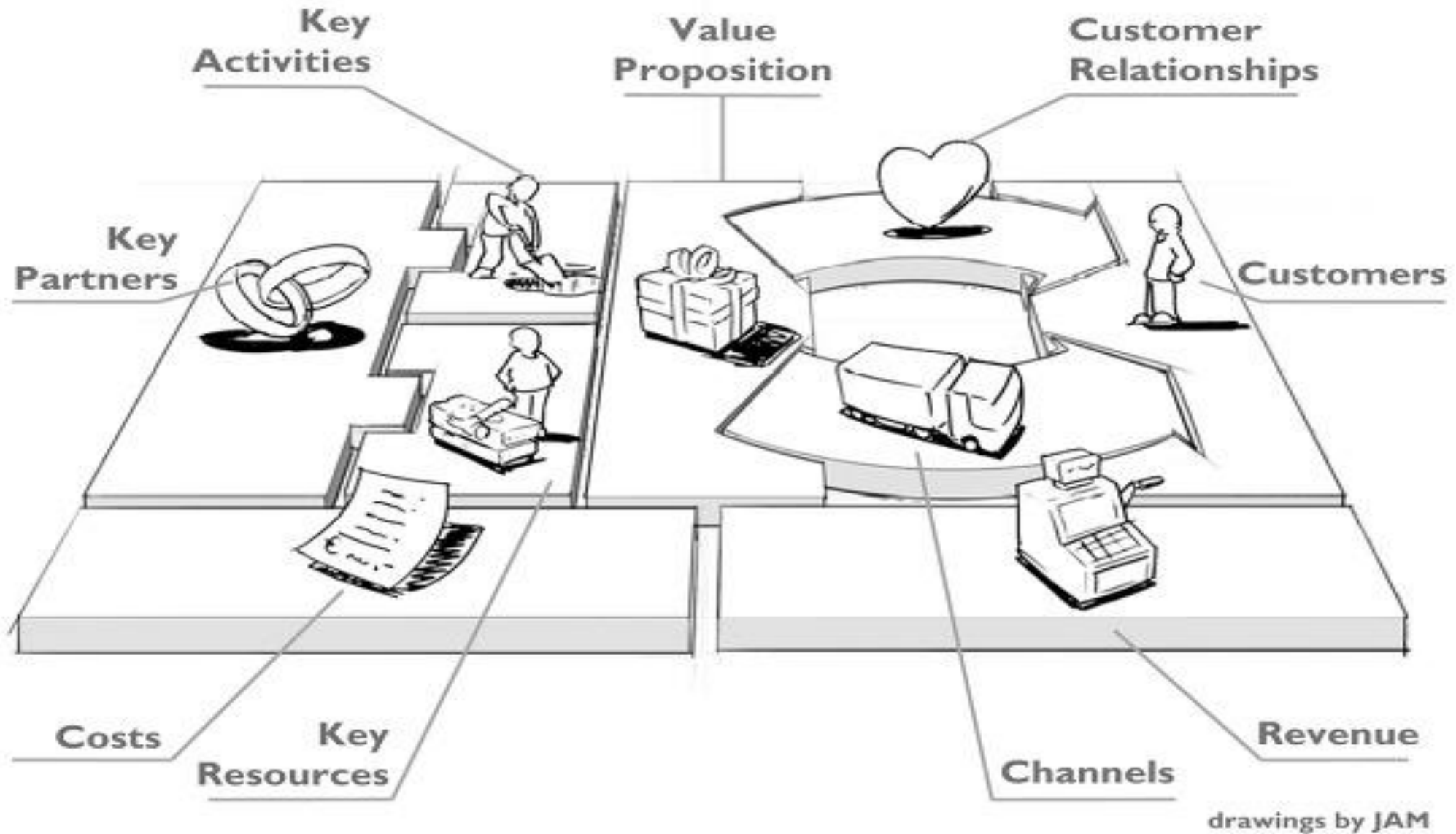
디자인 사고(Design Thinking)



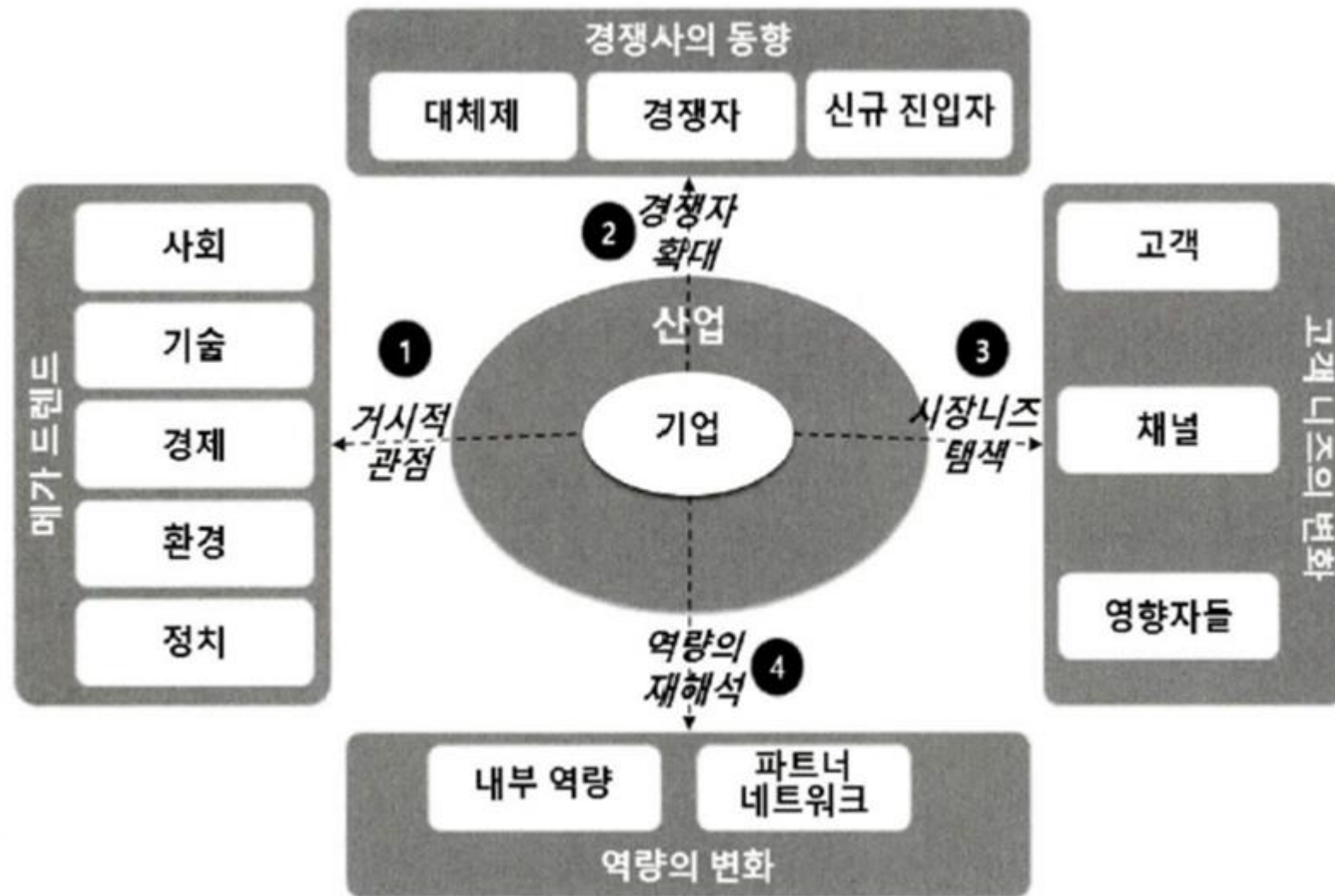
하향식 접근 방식(Top Down Approach)



비즈니스 모델 기반 문제 탐색- 하향식 접근법 1단계



분석 기회 발굴의 범위 확장-하향식 접근법 1단계



거시적 관점의 메가트렌드

STEEP 분석

Social (사회적분석)

고령화사회진입
출산율 감소
1인 가구 증가
결혼 연령 증가
신기술의 도입
세계화
(예시 : 직구족 증가)
라이프스타일 변화

Technological (기술적분석)

4차 산업혁명
- 무인 결제 시스템
- AI
- 인공지능
- 자율주행 자동차
- IOT
- 스마트폰 대중화

Economic (경제적분석)

물가상승률
환율의 변동
원자재가격 변동성
미중 무역전쟁
특정국가 부도위기
GNP/GDP
청년 실업률

Ecological (생태적분석)

지구 온난화
엘리뇨/라니냐
환경오염(미세먼지)
기후
농산물
물
지하자원

Political (정치적/법적)

시장의 규제
(규제 완화)
무역협정
세율의 변동
정부의 간섭 정도
법률의 통과

경제적 확대 관점

- 위협이 될 수 있는 상황에 대한 분석 기회 발굴의 폭을 넓혀서 탐색

영역	내용	예
대체재 (Substitute)	융합적인 경쟁 환경에서 현재 생산을 수행하고 있는 제품, 서비스를 온라인으로 제공하는 것에 대한 탐색 및 잠재적 위험을 파악	오프라인 제공 서비스 -> 온라인 제공에 대한 탐색 및 잠재적 위험 파악
경쟁자 (Competitor)	현재 생산하고 있는 제품, 서비스의 주요 경쟁자에 대한 동향에 파악하여 이를 고려한 분석 기회를 도출	식별된 주요 경쟁사의 제품 서비스 카탈로그 및 전략 분석을 통한 잠재적 위험 파악
신규 진입자 (New Entrant)	향후 시장에 대해서 파괴적인 역할을 수행할 수 있는 신규 진입자에 대한 동향을 파악하여 이를 고려한 분석 기회를 도출	새 제품에 대한 클라우드 소싱 서비스인 킥스타터의 유사 제품을 분석하고 잠재적 위험 파악

시장의 니즈 탐색 관점

- 의사결정에 영향을 미치는 영향자들(Influencer)에 대한 폭넓은 관점을 바탕으로 분석 기회를 탐색.

영역	내용	예
고객 (Customer)	고객의 구매 동향 및 고객의 컨텍스트를 더욱 깊게 이해하여 제품, 서비스의 개선에 필요한 분석 기회를 도출	철강 기업의 경우 조선 산업과 자동차 산업의 동향 및 주요 거래선의 경영 현황 등을 파악하고 분석 기회 도출
채널 (Channel)	영업사원, 직판 대리점, 홈페이지 등의 자체적으로 운영하는 채널 뿐만 아니라 최종 고객에게 상품, 서비스를 전달하는 것에 가능한 경로를 파악하여 해당 경로에 존재하는 채널별로 분석기회를 확대하여 탐색 한다.	은행의 경우 인터넷 전문은행 등 온라인 채널의 등장에 따른 변화에 대한 전략 분석 기회 도출
영향자 (Influencer))	기업 의사결정에 영향을 미치는 주주, 투자자, 협회 및 기타 이해 관계자의 주요 관심사항에 대해서 파악하고 분석 기회를 탐색 한다.	M&A 시장 확대에 따른 유사업종의 신규 기업 인수 기회 탐색.

역량의 재해석 관점

- 현재 해당 조직 및 기업이 보유한 역량 뿐만 아니라 해당 조직의 비즈니스에 영향을 끼치는 파트너 네트워크를 포함한 활용 가능한 역량을 토대로 폭넓은 분석 기회를 탐색한다.

영역	내용	예
내부 역량 (Competency)	지적 재산권, 기술력 등 기본적인 것 뿐만 아니라 중요하면서도 자칫 간과하기 쉬운 지식, 기술, 스킬 등의 노하우와 인프라적인 유형 자산에 대해서 재해석하고 해당영역에서 분석 기회를 탐색한다.	자사 소유 부동산을 활용한 부가가치 창출 기회 발굴
파트너와 네트워크 (Partners & Network)	자사가 직접 보유하고 있지는 않지만 밀접한 관계를 유지하고 있는 관계사와 공급사 등의 역량을 활용해 수행할 수 있는 기능을 파악해보고 이에 대한 분석 기회를 추가적으로 도출한다.	수출입, 통관, 노하우를 활용한 추가 사업 기회 탐색

데이터 분석 기반 문제 탐색-하향식 접근법 2단계

- 데이터 분석 문제의 정의 및 요구사항 : 분석을 수행하는 당사자뿐만 아니라 해당 문제가 해결되었을 때 효용을 얻을 수 있는 최종 사용자(End User) 관점에서 이루어져야 한다.
- 데이터 분석 문제가 잘 정의 되었을 때 필요한 데이터의 정의 및 기법 발굴이 용이하기 때문에 **가능한 정확하게 분석의 관점으로 문제를 재정의할 필요**가 있다.

비즈니스 문제
고객 이탈 증대
예상치 많은 설비 장애로 인한 판매량 감소
기존 판매 정보 기반 영업사원의 판단 시 재고 관리 및 적정 가격 판매 어려움



분석 문제
고객의 이탈에 영향을 미치는 요인을 식별하고, 이탈 가능성을 예측
설비의 장애를 이끄는 신호를 감지하여 설비 장애 요인으로 식별하고, 장애 발생 시점 및 가능성을 예측
내부 판매 정보 외의 수요 예측을 수행할 수 있는 인자의 추출 및 모델링을 통한 수요 예측

데이터 분석 기반 문제 탐색- 하향식 접근 3단계

- 데이터 **분석 문제를 해결하기 위한 다양한 방안이 모색**된다.
- 분석역량을 기존에 가지고 있는 지의 여부를 파악하여 보유하고 있지 않은 경우에는 교육이나 전문인력 채용을 통한 역량을 확보하거나 분석 전문업체를 활용하여 **과제를 해결하는 방안**에 대해 사전 검토를 수행한다.

		분석 역량 (Who)	
		확보	미확보
분석 기법 및 시스템 (HoW)	기존 시스템	기존 시스템 개선 활용	교육 및 채용을 통한 역량 확보
	신규 도입	시스템 고도화	전문 업체 Sourcing

데이터 분석 기반 문제 탐색- 하향식 접근 4단계

- 도출된 분석 문제나 가설에 대한 대안을 과제화하기 위해서는 **다각적인 타당성 분석이 수행**되어야 한다.
- **경제적 타당성 : 비용대비 편익 분석 관점의 접근이 필요.**
- 데이터 및 기술적 타당성 : 데이터 분석에는 데이터 존재 여부, 분석 시스템 환경 그리고 분석 역량이 필요

분석 역량의 경우 실제 프로젝트 수행 시, 걸림돌이 되는 경우가 많기 때문에 기술적 타당성 분석 시 역량 확보 방안을 사전에 수립하고 이를 **효과적으로 평가하기 위해서는 비즈니스 지식과 기술적 지식이 요구됨.**

- 1) 평가 과정을 거쳐 가장 우월한 대안을 선택.
- 2) 도출한 데이터 분석 문제 및 선정된 솔루션 방안을 포함
- 3) 분석과제 정의서의 형태로 명시하는 후속작업을 시행
- 4) 프로젝트 계획의 입력물로 활용됨.

