



프로젝트 방법론

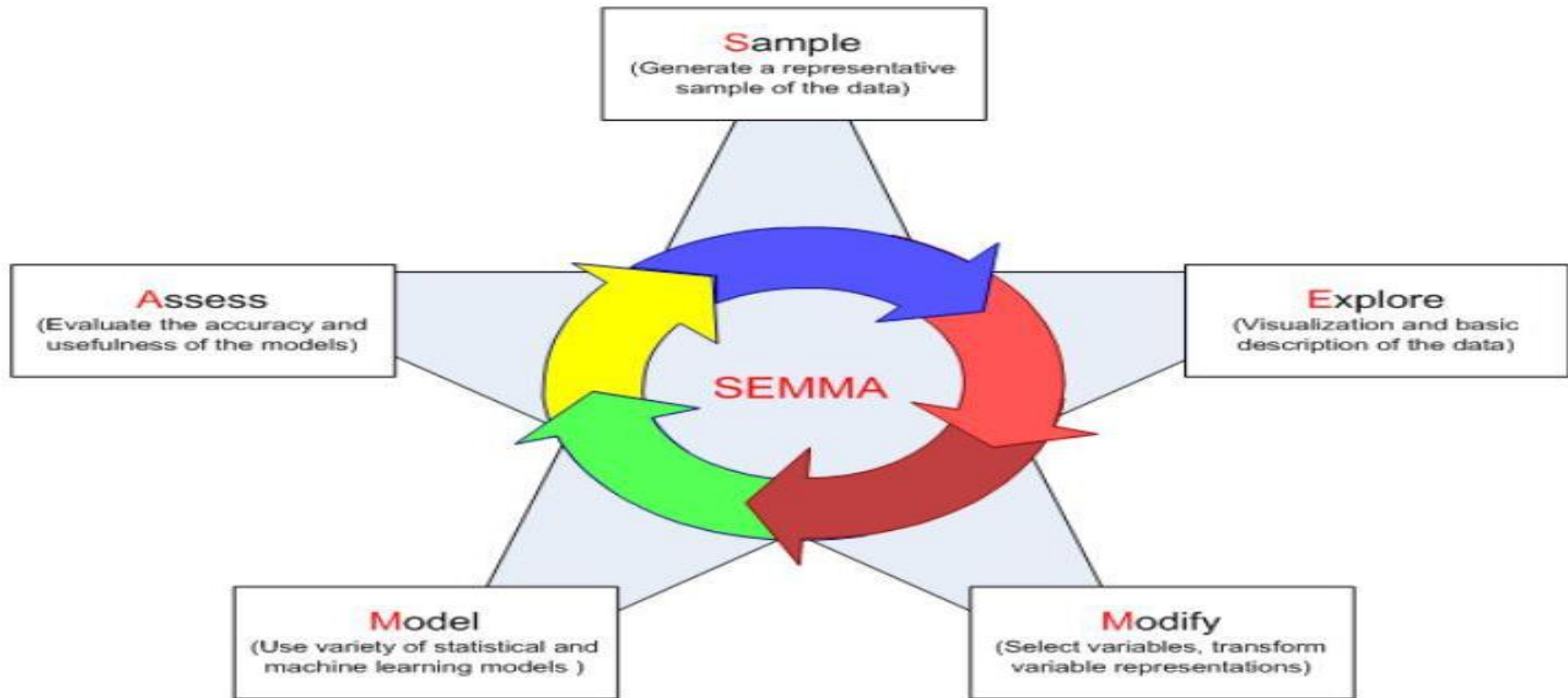


◆ 프로젝트 방법론

➤ SEMMA

➤ KMM

➤ CRISP-DM



SEMMA (Sampling Explore Modify Model)	개념	- SAS사가 주도하여 만들어진 데이터마이닝방법론 - 기술중심, 통계중심의 방법론
	특징	자사 기술로 데이터 마이닝기능을 구성하여 쉽게 적용하게 함
	주요단계	Sample – Explore – Modify – Model – Assess
	제품	IBM 클레멘타인

마이닝 단계	설명	세부요소/산출물
Sampling	분석데이터 생성 모델 평가 위한 데이터 준비	통계적 추출 조건추출
Modify	분석데이터 변환 데이터 정보 표현 극대화 (가시화) 변수 생성, 선택, 변형	수량화 표준화 변환 그룹화
Modeling	모델구축 패턴발견 모델링과 알고리즘의 적용	Neural Network Decision Tree Logistic Regression 통계기법
Assessment	모델평가 및 검증 서로 다른 모델 동시 비교 Next Step 결정	Report FeedBack 모델검증 자료

Knowledge Discovery in Database)

개념

1996년 Fayyad가 체계적으로 정리한 데이터
마이닝 프로세스 기술과 데이터베이스를 중심으로
인사이트 발굴을 위한 5개 프로세스를 적용한
방법론

