

# I.빅데이터 분석 기획

## 02. 데이터 분석 계획

### 2.1 분석 방안 수립

-	KeyWord
분석 문제	분석문제, 하향식 접근방식, 상향식 접근방식, 디자인사고
데이터 분석 방안	빅데이터 분석 방법론, KDD, CRISP-DM, SEMMA

=====

#### 1) 분석 로드맵 설정

##### (1) 분석 로드맵 개념

- 단계별로 목표 명확히/ 선·후행 단계 고려/ 단계별 추진내용 정렬

##### (2) 분석 로드맵 단계

- 데이터 분석체계 도입: 분석기회 발굴 / 과제 정의 / 로드맵 수립
- 데이터 분석 유효성 검증: 알고리즘 설계 / 아키텍처 설계 / 파일럿 수행
- 데이터 분석 확산 및 고도화: 변화 관리 / 시스템 구축 / 유관 시스템 고도화

=====

#### 2) 분석 문제 정의

##### (1) 분석 문제의 의미

- 과제: 처리해야할 문제, 이슈
- 분석: 과제와 관련된 현상, 원인, 해결방안에 대한 자료를 수집 및 분석하여 의사결정에 활용하는 활동
- 분석 문제: 기대 상태(목표)와 현재 상태(현상)를 동일한 수준으로 맞추는 과정
  - 제약조건 파악/ 잠재원인 진단/ 관련 데이터 수집, 가공, 분석하는 과정
  - 목표(Should be)와 현상(As-Is)의 차이(Gap)/ 해결해야 하는 사항
- 분석 문제 발굴: 하향식, 상향식 접근방식을 반복 수행하면서 상호보완하여 발굴

##### (2) 하향식 접근방식(Top Down Approach)

- 분석 과제 정해져 있음 -> 해결방안 탐색
- 하향식 분석 과제 발굴 절차
  - 문제 탐색 -> 문제 정의 -> 해결방안 탐색 -> 타당성 검토 -> 선택
  - 문제 탐색: 비즈니스 모델기반 문제 탐색 / 분석기회 발굴범위 확장 / 외부 참조 모델기반 문제탐색 / 유스케이스 정의
  - 문제 정의: 사용자 관점에서 정의 / 필요 데이터 및 기법 정의

- 해결방안 탐색: 분석기법(How) 및 역량(Who)에 따라 다양하게 탐색 -> 기존 시스템 / 교육 및 채용 / 시스템 고도화 / 아웃소싱
- 타당성 검토: 경제적 / 데이터 및 기술적 / 운영적 타당성 평가
- 선택: 최적 대안을 선택하여 프로젝트화

### (3) 상향식 접근방식(Bottom Up Approach)

- 문제 정의 자체가 어려움 -> 데이터 기반으로 지속적으로 문제 개선
- 기존 하향식의 한계 극복을 위한 방법
- 디자인 사고 접근법(Design Thinking)
  - 인간에 대한 관찰과 공감을 바탕으로 다양한 대안을 찾는 확산적 사고
  - 주어진 상황에 대한 최선의 방법을 찾는 수렴적 사고의 반복을 통해 혁신적 결과를 도출하는 창의적 문제 해결 방법
- 특징: 비지도 학습 방법/ 프로토타이핑 접근법
  - 비지도 학습: 데이터 자체의 결합, 연관성, 유사성 등을 중심으로 데이터 상태 분석
  - 프로토타이핑: 시행착오를 통한 문제해결

### (4) 대상별 분석 기획 유형

- 분석 대상, 방법에 따라 4가지로 나눔 -> 최적화/ 솔루션/ 통찰/ 발견

분석 대상(What)	분석 방법(How)	분석 기획 유형
알고 있음	알고 있음	최적화
알고 있음	모름	솔루션
모름	알고 있음	통찰
모름	모름	발견

### (5) 과제 추진 시 고려해야 하는 우선순위 평가 기준

- 시급성 & 난이도
- 시급성: 전략적 중요도 / 목표 가치(KPI) -> Value
- 난이도: 데이터 관련 비용 / 분석 적용 비용 / 분석 수준 -> Volume / Variety / Velocity
- 가장 우선적위 높은 과제는 시급성↑, 난이도↓
- 가장 우선순위 낮은 과제는 시급성↓, 난이도↑

=====

## 3) 데이터 분석 방안

### (1) 빅데이터 분석 방법론 개념

- 빅데이터 분석 방법론
  - 문제를 정의하고 답을 도출하기 위한 체계적인 절차와 방법
  - 구성요소: 절차 / 방법 / 도구 / 기법 / 템플릿 / 산출물

