

빅데이터의 이해와 활용

Understanding and Using Big Data

12

빅데이터 서비스 기획과 도입 사례

학습 내용

- 01 빅데이터 분석업무를 위한 지식과 기술
- 02 빅데이터 서비스 기획과 도입 사례

학습 목표

- 데이터사이언스의 역할과 직무를 설명할 수 있다.
- 빅데이터 분석시스템을 설명할 수 있다.
- 빅데이터 처리 과정을 나열할 수 있다.
- 빅데이터 분석기획을 위한 기반 지식을 적용할 수 있다.
- 업무 도메인별 분석업무의 차이점을 구분할 수 있다.

주요 용어

통계학

- 산술적 방법을 기초로 하여, 주로 다량의 데이터를 관찰하고 정리 및 분석하는 방법을 연구하는 수학의 한 분야
- 근대 과학으로서의 통계학은 19세기 중반 벨기에의 케틀레가 독일의 “국상학(國狀學, Staatenkunde, 넓은 의미의 국가학)”과 영국의 “정치 산술(Political Arithmetic, 정치 사회에 대한 수량적 연구 방법)”을 자연과학의 “확률 이론”과 결합하여, 수립한 학문에서 발전되었음

〈출처 : 위키백과〉

주요 용어

데이터분석

- 유용한 정보를 발굴하고 결론 내용을 알리며 의사결정을 지원하는 것을 목표로 데이터를 정리, 변환, 모델링하는 과정
- 여러 면과 접근 방식이 있으며 다양한 이름의 다양한 기술을 아우르며 각기 다른 비즈니스, 과학, 사회과학 분야에 사용됨
- 오늘날 비즈니스 부문에서 데이터 분석은 의사 결정을 더 과학적으로 만들어주고 비즈니스를 더 효율적으로 운영할 수 있도록 도와주는 역할을 함

〈출처 : 위키백과〉

사전 퀴즈

01

데이터 사이언티스트는 데이터 처리와 관련된 IT 영역 업무를 수행하고, 수학 · 확률모델, 머신러닝 등과 같은 분석적 영역을 학습하지만, 비즈니스 컨설팅 영역은 업무전문가에게 그 역할을 맡겨야 한다.

Q1

Q2

O

X

사전 퀴즈

01

데이터 사이언티스트는 데이터 처리와 관련된 IT 영역 업무를 수행하고, 수학 · 확률모델, 머신러닝 등과 같은 분석적 영역을 학습하지만, 비즈니스 컨설팅 영역은 업무전문가에게 그 역할을 맡겨야 한다.

Q1

Q2

O

X

정답

X

해설

데이터 사이언티스트는 데이터 처리와 관련된 IT 영역 업무를 수행하고, 수학 · 확률모델, 머신러닝 등과 같은 분석적 영역을 학습하며, 비즈니스 컨설팅 영역은 업무전문가와 소통할 수 있도록 커뮤니케이션, 프리젠테이션, 스토리텔링 및 시각화 등의 기능들에 대해 익숙해져야 합니다.

사전 퀴즈

02

Q1

Q2

분석할 데이터 셋(집합)에서 개인을 식별할 수 있는 요소를 전부 또는 일부를 삭제하거나 다른 값으로 대체하는 등의 방법으로 개인을 알아볼 수 없도록 하는 기술을 “개인정보 비식별 기술”이라고 한다.



정답

0

해설

분석할 데이터 셋(집합)에서 개인을 식별할 수 있는 요소를 전부 또는 일부를 삭제하거나 다른 값으로 대체하는 등의 방법으로 특정 개인을 알아볼 수 없도록 하는 기술을 “개인정보 비식별 기술”이라고 합니다. 비식별 기술에는 데이터마스킹, 가명처리, 총계처리, 데이터값 삭제, 데이터 범주화 등이 있습니다.