상향식 접근법(Bottom up Approach)

상향식 접근법 (Bottom up Approach)	지도 학습 (Supervised Learning)	비지도학습 (Unsupervised Learning)	프로토타이핑 접근법 (Prototype)
<ul style="list-style-type: none">• 분석을 대상을 모른다.• 원천 데이터로부터 통찰과 지식을 얻는 방법• 분석부터 하고, 그 결과로부터 가치 있는 문제를 도출함.	<ul style="list-style-type: none">• 정답 / 라벨이 있는 데이터<ul style="list-style-type: none">• 범주형 - 분류• 연속형 - 회귀• 의사결정트리, 인공신경망 등	<ul style="list-style-type: none">• 정답 / 라벨이 없는 데이터• 비슷한 특징을 가진 데이터끼리 군집화<ul style="list-style-type: none">• 군집분석, 주성분 분석등	<ul style="list-style-type: none">• 시행착오 해결법• 분석 먼저 하고 결과를 확인<ul style="list-style-type: none">• 조금씩 개선해 나감• 문제 정의가 불명확하고, 문제인 경우

상향식 접근법(Bottom up Approach)

기업 고객 분류



상향식 접근

(Bottom-up Approach)

즉시 도입 가능
(구매 프로세스 한 단계)

거래 액수 작음
(개인, 팀 단위)

서비스 적극적 수용

구전 효과 높음
확산 속도 빠름

슬랙, 드롭박스, 트렐로



하향식 접근

(Top-down Approach)

일정 기간 소요
(구매 프로세스 다단계, 약 3-6개월 소요)

거래 액수 큼
(기업 단위)

서비스 소극적 수용

기업별 맞춤 전략 및 대응 요구
지속적 관계 유지 필요
(세일즈 역할중요)

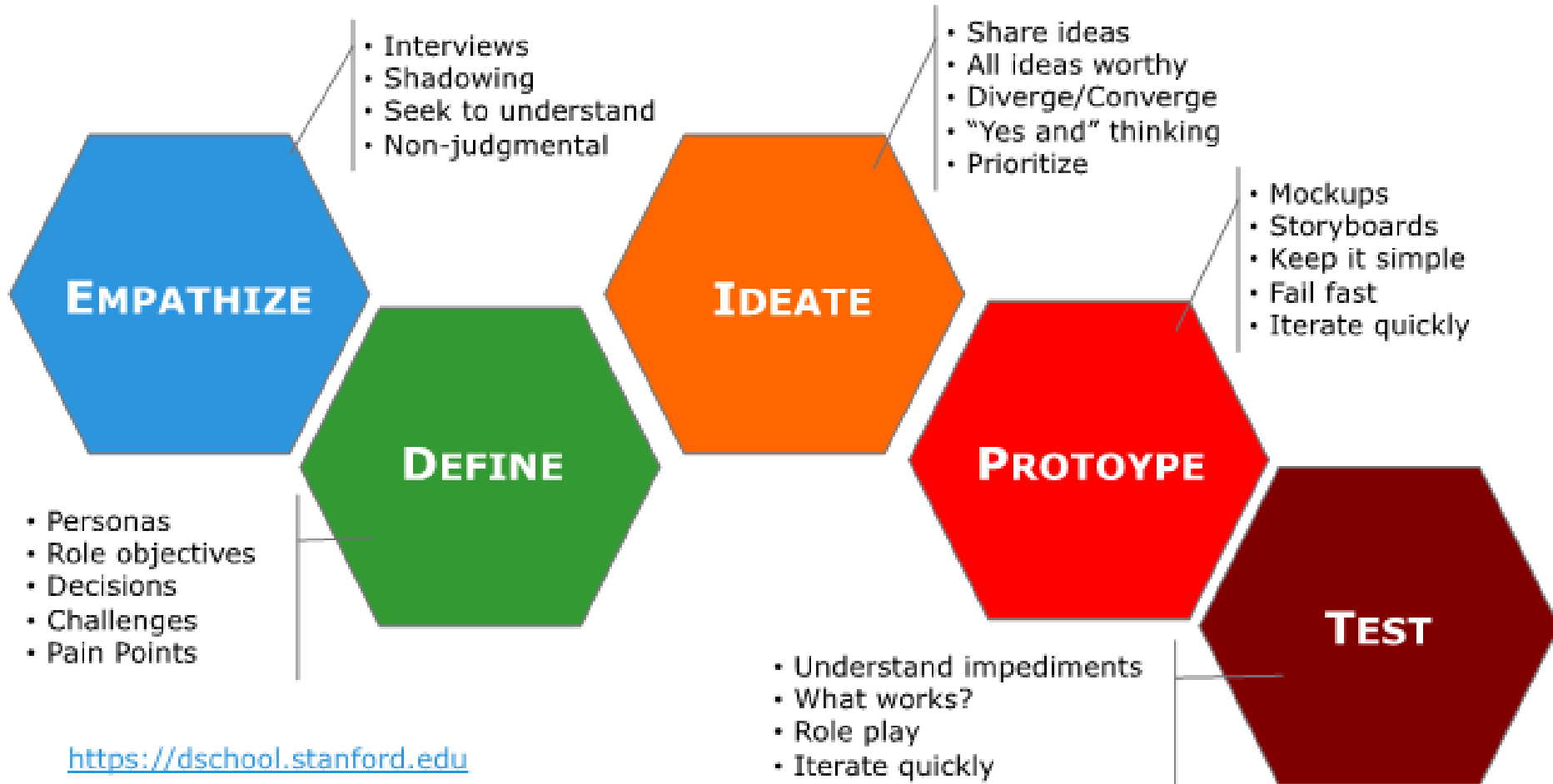
그룹웨어, ERP 시스템

하향식 접근 방법의 한계

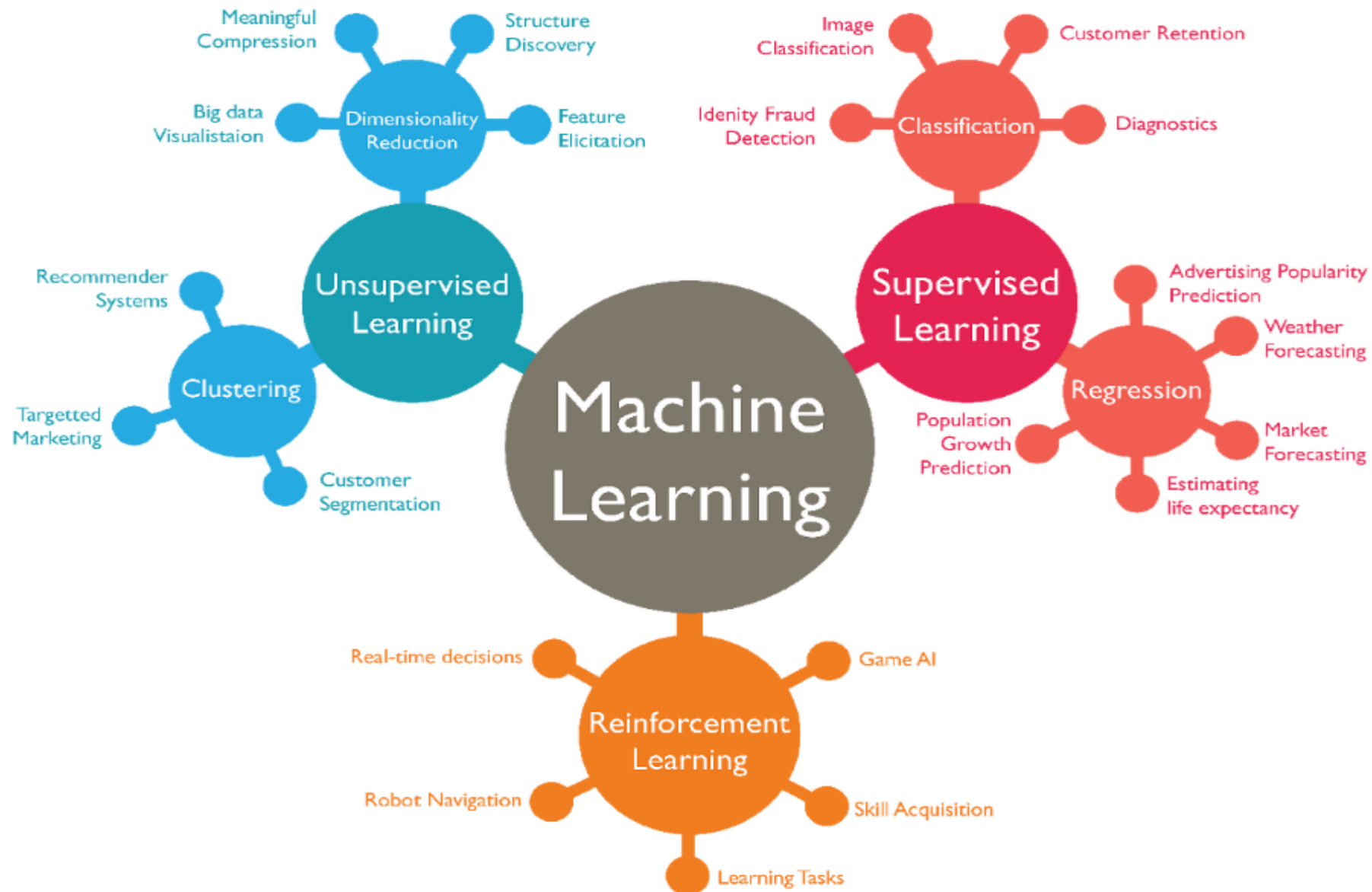
- 솔루션 도출에는 유효하지만 새로운 문제 탐색의 한계
- 논리적인 단계별 접근법 기반의 문제해결 방식은 복잡하고 다양한 환경에서 발생하는 문제에는 비적합
- 이를 해결하기 위해 스탠포드 대학의 d.school(Institute of Design at Stanford)은 디자인 사고(Design Thinking) 접근법을 통해 전통적인 분석적 사고를 극복하려고 함
- 통상적인 관점에서는 'Why'를 강조하지만, 있는 그대로 인식하는 'What' 관점 필요
- 이와 같은 점을 고려하여 d.school에서는 첫단계로 Empathize(감정이입)을 강조

하향식 접근 방법의 한계

Stanford d.school Design Thinking Process



Machine learning

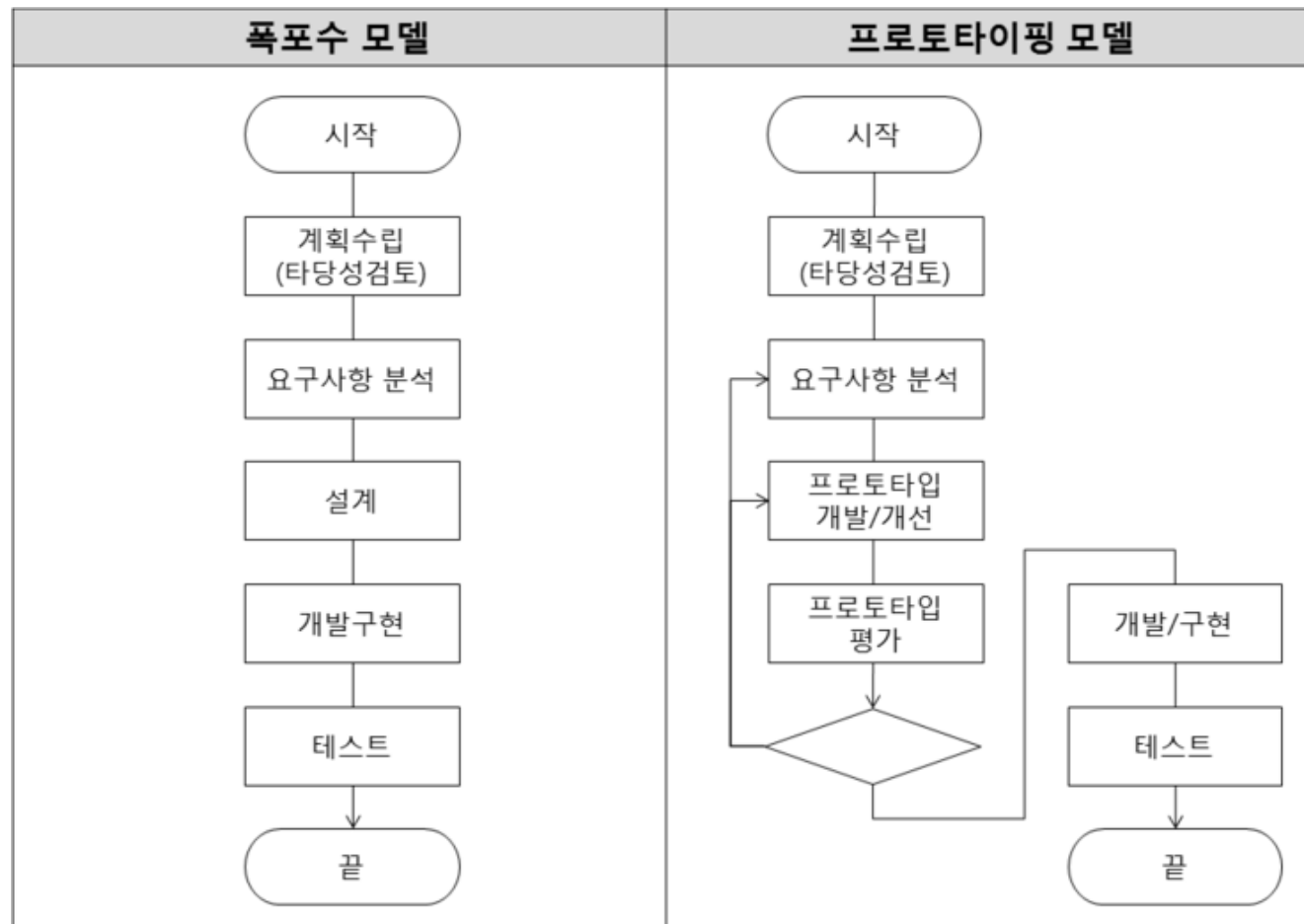


통계적 분석

- 인과관계 분석을 위해 가설을 설정하고 이를 검정하기 위해 모집단으로 부터 표본을 추출하고 **그 표본을 이용한 가설 검정을 실시하는 방식으로 문제 해결**
- 빅데이터 환경에서는 이와 같은 **논리적인 인과관계 분석뿐만 아니라 상관관계 분석 또는 연관 분석을 통하여 다양한 문제 해결에 도움을 받을 수 있음.**
- **인과관계에서 상관관계 분석으로의 이동이 빅데이터 분석에서의 주요 변화**
- 다량의 데이터 분석을 통해서 “왜” 그러한 일이 발생하는지 역으로 추적하면서 문제를 도출하거나 재정의 할 수 있는 것이 **상향식 접근법.**