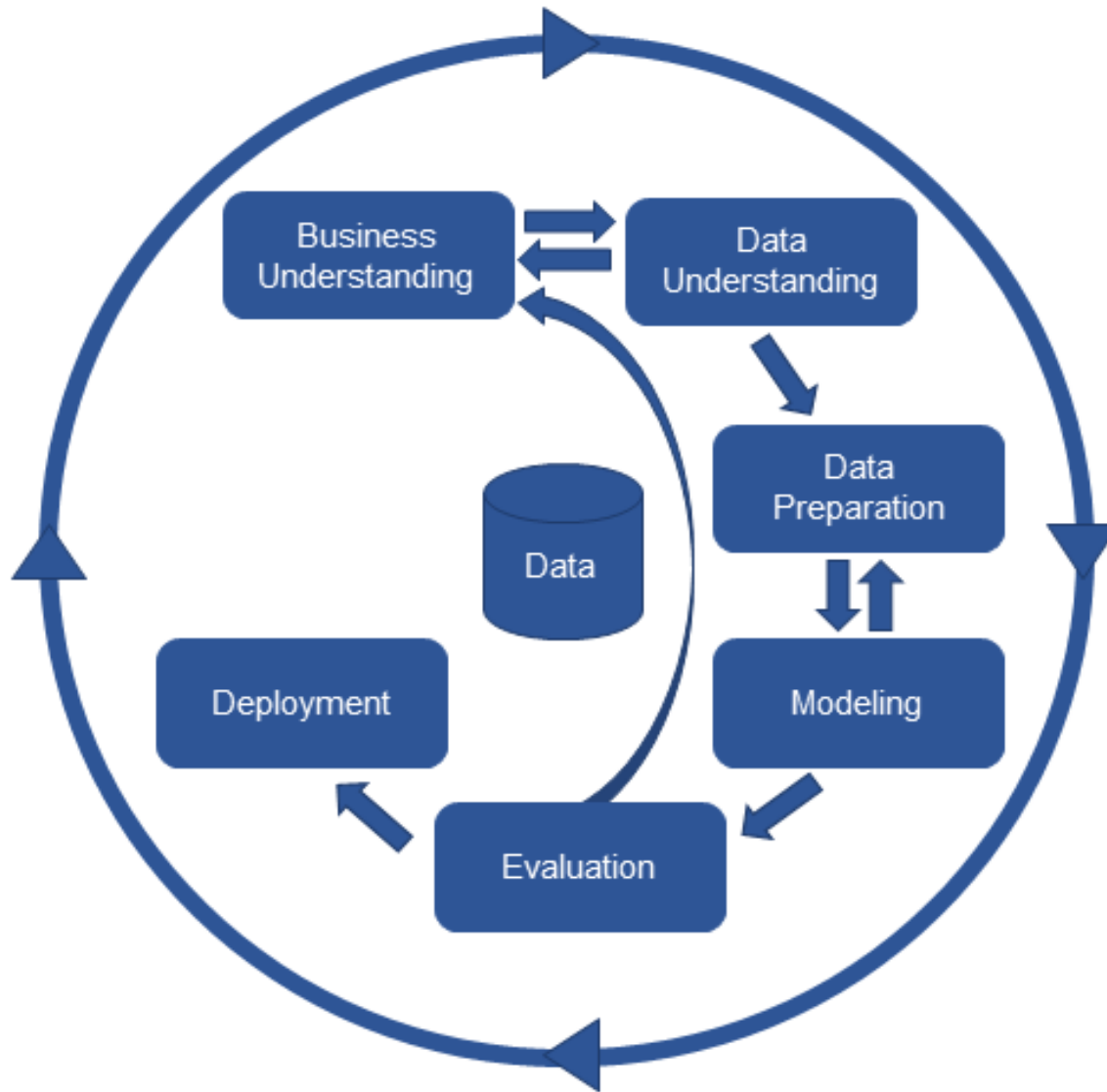
주요단계

- 추출(데이터타겟 select)
- Pre processing(전처리)
- Transformation(변환)
- Data Mining(데이터마이닝)
- Interpretation/Evaluation(해석/평가)