### (2) 빅데이터 분석 방법론 계층

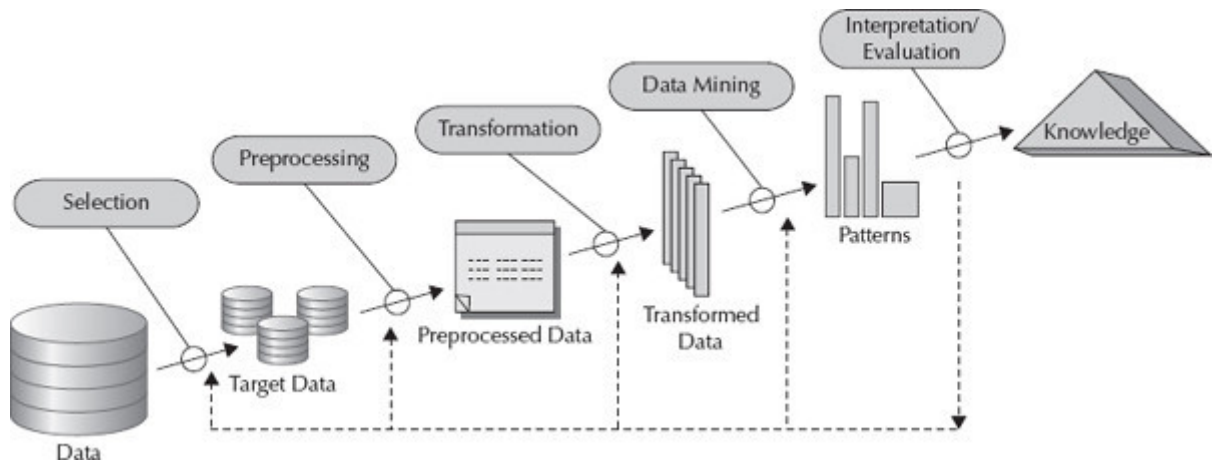
- 단계 / 태스크 / 스텝
- 단계(Phase)
  - 프로세스 그룹을 통해 완성된 단계별 산출물 생성
  - 기준선으로 설정 관리 / 버전 관리 등으로 통제
- 태스크(Task)
  - 단계를 구성하는 단위 활동
  - 품질 검토의 항목이 될 수 있음
- 스텝(Step)
  - 단위 프로세스: 입력자료, 처리 및 도구, 출력자료로 구성

### (3) 빅데이터 분석 방법론의 분석 절차

- 분석 기획 -> 데이터 준비 -> 데이터 분석 -> 시스템 구현 -> 평가 및 전개
- 분석 기획
  - 비즈니스 이해 및 범위 설정 / 프로젝트 정의 및 계획 수립 / 프로젝트 위험 계획 수립
  - 위험대응 방법: 회피 / 전가 / 완화 / 수용
- 데이터 준비
  - 필요 데이터 정의 / 데이터 스토어 설계 / 데이터 수집 및 정합성 검증
  - 데이터 스토어: DB에 들어가는 데이터 이외에 단순 파일, 이메일 등의 단순한 스토어 타입들을 포함하는 저장소
- 데이터 분석
  - 데이터 준비 / 텍스트 분석 / 탐색적 분석(EDA) / 모델링 / 모델 평가 및 검증 / 모델 적용 및 운영 방안 수립
- 시스템 구현
  - 설계 및 구현 / 시스템 테스트 및 운영
  - 시스템 및 데이터 아키텍처 설계, 사용자 인터페이스 설계 진행
- 평가 및 전개
  - 모델 발전 계획 수립 / 프로젝트 평가 보고
  - 모델 생명주기 설정, 주기적인 평가 및 유지보수