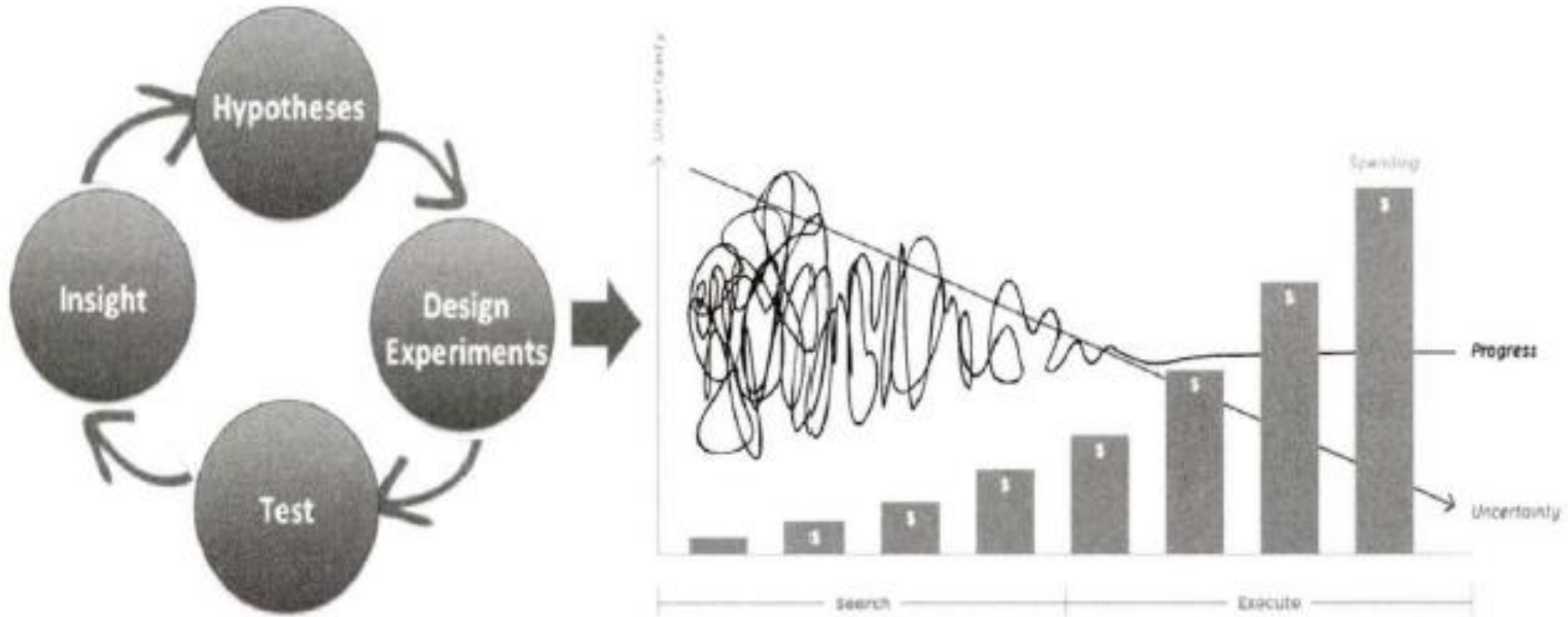
프로토타이핑 모델

- 사용자가 요구사항이나 데이터를 정확히 규정하기 어렵고 데이터 소스도 명확히 파악하기 어려운 상황에서 일단 분석을 시도해 보고 그 결과를 확인해 가면서 반복적으로 개선해 나가는 방법



프로토타이핑 모델

- 사용자가 요구사항이나 데이터를 정확히 규정하기 어렵고 데이터 소스도 명확히 파악하기 어려운 상황에서 일단 분석을 시도해 보고 그 결과를 확인해 가면서 반복적으로 개선해 나가는 방법



빅데이터 분석 환경에서 프로토타이핑의 필요성

1. 문제에 대한 인식 수준

: 문제 정의가 불명확하거나 이전에 접해보지 못한 새로운 문제일 경우 사용자 및 이해관계자는 **프로토타입을 이용하여 문제를 이해하고, 이를 바탕으로 구체화함.**

2. 필요 데이터 존재 여부의 불확실성

: 문제 해결을 위해 필요한 데이터의 집합이 모두 존재하지 않을 경우, 그 데이터의 수집을 어떻게 할 것인지 또는 그 데이터를 다른 데이터로 대체할 것인지 등에 대한 **사용자와 분석자간의 반복적이고 순환적인 협의 과정이 필요.**

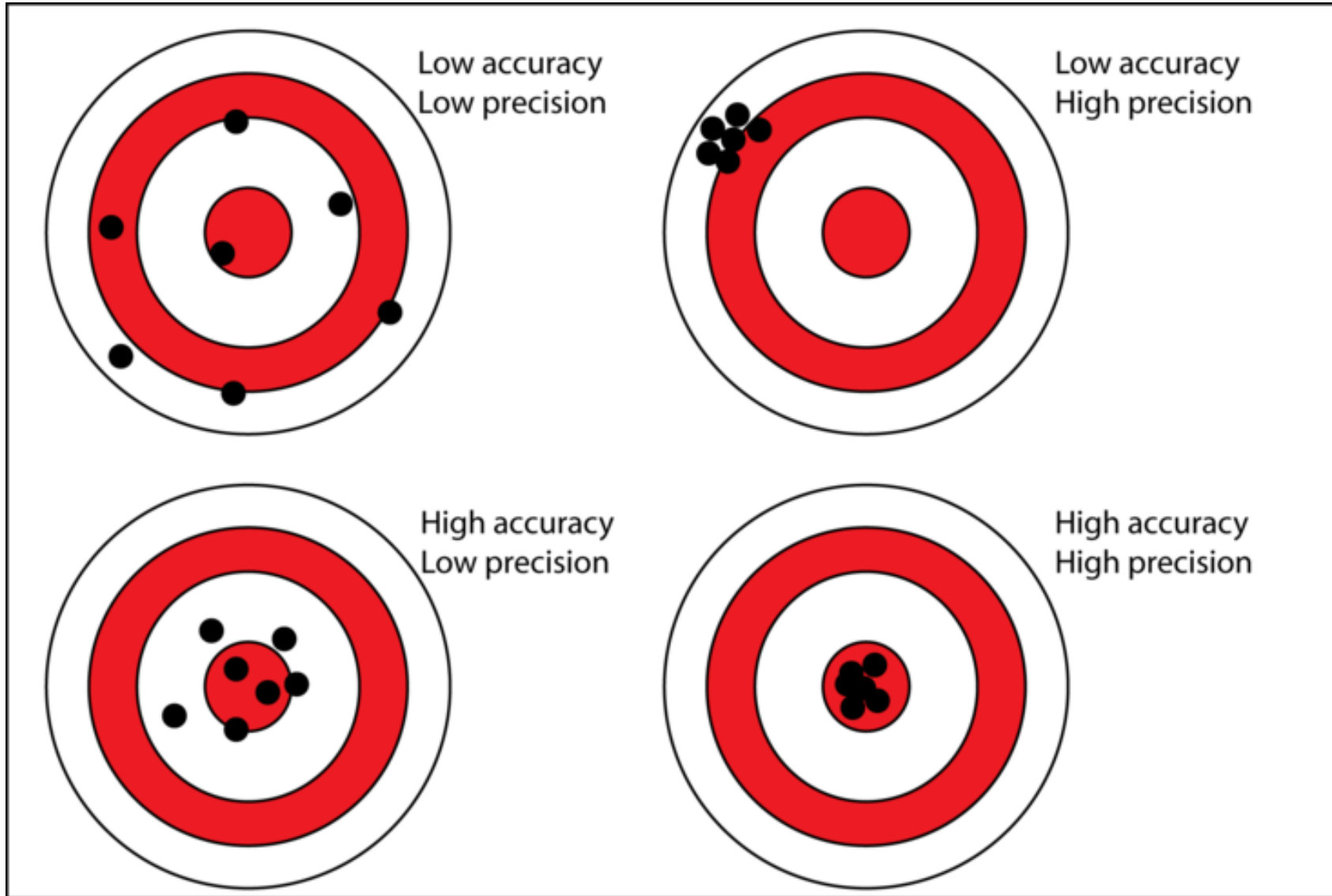
3. 데이터 사용 목적의 가변성

: 조직에서 보유 중인 데이터라 하더라도 **기존의 데이터 정의를 재검토하여 데이터의 사용 목적과 범위를 확대**

분석 프로젝트 관리 방안



분석 프로젝트 관리 방안



분석 프로젝트 관리 방안

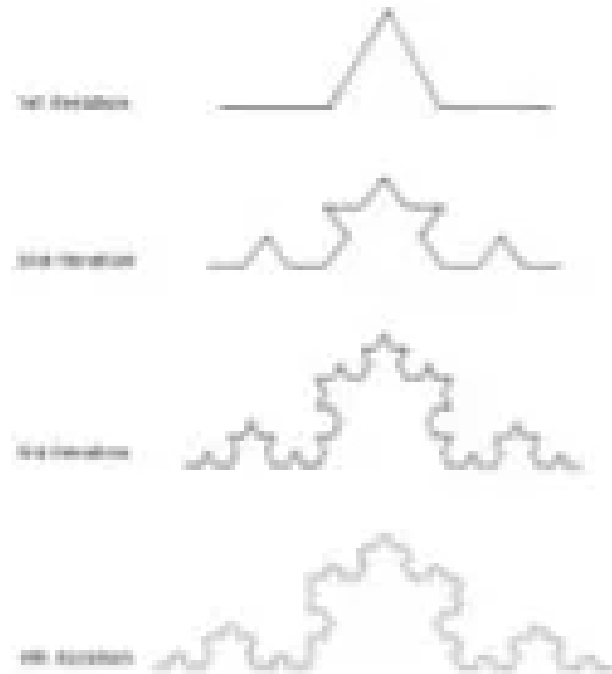
		예 측	
		Positive	Negative
실제 결과	Positive	TP (True Positive)	FN (False Negative)
	Negative	FP (False Positive)	TN (True Negative)

- 정밀도(Precision) : Positive로 예측한 것 중에서, 실제로 True인 비율. $TP / (TP+FP)$
- 재현율(Recall/Sensitivity) : 실제로 True인 것 중에서 예측도 Positive로 한 비율. $TP / (TP+FN)$

분석 프로젝트 특성

- 분석가 : 데이터의 원천을 다루는 데이터 영역과 결과를 활용할 비즈니스 영역의 중간에서 분석 모델을 통한 조율을 수행하는 조정자의 역할이 핵심.
- 분석 프로젝트는 도출된 결과의 재해석을 통한 지속적인 반복 및 정교화가 수행되는 경우가 대부분이므로, 프로토타이핑 방식의 애자일(Agile) 프로젝트 관리방식에 대한 고려도 필요.

The Data Scientist as builder



분석 프로젝트 관리 방안

관리 영역	주요 프로세스
통합관리 (Integration Management)	프로젝트 헌장 개발, 프로젝트 관리 계획 수립, 프로젝트 실행 지시 및 관리, 프로젝트 작업 감시 및 통제, 통합 변경 통제, 프로젝트 종료 관리 등
범위관리 (Scope Management)	프로젝트 범위 계획, 범위 정의, 작업분류체계(WBS) 작성, 범위 검증, 범위 통제 관리 등
일정관리 (Time Management)	작업 정의, 작업 순서 배열, 작업별 자원 산정, 작업 기간 산정, 일정 개발, 일정 통제 등
비용관리 (Cost Management)	자원 계획, 비용 산정, 비용 예산 및 비용 통제 등
품질관리 (Quality Management)	품질 계획, 품질 보증, 품질 관리 등
인적자원관리 (Human Resource Management)	조직 계획, 인적 자원 획득, 프로젝트 팀 확보, 프로젝트 팀 개발, 프로젝트 팀 관리 등
위험관리 (Risk Management)	위험 관리 계획, 위험 식별, 정성적 위험 분석, 정량적 위험 분석, 위험 대응 계획, 위험 감시 통제 등
의사소통관리 (Communication Management)	의사소통 계획, 정보 배포, 진척 관리, 종료 절차 등
조달관리 (Procurement Management)	획득계획, 공급자 유치 계획, 공급자 선정, 계약 관리, 계약 종료 등

분석 마스터 플랜

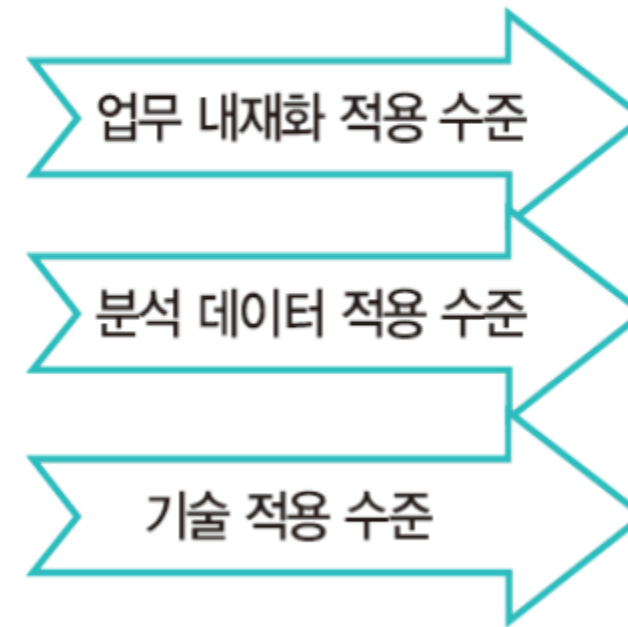
- 데이터 기반 구축을 위해서 분석 과제를 대상으로 전략적 중요도, 비즈니스 성과 및 ROI, 분석과제의 실용 용이성 등 다양한 기준을 고려해 **적용 우선순위**를 설정