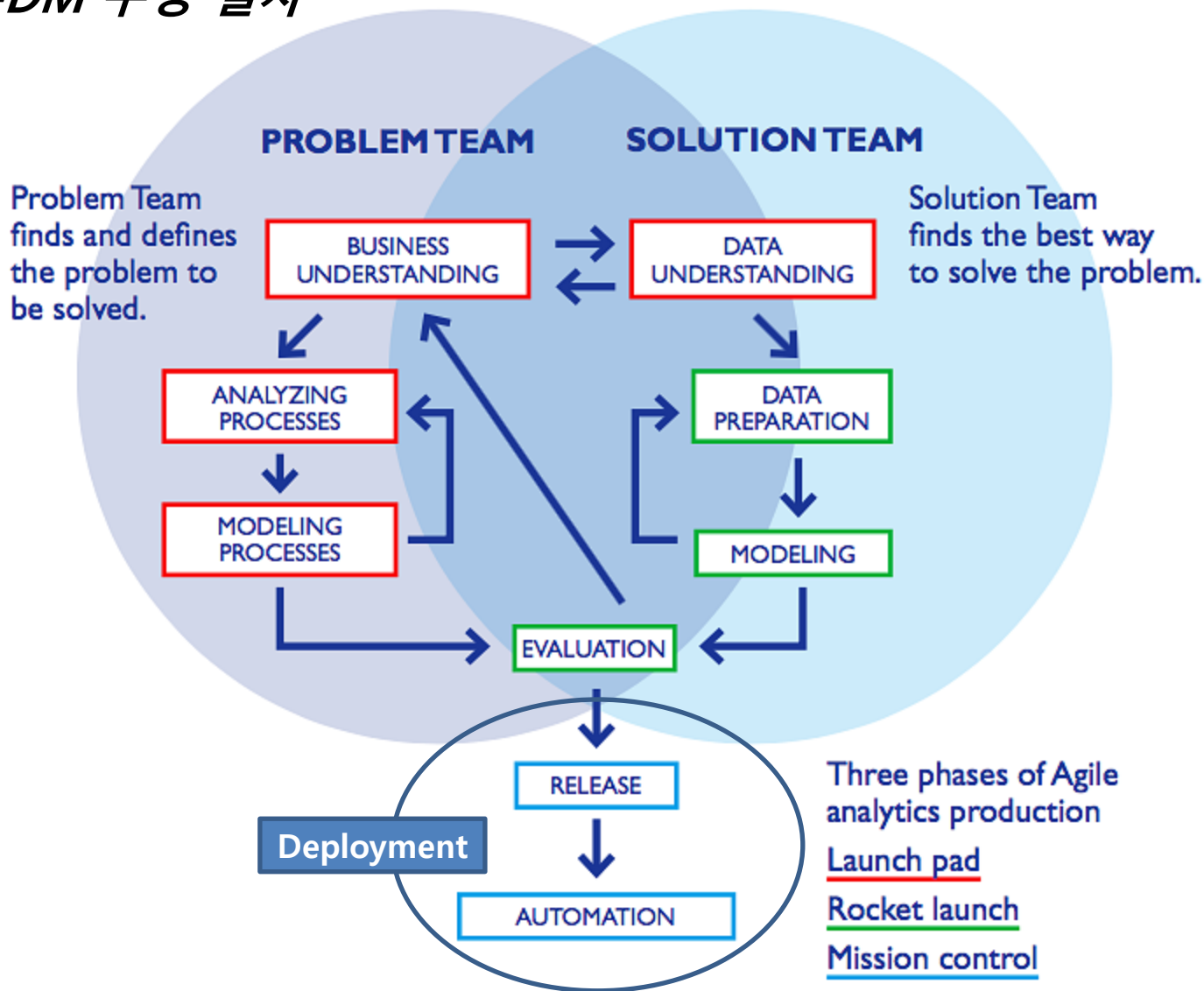
마이닝 단계	설명/절차
Selection (Data Set 선택)	비즈니스 이해 프로젝트 목표설정 데이터선택 데이터 셋 생성
Preprocessing (데이터전처리)	잡음(Noise), 이상치(Outlier), 결측치(Missing Value) 식별 및 처리, 정제
Transformation (데이터변환)	변수생성, 선택, 차원축소
Data Mining (데이터마이닝)	데이터 마이닝
Interpretation/Evaluation (마이닝 결과 해석 및 평가)	결과해석 및 평가 분석목적과의 일치성 도출

- ◆ CRISP-DM (Cross-Industry Standard Process for Data Mining)은 전 세계에서 가장 많이 사용되는 데이터 마이닝 표준 방법론
- ◆ 데이터 마이닝 프로젝트를 계획하기 위한 체계적인 접근 방식을 제공하는 강력하고 검증된 방법론
- ◆ 강력한 실용성, 유연성과 첨단 비즈니스 문제를 해결하기 위해 분석을 사용할 때의 유용성이 뛰어남

방법론	구분	내용
CRISP-DM (Cross Industry Standard Process For Data Mining)	개념	비즈니스의 이해를 바탕으로 데이터 분석 목적의 6단계로 진행되는 데이터마이닝 방법론
	현황	<ul style="list-style-type: none"> - 1996년 유럽연합의 ESPRIT프로젝트에서 시작한 방법론으로 1997년 SPSS 등 참여 - 1999년 1st 버전 발표. 이후 SIG그룹 (Special Interest Group)에서 모델 업그레이드 논의 - 2009년 IBM, SPSS가 인수
	주요고객	크라이슬러, OHRA(네덜란드 보험사), SPSS
	주요단계	<ul style="list-style-type: none"> - Business Understand - Data Understand - Data Preparation - Modeling - Evaluation - Deployment



CRISP-DM 수행 절차



CRISP-DM의 프로세스 단계

① Business Understanding

- 업무 목표 수립
- 현재 상황 평가
- 마이닝 목표 수립
- 프로젝트 계획 수립

② Data Understanding

- 초기 데이터 수집
- 데이터 기술
- 데이터 탐색
- 데이터 품질 검증

③ Data Preparation

- 데이터 설정
- 데이터 선택
- 데이터 정제
- 데이터 생성
- 데이터 통합
- 데이터 형식 적용

④ Modeling

- 모델링 기법 선택
- 테스트 설계 생성
- 모델 생성
- 모델 평가

⑤ Evaluation

- 결과 평가
- 프로세스 재검토
- 향후 단계 결정

⑥ Deployment

- 전개 계획 수립
- 모니터링/유지 보수 계획 수립
- 최종 보고서 작성
- 프로젝트 재검토

마이닝 단계	General Task	Output	설명
Business Understand	Determine Business Objectives	Success Criteria	비즈니스 관점 이해
	Assess Situation	Costs and Benefits	영향평가 및 상황조사
	Determine Data Mining Goals	Goals and Criteria	기술관점 목표수립
	Produce Project Plan	Initial Assessment	프로젝트 세부계획 수립
Data Understand	Collect Initial Data	Initial Data Report	초기데이터 수집/확보
	Describe Data	Describe Report	수집데이터 특성 확인
	Explore Data	Explore Report	목표 부합 데이터 추출
	Verify Data Quality	Quality Report	데이터 품질 검사/검증

