머신러닝 연구자와 학생을 위한
머신러닝 프로세스 간소화 통합 관리 분석 기획안

Y2023011 윤요섭

Problem Definition & Biz Objective

머신러닝 학습 및 연구 과정에서 발생하는 복잡한 프로세스와 관리의 어려움으로 인한 연구 및 학습 효율성 저하를 문제로 정의하고, 간편한 머신러닝 프로세스와 통합 관리로 학습 및 연구 과정에서 시간과 비용을 절약하고 진입장벽을 낮춥니다.

High Entry Barrier

머신러닝 기술에 대한 전문적인 지식과 풍부한 경험이 필요하여 높은 진입장벽 존재

Root Cause of Problem

머신러닝 프로세스의 복잡성과 다양한 지식 요구

개인이나 소규모 그룹에서 쉽게 구축하고 진행하는데 있어 많은 어려움 존재

Problem

머신러닝 학습 및 연구 과정에서 발생하는 복잡한 프로세스와 관리의 어려움

Time and Knowledge Constraints

다양한 지식과 기술들을 익히는데 부담과 제약이 있음

Inefficient Management

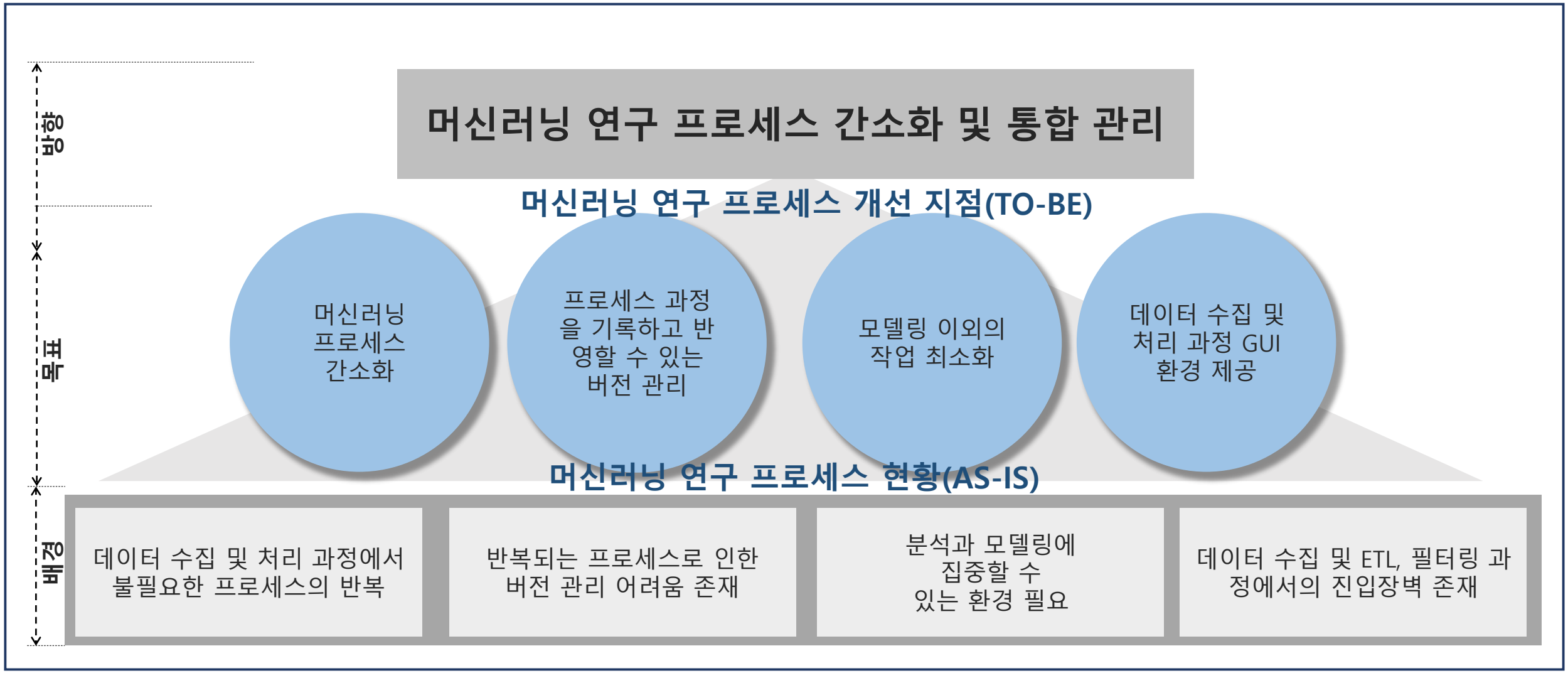
데이터 수집, ETL, 전처리, 데이터 분석, 모델링 과정 반복시 효율적인 관리의 어려움