우선순위 고려 요소



적용 우선순위
결정

적용 범위/방식 고려 요소



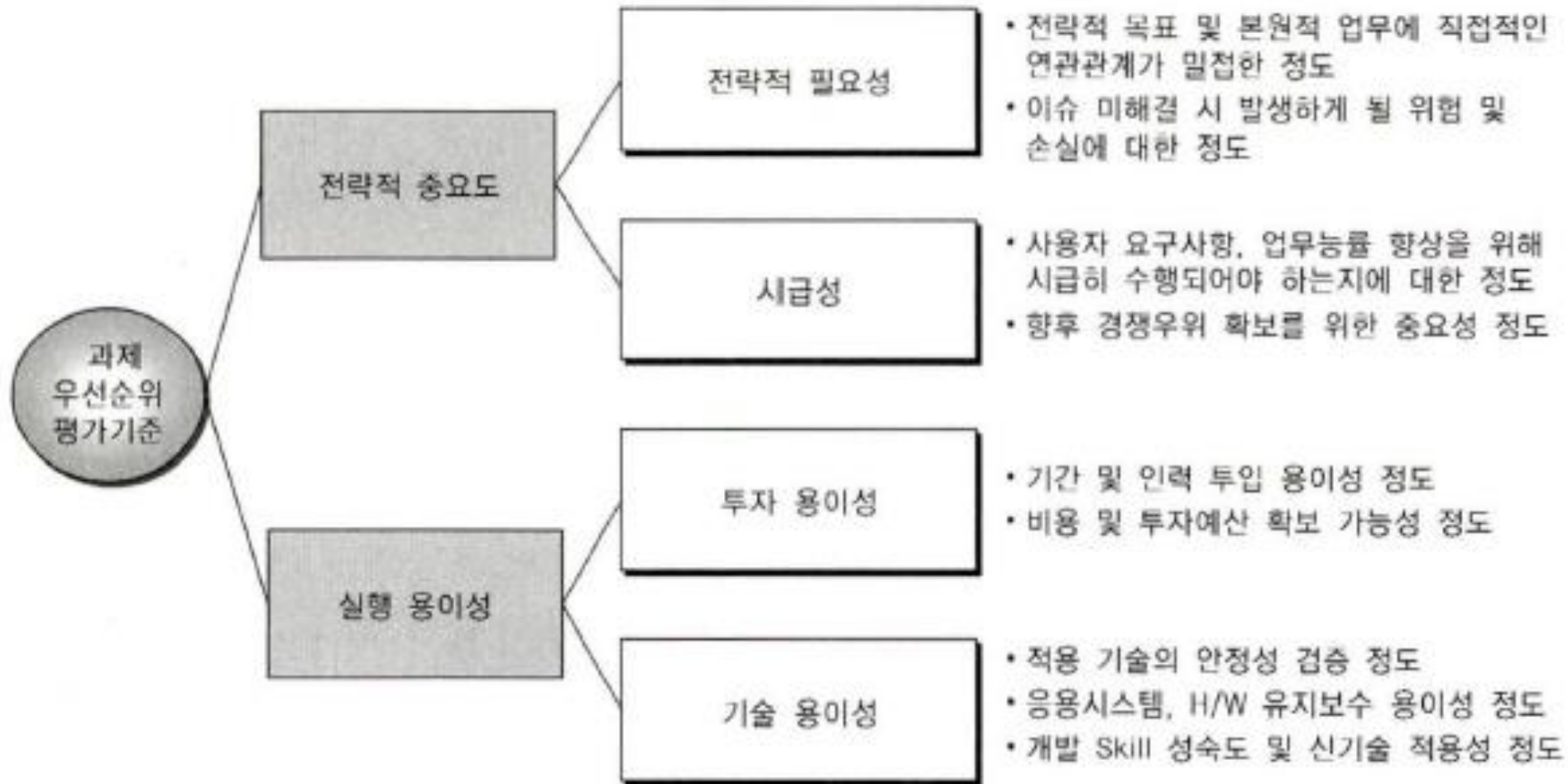
분석 구현
로드맵 수립

마스터플랜 수립 개요 (★기출)

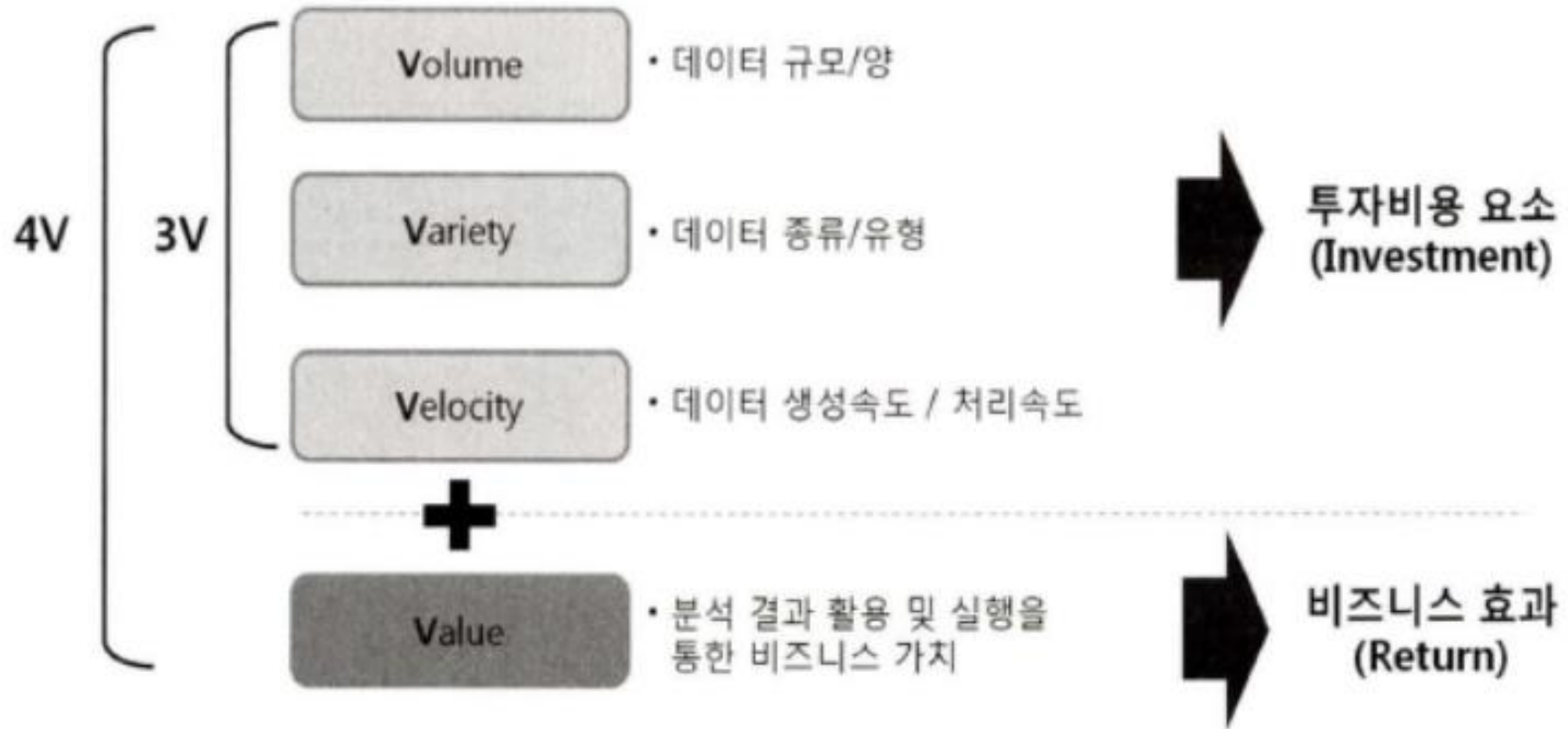
분석 마스터 플랜

ISP(Information Strategy Planning) 정보 전략 계획	분석 마스터 플랜
<ul style="list-style-type: none">• 중장기 로드맵을 정의하기 위한 정보전략계획• 조직 내, 외부 환경을 분석하여 기회나 문제점을 도출• 사용자의 요구사항을 분석하여 시스템 구축 우선순위 결정• 정보기술 또는 정보 시스템을 전략적으로 활용하기 위하여 <p>중장기 마스터플랜을 수립하는 절차</p>	<ul style="list-style-type: none">• 일반적인 ISP 방법론을 활용하여 필요한 데이터 분석 과제를 빠짐 없이 도출한 후 과제의 우선순위를 결정하고 단기 및 중,장기로 나누어 계획을 수립함.