마이닝 단계	General Task	Output	설명
Data Preparation	Select Data	Rationale	분석대상 데이터선별
	Clean Data	Cleaning Report	데이터정제, 품질확보
	Construct Data	Attributes/Records	분석데이터 구조화
	Integrate Data	Merged Data	데이터통합
	Format Data	Reformatted Data	모델적합목적 형식보완
Modeling	Select Modeling Techniques	Tech and Assumption	데이터모델링 기법선택
	Generate Test Design	Test Design	품질검증/유효성검사
	Build Model	Parameter Setting	유효 모델링 생성

마이닝 단계	General Task	Output	설명
Evaluation	Evaluate Results	Approved Models	측정 / 모델링 결과 평가
	Review Process	Review Report	중요항목 도출/확인
	Determine Next Steps	List of Actions	다음단계 결정
Deployment	Plan Deployment	Plan Report	전개전략 수립
	Plan Monitoring and Maintenance	Detailed Plan	지속/유지전략 수립
	Produce Final Report	Final Report	최종보고서 작성
	Review Project	Final Documentation	최종보고서 검토

- ◆ CRISP-DM 프로세스의 첫번째 단계는 비즈니스 관점에서 달성하고자하는 것을 이해하는 것입니다. 당신의 조직은 적절한 균형을 이루어야하는 경쟁 목표와 제약 조건을 가지고 있을 수 있습니다. 이 단계의 목표는 프로젝트 결과에 영향을 줄 수 있는 중요한 요소를 밝히는 것입니다. 이 단계를 무시하면 잘못된 질문에 대한 대답을 산출하는 데 많은 노력을 기울일 수 있습니다.

➤ 목표 설정(Set objectives)

- 이는 비즈니스 관점에서 주요 목표를 설명하는 것을 의미합니다. 또한 다른 관련 질문에 답할 수 있습니다. 예를 들어, 주요 목표는 고객이 경쟁 업체로 이동하기 쉬운 시기를 예측하여 현재 고객을 유지하는 것입니다. 관련 비즈니스 질문은 "사용된 채널이 고객의 유지 여부에 영향을 줍니까?" 또는 "ATM 요금을 낮추면 이탈하는 고가치 고객의 수가 크게 줄어 듭니까?"

➤ 프로젝트 계획 수립(Produce project plan)

- 여기서는 데이터 마이닝과 비즈니스 목표를 달성하기 위한 계획을 설명합니다. 이 계획에는 도구와 기법의 초기 선택을 포함하여 나머지 프로젝트 동안 수행할 단계가 지정되어야 합니다.

➤ 비즈니스 성공 기준(Business success criteria)

- 여기서는 프로젝트가 비즈니스 관점에서 성공했는지 여부를 결정하는 데 사용할 기준을 정의합니다. 이상적으로는 구체적이고 측정 가능해야 합니다 (예 : 고객 이탈을 특정 수준으로 줄이는 것). 그러나 때때로 "관계에 대한 유용한 통찰력"과 같은 주관적인 기준이 필요할 수도 있습니다. 주관적인 판단을 내리는 사람이 누구인지 분명히 해야 합니다.

● 현재 상황 평가(Assess the current situation)

여기에는 데이터 분석 목표와 프로젝트 계획을 결정할 때 고려해야 할 모든 리소스, 제약 조건, 가정 및 기타 요인에 대한 보다 자세한 사실 조사가 포함됩니다.

■ 인벤토리 자원 목록(Inventory of resources)

- 인력 (비즈니스 전문가, 데이터 전문가, 기술 지원, 데이터 마이닝 전문가)
- 데이터 (고정된 추출, 실행 운영시스템 접근, 저장소 접근, 운영 데이터에 대한 접근)
- 컴퓨팅 리소스 (하드웨어 플랫폼)
- 소프트웨어 (데이터 마이닝 도구, 기타 관련 소프트웨어)

■ 요구 사항, 가정 및 제약(Requirements, assumptions and constraints)

- 완료 일정, 결과의 이해 및 품질, 데이터 보안 문제 및 법적 문제를 포함하여 프로젝트의 모든 요구 사항을 나열하십시오.
- 데이터를 사용할 수 있는지 확인하십시오.
- 프로젝트가 가정한 것을 나열하십시오.
이는 데이터 마이닝 중에 확인할 수 있는 데이터에 대한 가정 일 수도 있지만 프로젝트와 관련된 비즈니스에 대한 검증되지 않은 가정을 포함 할 수도 있습니다. 특히 후자가 결과의 유효성에 영향을 미칠 경우 이를 열거하는 것이 중요합니다.
- 프로젝트의 제약 조건을 나열하십시오.
이것들은 자원의 가용성에 대한 제약 일 수 있지만, 모델링에 사용하는 것이 실제적인 데이터 세트의 크기와 같은 기술적 제약을 포함 할 수도 있다.