Biz Objective

머신러닝 프로세스 복잡성 및 통합 관리를 통해 체계적이고 효율적인 방법으로 연구 수행 지원

Desired results

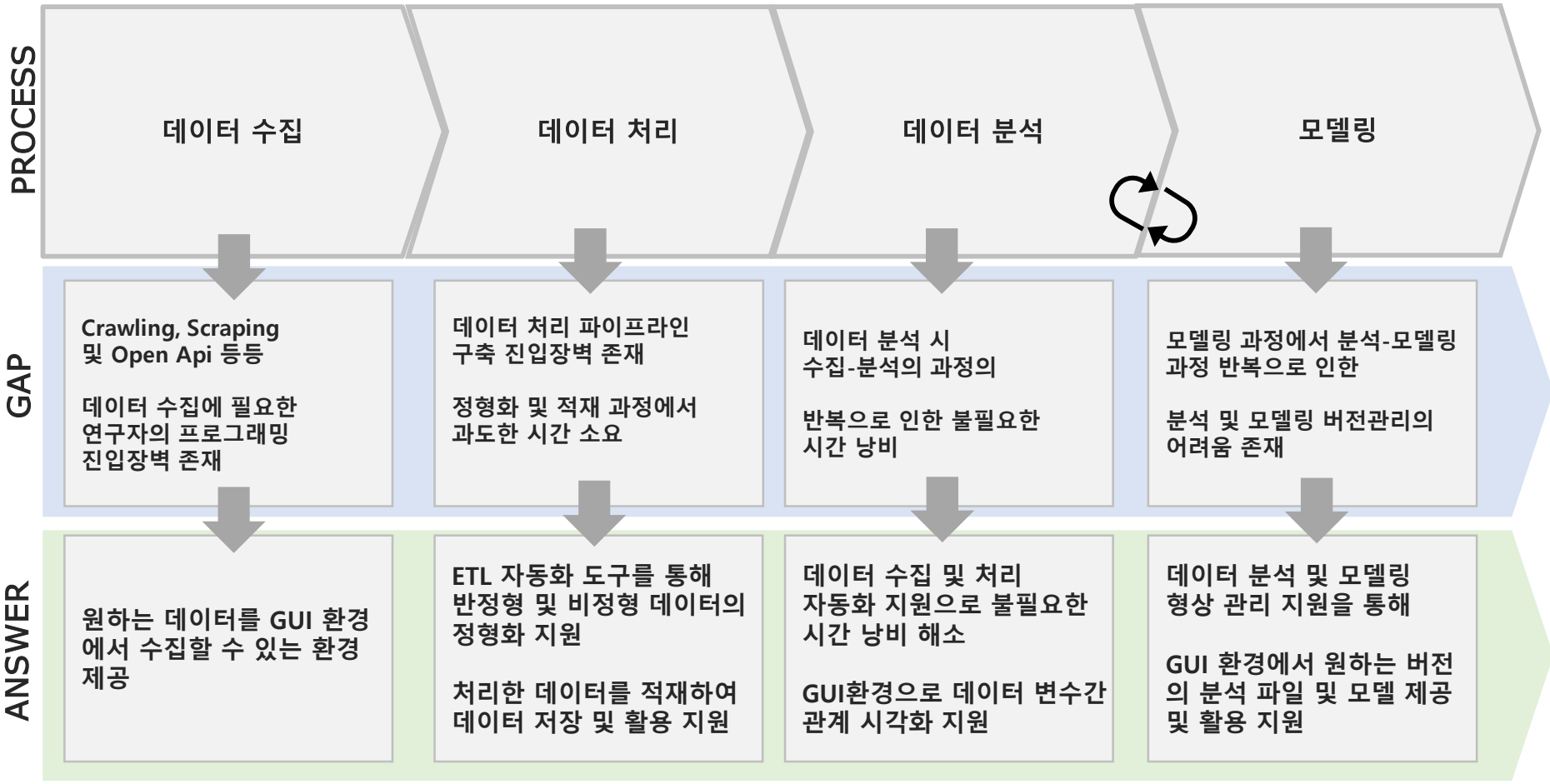
머신러닝 프로젝트 진행 시, 데이터 수집부터 모델링까지 프로세스에 있어서 분석 및 모델링 외의 불필요한 리소스 낭비를 제거하고 간소화합니다. 이를 통해 학생과 연구자들에게 분석 및 모델링에 집중할 수 있는 환경 제공합니다.