수행 과제 도출 및 우선순위 평가



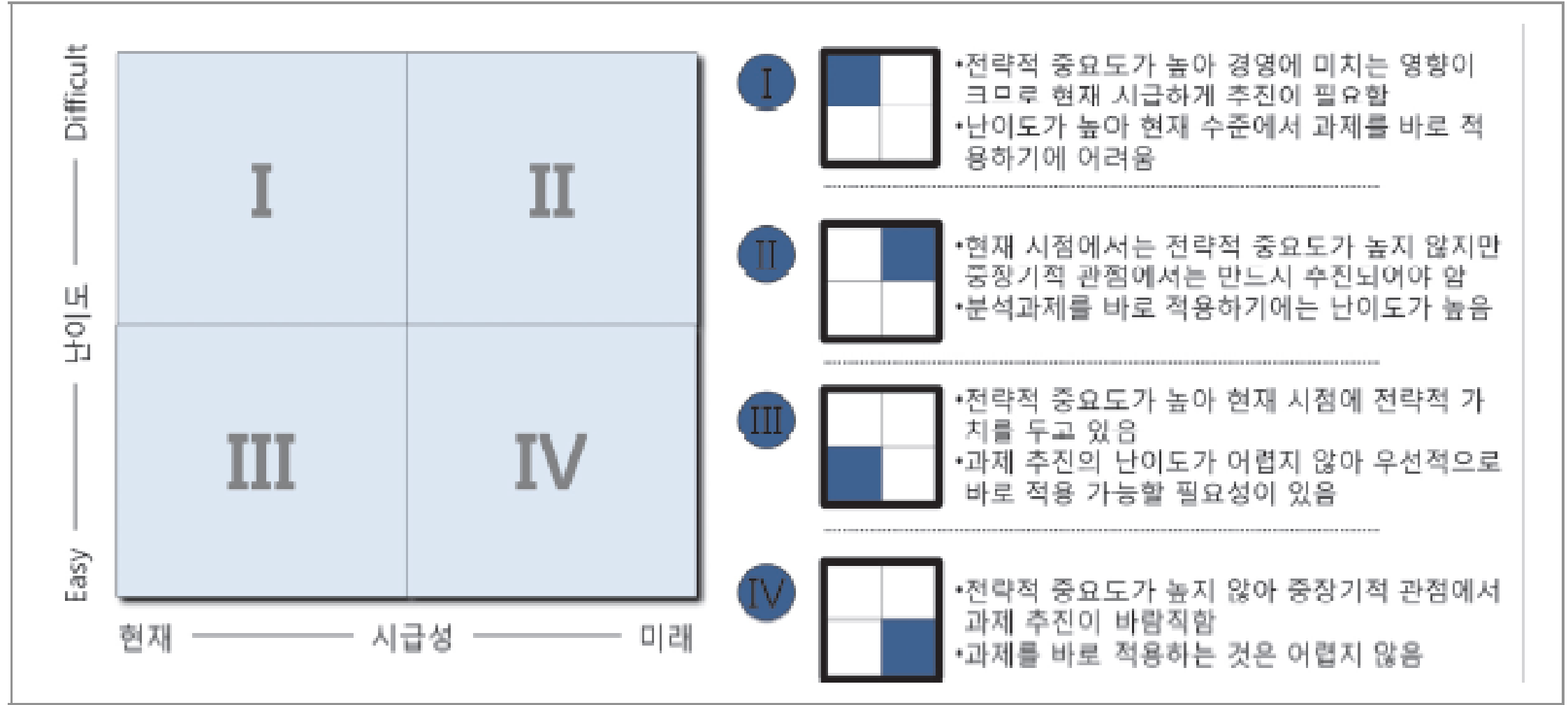
ROI 관점에서 빅데이터의 핵심 특징



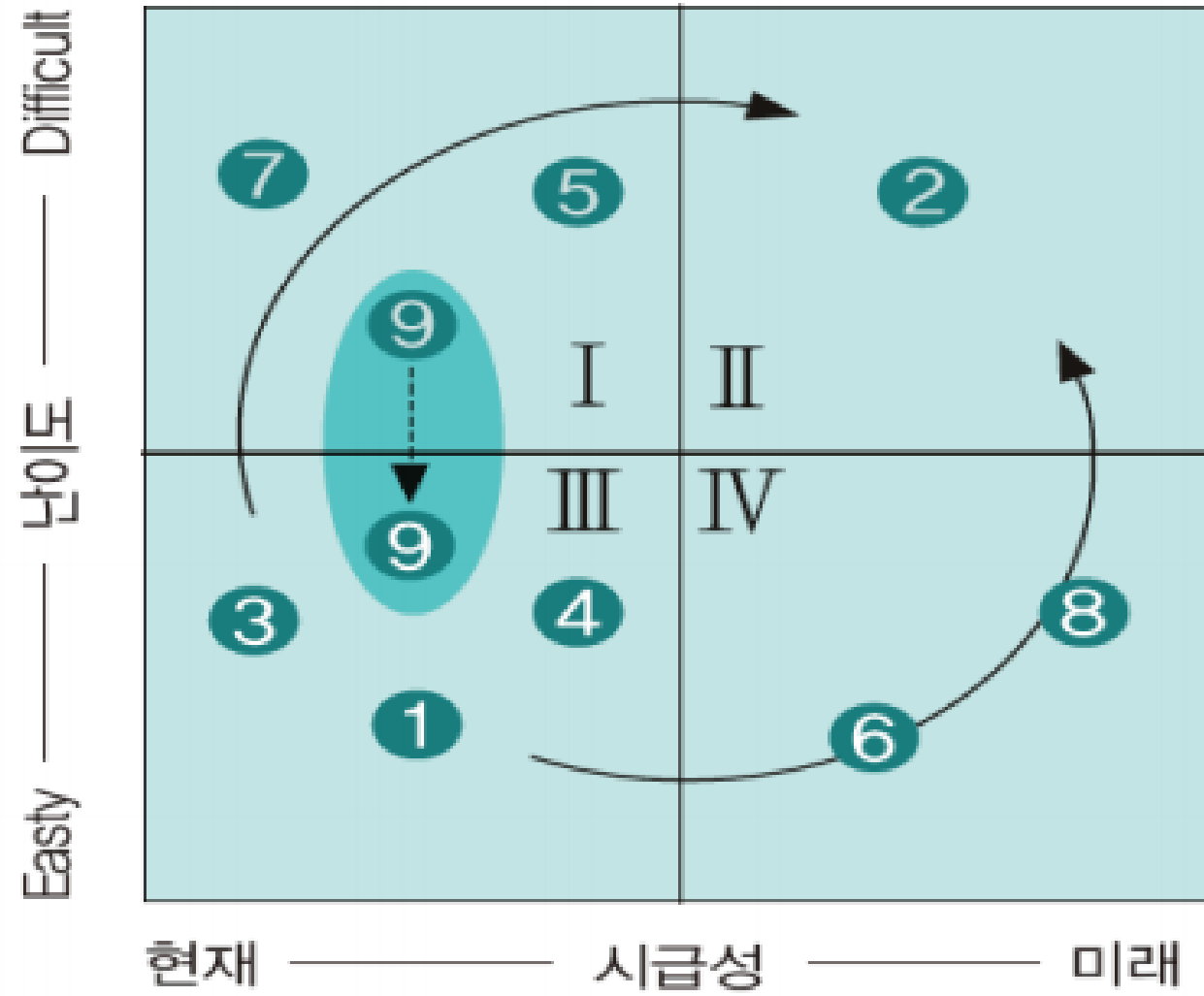
ROI 관점에서 빅데이터의 핵심 특징



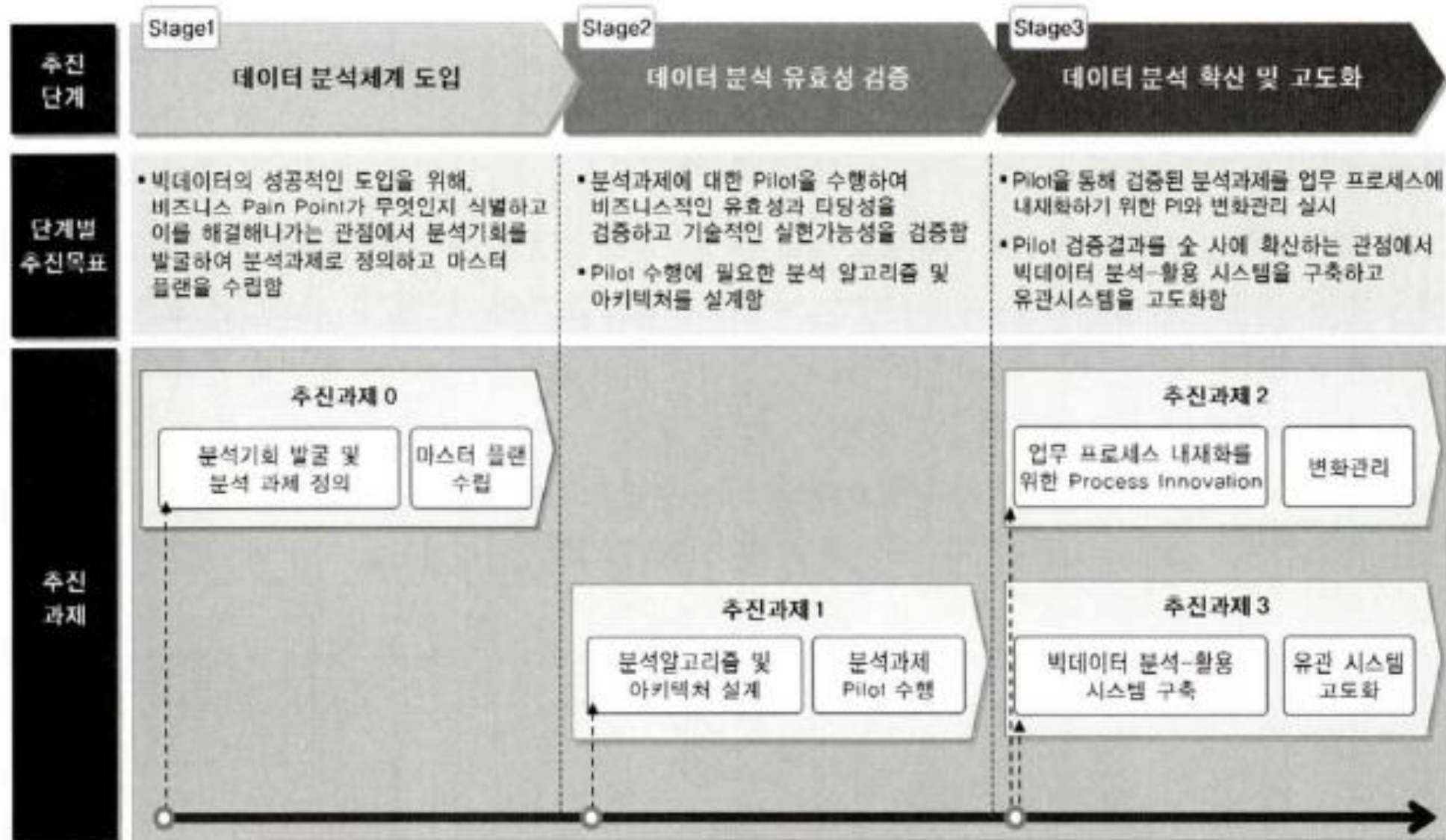
분석 마스터 플랜



분석 과제 우선순위 선정 조정



로드맵 수립



이행계획 수립



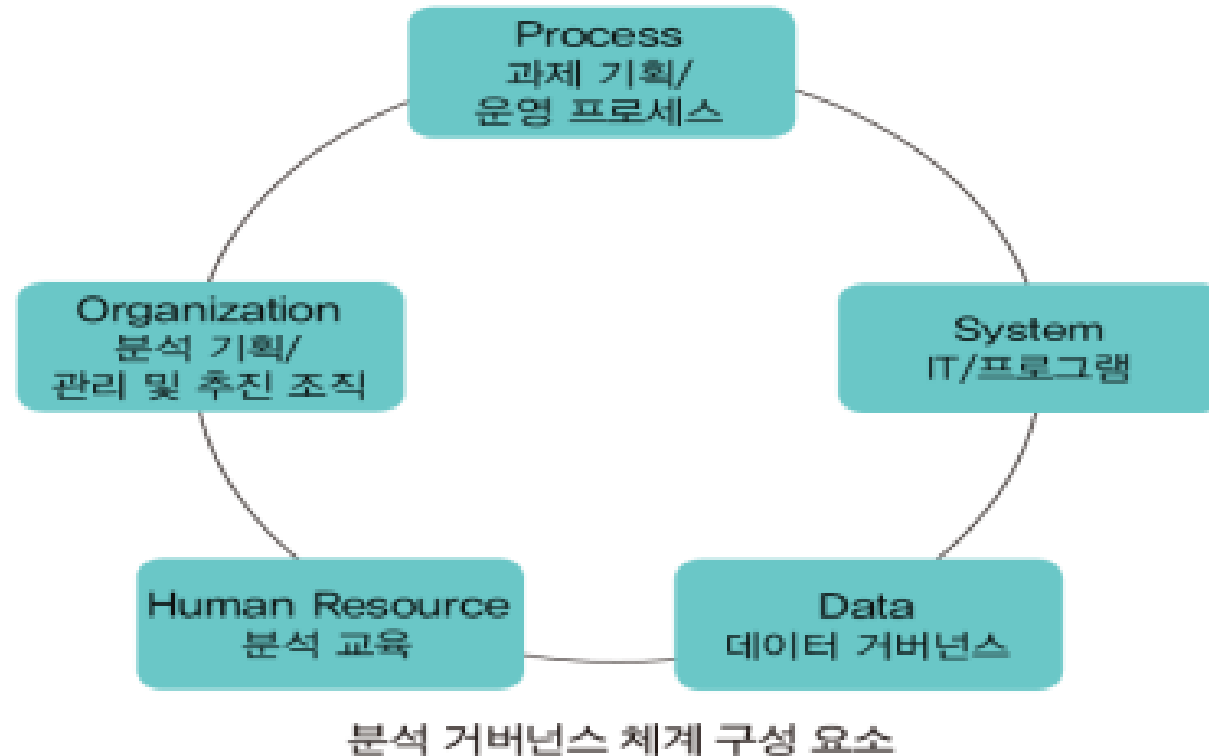
순차적 구현



Refine Analytics Model

분석 거버넌스 체계 수립

- 기업에서 데이터를 이용한 의사결정이 강조될수록 데이터 분석과 활용을 위한 체계적인 관리가 중요해 짐.



분석 준비도

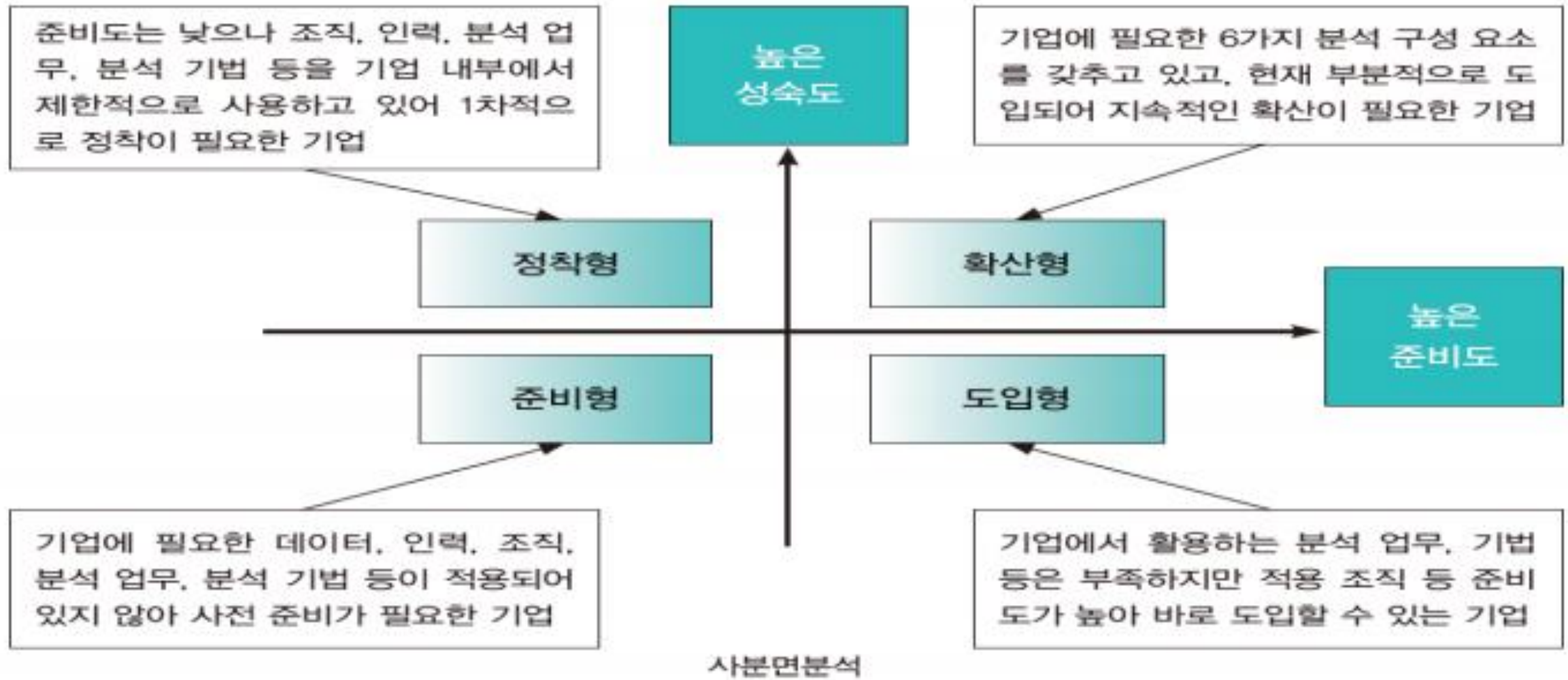
분석 업무 파악	인력 및 조직	분석 기법
<ul style="list-style-type: none">• 발생한 사실 분석 업무• 예측 분석 업무• 시뮬레이션 분석 업무• 최적화 분석 업무• 분석 업무 정기적 개선	<ul style="list-style-type: none">• 분석 전문가 직무 존재• 분석 전문가 교육 훈련 프로그램• 관리자층의 기본적인 분석 능력• 전사 분석업무 총괄 조직 존재• 경영진 분석 업무 이해 능력	<ul style="list-style-type: none">• 업무별 적합한 분석기법 사용• 분석 업무 도입 방법론• 분석기법 라이브러리• 분석기법 효과성 평가• 분석기법 정기적 개선
분석 데이터	분석 문화	IT 인프라
<ul style="list-style-type: none">• 분석업무를 위한 데이터 충분성• 분석업무를 위한 데이터 신뢰성• 분석업무를 위한 데이터 적시성• 비구조적 데이터 관리• 외부 데이터 활용 체계• 기준데이터 관리(MDM)	<ul style="list-style-type: none">• 사실에 근거한 의사결정• 관리자층의 데이터 중시• 회의 등에서 데이터 활용• 경영진의 직관보다 데이터• 데이터 공유 및 협업 문화	<ul style="list-style-type: none">• 운영시스템 데이터 통합• EAI, ETL 등 데이터유통체계• 분석 전용 서버 및 스토리지• 빅데이터 분석 환경• 통계 분석 환경• 비주얼 분석 환경

분석 성숙도 모델

- 소프트웨어 공학에서는 시스템 개발 업무 능력과 조직의 성숙도를 파악하기 위해 CMMI(Capability Maturity Model Integration) 모델을 기반으로 조직의 성숙도를 평가한다.

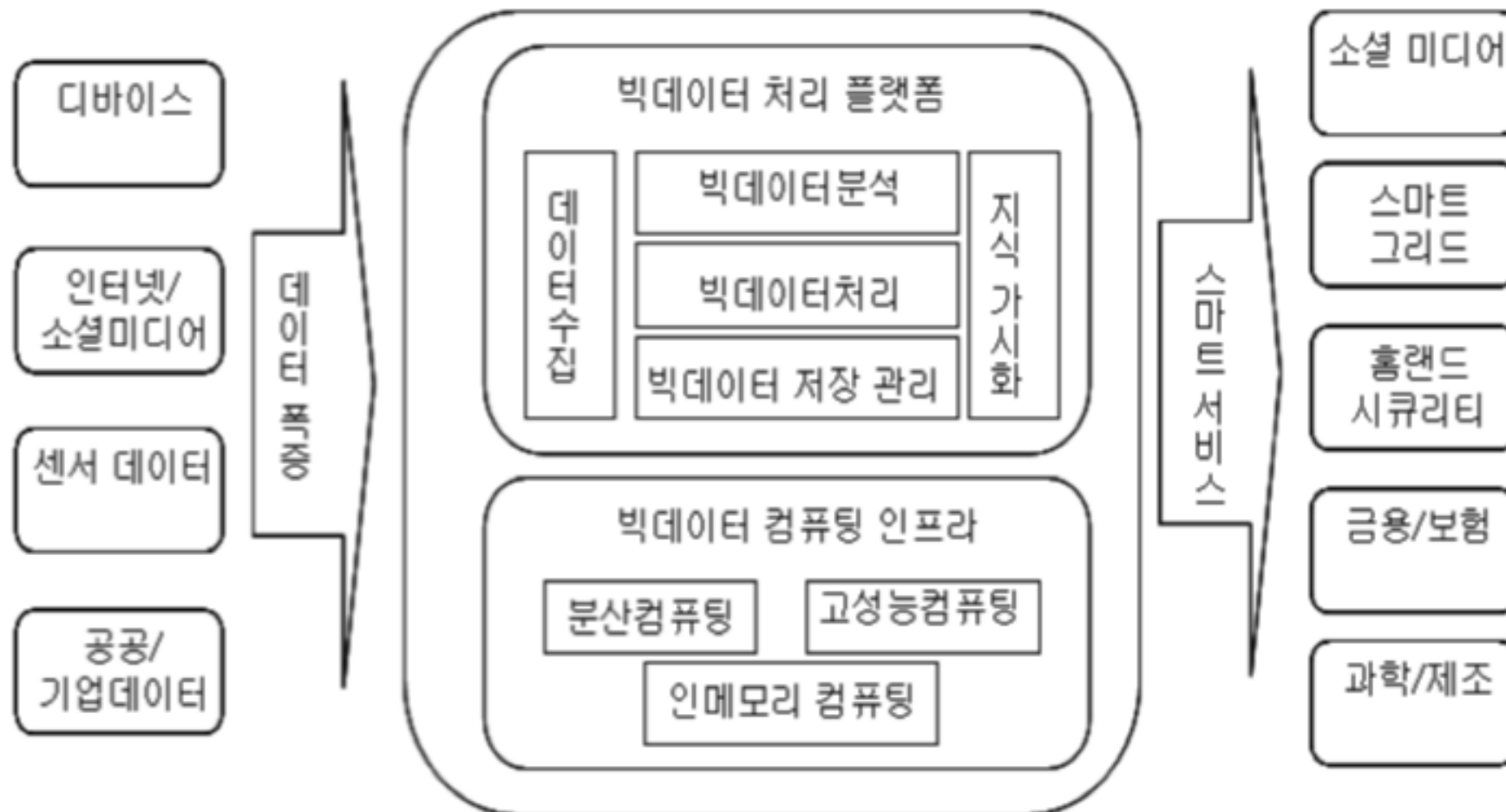
단계	도입 단계	활용 단계	확산 단계	최적화 단계
설명	분석을 시작하여 환경과 시스템을 구축	분석 결과를 실제 업무에 적용	전사 차원에서 분석을 관리하고 공유	분석을 진화시켜서 혁신 및 성과 향상에 기여
비즈니스 부문	✓ 실적분석 및 통계 ✓ 정기보고 수행 ✓ 운영 데이터 기반	✓ 미래 결과 예측 ✓ 시뮬레이션 ✓ 운영 데이터기반	✓ 전사 성과 실시간 분석 ✓ 프로세스혁신 3.0 ✓ 분석규칙 관리 ✓ 이벤트 관리	✓ 외부환경 분석 활용 ✓ 최적화 업무 적용 ✓ 실시간 분석 ✓ 비즈니스 모델 진화
조직·역량 부문	✓ 일부 부서에서 수행 ✓ 담당자 역량에 의존	✓ 전문 담당부서에서 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
IT 부문	✓ 데이터 웨어하우스 ✓ 데이터 마트 ✓ ETL/EAI ✓ OLAP	✓ 실시간 대시보드 ✓ 통계분석 환경	✓ 빅데이터 관리 환경 ✓ 시뮬레이션·최적화 ✓ 비주열 분석 ✓ 분석 전용 서버	✓ 분석 협업 환경 ✓ 분석 Sandbox ✓ 프로세스 내재화 ✓ 빅데이터 분석

분석 진단 결과



빅데이터 플랫폼

- 단순한 분석 응용 프로그램뿐만 아니라 분석 서비스를 위한 응용프로그램이 실행될 수 있는 기초를 이루는 컴퓨터 시스템을 의미함.
- 분석플랫폼이 구성되어 있는 경우에는 새로운 데이터 분석 니즈가 존재할 경우 개별적인 분석시스템을 추가하는 방식이 아닌 서비스를 추가적으로 제공하는 방식의 확장성을 높일 수 있다.