- 위험 및 우연성(Risks and contingencies)

- 프로젝트를 지연 시키거나 실패하게 만들 위험 요소 또는 이벤트를 나열하십시오.
- 해당되는 비상 계획을 열거하십시오
- 이런 위험이나 사건이 발생하면 어떤 조치를 취할 것입니까?

- 용어(Terminology)

- 프로젝트와 관련된 용어집을 작성하십시오. 일반적으로 두가지 구성 요소가 있습니다.
- 프로젝트에서 사용할 수 있는 비즈니스 이해와 관련된 비즈니스 용어집입니다.
- 이 용어집을 작성하는 것에 유용한 방법은 "지식 추출"과 교육 훈련입니다.
- 문제가 되는 비즈니스 문제와 관련된 예제로 설명된 데이터 마이닝 용어집.

- 비용과 효과(Costs and benefits)

- 프로젝트의 비용-효과 분석을 구축하여 프로젝트의 비용과 비즈니스에 대한 잠재적 효과를 비교합니다.
- 이 비교는 가능한 한 구체적이어야 합니다. 예를 들어, 장사가 이루어지는 상황에서 재무적 측정치를 사용해야 합니다.

- 데이터 마이닝 목표 결정(Determine data mining goals)

- 비즈니스 목표는 비즈니스 용어의 목표를 기술합니다.
- 데이터 마이닝 목표는 기술적인 측면에서 프로젝트 목표를 기술합니다.
- 예를 들어 비즈니스 목표는 "기존 고객에 대한 카탈로그 판매 증가"일 수 있습니다.
- 데이터 마이닝 목표는 "지난 3년 동안 구매한 경우 고객이 구매할 상품 수, 인구 통계학적 정보 (연령, 급여, 도시 등) 및 항목의 가격" 일 수 있습니다.

- 비즈니스 성공 기준(Business success criteria)

- 비즈니스 목표 달성을 가능하게 하는 프로젝트의 의도된 결과를 설명합니다.

- 데이터 마이닝 성공 기준(Data mining success criteria)

- 특정 수준의 예측 정확도 또는 일정 수준의 "상승"을 가진 구매 경향 프로파일과 같이 전문 용어로 프로젝트의 성공적인 결과에 대한 기준을 정의합니다.
- 비즈니스 성공 기준, 주관적인 판단을 하는 사람이 식별해야 하는 주관적인 용어로 이것을 기술할 필요가 있을 수 있습니다.

- **프로젝트 계획 수립(Produce project plan)**

- 데이터 마이닝 목표를 달성하고 비즈니스 목표를 달성하기 위한 계획을 기술.
- 계획에는 도구와 기법의 초기 선택을 포함하여 나머지 프로젝트 동안 수행할 단계가 지정.

- **프로젝트 계획(Project plan)**

- 프로젝트에서 실행될 단계와 해당 기간, 필요한 자원, 입력, 출력 및 종속성을 나열.
- 가능한 경우 모델링 및 평가 단계를 반복하는 등 데이터 마이닝 프로세스에서 대규모 반복을 명시적으로 시도
- 프로젝트 계획의 일환으로, 시간 일정과 위험 간의 의존성을 분석
- 프로젝트 계획에서 이러한 분석의 결과를 명시적으로 표시하고, 위험이 나타난 경우 조치와 권장 사항을 제시
- 이 단계에서 어떤 평가 전략이 평가 단계에서 사용될 것인지 결정
- 각 단계가 끝나면 진행 상황과 업적을 검토하고 이에 따라 프로젝트 계획을 업데이트
- 이러한 업데이트에 대한 특정 검토 지점은 프로젝트 계획의 일부 여야합니다.

- **도구 및 기법의 초기 평가(Initial assessment of tools and techniques)**

- 첫번째 단계가 끝나면 도구 및 기법에 대한 초기 평가를 수행
- 프로세스의 각 단계에 대한 다양한 방법을 지원하는 데이터 마이닝 도구를 선택
- 도구와 기술을 선택하면 전체 프로젝트에 영향을 미칠 수 있으므로 프로세스 초기에 도구와 기술을 평가하는 것이 중요

- CRISP-DM 프로세스의 두번째 단계에서는 프로젝트 리소스에 정리된 데이터를 획득
 - 이 초기 수집에는 데이터를 이해하는 데 필요한 경우 데이터 로드가 포함
 - 데이터를 이해하는 데 특정 도구를 사용하는 경우 이 도구에 데이터를 로드
 - 여러 데이터 소스를 얻은 경우 이러한 데이터 소스를 언제 어떻게 통합 할 것인지 고려
- 초기 데이터 수집 보고서(Initial data collection report)
 - 수집된 데이터 소스를 위치, 수집 방법 및 발생한 문제와 함께 나열합니다.
 - 발생한 문제 및 발생한 모든 해결책을 기록하십시오.
 - 이것은 이 프로젝트의 향후 복제와 유사한 미래 프로젝트의 실행 모두에 도움이 될 것입니다.
 - 데이터 설명(Describe data)
 - 획득한 데이터의 "gross"또는 "surface"특성을 조사하고 결과를 리포트하십시오.
 - 데이터 설명 리포트(Data exploration report) – 형식, 수 (예 : 각 테이블의 레코드 및 필드 수), 필드의 ID 및 발견된 기타 surface 피처를 포함하여 수집된 데이터를 설명하십시오. 획득한 데이터가 요구 사항을 충족시키는 지 여부를 평가하십시오.