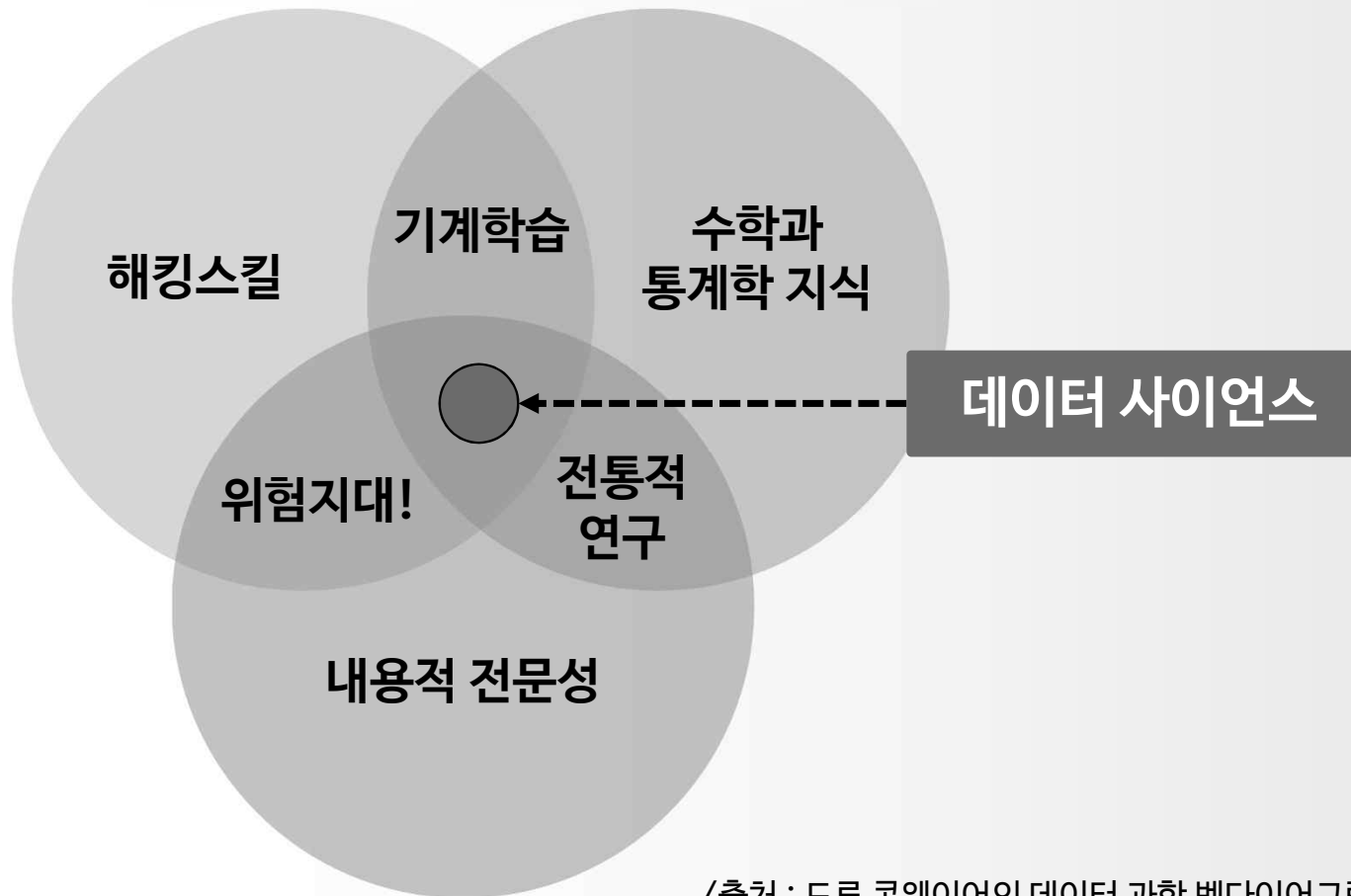
01

빅데이터 분석 업무를 위한 지식과 기술

- 1) 데이터 사이언스의 이해
- 2) 데이터 사이언스 학습체계

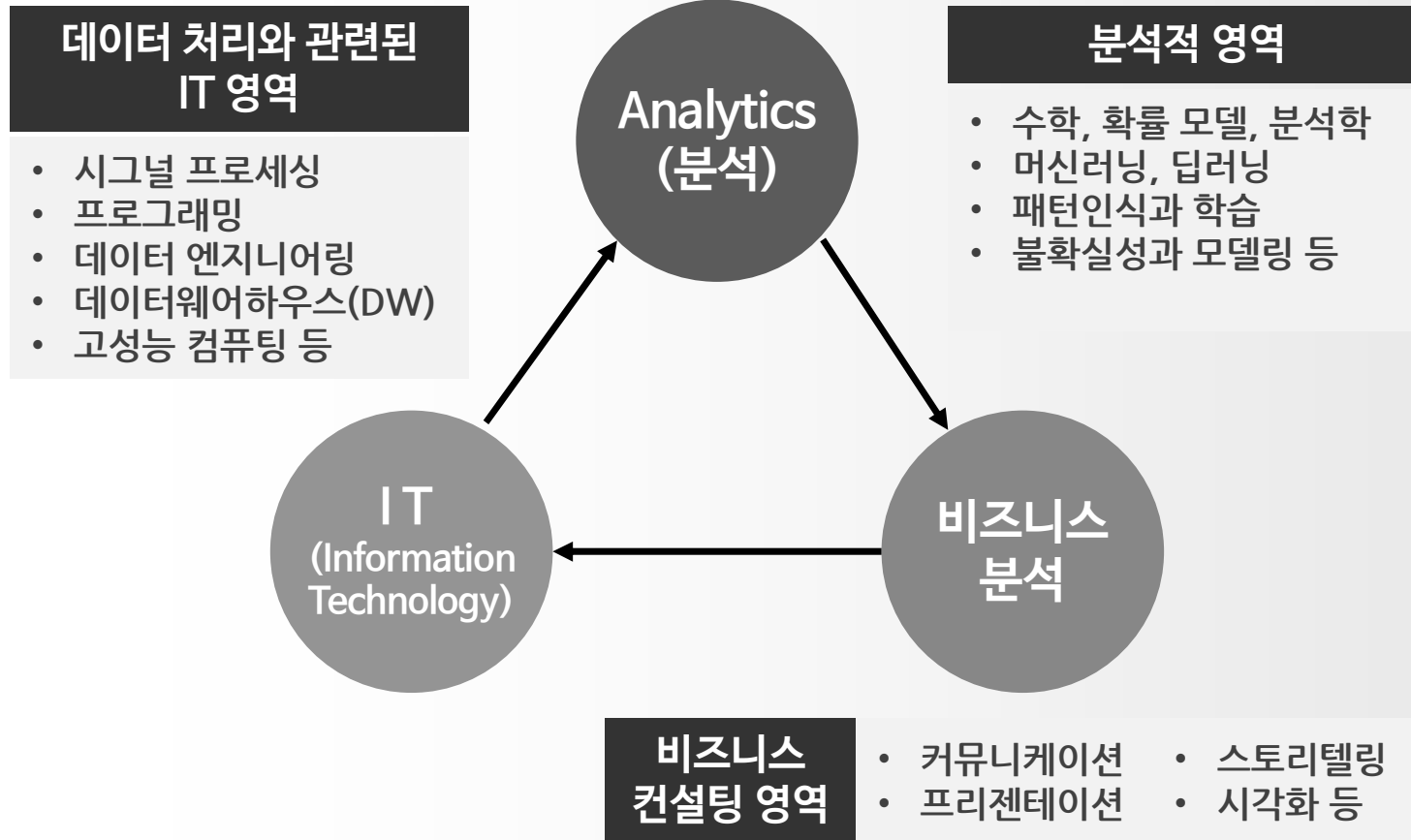
1) 데이터 사이언스의 이해

● 데이터 사이언스란



1) 데이터 사이언스의 이해

● 데이터 사이언티스트의 영역



〈출처 : 데이터분석전문가가이드, K-data(2014)〉

1) 데이터 사이언스의 이해

- 데이터 분석을 위한 다양한 직업

- ▶ 데이터 분석 직업군



〈출처 : 링크드인〉



1) 데이터 사이언스의 이해

- 데이터 분석을 위한 다양한 직업

- ▶ 데이터 분석 직업군

01

Chief Actuary of Geospatial Analytics and Modeling

- 보험모델과 지리공간 데이터 분석 담당 최고책임자

02

Chief Analytics & Algorithms Officers

- 데이터분석과 알고리즘 담당 최고책임자

03

Chief Credit & Analytics Officers

- 신용 및 데이터 분석 담당 최고책임자

〈출처 : 링크드인〉

1) 데이터 사이언스의 이해

- 데이터 분석을 위한 다양한 직업

- ▶ 데이터 분석 직업군

- 04 Chief Research & Analytics Officers

- 연구 및 분석 담당 최고책임자

- 05 Chief Scientist, Global Head of Analytics Head of Analytics

- 데이터분석 글로벌 책임자 겸 최고위 과학자

〈출처 : 링크드인〉

1) 데이터 사이언스의 이해

- 데이터 분석을 위한 다양한 직업

- ▶ 데이터 분석 직업군

06 Director - BI & Analytics

- BI 및 분석 담당 임원

07 Director - Fraud Analytics & R&D











- R&D 및 사기데이터 분석 담당 임원

〈출처 : 링크드인〉

1) 데이터 사이언스의 이해

● 데이터 분석가와 데이터 사이언티스트

▶ 데이터 분석가와 데이터 사이언티스트의 분석범위

구분		데이터 분석가	데이터 사이언티스트
기초통계	아는 것을 확인하고 이해		
추정 및 예측 분석(관계)	무엇을 모르는지 아는 단계		
추정 및 예측 분석(원인)	무엇을 모르는지 아는 단계		
예측 최적화 분석	무엇을 모르는지 모르는 단계		
기계학습	무엇을 모르는지 모르는 단계		