GAP Analysis

머신러닝 프로젝트에서 데이터 수집-모델링 프로세스의 AS-IS와 TO-BE의 차이를 명확히 이해하고 GAP Analysis를 통해, Mapping & GAP 처리 방안을 제시합니다.

데이터 수집-모델링 프로세스



Results Description

1. Undesired Results (R1)

- 머신러닝 프로젝트에서 연구자가 데이터 분석과 모델링에 집중할 수 있는 환경의 부재
- 데이터 수집부터 모델링까지 반복되는 프로세스에서 관리의 어려움 존재

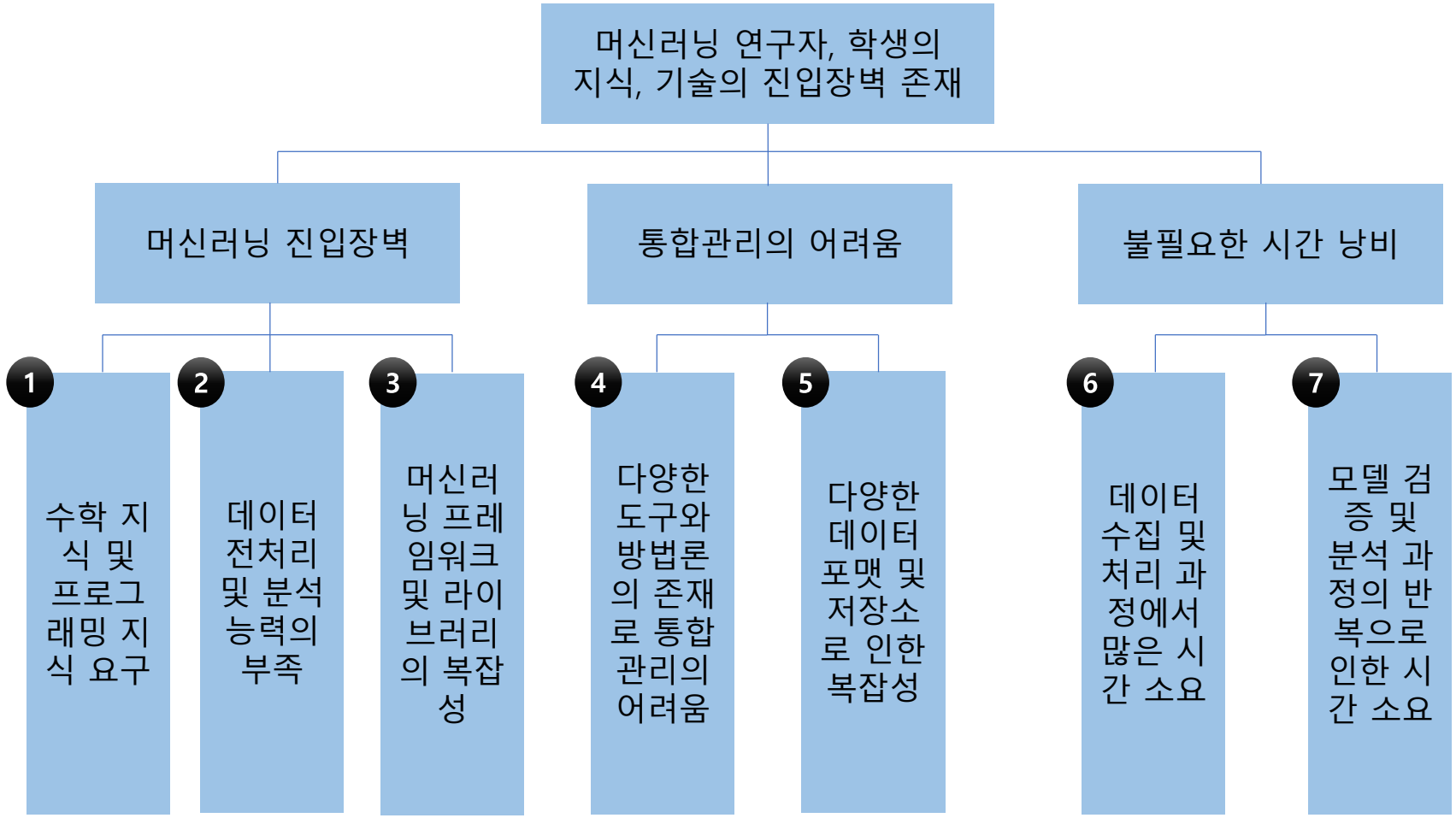
2. Desired Results (R2)

- 머신러닝 프로젝트에서 연구자가 분석과 모델링에 집중할 수 있는 환경 구축
- 데이터 수집부터 모델링 버전 및 형상 관리 지원

"R1과 R2의 GAP 분석을 통한 ANSWER 도출"

Problem Analysis

Problem으로 부터 Main point 도출, Sub issue 세분화를 통해 문제를 구조화하고 분석합니다. Process 진행 시, Problem에 대한 Sub Issue Key-point를 기반으로 인사이트를 도출하고 Answer를 도출합니다.