- 데이터 탐색(Explore data)

- 이 단계에서는 쿼리, 데이터 시각화, 리포팅 기술을 사용하여 데이터 마이닝 질문을 처리
 - ✓ 주요 속성의 배포 (예 : 예측 작업의 대상 속성)
 - ✓ 쌍 또는 적은 수의 속성 사이의 관계
 - ✓ 단순 집계 결과
 - ✓ 중요한 하위 집단의 특성
 - ✓ 간단한 통계 분석
- 이러한 분석은 데이터 마이닝 목표를 직접적으로 처리 할 수 있음
- 데이터 기술과 품질 보고서에 기여하거나 수정하고, 추후 분석에 필요한 변환 및 기타 데이터 준비 단계에 투입 할 수 있음
- 데이터 탐색 보고서
 - ✓ 첫 번째 결과 또는 초기 가설과 나머지 프로젝트에 미치는 영향을 포함하여 데이터 탐색 결과를 설명함
 - ✓ 적절한 경우 여기에 그래프와 플롯을 포함시켜 흥미로운 데이터 하위 집합에 대한 추가 조사를 제안하는 데이터 특성을 나타낼 수 있음

- 데이터 품질 확인(Verify data quality).

- 데이터가 완료 되었습니까 (필요한 모든 경우가 포함됩니까?)?
- 그것이 맞습니까? 아니면 오류를 포함하고 있습니까? 오류가있는 경우 얼마나 많습니까?
- 데이터에 누락된 값이 있습니까? 그렇다면 어떻게 표현되며, 어디에서 발생하며, 얼마나 흔하게 나타 납니까?

- 데이터 품질 보고서(Data quality report)

- 데이터 품질 검증 결과를 나열하십시오.
- 품질 문제가 있는 경우 가능한 해결책을 제안하십시오.
- 데이터 품질 문제에 대한 해결책은 일반적으로 데이터나 비즈니스 지식에 크게 의존



● 데이터 선택(Select your data)

- 이것은 분석을 위해 사용할 데이터를 결정하는 프로젝트의 단계
- 이러한 결정을 내리는 데 사용할 수 있는 기준에는 데이터 마이닝 목표와의 관련성, 데이터 품질 및 데이터 볼륨 또는 데이터 유형에 대한 제한과 같은 기술적 제약이 포함
- 데이터 선택은 테이블의 레코드(행) 선택뿐만 아니라 속성(열)의 선택을 다룸
- 포함/배제의 이론적 근거(Rationale for inclusion/exclusion)
 - ✓ 데이터를 포함/제외하고 이러한 결정에 대한 이유를 설명합니다.

● 데이터 정제(Clean your data)

- 이 작업에는 선택한 분석 기술에 필요한 수준으로 데이터 품질을 높이는 작업이 포함
- 여기에는 데이터의 깨끗한 부분 집합 선택, 적절한 기본값의 삽입 또는 모델링을 통해 누락된 데이터를 추정하는 등보다 공격적인 기법이 포함될 수 있음
- 데이터 정제 보고서(Data cleaning report)
 - ✓ 데이터 품질 문제를 해결하기 위해 취한 조치와 조치를 설명
- 정제 목적으로 만들어진 데이터의 변형과 분석 결과에 미칠 수 있는 영향을 파악

● 필수 데이터 구성(Construct required data)

- 이 작업에는 파생 된 특성 또는 전체 새 레코드의 생성 또는 기존 특성의 변환 된 값과 같은 구성적인 데이터 준비 작업이 포함
- 파생 속성(Derived attributes) – 동일한 레코드의 하나 이상의 기존 속성으로 구성된 새로운 속성입니다. 예를 들어 length 및 width 변수를 사용하여 영역의 새 변수를 계산할 수 있음
- 생성 된 레코드(Generated records) – 완전히 새로운 레코드의 생성을 설명
- 예를 들어 작년에 구매하지 않은 고객에 대한 레코드를 작성해야 할 수 있습니다.
- 원시 데이터에 그러한 레코드를 가질 이유는 없었지만, 모델링 목적을 위해 특정 고객이 구매를하지 않았다는 사실을 명시 적으로 나타내는 것이 합리적 일 수 있음

● 데이터 통합(Integrate data)

- 이는 여러 데이터베이스, 테이블 또는 레코드에서 정보를 결합하여 새로운 레코드 값을 만듭니다.
- 병합된 데이터(Merged data)
 - ✓ 병합 테이블은 동일한 오브젝트에 대한 서로 다른 정보가 있는 두개 이상의 테이블을 결합하는 것을 말합니다.
 - ✓ 예를 들어, 소매 체인에는 각 매장의 일반적인 특성(예 : 매장 공간, 물 유형)에 대한 정보가 있는 테이블 하나, 요약된 판매 데이터가 있는 또 다른 테이블(예 : 이익, 전년 대비 매출 변동율) 및 주변 지역의 인구 통계에 관한 정보 테이블을 합칩니다.
 - ✓ 각 테이블에는 각 상점에 대해 하나의 레코드가 들어 있습니다.
 - ✓ 이 테이블은 소스 테이블의 필드를 결합하여 각 상점에 대해 하나의 레코드가 있는 새 테이블로 병합될 수 있습니다.
- 집계(Aggregation)
 - ✓ 집계는 여러 개의 레코드나 테이블의 정보를 요약하여 새 값을 계산하는 작업을 말합니다.
 - ✓ 예를 들어 각 구매에 대해 하나의 레코드가 있는 고객 구매 테이블을 구매 수, 평균 구매 금액, 신용 카드로 부과된 주문 비율 등의 필드가 있는 각 고객에 대한 레코드가 있는 새 테이블로 변환합니다.