〈0 % 노력〉



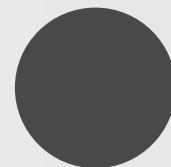
〈25 % 노력〉



〈50 % 노력〉



〈75 % 노력〉

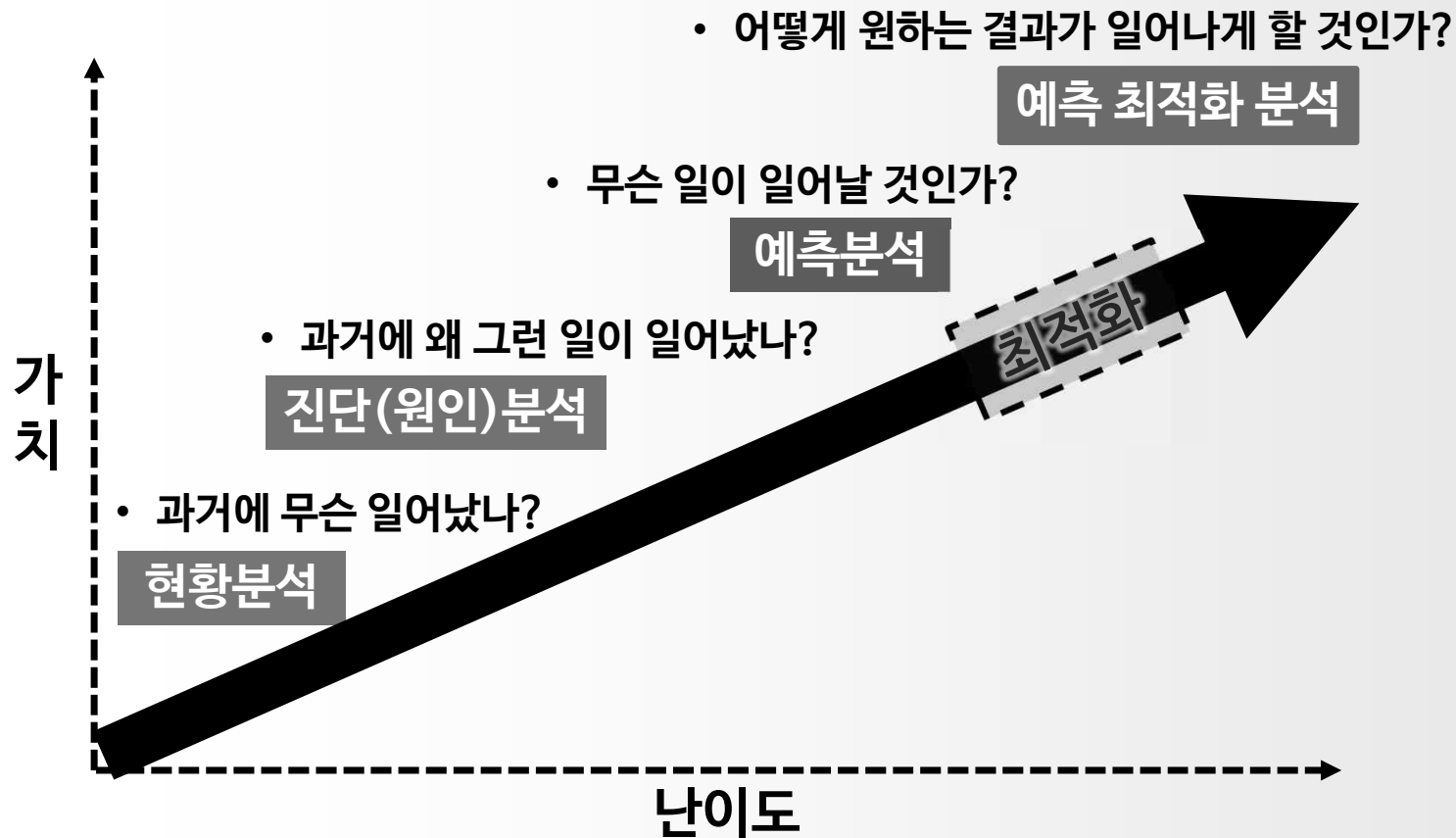


〈100 % 노력〉

〈출처 : 재가공, 데이터과학 무엇을 하는가, 김옥기(2015)〉

1) 데이터 사이언스의 이해

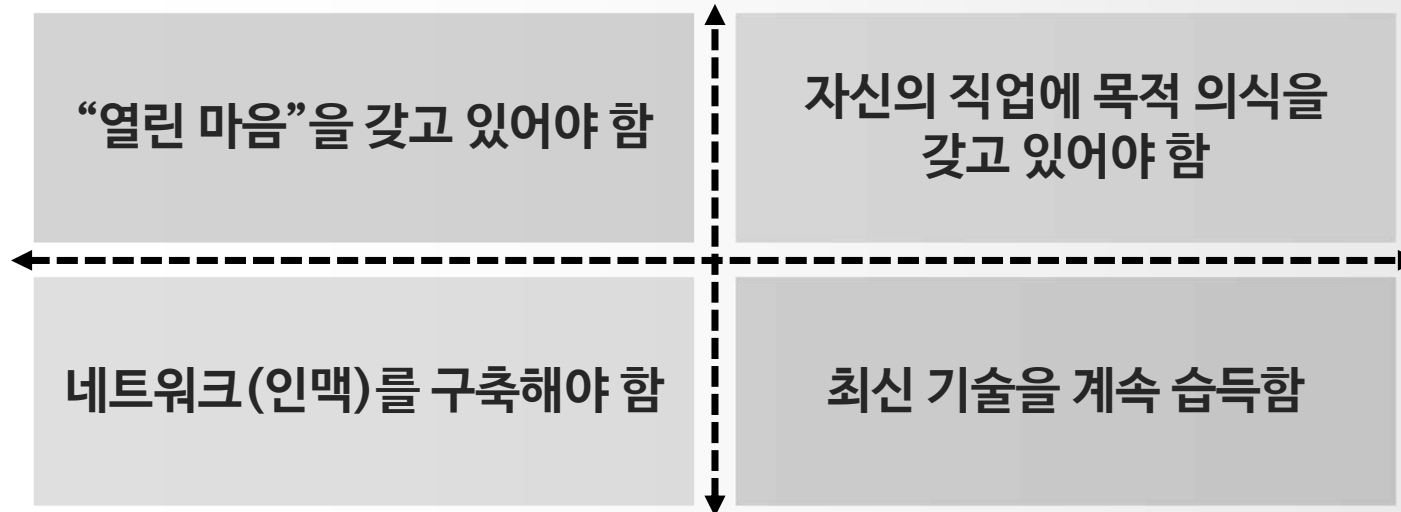
● 단계별 분석기술 비교



〈출처 : 가트너〉

1) 데이터 사이언스의 이해

● 데이터 사이언티스트에게 필요한 자질



〈출처 : IDG, <http://www.ciokorea.com/print/26909>〉

1) 데이터 사이언스의 이해

● 데이터 사이언티스트에게 요구되는 태도

지식 습득에
대한 갈망

폭넓은
기술 활용

협업을
위한 자신의
태도

도달
가능한 능력



논리학, 심리학, 통계학, 경영학, 컴퓨터 공학 등



관련된 학문의 동향 주시하기



융합을 위한 필요한 지식을 탐구하기

1) 데이터 사이언스의 이해

● 데이터 사이언티스트에게 요구되는 태도

지식 습득에
대한 갈망

폭넓은
기술 활용

협업을
위한 자신의
태도

도달
가능한 능력



컴퓨터 활용, 프로그래밍, 분석방법론 등



빅데이터 수집, 저장, 분석, 표현 기술 습득



업무 시작 전 두 단계 깊이 있는 기술을 찾기

1) 데이터 사이언스의 이해

● 데이터 사이언티스트에게 요구되는 태도

지식 습득에
대한 갈망

폭넓은
기술 활용

협업을
위한 자신의
태도

도달
가능한 능력



지식과 기술을 배우려는 자세가 필요함



함께 일하는 사람이 서로가 다름을 인정하기

1) 데이터 사이언스의 이해

- 데이터 사이언티스트에게 요구되는 태도

지식 습득에
대한 갈망

폭넓은
기술 활용

협업을
위한 자신의
태도

도달
가능한 능력



어떤 일이든지 프로페셔널이 되기에 오래 걸림



1만 시간의 법칙이 중요함

1) 데이터 사이언스의 이해

● 데이터 사이언티스트에게 요구되는 태도

지식 습득에
대한 갈망

폭넓은
기술 활용

협업을
위한 자신의
태도

도달
가능한 능력



효과적이며 효율적인 단계별 개인능력을 지향하기



개인의 능력과 협업의 생산성 및 조화를 존중하기

1) 데이터 사이언스의 이해

● 데이터 사이언티스트에게 요구되는 태도

지식 습득에
대한 갈망

폭넓은
기술 활용

협업을
위한 자신의
태도

도달
가능한 능력



멘토를 통한 자신의 서로 다른 능력 키우기



다양한 방법을 통한 최고의 전문가 찾기

2) 데이터 사이언스 학습체계

- 빅데이터 사업의 성공 조건과 직무능력

- ▶ 빅데이터 분석 사업의 필수 성공 조건

분석모델의 명확성 및 사전정의

- 사용자가 분석결과 및 사용목적 등을 명확히 이해해야 함
- 분석대상 데이터 용량과 분석시간을 충분히 확보하여
분석대상 데이터를 고품질로 유지해야 함(저품질)

고품질 데이터
INPUT



고품질 결과
OUTPUT

2) 데이터 사이언스 학습체계

- 빅데이터 사업의 성공 조건과 직무능력

- ▶ 빅데이터 분석 사업의 핵심 성공 요인

01 분석목적, 사용자, 활용 방안 우선 정의

02 데이터 볼륨 보다 가치 창출 관점에서의 고민 전제

03 업무 전문가의 참여 필수

04 분석 목적에 따른 모델 정의

2) 데이터 사이언스 학습체계

- 빅데이터 사업의 성공 조건과 직무능력

- ▶ 빅데이터 분석 사업의 핵심 성공 요인

05

분석 대상 데이터 포맷과 서비스 방안에 따른 분석
인프라 검토

06

분석 시스템 구축 후 주기적으로 모델 변수 모니터링 및
정제

2) 데이터 사이언스 학습체계

- 빅데이터 사업의 성공 조건과 직무능력

- ▶ 빅데이터 분석 사업의 핵심 성공 요인

“ 작게, 그리고 적게 시작하라! ”

작은 모델의
성공 사례 경험 및 공유

성공 사례를 바탕으로
분석 대상/영역 점진적 확대



2) 데이터 사이언스 학습체계

- 빅데이터 국가직무능력

- ▶ 2017년 빅데이터 국가직무능력 표준 변경사항

“ 빅데이터 관련 국가직무능력표준 ”
(NCS : National Competence Standards)



초기 빅데이터 분석 시스템 구축 차원의 빅데이터 분석
모듈만 개발



비즈니스 모델의 중요성이 인식되면서 2017년 빅데이터
기획 모듈 추가

2) 데이터 사이언스 학습체계

● 빅데이터 국가직무능력

▶ 2017년 빅데이터 국가직무능력 표준 변경사항

2013년 빅데이터
국가직무능력 표준

2017년 빅데이터
국가직무능력 표준 추가사항

- 빅데이터 분석기획
- 빅데이터 수집, 빅데이터 저장
- 빅데이터 처리, 분석용 데이터 탐색
- 통계기반 데이터 분석
- 머신러닝 기반 데이터 분석
- 텍스트마이닝 기반 데이터 분석

2) 데이터 사이언스 학습체계

● 빅데이터 국가직무능력

▶ 2017년 빅데이터 국가직무능력 표준 변경사항

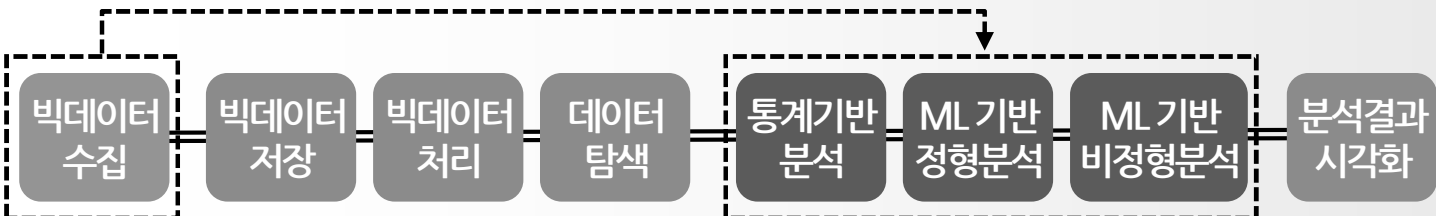
2013년 빅데이터
국가직무능력 표준

2017년 빅데이터
국가직무능력 표준 추가사항

- 빅데이터 서비스 기획
- 빅데이터 환경 분석, 빅데이터 분석 기획
- 빅데이터 기술 플랫폼 기획
- 빅데이터 성과 관리 기획
- 빅데이터 활용 기획, 빅데이터 운영 기획