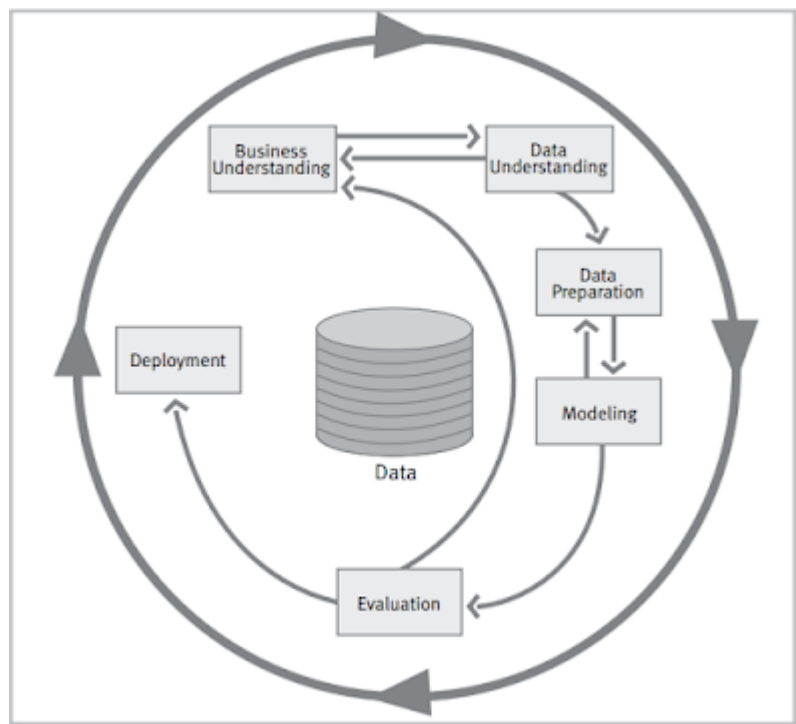
### (4) 분석 방법론 유형

- KDD 분석 방법론(Knowledge Discovery in Databases)
  - 1996, Fayyad, 프로파일링 기술 기반
  - 통계적 패턴, 지식을 찾기 위해 정리한 방법론
  - 데이터로부터 목표 데이터 -> 전처리 데이터 -> 변환 데이터 -> 패턴 -> 지식을 생성
  - 분석 절차: 데이터셋 선택 -> 전처리 -> 데이터 변환(변수 찾고 차원 축소) -> 데이터 마이닝 -> 결과 평가



• CRISP-DM 분석 방법론 (Cross Industry Standard Process for Data Mining)

- 1996, 유럽연합 ESPRIT 프로젝트 / 1997, SPSS 참여, 현재 중단
- 비즈니스 이해를 바탕으로 데이터 분석
- 방법론 구성
  - 단계(최상위 레벨) / 일반화 태스크 / 세분화 태스크 / 프로세스 실행
  - 일반화 태스크: 단일 프로세스를 완전하게 수행하는 단위
  - 세분화 태스크: 일반화 태스크를 구체적으로 수행하는 레벨
- 분석 절차: 업무 이해 -> 데이터 이해 -> 데이터 준비 -> 모델링 -> 평가 -> 전개
  - 데이터 이해: 데이터 수집 / 속성 이해 / 문제점 식별 / 숨겨진 인사이트 발견 / 품질 확인
  - 데이터 준비: 데이터 정제 및 생성 -> 분석 가능한 상태로 만들
  - 단계 간 피드백을 통해 단계별 완성도 높임



• SEMMA 분석 방법론 (Sampling Exploration Modification Modeling Assessment)

- SAS사가 주도한 통계 중심의 5단계
- 샘플링 → 탐색 → 수정 → 모델링 → 검증
- 샘플링: 통계적, 조건 추출 통해 분석 데이터 생성

- 탐색
  - 분석 데이터 탐색 / 오류 검색 / 현황
  - 기초 통계/ 그래프 탐색/ 요인별 분할표/ 클러스터링/ 상관분석 등
- 수정
  - 데이터 수정 및 변환 / 정보의 표현 극대화 / 변수 생성, 선택, 변형
  - 수량화/ 표준화/ 변환/ 그룹화 등
- 모델링
  - 전통적 통계를 이용한 모델 구축 / 숨겨진 패턴 발견
  - 신경망/ 의사결정나무/ 로지스틱 회귀분석 등
- 검증
  - 모델 평가 검증 / 서로 다른 모델 동시 비교

## 2.2 분석 작업 계획

**Keyword** 데이터 확보 계획, 빅데이터 분석 절차, WBS설정

### 1) 데이터 확보 계획

#### (1) 데이터 획득 방안 수립

- 내외부 다양한 시스템으로부터 데이터를 수집하기 위한 구체적인 방안을 수립

#### (2) 데이터 확보 계획 수립 절차

- 목표 정의 → 요구사항 도출 → 예산안 수립 → 계획 수립
- 목표 정의: 성과 목표, 지표 설정
- 요구사항 도출: 데이터 및 기술 지원 등과 관련된 요구사항 (확보 및 관리 계획, 정제 수준, 저장 형태 등)
- 예산안 수립: 데이터 확보, 구축, 정비, 관리 예산
- 계획 수립: 인력 투입 방안/ 일정 관리/ 위험 및 품질 관리

### 2) 분석 절차 및 작업 계획

#### (1) 빅데이터 분석 절차 | 문제 인식 → 연구조사 → 모형화 → 자료 수집 → 자료 분석 → 분석결과 공유

문제 인식: 비즈니스 문제와 기회 인식/ 가설 형태로 문제 정의 연구조사: 각종 문헌 조사/ 조사 내용을 해결방안에 적용 모형화: 복잡한 문제를 분리하고 단순화하는 과정/ 문제를 변수 간의 관계로 정의 자료 수집: 데이터 수집 및 변수 측정 자료 분석: 자료에서 의미 찾기/ 변수들 간 관계 분석 분석결과 공유: 변수 간 관련성을 포함한 분석결과 제시/ 가시화

#### (2) 빅데이터 분석 작업 WBS 설정 (WBS; Work Breakdown Structure)

- 데이터 분석 과제 정의: 사전 준비 단계
- 데이터 준비 및 탐색: 세부 일정 만드는 단계
- 데이터 분석 모델링 및 검증
- 산출물 정리

- 프로세스 혁신(Process Innovation)
  - 일하는 방식의 혁신
  - 프로세스를 재정립하는 것 이외에 경영관리체계를 확립하는 활동