◆ 모델링 기법 선택(Select modeling technique)

- 모델링의 첫번째 단계로서 사용할 실제 모델링 기법을 선택하게 됩니다.
- 비즈니스 이해 단계에서 이미 도구를 선택했더라도 이 단계에서는 C5.0이나 의사 결정 트리(decision-tree) 작성 또는 역전파(back propagation)가 있는 신경망(neural network) 생성과 같은 특정 모델링 기술 사용을 선택하게됩니다.
- 여러 기술이 적용되는 경우 각 기술에 대해 이 작업을 개별적으로 수행하십시오.
- 모델링 기법(Modelling technique)
 - 사용할 실제 모델링 기법을 문서화합니다.
- 모델링 가정(Modelling assumptions)
 - 많은 모델링 기술은 데이터에 대해 특정 가정을합니다.
 - 예를 들어 모든 속성이 균일한 분포를 가지며 허용된 값이 누락되지 않고 클래스 속성이 기호(symbolic)여야 하는 것 등입니다.

◆ 테스트 디자인 생성(Generate test design)

- 실제로 모델을 작성하기 전에 모델의 품질(quality)과 유효성(validity)을 테스트하기 위한 프로시저(procedure) 또는 메커니즘을 생성해야합니다.
- 예를 들어 분류(classification)와 같은 지도 데이터 마이닝 작업(supervised data mining task)에서 데이터 마이닝 모델의 품질 측정으로 오류율(error rate)을 사용하는 것이 일반적입니다.
- 따라서 일반적으로 데이터 세트를 학습과 테스트 세트로 분리하고, 학습 세트에서 모델을 작성하고, 별도의 테스트 세트에서 품질을 평가합니다.

◆ 테스트 설계(Test design)

- 모델을 학습, 테스트 및 평가하기 위한 계획을 기술하십시오. 계획의 주요 구성 요소는 사용 가능한 데이터 집합을 학습, 테스트 및 유효성 검사 데이터 집합으로 나누는 방법을 결정하는 것입니다.

◆ 모델 빌드(Build model)

- 준비된 데이터 세트에서 모델링 도구를 실행하여 하나 이상의 모델을 만듭니다.
- 파라미터 설정(Parameter settings)
 - 어떤 모델링 도구로도 조정할 수 있는 파라미터가 많습니다.
 - 파라미터 설정의 선택에 대한 근거와 함께 파라미터와 선택한 값을 나열하십시오.
- 모델(Models)
 - 모델에 대한 보고서가 아니라 모델링 도구에서 생성한 실제 모델입니다.
- 모델 설명(Model descriptions)
 - 결과 모델을 설명하고, 모델의 해석에 대해 보고하고, 그 의미와 관련하여 문제점을 문서화합니다.



◆ 모델 평가(Assess model)

- 도메인 지식, 데이터 마이닝 성공 기준 및 원하는 테스트 설계에 따라 모델을 해석하십시오.
- 기술적 모델링과 검색 기술의 적용을 판단한 다음 나중에 비즈니스 문맥(business context)에서 데이터 마이닝 결과를 논의하기 위해 비즈니스 분석가와 도메인 전문가에게 문의하십시오.
- 이 작업에서는 모델만 고려하고 평가 단계에서는 프로젝트 과정에서 생성된 다른 모든 결과도 고려합니다.
- 이 단계에서 모델을 평가하고 기준에 따라 평가해야 합니다. 비즈니스 목표와 비즈니스 성공 기준을 고려해야 합니다. 대부분의 데이터 마이닝 프로젝트에서 한가지 기법이 두 번 이상 적용되고 데이터 마이닝 결과가 여러 가지 기법으로 생성됩니다.
- 모델 평가(Model assessment)
 - 이 작업의 결과를 요약하고 생성된 모델의 품질 (예 : 정확도)을 나열하고 서로 관련하여 품질을 평가합니다.
- 수정된 파라미터 설정(Revised parameter settings)
 - 모델 평가에 따라 파라미터 설정을 수정하고 다음 모델링 실행을 위해 파라미터 설정을 조정하십시오. 최고의 모델을 찾았다고 강력하게 믿을때까지 모델 구축과 평가를 반복하십시오. 그러한 모든 작업과 평가를 문서화하십시오.