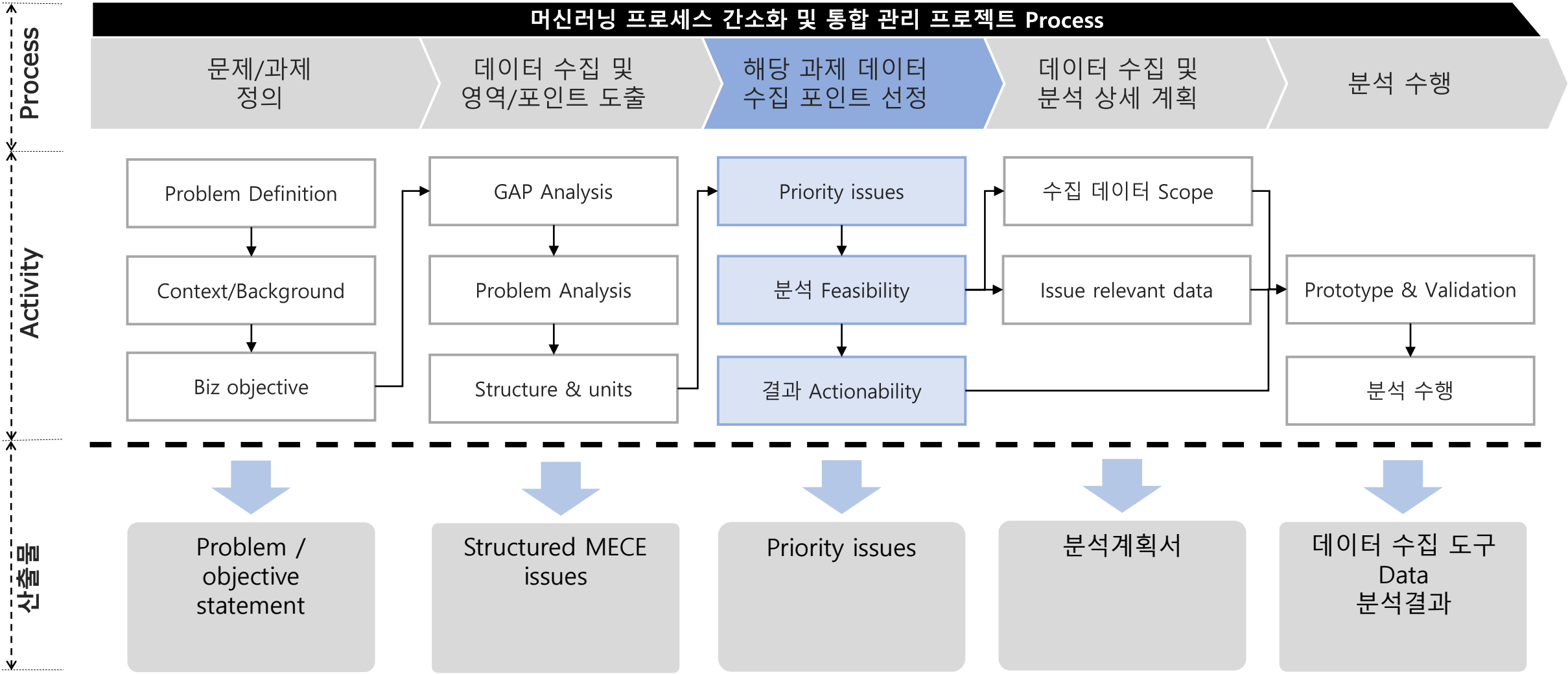


Sub Issue Key-point

1. 데이터 분석 및 모델링을 잘 하기 위해서는 수학적 지식과 프로그래밍 지식 필요
2. 데이터 형식에 대한 이해와 데이터를 탐색하고 분석하는 과정에 대한 이해가 필요
3. 다양한 머신러닝 프레임워크를 Situation에 맞게 활용할 수 있는 역량 필요
4. 지속적인 개발 및 테스트 과정에서 자동화 도구를 통해 버전 관리 필요
5. 데이터 포맷 변환 및 ETL 작업에 대한 이해 필요
6. 데이터 수집 및 처리 자동화 도구 필요
7. 모델 검증 및 반복 과정에서 평가지표 설정 및 모니터링 필요

Comprehensive Plan

Analytical framework을 통해 Item/activity 발굴하고, Process & step framework에 따라 분석을 진행합니다.



Analysis Plan

머신러닝 프로세스 간소화 및 통합관리를 한다면 연구자 및 학생의 노력을 최소화 할 수 있습니다. 이를 통해 데이터 분석과 모델링 작업의 효율성과 정확성을 개선 할 수 있는 기대효과를 도출할 수 있습니다.