2) 데이터 사이언스 학습체계

● 데이터분석업무 플로우



빅데이터 수집	1. 데이터 수집 계획 수립하기	2. 빅데이터 수집 시스템 구성하기	3. 내부 데이터 수집하기	4. 외부 데이터 수집하기
	5. 데이터 변환하기	6. 수집 데이터 검증하기		
빅데이터 저장	1. 빅데이터 저장 계획 수립하기	2. 빅데이터 저장 모델 설계하기	3. 빅데이터 저장 관리시스템 구성하기	4. 빅데이터 적재하기
	5. 빅데이터 운영하기			
빅데이터 처리	1. 빅데이터 처리 계획 수립하기	2. 빅데이터 처리 시스템 구성하기	3. 분산병렬 수행하기	4. 실시간 수행하기
데이터 탐색	1. 기본통계 확인하기	2. 데이터 분포 분석하기	3. 변수간 관계 확인하기	4. 데이터 정제하기
	5. 데이터 변화 적재하기	6. 분석용 데이터 지표도출하기		
통계기반 데이터분석	1. 가설 설정하기	2. 빅데이터 모델 개발하기	3. 빅데이터 모델 평가 검증하기	4. 빅데이터 모델 운영방안 마련하기
머신 러닝 기반 정형 데이터 분석	1. 머신러닝 수행방법 계획하기	2. 데이터셋 분할하기	3. 지도학습 모델적용하기	4. 자율학습 모델 적용하기
	5. 모델 성능 평가하기	6. 학습결과 적용하기		
머신 러닝 기반 비정형 데이터 분석	1. 텍스트마이닝 기반 텍스트 분석 실행방법 계획하기	2. 텍스트 변환하기	3. 단어사전 구축하기	4. 텍스트 분류 결과 분석하기
	5. 정형 데이터 결합 분석 수행하기	6. 사회연결망 분석하기	7. 영상 인식하기	8. 음성 인식하기
빅데이터 분석 결과 시각화	1. 분석 결과 스토리텔링하기	2. 분석 정보 시각화하기	3. 분석 정보 시각표현하기	

<출처 : 국가직무능력표준(2017)>

2) 데이터 사이언스 학습체계

● 빅데이터 수집을 위한 지식과 기술



빅데이터 저장모델 설계하기 위한 필요 지식(예시)

- 빅데이터 수집 방침 및 목표
- 수집 대상 데이터 특성
- 데이터 유형별 수집 기술 종류
- 현행 개선 요구, 리스크, 이슈, 제약 사항 식별
- 개인정보보호 및 권한 관리

빅데이터 저장모델 설계하기 위한 필요 기술(예시)

- 빅데이터 분석 대상 업무 도메인
- NoSQL (Not Only SQL) 논리적 구조
- 분산파일 시스템 논리적 구조
- 관계형 데이터베이스 논리적 구조
- 데이터 분석 및 설계 방법론
- 배치 처리 방법론

2) 데이터 사이언스 학습체계

● 빅데이터 수집을 위한 지식과 기술



빅데이터 저장모델 설계하기 위한 필요 지식(예시)

- 데이터 유형별 데이터 수집 기술
- 데이터 수집 요건 분석 및 설계 능력
- 정보 시스템 및 데이터 인터페이스 설계 기술
- 고가용성/부하 분산 방식의 데이터 전송 채널 구축 기술

빅데이터 저장모델 설계하기 위한 필요 기술(예시)

- 수집 어댑터 및 플러그인 개발 능력
- 수집 시스템 구성 요소들의 구동/설정 배포/상태 모니터링
- 운영을 위한 관리 도구 구축 기술
- 데이터 필터링, 정제 기능 구축 능력

2) 데이터 사이언스 학습체계

● 빅데이터 저장에 위한 지식과 기술



빅데이터 저장모델 설계하기 위한
필요 지식(예시)

빅데이터 저장모델 설계하기 위한
필요 기술(예시)

- 반정형 데이터 저장 NoSQL (Not Only SQL) 용량 산정 능력
- 정형 데이터 저장 관계형 데이터베이스 용량 산정 능력
- 관계형 데이터베이스 테이블 모델링 설계 및 구축 기술

2) 데이터 사이언스 학습체계

● 빅데이터 저장을 위한 지식과 기술



빅데이터 저장모델 설계하기 위한 필요 지식(예시)

- 관계형 데이터 모델링 기법
- 데이터웨어하우스 설계 기법
- 텍스트 저장모델 설계 기법
- 데이터 모델 표현 기법
- NoSQL(Not Only SQL) 설계 기법

빅데이터 저장모델 설계하기 위한 필요 기술(예시)

- EDA(Exploratory Data Analysis) 기법
- 머신러닝 기법
- 텍스트 마이닝 기법

〈출처 : 국가직무능력표준(2017)〉

2) 데이터 사이언스 학습체계

● 빅데이터 저장에 위한 지식과 기술



빅데이터 저장모델 설계하기 위한 필요 지식(예시)

- 데이터 요구사항 분석능력
- 커뮤니케이션을 통한 요구사항 도출 능력
- 비즈니스 도메인 분석기술

빅데이터 저장모델 설계하기 위한 필요 기술(예시)

- 데이터 요구사항과 기능 요구사항 연계 능력
- 데이터 저장모델 설계도구 활용 능력

〈출처 : 국가직무능력표준(2017)〉

2) 데이터 사이언스 학습체계

● 빅데이터 처리를 위한 지식과 기술



빅데이터 처리시스템 구성하기를
위한 필요 지식(예시)

- 빅데이터 처리 플랫폼
- 빅데이터 처리 인프라스트럭처
- 클라우드 IaaS 서비스와 클라우드 솔루션

빅데이터 처리시스템 구성하기를
위한 필요 기술(예시)

- 클러스터 데이터베이스
- 빅데이터 처리 플랫폼 검증 방법
- 빅데이터 처리 개선 사항 도출

2) 데이터 사이언스 학습체계

● 빅데이터 처리를 위한 지식과 기술



빅데이터 처리시스템 구성하기를
위한 필요 지식(예시)

- 빅데이터 처리 플랫폼 분석 기술
- 빅데이터 처리 플랫폼 설치 및 설정 능력
- 클라우드 인프라 구축 및 설정 능력

빅데이터 처리시스템 구성하기를
위한 필요 기술(예시)

- 분산 병렬 컴퓨팅 솔루션 설치 및 설정 능력
- 오픈소스 소프트웨어 사용 능력

2) 데이터 사이언스 학습체계

● 빅데이터 통계기반분석을 위한 지식과 기술



빅데이터 모델 운영방안 마련하기를
위한 필요 지식(예시)

- 빅데이터 분석 방법론
- 정보시스템 구축 방법론

빅데이터 모델 운영방안 마련하기를
위한 필요 기술(예시)

- 정보시스템 생명주기
- 분석 모델 생명주기

2) 데이터 사이언스 학습체계

● 빅데이터 통계기반분석을 위한 지식과 기술



빅데이터 모델 운영방안 마련하기를
위한 필요 지식(예시)

- 모델 알고리즘 설명서 작성 기술
- 시스템 설계 작성 기술
- UML(Unified Modeling Language) 기술

빅데이터 모델 운영방안 마련하기를
위한 필요 기술(예시)

- SQL(Structured Query Language) 기술
- 정보시스템 개발언어 활용 기술

2) 데이터 사이언스 학습체계

● 분석 알고리즘

알고리즘	주 활용분야
업리프트모델링	단계적 추정, 예측분석
생존분석	의료, 통계, 설비분야 사건예측 (이탈, 사망, 고장 등)
회귀분석	예측분석, 추정분석(매출, 신용점수 등)
시각화	원인과 관계분석
기초통계	기초통계 현황파악
부스팅	분류분석(일종의 앙상블기법)
시계열/순열 분석	시간상의 예측(이자율, 예산 등 예측)
요인분석	요인분석, 차원축소
이상치 감지기법	이상치분석
텍스트 마이닝	시대경향 및 감성분석

〈출처 : 재가공, 데이터과학 무엇을 하는가, 김옥기(2015)〉

2) 데이터 사이언스 학습체계

● 분석 알고리즘

알고리즘	주 활용분야
의사결정나무	분류(종속 변수 존재)
신경회로망	분류분석
군집분석	독립변수들만의 분류, 그룹화, 프로파일링
양상불기법	분류, 예측, 최적화 등의 결합분석
소셜네트워크 분석	관계망 분석 (사람관계, 통화관계, 소셜네트워크 등)
연관법칙	관계분석, 장바구니분석 등
서포트벡터머신	분류분석
주성분분석	
추천	

〈출처 : 재가공, 데이터과학 무엇을 하는가, 김옥기(2015)〉

2) 데이터 사이언스 학습체계

● 빅데이터 ML기반 정형분석을 위한 지식과 기술



지도학습 모델 적용하기를
위한 필요 지식(예시)

- 현업 업무 및 수행 프로세스
- 통계학 기반의 분석방법론
- 선형대수 및 미적분, 최적화 등의 수학기론
- 분류 모델을 위한 지도학습 머신러닝 기법