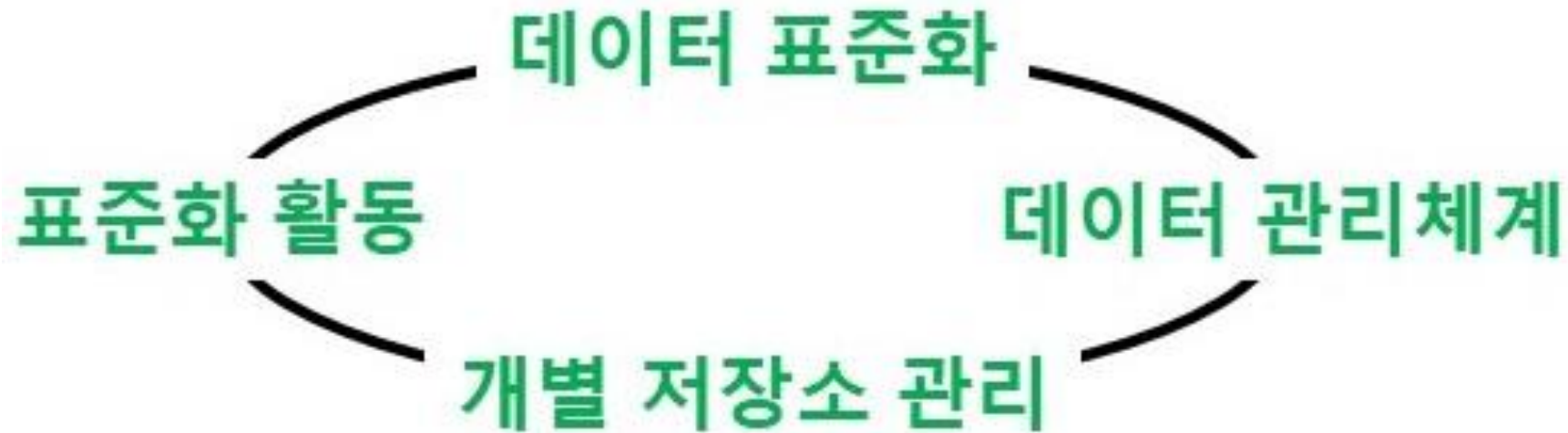


데이터 거버넌스 체계 수립

- 전사 차원의 모든 데이터에 대하여 정책 및 지침, 표준화, 운영조직 및 책임 등의 표준화된 관리 체계를 수립하고 운영을 위한 프레임 워크 및 저장소를 구축하는 것.
- 마스터 데이터(Master data) : 마스터 파일의 내용을 뜻하기도 하며 자료가 자주 변하지 않아 자료 처리 운용에 기본자료로 제공되는 자료의 집합을 말함.
- 메타 데이터(Meta data) : 데이터에 관한 구조화된 데이터로, 다른 데이터를 설명해 주는 데이터.
- 데이터 사전(Data Dictionary) : DBMS를 효율적으로 사용하기 위해 데이터베이스에 저장된 정보를 요약한 것으로 데이터 해석상의 오류 또는 어려움을 제거하기 위해 사용됨.
- 데이터 거버넌스 구성요소

원칙(Principle)	조직(Organization)	프로세스(Process)
<ul style="list-style-type: none">• 데이터를 유지, 관리하기 위한 지침과 가이드<ul style="list-style-type: none">• 보안, 품질 기준, 변경 관리	<ul style="list-style-type: none">• 데이터를 관리할 조직의 역할과 책임• 데이터 관리자, 데이터베이스 관리자, 데이터 아키텍처	<ul style="list-style-type: none">• 데이터 관리를 위한 활동과 체계• 작업 절차, 모니터링 활동, 측정 활동

데이터 거버넌스 체계



- 데이터 표준화 - 데이터 표준 용어 설정, 명명규칙 수립, 메타 데이터 구축, 데이터 사전 구축 등의 업무로 구성
- 데이터 관리 체계 - 데이터 정합성 및 활용의 효율성을 위하여 표준 데이터를 포함한 메타 데이터와 데이터 사전의 관리 원칙을 수립
- 데이터 저장소 관리 - 메타 데이터 및 표준 데이터를 관리하기 위한 전사 차원의 저장소를 구성
- 표준화 활동 - 데이터 거버넌스 체계를 구축한 후 표준 준수 여부를 주기적으로 점검하고 모니터링 실시.

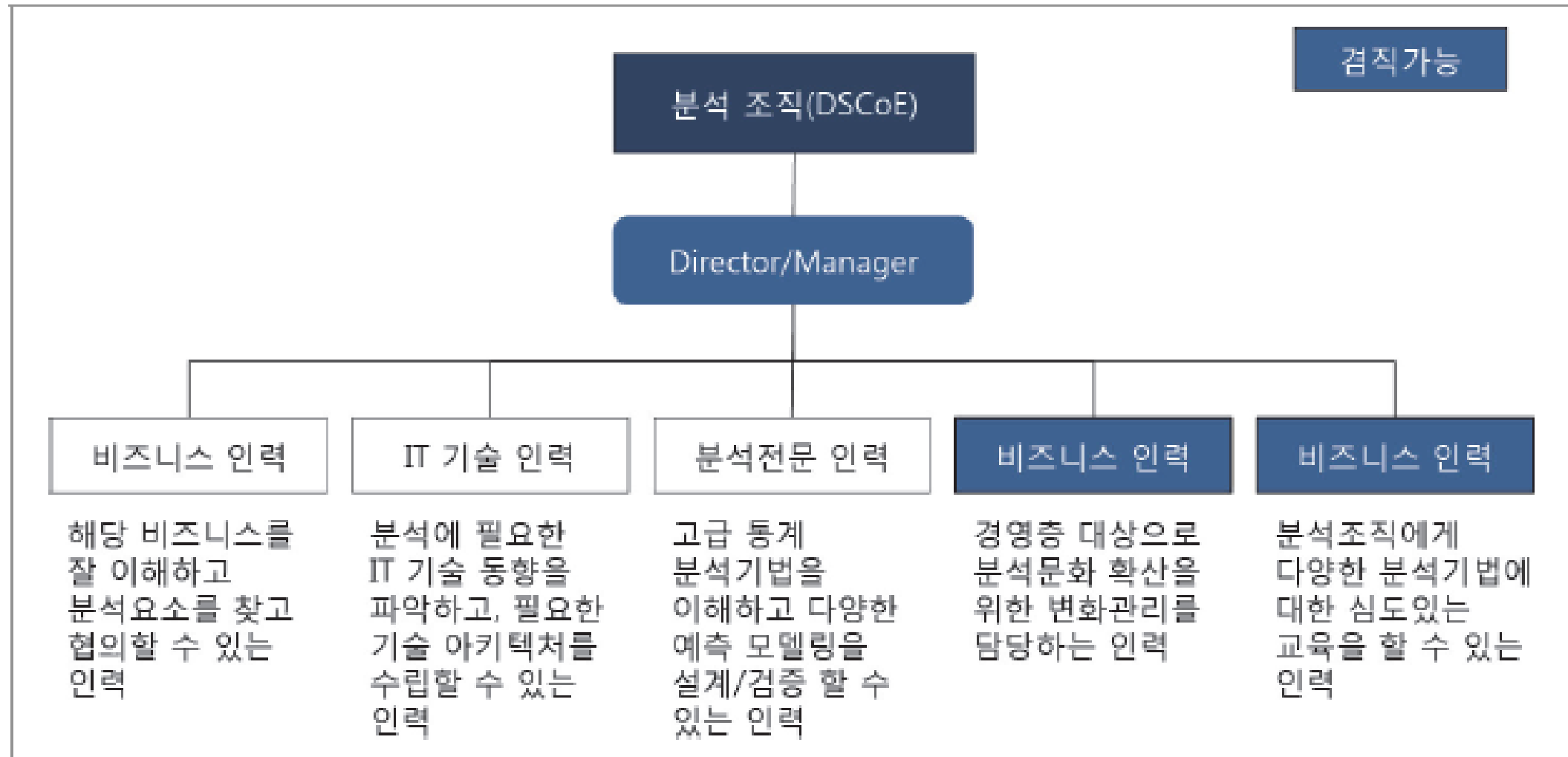
데이터 조직 및 인력 방안 수립



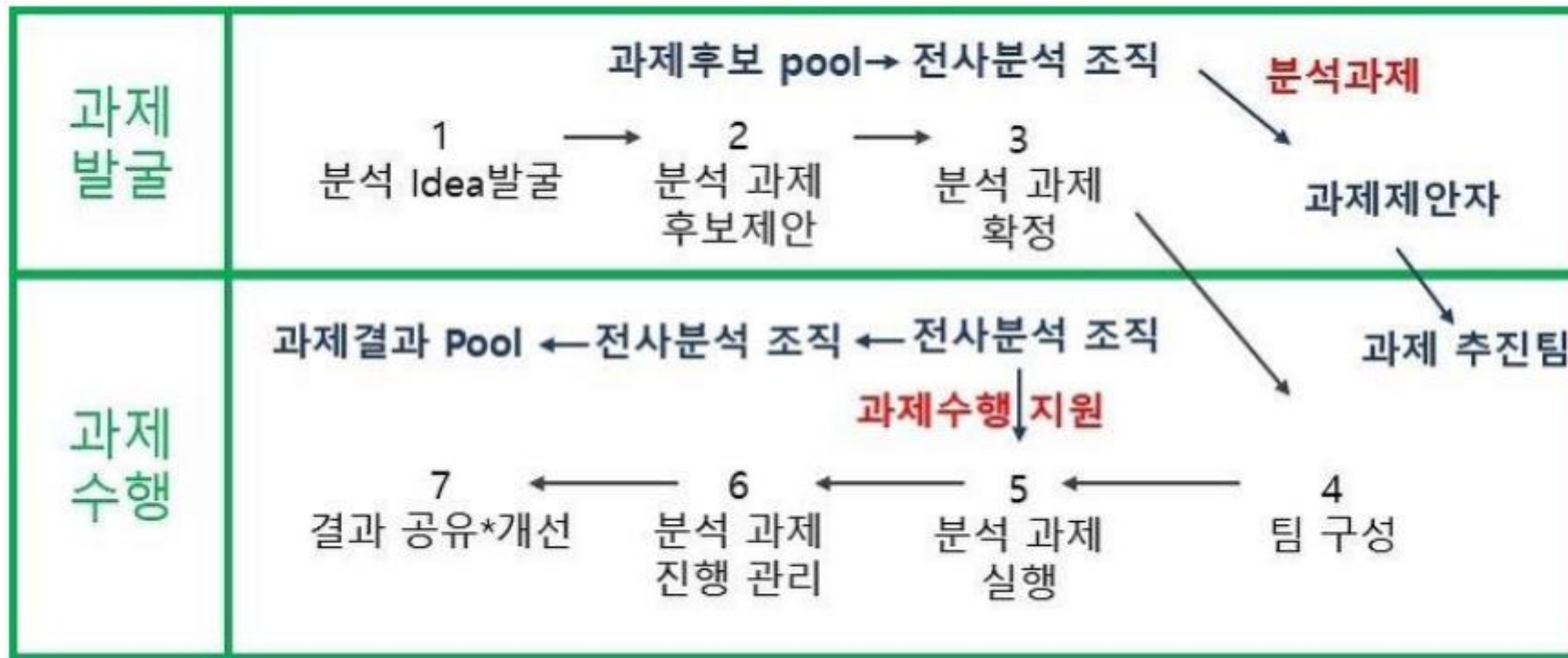
※ DSCoE : Data Science Center of Excellence

집중구조	기능구조	분산구조
<ul style="list-style-type: none"> 전사 분석업무를 별도의 분석전담 조직에서 담당 전략적 중요도에 따라 분석조직이 우선순위를 정해서 진행 가능 이중화/이원화 가능성 높음 	<ul style="list-style-type: none"> 일반적인 분석 수행 구조 별도 분석조직이 없고 해당 업무 부서에서 분석 수행 <ul style="list-style-type: none"> 전사적 핵심분석이 어려움 과거 실적에 국한된 분석 수행 가능성 높음 	<ul style="list-style-type: none"> 분석조직 인력들을 현업부서로 직접 배치하여 분석 업무 수행 <ul style="list-style-type: none"> 전사차원의 우선순위 수행 분석결과에 따른 신속한 액션 가능 부서 분석업무와 역할분담을 명확히 해야함 (업무과다 이원화 가능성)

분석조직의 인력구성



분석 과제 관리 프로세스



- 빅데이터 시대 변화에 적극적인 대응 방법

- : 기업에 맞는 적합한 분석 업무 시행

- : 분석 조직 및 인력에 대한 지속적인 교육 및 훈련 실시

- : 경영층이 사실기반 의사결정 할 수 있는 문화 정착

- : 지속적인 변화관리를 계획하고 수행, 업무 수행자에 대한 분석기회 발굴, 구체화, 시나리오 작성법 등

연습문제

1. 다음 중 분석 주체 유형을 분류할 때 데이터 분석 방법과 다양한 분석 구조의 활용은 충분히 이해하고 있으나, 조직 내 분석 대상이 무엇인지 인지하지 못하는 유형은 무엇인가?