"Identifying problems, formulating hypotheses, and archieving expected outcomes"

Sub Issue Task

1. 수학 지식 및 프로그래밍 지식 요구
2. 데이터 전처리 및 분석능력의 부족
3. 머신러닝 프레임워크 및 라이브러리의 복잡성
4. 다양한 도구와 방법론의 존재로 통합 관리의 어려움
5. 다양한 데이터 포맷 및 저장소로 인한 복잡성
6. 데이터 수집 처리의 과정에서 많은 시간 소요
7. 모델 검증 및 분석 과정의 반복으로 인한 시간 소요

Detailed Hypothesis

	IF	THEN
1	수학 지식 및 프로그래밍 지식 지원하는 Framwork 활용	모델링 정확도 및 효율성 개선
2	데이터 전처리 및 분석을 Optional하게 Custormizing	데이터 탐색 시간 단축
3	다양한 머신러닝 프레임워크 및 라이브러리를 Situation에 맞게 선택 지원	다양한 데이터 분석 문제에 대한 해결책 제시 가능
4	자동화 도구를 통해 지속적인 개발 및 테스트 과정 지원 및 버전 관리	분석 및 모델링 효율적인 관리 가능
5	데이터 포맷 변환 및 ETL 작업 지원	데이터 처리 과정에서 정확성과 원하는 데이터 적재까지 소요시간 단축
6	데이터 수집 및 처리 Optional하게 Custormizing하여 needs에 맞게 지원	데이터 수집 및 처리의 소요시간을 단축
7	평가지표 설정 및 모니터링 지원	모델 검증 및 분석 과정 반복으로 인한 소요 시간 단축

Improving Efficiency

- Framwork 선택 및 구축 탐색 시간 개선
- Version Life Cycle 지원을 통한 버전 관리 효율성 증대
- 데이터 처리 UX/CX 개선
- 고객여정에 대한 통합된 사이클 개선을 통한 CX 개선

데이터 수집, 분석 Process

가설에 필요한 데이터만을 수집하여, 분석하여 목표를 달성하고 필요한 기능을 분석하고 구현합니다.