지도학습 모델 적용하기를
위한 필요 기술(예시)

- 예측 모델을 위한 지도학습 머신러닝 기법
- 최신 머신러닝 기법 발전 트렌드와 연구 동향

2) 데이터 사이언스 학습체계

● 빅데이터 ML기반 정형분석을 위한 지식과 기술



지도학습 모델 적용하기를
위한 필요 지식(예시)

- 문제 핵심 파악 및 이슈 도출 능력
- 문제 및 이슈를 실제로 검증 가능한 일련의 명확한 가설로 만들어 낼 수 있는 능력
- 머신러닝 알고리즘을 실제로 구현해낼 수 있는 프로그래밍 기술

지도학습 모델 적용하기를
위한 필요 기술(예시)

- 통계 및 데이터 분석에 특화된 언어를 다룰 수 있는 기술
- 기존의 알고리즘을 융합하거나 새로운 알고리즘을 설계하고 구현할 수 있는 능력
- 자동화 처리 및 시스템 구현 위한 개발 능력

2) 데이터 사이언스 학습체계

● 빅데이터 분석결과 시각화를 위한 지식과 기술



분석정보 시각화를 위한 필요 지식(예시)

- 현업 시각화 기술 및 알고리즘
- 데이터 분석 대상 업무
- 시각화 정보 구조 및 구성 요소
- 시각화 구축의 목적과 사용 의도
- 시각화 표현 기법
- UX(User Experience) 설계 방법론

분석정보 시각화를 위한 필요 기술(예시)

- 효과적인 컬러 및 폰트 사용법
- 시각화 디자인 원칙



02

빅데이터 서비스 기획과 도입 사례

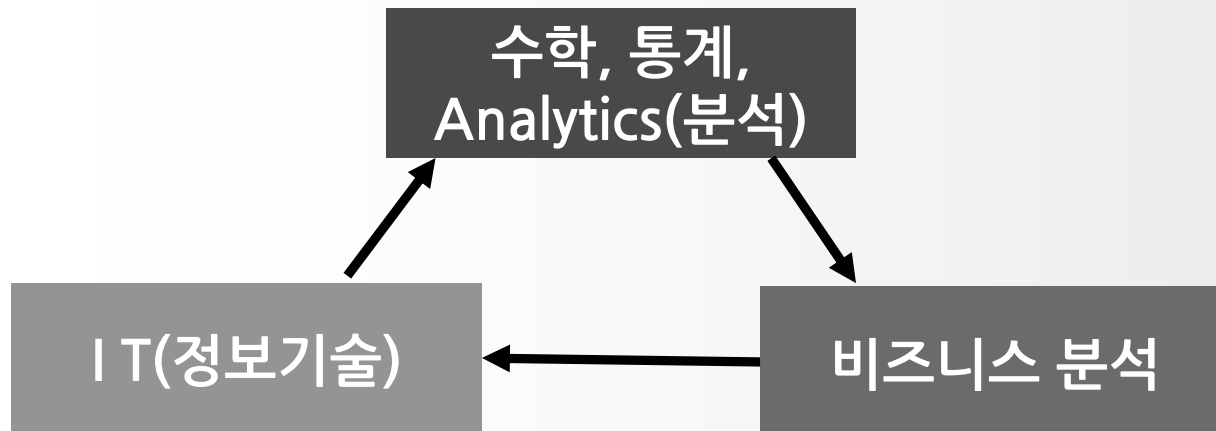
- 1) 빅데이터 분석 기획
- 2) 빅데이터 분석시스템 도입사례

1) 빅데이터 분석 기획

- 분석기획의 특징

빅데이터 분석기획

해당 문제영역에 대한 전문성 역량 및 수학/통계학적 지식을 활용한 분석역량과 분석 도구인 데이터 및 프로그래밍 기술 역량에 대한 균형 잡힌 시각을 가지고 방향성 및 계획을 수립해야 한다는 것



1) 빅데이터 분석 기획

● 분석기획의 특징

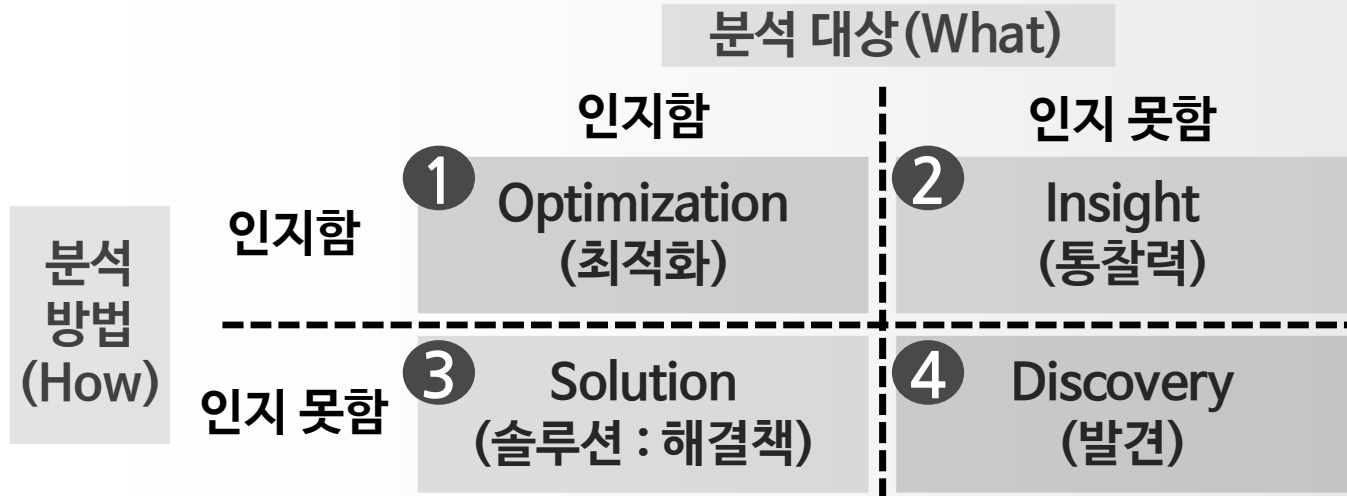
빅데이터 분석기획

- 실제 분석을 수행하기에 앞서 수행할 과제를 정의하고
의도했던 결과를 도출할 수 있도록 이를 적절하게 관리할 수
있는 방안을 사전에 계획하는 일련의 작업
- 분석과제 및 프로젝트를 직접 수행하는 것은 아니지만,
어떤 목표(What)를 달성하기 위하여(Why) 어떠한 데이터를
가지고 어떤 방식으로(How) 수행할 지에 대한 일련의
계획을 수립하는 작업이기 때문에 성공적인 분석결과를
도출하기 위한 중요한 사전 작업



1) 빅데이터 분석 기획

● 분석대상과 방법(분석주제 유형)



01

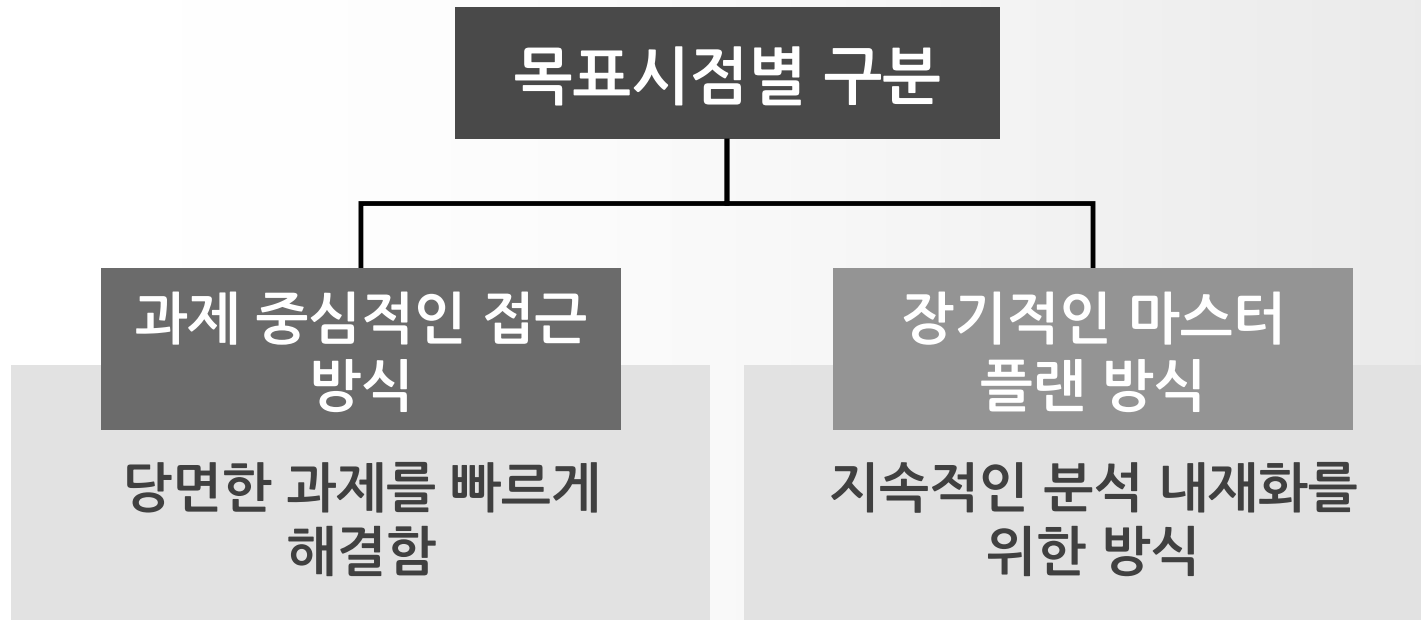
분석대상을 알고 있고, 분석방법도 알고 있는 경우,
개선을 통한 최적화 형태로 분석함