- ① 최적화
- ② 솔루션
- ③ 발견
- ④ 통찰



연습문제

2. 빅데이터를 다각적으로 분석하여 인사이트를 도출하는 데이터 사이언티스트(Data Scientist)의 요구 역량으로 가장 부적절한 것은 무엇인가?

- ① 통찰력 있는 분석 능력
- ② 인공지능 분야 최적화 능력
- ③ 다양한 분야를 아우르는 협업 능력
- ④ 설득력 있는 전달 능력



연습문제

3. 다음은 빅데이터와 인공지능의 관계에 대한 설명이다. 가장 옳지 않은 것은?

- ① 상호보완 관계로 빅데이터는 인공지능 구현 완성도를 높여주고, 빅데이터는 인공지능을 통해 문제해결 완성도를 높인다.
- ② 빅데이터 기술이 주목을 받는 이유는 정보처리능력이 중심이 아니라 우수한 정보처리를 바탕으로 의미 있는 솔루션을 도출할 수 있다는 점이 빅데이터가 주목받는 이유이다.
- ③ 인공지능의 암흑기를 지나 빅데이터를 통해 자체 알고리즘을 가지고 스스로 학습하는 지도 학습 기술로 특정 분야에서 인간의 지능을 뛰어넘는 능력을 갖추게 되었다.
- ④ 빅데이터 목표가 인공지능 목표와 부합하고, 인공지능 판단을 위해서는 빅데이터와 같은 기술이 필수이므로, 빅데이터는 인공지능을 위한 기술이 될 가능성이 크다.



연습문제

4. 아래 (가)와 (나) 순서대로 들어갈 내용으로 적절한 것은?

분석은 분석 대상(What) 및 분석 방법(How)에 따라서 4자리로 나눌 수 있다. 분석 대상이 명확하게 무엇인지 모르는 경우에는 기존 분석 방법을 활용하여 (가)을(를) 도출해 냄으로써 문제의 도출 및 해결에 기여하거나 (나) 접근법으로 분석 대상 자체를 새롭게 도출할 수 있다.

- ① 최적화 - 통찰
- ② 솔루션 - 통찰
- ③ 통찰 - 발견
- ④ 발견 - 솔루션



연습문제

5. 빅데이터 분석은 분석의 대상과 방법에 따른 분류 중에서 분석의 대상은 인지(Know)하고 있으나 방법을 모르는 경우(Un-Known)에 사용하는 유형으로 가장 적절한 것은?

- ① 솔루션 유형
- ② 최적화 유형
- ③ 통찰 유형
- ④ 발견 유형



6. 분석 기획 고려사항 중 장애요소에 대한 설명으로 부적절한 것은?

- ① 비용대비 효과의 적정한 비용
- ② 분석 모형의 안정적 성능 확보
- ③ 이해도 높은 모형보다는 복잡하고 정교한 모형
- ④ 조직역량으로 내재화를 위한 변화 관리



연습문제

7. 다음 분석 방법론 중 순차적으로 진행되면서 이전 단계가 완료된 후 다음 단계로 진행하는 하향식(Top Down)으로 진행되는 특징을 가지고 있는 모델은 ?

- ① 프로토타입(Prototype) 모델
- ② 폭포수 모델(Waterfall) 모델
- ③ 나선형(Spiral) 모델
- ④ 애자일(Agile) 모델



연습문제

8. 분석과제를 발굴하기 위한 상향식 접근법(Bottom Up Approach)에 대한 설명 중 적절한 것은?

- ① 일반적으로 상향식 접근 방식의 데이터 분석은 지도 학습(Supervised learning)방법에 의해 수행된다.
- ② 상향식 접근방법은 문제의 구조가 분명하고 문제 정의가 명확한 경우에 적합하다.
- ③ 디자인 사고 프로세스의 수렴단계(Converge)해당된다.
- ④ 인사이트를 도출 후 반복적인 시행착오를 통해서 수행하며 문제를 도출하는 과정을 말한다.



연습문제

9. 다음 중 CRISP-DM 방법론의 모델링 단계에서 수행하는 태스크(task)로 적절하지 않은 것은 무엇인가?

- ① 모델 테스트 계획 설계
- ② 모델 평가
- ③ 모델링 기법 선택
- ④ 모델 적용성 평가



10. CRISP-DM의 업무 이해(Business Understanding) 주요 구성은 무엇인가?

- ① 업무 목적 파악 - 상황 파악 - 데이터 마이닝 목표 설정 - 프로젝트 계획 수립
- ② 상황 파악 - 업무 목적 파악 - 데이터 마이닝 목표 설정 - 프로젝트 계획 수립
- ③ 업무 목적 파악 - 프로젝트 계획 수립 - 데이터 마이닝 목표 설정 - 상황 파악
- ④ 업무 목적 파악 - 상황 파악 - 프로젝트 계획 수립 - 데이터 마이닝 목표 설정



연습문제

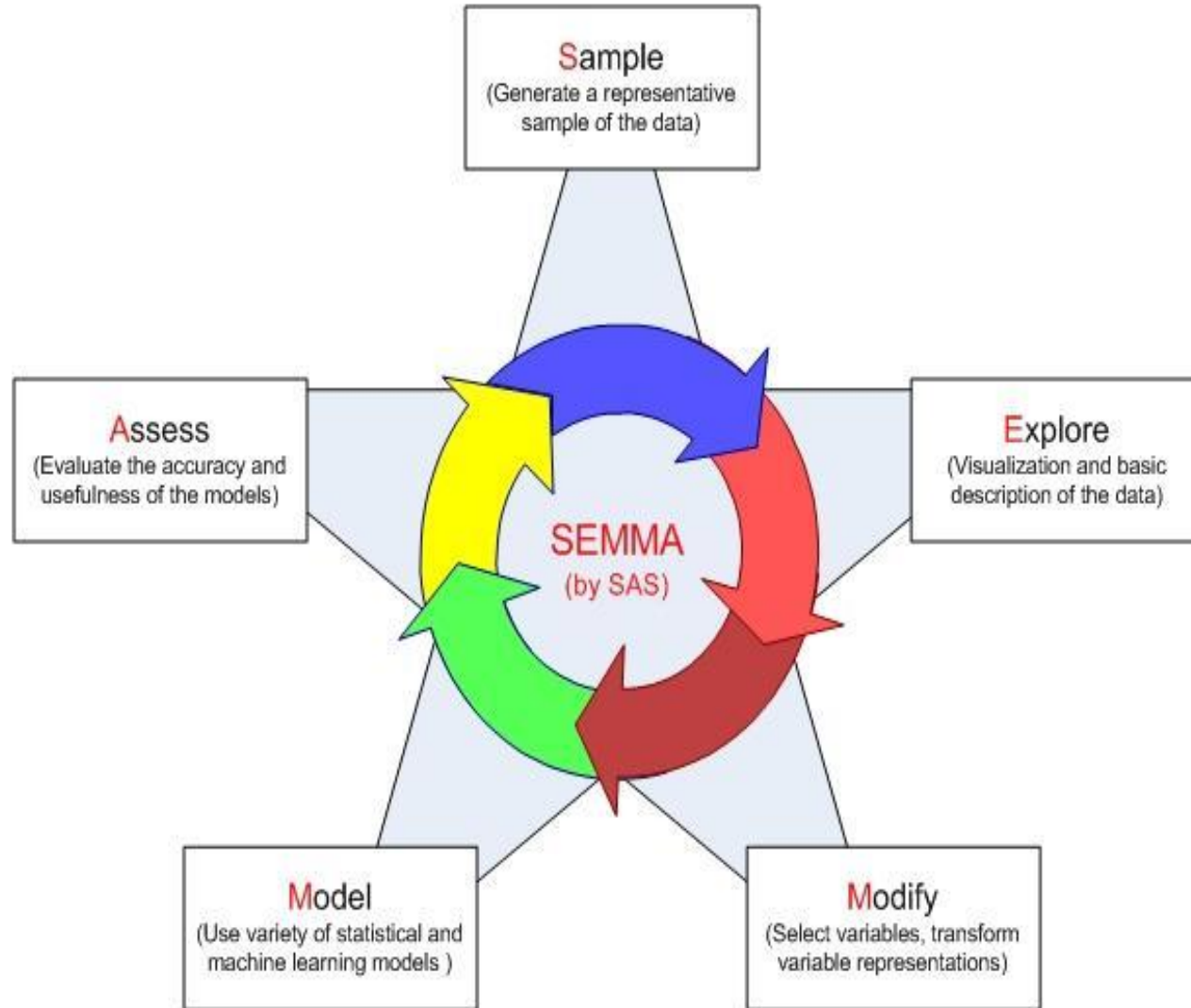
11. 다음이 설명하는 분석 방법론은 무엇인가?

- 1997년 Fayyad가 프로파일링 기술을 기반으로 통계적 패턴이나 지식을 찾기 위해 체계적으로 정리한 방법론.
- 분석 절차는 데이터 세트 선택 데이터 전처리, 데이터 변환, 데이터 마이닝, 데이터 마이닝 결과 평가의 5단계이다.

- ① KDD
- ② CRISP-DM
- ③ SEMMA
- ④ K-means



Sampling Exploration Modification Modeling Assessment, SEMMA



단계	상세
Sampling	<ul style="list-style-type: none"> ■ 분석 데이터 생성 <ul style="list-style-type: none"> ■ 통계적 추출 ■ 조건 추출 ■ 비용 절감 및 모델 평가를 위한 데이터 준비
Explore	<ul style="list-style-type: none"> ■ 분석 데이터 탐색 <ul style="list-style-type: none"> ■ 기초 통계, 그래픽적 탐색 ■ 요인별 분할표 ■ Clustering ■ 변수 유의성 및 상관분석 ■ 데이터 조감을 통한 데이터 오류 검색 ■ 모델의 효율 증대 ■ 데이터 현황을 통해 비즈니스 이해, 아이디어를 위해 이상현상, 변화 등을 탐색
Modify	<ul style="list-style-type: none"> ■ 분석 데이터 수정/변환 <ul style="list-style-type: none"> ■ 수량화, 표준화, 각종 변환, 그룹화 ■ 데이터가 지닌 정보의 표현 극대화 ■ 최적의 모델을 구축할 수 있도록 다양한 형태로 변수를 생성, 선택, 변형
Modeling	<ul style="list-style-type: none"> ■ 모델 구축 <ul style="list-style-type: none"> ■ Neural Network ■ Decision Tree ■ Logistic Regression ■ 전통적 통계 ■ 데이터의 숨겨진 패턴 발견 ■ 하나의 비즈니스 문제 해결을 위해 특수의 모델과 알고리즘 적용 가능
Assessment	<ul style="list-style-type: none"> ■ 모델 평가 및 검증 <ul style="list-style-type: none"> ■ 텍스트 교본 ■ Feedback ■ 모델의 검증 ■ 서로 다른 모델을 동시에 비교 ■ 추가 분석 수행 여부 결정

12. 프로토타이핑(Prototyping) 접근법에 대한 설명으로 가장 적절한 것은?

- ① 문제가 정형화되어 있고 문제해결을 위한 데이터가 완벽하게 조직에 존재하는 경우 효과적이다.
- ② 신속하게 해결책이나 모형을 제시함으로써 이를 바탕으로 문제를 좀 더 명확하게 인식하고 필요한 데이터를 식별하여 구체화할 수 있게 하는 유용한 상향식 접근 방법이다.
- ③ 문제가 주어지고 이에 대한 해법을 찾기 위하여 각 과정이 체계적으로 단계화되어 수행하는 방식이다.
- ④ 문제 정의가 불명확하거나 이전에 접하지 못한 새로운 문제일 경우에는 적용하기 어렵다.



13. 프로토타이핑 프로세스에 대한 설명 중 적절하지 않은 것은?

- ① 가설의 생성
- ② 디자인에 대한 실험
- ③ 실제 환경에서의 테스트
- ④ 반복적으로 위험분석을 수행하여 위험관리 개선



연습문제

14. 빅데이터 분석 방법론의 분석 기획 단계 순서가 바르게 연결된 것은?

- ① 프로젝트 범위 설정 - 데이터 분석 프로젝트 정의 - 프로젝트 수행계획 수립 - 데이터 분석 위험 식별
- ② 프로젝트 범위 설정 - 데이터 분석 프로젝트 정의 - 데이터 분석 위험 식별 - 프로젝트 수행계획 수립
- ③ 데이터 분석 위험 식별 - 프로젝트 범위 설정 - 프로젝트 수행계획 수립 - 데이터 분석 프로젝트 정의
- ④ 데이터 분석 위험 식별 - 프로젝트 범위 설정 - 데이터 분석 프로젝트 정의 - 프로젝트 수행계획 수립



연습문제

15. 다음 중 기업에서 데이터에 기반한 의사결정을 방해하는 요소들로 구성된 것은?

- ① 바이어스, 비편향적 사고
- ② 프레이밍 효과, 고정관념
- ③ 프레이밍 효과, 직관력
- ④ 직관력, 비편향적 사고



연습문제

16. 비즈니스 모델 캔버스는 문제 탐색단계에서 사용되는 도구이다. 이를 구성하는 단위로 적절한 것은?

- ① 영업 - 고객 - 제도 - 업무 - IT 인프라
- ② 업무 - 제품 - 고객 - 규제와 감사 - 자원인프라
- ③ 프로세스 - 제품 - 고객 - 영업 - IT인프라
- ④ 비용 - 수입 - 고객 - 상품 - 감사



연습문제

17. 데이터 분석에서는 하향식 접근 방법 방식과 상향식 접근 방식으로 분석과제를 발굴하게 되는데, 다음 중 하향식 접근 방식의 단계에서 타당성 평가에 대한 설명으로 부적절한 것은?

- ① 도출된 분석 문제에 대한 대안을 과제화하기 위해서는 다각적 타당성 검토가 필요하다.
- ② 경제적 타당성은 비용대비 효익의 관점에서 평가한다.
- ③ 데이터 타당성 확보를 위하여 문제발생 포인트에 대한 데이터 확보가 중요하다.
- ④ 기술적 타당성 분석 시 적용 가능한 요소기술 확보 방안에 대한 사전 고려가 필요하다.



연습문제

18. 하향식 데이터 분석 기획에서 문제 탐색 단계에 대한 설명으로 가장 부적절한 것은?

- ① 빠짐없이 문제를 도출하고 식별하는 것이 중요
- ② 문제를 해결함으로써 발생하는 가치에 중점을 두는 것이 중요
- ③ 비즈니스 모델 캔버스는 문제 탐색 도구로 활용
- ④ 문제 탐색은 유즈케이스 활용보다는 새로운 이슈탐색이 우선



19. 하향식 접근방법의 문제탐색 관련한 거시적 관점의 요인이 아닌 것은?

- ① 사회(Social)
- ② 기술(Technology)
- ③ 환경(Environment)
- ④ 채널(Channel)



연습문제

20. 인간에 대한 관찰과 공감을 바탕으로 다양한 대안을 찾는 확산적 사고와 주어진 상황에 대한 최선의 방법을 찾는 수렴적 사고의 반복을 통해 과제를 발굴하는 상향식 접근 방법은 무엇인가?