◆ 결과 평가(Evaluate results)

- 이전 평가 단계에서는 모델의 정확성(accuracy) 및 일반성(generality)과 같은 요인을 처리했습니다.
- 이 단계에서 모델이 비즈니스 목표를 충족시키는 정도를 평가하고 이 모델이 불완전한 사업적 이유가 있는지 확인해야 합니다.
- 또 다른 옵션은 시간과 예산 제약이 허락하는 경우 실제 응용 프로그램에서 테스트 응용 프로그램의 모델을 테스트하는 것입니다.
- 평가 단계에서는 생성한 다른 데이터 마이닝 결과도 평가합니다. 데이터 마이닝 결과는 원래 비즈니스 목표와 필연적으로 원래 비즈니스 목표와 관련이 없는 기타 모든 결과에 반드시 관련되는 모델을 포함하지만 향후 방향에 대한 추가 과제, 정보 또는 힌트를 공개 할 수도 있습니다.
- 데이터 마이닝 결과 평가(Assessment of data mining results)
 - 프로젝트가 이미 초기 비즈니스 목표를 충족하는지 여부에 관한 최종 진술을 포함하여 비즈니스 성공 기준에 따라 평가 결과를 요약합니다.
- 모델 승인(Approved models)
 - 비즈니스 성공 기준과 관련하여 모델을 평가 한 후 선택한 기준을 충족하는 생성 된 모델이 승인 된 모델이됩니다.

◆ 프로세스 검토(Review process)

- 이 시점에서 결과 모델은 만족스럽고 비즈니스 요구를 만족시키는 것으로 보입니다.
- 어쨌든 간과된 중요한 요소나 작업이 없는지 확인하기 위해 보다 철저하게 검토하는 것이 좋습니다.
- 이 리뷰에는 품질 보증 문제(quality assurance issue)도 포함됩니다. 예를 들어 모델을 올바르게 빌드 했습니까? 우리는 사용을 허용하고 향후 분석에 사용할 수 있는 속성만 사용했습니까?
- 프로세스의 검토
 - 프로세스 검토를 요약하고 놓치거나 반복해야 할 활동을 강조합니다.

◆ 다음 단계 결정(Determine next steps)

- 평가 결과 및 프로세스 검토 결과에 따라 이제 진행 방법을 결정할 수 있습니다.
- 이 프로젝트를 완료하고 배포로 이동하거나 추가 반복을 시작하거나 새로운 데이터 마이닝 프로젝트를 설정할지 결정합니다. 이는 남은 자원과 예산을 고려하여야 합니다.
- 가능한 조치(possible action) 목록
 - 잠재적인 추가 조치와 각 옵션에 대한 이유를 나열하십시오.
- 결정(Decision)
 - 합리적인 근거와 함께 진행 방법에 대한 결정을 설명하십시오.

◆ 배포 계획(Plan deployment)

- 배포 단계에서는 평가 결과를 가져와 배포 전략을 결정합니다.
- 관련 모델을 작성하기 위한 일반적인 절차가 확인된 경우 이 절차는 이후의 배포를 위해 여기에 문서화되어 있습니다.
- 배포가 비즈니스 성공에 절대적으로 중요하기 때문에 비즈니스 이해 단계에서도 배포 방법과 방법을 고려하는 것이 좋습니다.
- 예측 분석이 비즈니스 운영 측면을 개선하는 데 실제로 도움이 되는 곳입니다.
- 배포 계획
 - 필요한 단계와 수행 방법을 포함하여 배포 전략을 요약합니다.

◆ 모니터링 및 유지 관리 계획(Plan monitoring and maintenance)

- 모니터링 및 유지 관리는 데이터 마이닝 결과가 일상적인 비즈니스 및 환경의 일부가 되는 경우 중요한 문제입니다.
- 유지 관리 전략을 세심하게 준비하면 불필요하게 오랜 기간 동안 데이터 마이닝 결과를 잘못 사용하는 것을 방지 할 수 있습니다.
- 데이터 마이닝 결과의 배치를 모니터링하기 위해 프로젝트는 상세한 모니터링 프로세스 계획을 필요로 합니다. 이 계획은 특정 배포 유형을 고려합니다.
- 모니터링 및 유지 관리 계획
 - 필요한 단계 및 수행 방법을 비롯하여 모니터링 및 유지 관리 전략을 요약합니다.



◆ 최종 보고서(Final report) 작성

- 프로젝트가 끝나면 최종 보고서를 작성합니다. 배포 계획에 따라이 보고서는 프로젝트 및 경험의 요약 일뿐 (진행중인 활동으로 아직 문서화되지 않은 경우)이거나 데이터 마이닝 결과의 최종 포괄적 인 프레젠테이션 일 수 있습니다.
- 최종 보고서 – 데이터 마이닝 참여에 대한 최종 보고서입니다. 여기에는 이전 결과물이 모두 포함되어 결과를 요약하고 구성합니다.
- 최종 발표(Final presentation)
 - 프로젝트가 끝나면 종종 회의가 열리게됩니다.

◆ 프로젝트 검토(Review project)

- 옳은 점과 잘못 된 점, 잘 된 점과 개선해야 할 점을 평가하십시오.
- 경험 문서(Experience documentation)
 - 프로젝트 중 얻은 중요한 경험을 요약합니다.
 - 예를 들어, 비슷한 상황에서 가장 적합한 데이터 마이닝 기술을 선택하는 데 있어서 발생한 함정, 오도된 접근법 또는 힌트가 이 문서의 일부일 수 있습니다.
 - 이상적인 프로젝트에서 경험 문서는 프로젝트의 이전 단계에서 개별 프로젝트 구성원이 작성한 보고서를 포함합니다.