02

분석대상을 알고 있고, 분석방법을 알지 못하는 경우,
해당 분석 주제에 대한 솔루션을 찾아내는 방식으로 수행함

1) 빅데이터 분석 기획

- 분석시점별 분석 기획 방안



1) 빅데이터 분석 기획

- 분석시점별 분석 기획 방안

당면한 분석주제의
해결

과제 단위

Speed & Test

Quick & Win

Problem Solving

1차 목표

과제의 유형

접근 방식

지속적 분석 문화
내재화

마스터 플랜 단위

Accuracy & Deploy

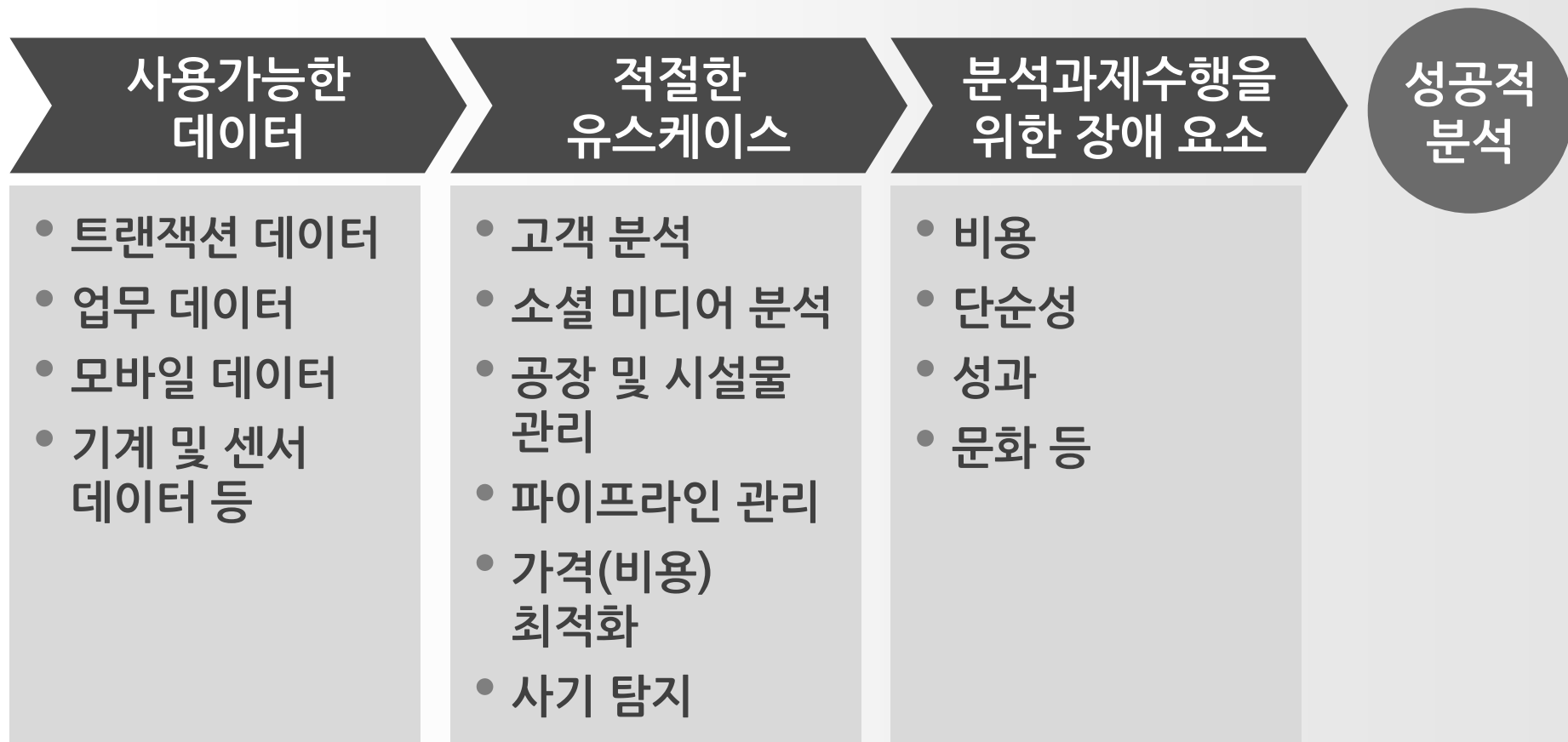
Long Term View

Problem Definition

〈출처 : 데이터분석전문가가이드, K-data(2014)〉

1) 빅데이터 분석 기획

● 분석 기획 시 고려사항



〈출처 : 데이터분석전문가가이드, K-data(2014)〉

1) 빅데이터 분석 기획

● 분석 기획 시 고려사항

사용가능한
데이터

적절한
유스케이스

분석과제 수행을
위한 장애 요소

- 분석을 위한 데이터 확보가 우선적임
- 데이터의 유형에 따라 적용 가능한 솔루션 및 분석 방법이 다르기 때문에 유형에 대한 분석이 선행적으로 이루어져야 함

〈출처 : 데이터분석전문가가이드, K-data(2014)〉

1) 빅데이터 분석 기획

● 분석 기획 시 고려사항

사용가능한
데이터

적절한
유스케이스

분석과제 수행을
위한 장애 요소

- 분석을 통해 가치가 창출될 수 있는 적절한 활용방안과 유스케이스 탐색이 필요함
- “바퀴를 재발명하지 마라”라는 격언처럼 기존에 잘 구현되어 활용되고 있는 유사 분석 시나리오 및 솔루션을 최대한 활용하는 것이 중요함

〈출처 : 데이터분석전문가가이드, K-data(2014)〉

1) 빅데이터 분석 기획

● 분석 기획 시 고려사항

사용가능한
데이터

적절한
유스케이스

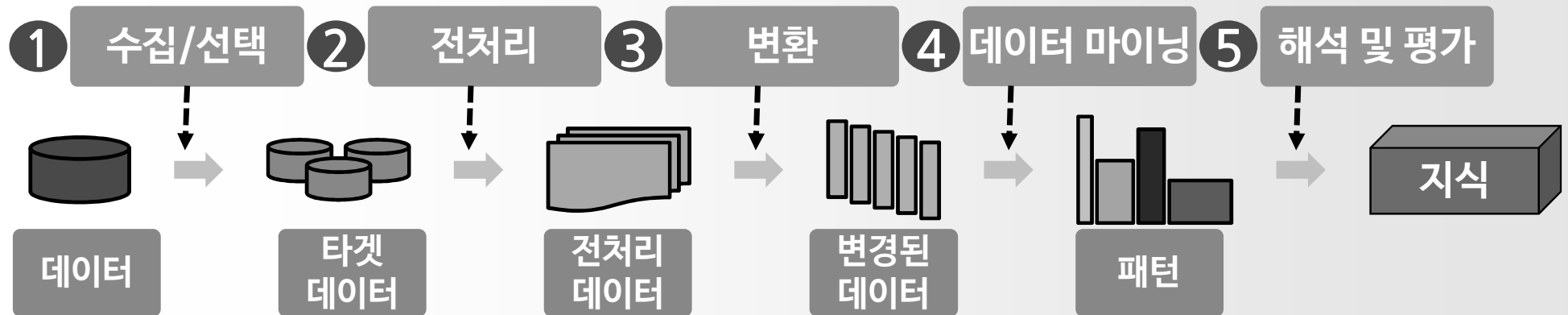
분석과제 수행을
위한 장애 요소

- 분석 수행 시 발생하는 장애 요소들에 대한 사전계획 수립이 필요함
- 일회성 분석으로 그치지 않고 조직의 역량으로 내재화하기 위해서는 충분하고 지속적인 교육 및 활용방안 등의 변화 관리(Change Management)가 고려되어야 함