- ① 비즈니스 사고
- ② 비즈니스 모델 캔버스
- ③ 디자인 사고
- ④ 탐색적 접근



연습문제

21. 다음 중 '분석과제 정의서'에 대한 설명으로 가장 적절한 것은?

- ① '분석과제 정의서'에는 소스 데이터, 데이터 입수 및 분석의 난이도, 분석방법 등에 대한 항목이 포함되어야 한다.
- ② '분석과제 정의서'는 프로젝트 수행하는 이해관계자가 프로젝트의 방향을 설정하고 성공여부를 판별할 수 없는 자료이다.
- ③ '분석과제 정의서'에는 분석모델에 적용된 알고리즘과 분석모델의 기반이 되는 Feature가 포함되어야 한다.
- ④ '분석과제 정의서'는 프로젝트 계획서를 작성하기 위한 중간 결과로써 구성 항목(Configuration Item)으로 도출할 필요는 없다.



연습문제

22. 다음 보기 빈칸에 공통으로 들어갈 알맞은 용어는?

현재의 비즈니스 모델 및 유사, 동종사례 탐색을 통해서 빠짐없이 도출한 분석 기회들을 구체적인 과제로 만들기 전에 ()로 표기하는 것이 필요하다. 풀어야 할 문제에 대한 상세설명 및 해당 문제 해결했을 때 발생하는 효과를 명시함으로써 향후 데이터 분석 문제로의 전환 및 적합성 평가에 ()을 활용하도록 한다.

- ① 분석과제 정의서
- ② 분석유즈 케이스
- ③ 분석프로젝트
- ④ 과제 관리 프로세스



23. 문제 탐색의 도구가 아닌 것은?

- ① 비즈니스 모델 기반 문제 탐색
- ② 탐색적 문제 발견
- ③ 외부 참조 모델 기반 문제 탐색
- ④ 분석 유즈 케이스



연습문제

24. 다음 중 분석 프로젝트 관리에 대한 설명으로 가장 부적절한 것은?

- ① 데이터 분석 모델의 품질을 평가하기 위해서 SPICE를 활용할 수 있다.
- ② 분석 프로젝트 관리는 KSA ISO 21500:2013를 가이드로 활용할 수 있다.
- ③ 분석 프로젝트의 일정계획 수립시 데이터 수집에 대한 철저한 통제와 관리가 필요
- ④ 분석 프로젝트의 최종 산출물이 보고서인지 또는 시스템인지에 따라 프로젝트 관리에 차이가 있다.



연습문제

25. 분석 마스터 플랜과 ISP의 관계에 대한 설명 중 적절하지 않은 것은?

- ① 분석 마스터 플랜은 ISP와는 다르게 인프라와 모델링에만 신경 쓴다.
- ② ISP와 같은 일반적인 IT 프로젝트는 과제의 우선순위 평가를 위해 전략적 중요도, 실행 용이성 등 기업에서 고려하는 중요 가치 기준에 따라 우선순위를 평가한다.
- ③ 빅데이터 3V는 투자비용 요소에 해당된다.
- ④ Values는 비즈니스 효과에 해당한다.



26. 다음 중 분석 성숙도 모델에 대한 설명 중 성격이 다른 것은?

- ① 분석 CoE 조직 운영
- ② 데이터 사이언스 그룹
- ③ 빅데이터 분석
- ④ 분석 샌드박스



연습문제

27. 다음 중 빅데이터 분석 과제에서 프로젝트 관리에 대한 설명으로 가장 부적절한 것은?

- ① 분석 과제는 분석 전문가의 상상력을 요구하므로 일정을 제한하는 일정계획은 적절하지 못한다.
- ② 분석 과제에는 많은 위험이 있어 사전에 위험을 식별하고 대응방안을 수립해야 한다.
- ③ 분석과제는 적용되는 알고리즘에 따라 범위가 변할 수 있어 범위관리가 중요하다.
- ④ 분석과제에서 다양한 데이터를 확보하는 경우가 있어 조달관리 또한 중요하다.



연습문제

28. 빅데이터 분석방법론 분석기획 단계에서 발생하는 산출물로 프로젝트에 참여하는 관계자들 이해를 일치시키기 위한 결과물을 무엇이라 하는가?

- ① 데이터 스토어
- ② SOW(State of Work)
- ③ 상세 알고리즘
- ④ WBS(work Breakdown Structure)



연습문제

29. 다음 중 데이터 분석을 위한 수준진단에서 ‘분석 준비도’의 분석 데이터의 진단 항목으로 가장 부적절 한 것은?

- ① 분석 업무를 위한 데이터 충실성, 신뢰성, 적시성
- ② 내부데이터 집중 활용 체계
- ③ 기준데이터 관리(MDM)
- ④ 비구조적 데이터 관리



연습문제

30. 아래에서 설명하는 데이터 거버넌스 체계 항목은 무엇인가?

데이터 표준 용어 설정, 명명 규칙 수립, 메타 데이터 구축, 데이터 사전 구축 등의 업무로 구성된 데이터 거버넌스 체계

- ① 데이터 표준화
- ② 데이터 관리 체계
- ③ 데이터 저장소 관리
- ④ 표준화 활동



연습문제

31. 데이터 분석을 위한 조직 구조 중 아래 보기에 해당하는 것은?

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

- ① 집중구조
- ② 기능구조
- ③ 분산 구조
- ④ 복합 구조



연습문제

32. 분석 마스터 플랜에 대한 설명으로 가장 부적절한 것은?

- ① 과제 우선순위 평가는 비즈니스 효과인 시급성과 투자비용 요소인 난이도에 근거하여 결정된다.
- ② 분석 마스터 플랜은 분석과제 도출, 우선순위 결정, 중장기 마스터 플랜 수립과제 도출 순으로 된다.
- ③ 과제별 데이터 분석 체계는 폭포수 방식도 있으나 반복적인 정렬과정을 통하여 과제의 완성도를 높이는 방식으로 많이 사용된다.
- ④ 분석 과제 로드맵은 과제의 우선순위를 고려하여 작성하되 과제별 선후관계를 감안하여 반복이 없는 계획을 작성한다.



연습문제

33. 마스터 플랜 수립 단계에서 다양한 분석 과제를 도출한 후 데이터 분석 구현을 위한 로드맵 수현 시 고려되어야 하는 요소들 중 성격이 다른 하나는?

- ① 기술 적용 수준
- ② 분석 데이터 적용수준
- ③ 전략적 중요도 수준
- ④ 업무 내재화 적용 수준



연습문제

34. 다음 중 분석 기회 발굴의 범위 확장 방법에 관한 설명으로 부적절한 것은?

- ① 거시적 관점의 메가트렌드에서는 현재의 조직 및 해당 산업에 폭넓게 영향을 미치는 사회, 경제적 요인을 사회, 기술, 경제, 환경, 정치 영역으로 나누어서 좀 더 폭 넓게 기회 탐색을 수행한다.
- ② 경쟁자 확대 관점에서는 현재 수행하고 있는 사업의 영역의 직접 경쟁사 및 제품, 서비스를 중심으로 현 상황에 대한 분석 기회 발굴의 폭을 넓혀서 탐색한다.
- ③ 시장의 니즈 탐색 관점에서는 현재 수행하고 있는 사업에서의 직접 고객뿐만 아니라 고객과 접촉하는 역할을 수행하는 채널 및 고객의 구매와 의사결정에 영향을 미치는 영향자들에 대한 폭넓은 관점을 바탕으로 분석 기회를 탐색한다.
- ④ 역량의 재해석 관점에서는 현재 해당 조직 및 기업이 보유한 역량뿐만 아니라 해당 조직의 비즈니스에 영향을 끼치는 파트너 네트워크를 포함한 활용 가능한 역량을 토대로 폭넓은 분석기회를 탐색한다.



연습문제

35. 빅데이터를 활용한 비즈니스는 기업에 많은 변화를 가져오고 있다. 다음 중 기업에서 이러한 변화를 수용하기 위한 중장기적 대응 방안으로 가장 거리가 먼 것은?

- ① 분석 조직 및 인력에 대한 교육과 훈련
- ② 데이터 기반의 의사결정문화 정착
- ③ 데이터 분석 도구 기반의 교육
- ④ 분석역량 강화를 위한 체계적인 계획 및 시행



연습문제

36. 빅데이터와 기존 정형데이터를 품질관리 측면에서 비교해 볼 때 빅데이터의 품질관리의 특성으로 적절하지 않은 것은?

- ① 데이터에 대한 소유권(Ownership)이 더 강하다.
- ② 생성된 데이터에 대한 데이터 정제가 어렵다
- ③ 개별 데이터에 대한 타당성 검증은 경우에 따라 불필요하다.
- ④ 데이터의 사용자의 오류는 무시할 수도 있다.



연습문제

37. 다음 중 분석과제 관리 프로세스에 대한 설명으로 가장 부적절한 것은?

- ① 과제발굴단계에는 분석 아이디어 발굴, 분석과제 후보제안, 분석과제 확정 프로세스가 있다.
- ② 분석과제로 확정된 분석과제를 풀(Pool)로 관리한다.
- ③ 분석과제 중에 발생한 시사점과 분석 결과물은 풀(Pool)로 관리하고 공유한다.
- ④ 과제수행단계에서는 팀 구성, 분석과제 실행, 분석과제 진행관리, 결과 공유 프로세스가 있다.



38. 분석 과제 수행시 고려하는 5요소가 아닌 것은?

- ① 데이터 분석 방법
- ② 데이터 크기
- ③ 데이터 복잡성
- ④ 스피드



연습문제

39. 다음 중 비즈니스 모델 분석의 상향식 접근방식에서의 특정 업무 영역의 주제 지향적 분석 기회를 발굴하는 절차가 옳은 것은?

- ① 프로세스 분류 → 분석요건 정의 → 분석요건 식별 → 프로세스 흐름 분석
- ② 프로세스 분류 → 프로세스 흐름 분석 → 분석요건 식별 → 분석 요건 정의
- ③ 분석요건 식별 → 프로세스 분류 → 프로세스 흐름 분석 → 분석 요건 정의
- ④ 분석요건 식별 → 분석요건 정의 → 프로세스 분류 → 프로세스 흐름 분석



연습문제

40. 다음 중 아래에서 설명하는 데이터 거버넌스의 체계 요소는 무엇인가?

- 메타 데이터 관리 원칙 수립
- 데이터 사전 관리 원칙 수립
- 데이터 생명 주기 관리 방안 수립

- ① 데이터 표준화
- ② 데이터 관리 체계
- ③ 데이터 저장소 관리
- ④ 표준화 활동



연습문제

41. 다음 중 빅데이터 분석에서 Self Service Analytics에 대한 설명으로 가장 부적절한 것은?

- ① Self Service Analytics에 포함되어야 하는 주요 기능은 BI도구, Ad hoc Report, OLAP, Visual Discovery, Machine learning등이다.
- ② Self Service Analytics의 성공적인 적용을 위해서는 Reference Method의 작성 및 공유, 표준 데이터의 활용, 데이터 거버넌스, 도구 사용에 대한 지속적인 교육이 필요하다.
- ③ Self Service Analytics를 수행하기 위해서는 R, Python 등의 데이터 분석 언어와 많은 통계적 지식을 필요로 한다.
- ④ Self Service Analytics는 산업용뿐만 아니라 OSS(Open Service Software)로도 구현되고, 분석의 성능 향상을 위하여 분산처리도 지원하고 있다.



연습문제

42. 다음 중 빅데이터 분석 방법론의 분석 기획 단계에서 프로젝트 위험 대응 계획을 수립할 때 예상되는 위험에 대한 대응 방법의 구분으로 부적절한 것은?

- ① 회피(Avoid)
- ② 관리(Manage)
- ③ 완화(Mitigate)
- ④ 수용(Accept)



43. 아래 (가)에 들어갈 용어를 기입하시오.

데이터 거버넌스 체계에서 데이터 저장소(Repository) 관리란 메타데이터 및 표준 데이터를 관리하기 위한 전사 차원의 저장소로 구성된다. 저장소는 데이터 관리 체계 지원을 위한 워크플로우 및 관리용 응용소프트웨어를 지원하고 관리 대상 시스템과의 인터페이스를 통한 통제가 이루어져야 한다. 또한 데이터 구조 변경에 따른 (가)도 수행되어야 효율적인 활용이 가능하다.



44. 아래 (가)에 들어갈 용어를 기입하시오.

분석 모델을 가동중인 운영시스템에 적용하기 위해서는 모델에 대한 상세한 “알고리즘 설명서” 작성이 필요하다. “알고리즘 설명서”는 “시스템 구현” 단계에서 중요한 입력 자료로 활용되므로 필요시 (가) 수준의 상세한 작성이 필요하다.



연습문제

45. 아래 (가), (나) 안에 들어갈 정확한 용어를 기입하시오.

데이터 거버넌스 체계는 데이터 저장소(Repository)는 메타데이터 및 표준 데이터를 관리하기 위한 전사 차원의 저장소로 구성된다. 저장소는 데이터 관리 체계 지원을 위한 (가) 및 관리용 응용소프트웨어를 지원하고 관리 대상 시스템과의 인터페이스를 통한 통제가 이루어져 한다. 또한 데이터 구조 변경에 따른 (나)도 수행되어야 효율적인 활용이 가능하다.



연습문제

46. KDD 분석 방법론에서 잡음, 이상치, 결측치를 식별하여 분석용 데이터셋을 선택하고 분석에 필요한 변수 등을 선정하는 단계와 유사한 CRISP-DM 방법 방법론은 단계는?



연습문제

47. 합리적 의사결정을 방해하는 요소로써 표현방식 및 발표자에 따라 동일한 사실에도 판단을 달리하는 현상을 무엇이라 하는가?



연습문제

48. 문제가 주어지고 이에 대한 해법을 찾기 위하여 각 과정이 체계적으로 단계화되어 수행하는 분석과제
발굴 방식을 무엇이라고 하는가?



연습문제

49. 아래는 여러 분석 방법론 중 하나에 대한 설명이다. () 안에 적절한 용어는?

() 모델은 반복을 통하여 점증적으로 개발하는 방법으로 처음 시도하는 프로젝트에 적용이 용이하지만, 반복에 대한 관리 체계를 효과적으로 갖추지 못한 경우 복잡도가 상승하여 프로젝트 진행이 어려울 수 있다.



연습문제

50. 아래 () 안에 공통적으로 들어갈 용어는?

기업 및 공공기관에서는 시스템의 중장기 로드맵을 정의하기 위한 ()을(를) 수행한다. ()은(는) 정보기술 또는 정보시스템을 전략적으로 활용하기 위하여 조직 내, 외부 환경을 분석하여 기회나 문제점을 도출하고 사용자의 요구사항을 분석하여 시스템 구축 우선순위를 결정하는 등 중장기 마스터플랜을 수립하는 절차이다.



연습문제

51. 데이터 분석 도입의 수준을 파악하기 위한 분석 준비도의 6가지 구성요소 중 하나로서 운영시스템 데이터 통합, 빅데이터 분석 환경, 통계분석 환경 등을 진단하는 구성요소는 무엇인가?





Thank you.

ADSP / 류영표 강사
ryp1662@gmail.com