〈출처 : 데이터분석전문가가이드, K-data(2014)〉

1) 빅데이터 분석 기획

● KDD 분석방법론



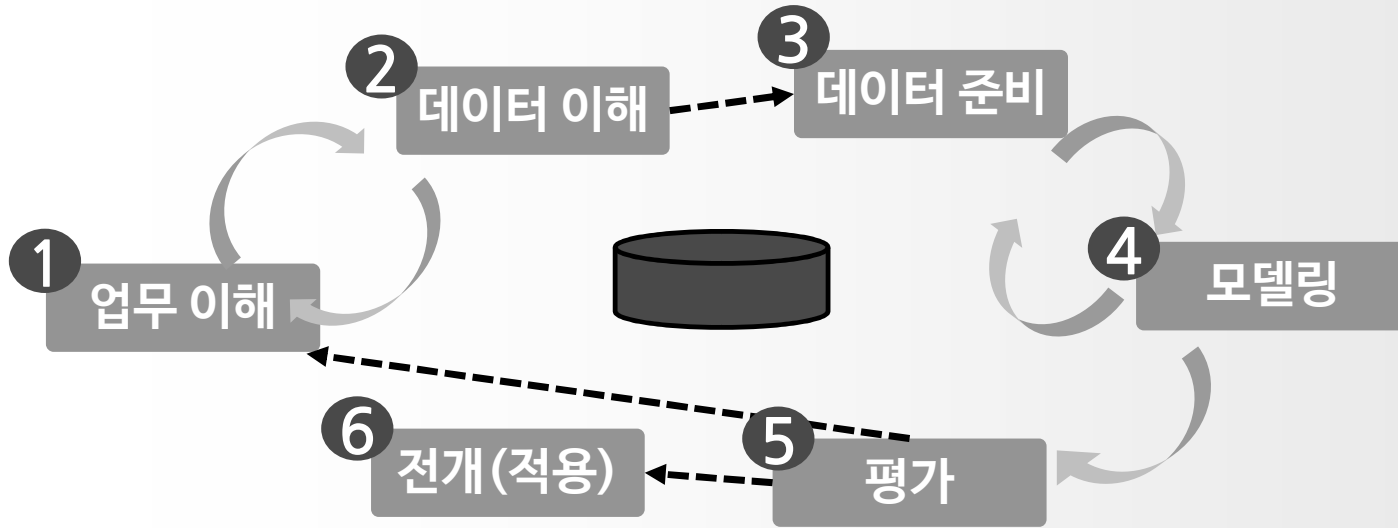
- 1**
 - 비즈니스 도메인에 대한 이해
 - 프로젝트 목표 설정
- 2**
 - 잡음, 이상치, 결측치 식별
 - 필요한 데이터 셋 준비
- 3**
 - 데이터마이닝을 위한 데이터 차원의 축소
 - 학습용 데이터와 검증용 데이터의 분리

- 4**
 - 데이터마이닝 기법 선택
 - 전처리와 필요한 변환 프로세스 추가
- 5**
 - 데이터마이닝 결과에 대한 해석과 평가
 - 분석목적과의 일치성 확인
 - 필요할 경우, 데이터마이닝 프로세스 반복 수행

〈출처 : 데이터분석전문가가이드, K-data(2014)〉

1) 빅데이터 분석 기획

● CRISP - DM 분석방법론



〈출처 : 데이터분석전문가가이드, K-data(2014)〉

1) 빅데이터 분석 기획

● CRISP - DM 분석방법론

업무 이해
(Business
Understanding)

- 프로젝트의 목적과 요구사항 이해
- 도메인 지식을 데이터분석을 위한 문제정의로 변경

데이터 이해
(Data
Understanding)

- 데이터 수집 및 속성의 이해
- 인사이트 발견

데이터 준비
(Data
Preparation)

- 데이터 편성
- 분석용 데이터 셋 선택

〈출처 : 데이터분석전문가가이드, K-data(2014)〉

1) 빅데이터 분석 기획

● CRISP - DM 분석방법론

모델링 (Modeling)

- 모델링 기법과 알고리즘 선택
- 파라미터 최적화
- 모델의 과적합 문제 확인

평가 (Evaluation)

- 모델링 결과가 프로젝트 목적에 부합하는지 평가

전개 (Deployment)

- 실무 적용 계획 수립 단계
- 모니터링과 모델 유지보수 계획 마련

〈출처 : 데이터분석전문가가이드, K-data(2014)〉

1) 빅데이터 분석 기획

● 빅데이터 분석 방법론 : 5단계

1. 기획 단계

- 비즈니스 이해 및 범위 설정
- 빅데이터 분석 과제
- 발굴 기획
- 내부 데이터 조사
- 프로젝트 위험계획 수립

2. 사전준비단계

- 프로젝트 착수 준비
- 데이터 정의 및 설계
- 현황 및 트렌드 분석
- 빅데이터 과제 도출
- 빅데이터 과제 선정
- 빅데이터 활용 계획 수립

3. 탐색 단계

- 내부 데이터 이관
- 분석용 마트 구현
- 탐색적 분석
- 분석모델 검토
- 모델링 및 성능평가

1) 빅데이터 분석 기획

● 빅데이터 분석 방법론 : 5단계

4. 분석 및 검증 단계

- 통계기반 데이터 분석
- 빅데이터 사례 기반
- 분석 머신 러닝 기반 분석
- 비즈니스 영향도 평가
- 설계 및 구현
- 시스템 테스트 및 구현

5. 이행(전개) 및 시각화 단계

- 분석 과제 타당성 검증
- 시각화 검토 및 구현
- 빅데이터 활용방안 도출
- 분석모델 고도화
- 프로젝트 평가 및 보고

〈출처 : 데이터분석전문가가이드, K-data(2014), 재가공〉

1) 빅데이터 분석 기획

- 빅데이터 분석 방법론 : 5단계

- ▶ 현업 부서(업무담당자) 역할

01 업무수요 제공

02 업무현안 및 문제점 제공

03 분석과제 발굴부터 현업부서 참여 필수

04 도출된 과제 후보들을 평가하여 빅데이터의 과제를 최종 결정

〈출처 : 데이터분석전문가가이드, K-data(2014)〉

1) 빅데이터 분석 기획

- 빅데이터 분석 방법론 : 5단계

- ▶ 정보화 부서(IT 담당자) 역할

01 아이디어 구현방안 제공

02 데이터 유무 확인

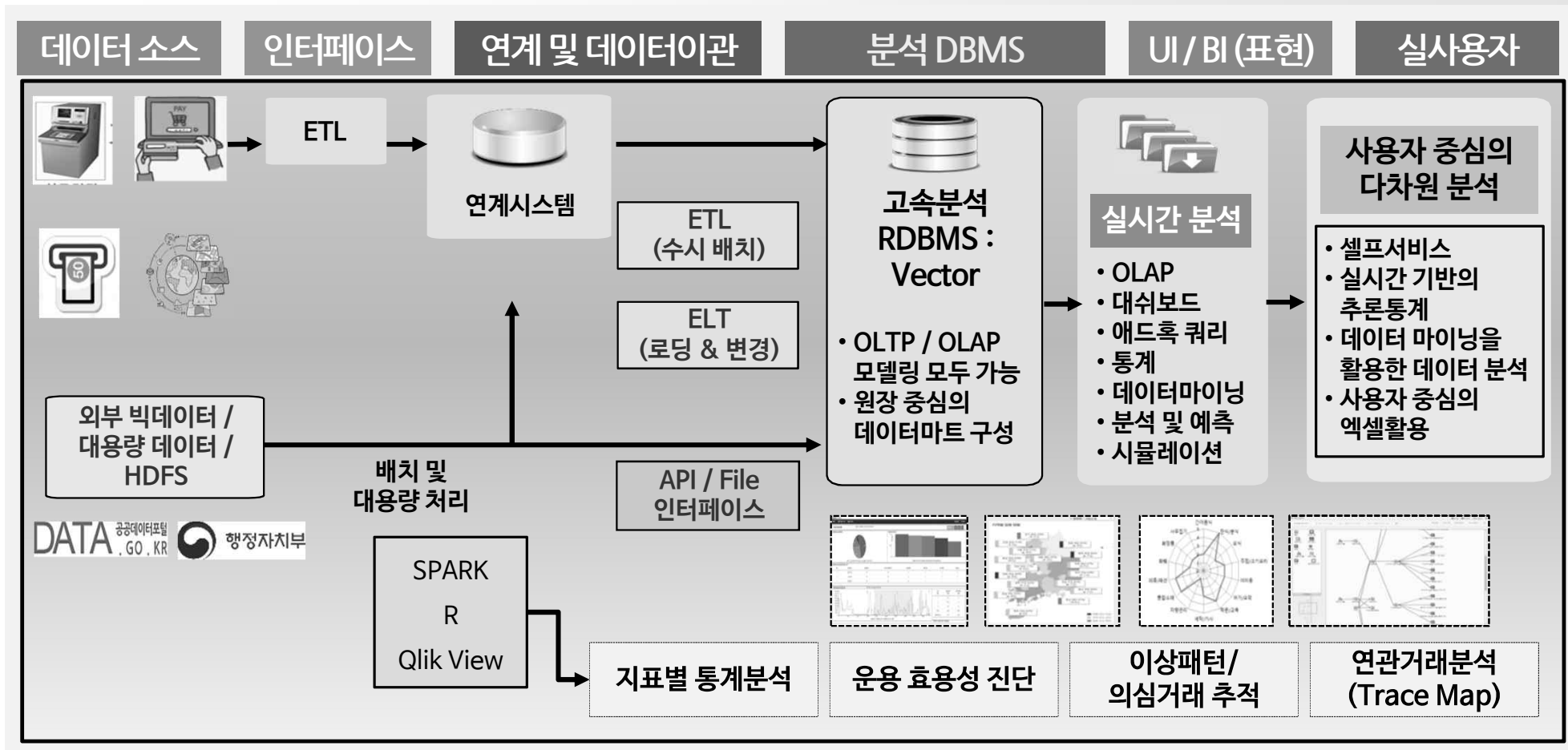
03 데이터 정보 제공

04 후보 과제를 평가하는 보조적 역할

〈출처 : 데이터분석전문가가이드, K-data(2014), 재가공〉

2) 빅데이터 분석시스템 도입사례

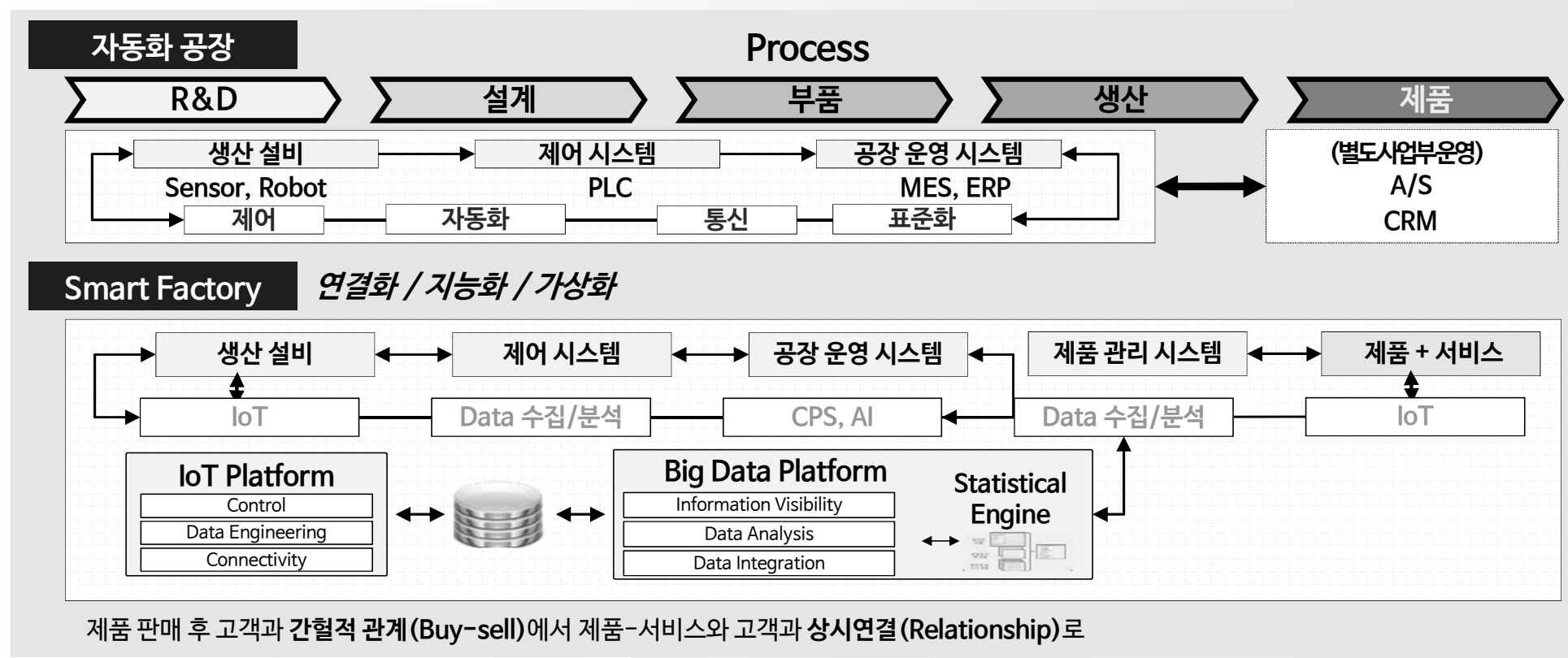
● 공공기관 사기탐지 시스템(예시)



2) 빅데이터 분석시스템 도입사례

- 빅데이터 활용 제조공정 개선시스템(예시)

- ▶ 자동화 공장과 스마트팩토리의 차이



〈출처 : 창조경제연구회(KCERN)〉

학습 평가

Q1

Q2

Q1

데이터 사이언티스트에게 요구되는
역량 중 소프트 스킬에는 어떤 항목들이
포함되어 있는가?

학습 평가

Q1

Q2

Q1

데이터 사이언티스트에게 요구되는
역량 중 소프트 스킬에는 어떤 항목들이
포함되어 있는가?

정답

통찰력 있는 분석, 설득력 있는 전달, 다분야간 협력

해설

소프트 스킬에는 통찰력 있는 분석(창의적 사고, 호기심,
논리적 비판), 설득력 있는 전달(스토리텔링,
비주얼라이제이션), 다분야간 협력(커뮤니케이션)이
해당됩니다.

학습 평가

Q1

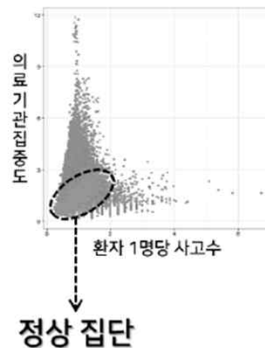
Q2

Q2

데이터 사이언티스트가 사용하는
최고의 분석도구인 분석 알고리즘의 종류에는
어떤 것들이 있는가?

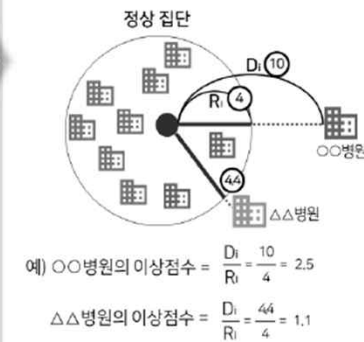
군집분석

- 의료기관 및 사고환자 데이터로 K 평균 군집화 알고리즘 적용하여 각각 정상 집단 정의
- 데이터의 유사성에 따라 주어진 데이터를 K개의 군집으로 묶음



이상점수 산출

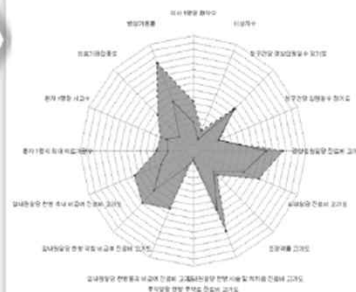
- 정상 집단의 중심점(평균)과 반경 R 산출
- 정상 집단의 중심점으로부터의 거리 D 산출
- R과 D로 이상점수(Anomaly Score, AS) 산출
- 이상점수가 높을수록 이상 패턴일 확률 높음



➡ ○○병원이 △△병원보다 이상 패턴일 확률 높음

이상패턴 탐지

- 모델을 생성할 때 이상점수 임계치 설정
- 정상 집단으로부터 멀리 떨어져있어 임계치 넘었을 경우 이상집단으로 판별
- 이상점수 임계치에 따라 이상집단 여부 결정



〈출처 : 2017년 빅데이터 성과공유대회, 행정안전부, 한국정보화진흥원〉

학습 평가

Q1

Q2

Q2

데이터 사이언티스트가 사용하는
최고의 분석도구인 분석 알고리즘의 종류에는
어떤 것들이 있는가?

정답

- 기초통계, 시계열 분석, 군집분석, 주성분분석
- 생존분석, 회기분석, 연관법칙, 요인분석
- 앙상블기법, 서포트벡터머신, 이상치 감지기법

정리 하기

빅데이터 분석업무의 성공적인 접근방법

- ✓ 빅데이터의 분석목적, 사용자 및 활용방안에 대하여 우선 정의를 내림
- ✓ 데이터 볼륨보다는 데이터를 통한 가치창출에 대해 깊은 관심을 가지고 고민을 더하여 수많은 아이디어를 도출해 냄
- ✓ 업무전문가가 필수적으로 참여하며, 분석 목적에 따른 모델을 정의한 후 검증하며, 분석대상에 따른 데이터를 정규화여, 데이터 서비스방안에 따라서 모델변수를 주기적으로 작성하고 이를 모니터링함

정리 하기

데이터 분석업무 절차

- ✓ 데이터 분석업무 플로우 절차
 - 빅데이터 수집과 정제
 - 빅데이터 저장
 - 빅데이터 처리
 - 빅데이터 탐색과 분포 확인
 - 통계기반 분석
 - 머신러닝기반 정형분석
 - 머신러닝기반 비정형분석
 - 분석결과 시각화 및 스토리텔링



- 다음 시간에 살펴 볼 내용 -

13강 빅데이터 마케팅

수고하셨습니다.