Detailed Hypothesis

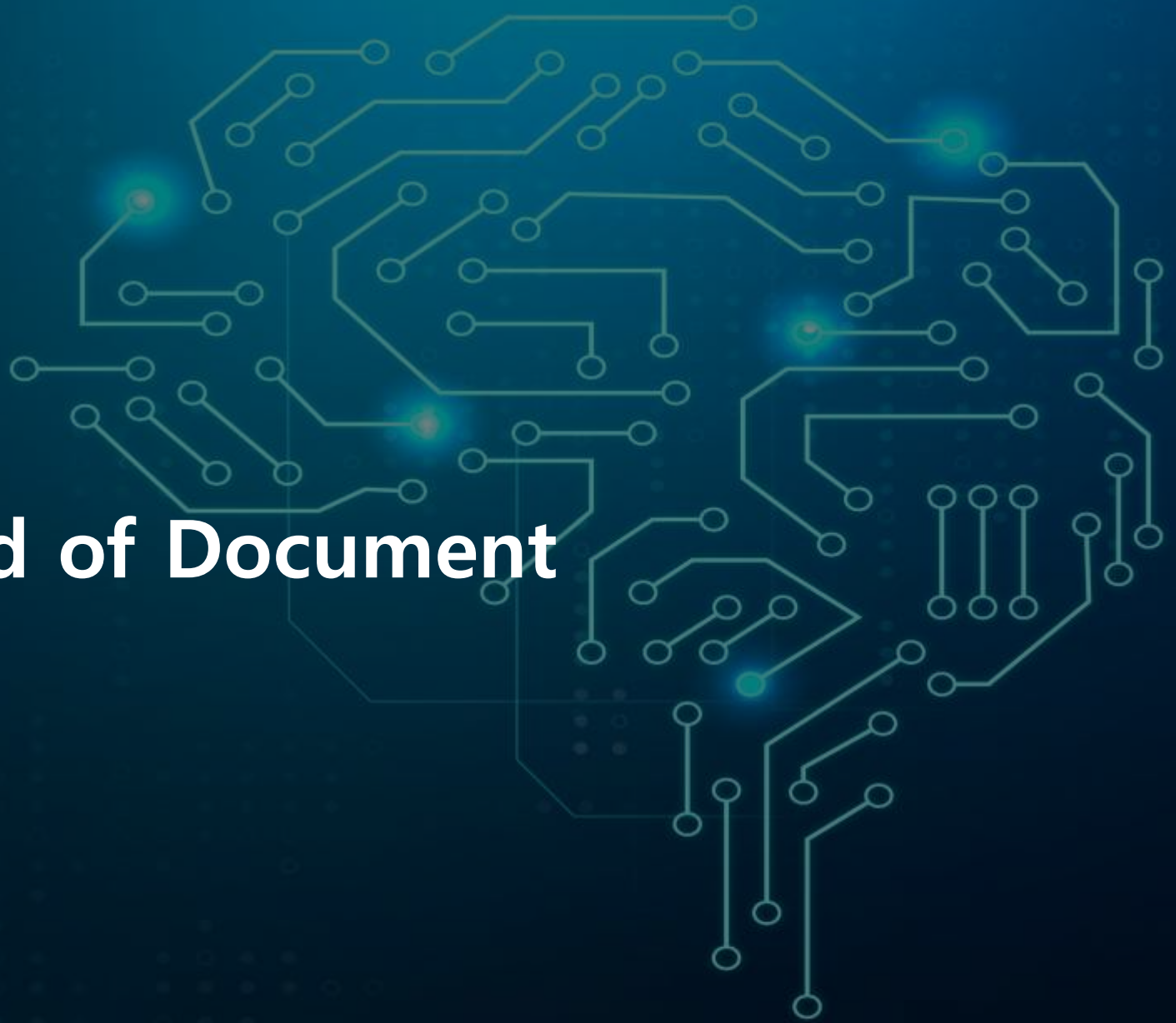
	IF	THEN
1	수학 지식 및 프로그래밍 지식 지원하는 Framwork 활용	모델링 정확도 및 효율성 개선
2	데이터 전처리 및 분석을 Optional하게 Custormizing	데이터 탐색 시간 단축
3	다양한 머신러닝 프레임워크 및 라이브러리를 Situation에 맞게 선택 지원	다양한 데이터 분석 문제에 대한 해결책 제시 가능
4	자동화 도구를 통해 지속적인 개발 및 테스트 과정 지원 및 버전 관리	분석 및 모델링 효율적인 관리 가능
5	데이터 포맷 변환 및 ETL 작업 지원	데이터 처리 과정에서 정확성과 원하는 데이터 적재까지 소요시간 단축
6	데이터 수집 및 처리 Optional하게 Custormizing하여 needs에 맞게 지원	데이터 수집 및 처리의 소요시간을 단축
7	평가지표 설정 및 모니터링 지원	모델 검증 및 분석 과정 반복으로 인한 소요 시간 단축

Data Collection

1. Situation별 데이터 처리 테스트를 위한 필터링 및 클렌징 되지 않은 Raw Data Sample
2. 데이터 전처리, 변수간 상관관계를 테스트 및 시각화 하기 위한 머신러닝 라이브러리에서 제공되는 데이터 셋
3. Sample Data 추출 후, Situation별 머신러닝 프레임워크 및 라이브러리 Data
4. Test Sample 만든 후, Stream, Batch 등 Situation별 포맷 변환 및 ETL 작업 테스트를 위한 정형화 되지 않은 Generator된 Data
5. 데이터 처리를 위한 정형/비정형/반정형 데이터
6. 평가지표를 비교 및 설정할 수 있는 Exporter Data

Data Analysis Method

1. Raw Data Sample기반으로 필터링 및 클렌징시 Situation에 맞는 데이터 추출 정확성 비교
2. 머신러닝 라이브러리를 비교하여 User의 Needs별로 전처리, 변수간 상관관계의 시각화 가능 여부 비교
3. User의 Needs별로 데이터 분석 및 모델링에서 Situation별로 분석 과정의 시간의 단축 여부 비교
4. 데이터의 정합성, 유효성 검사, User의 조건을 만족하는 데이터 Load 여부 체크
5. 비정형/반정형 데이터의 정형화 자동화 분석
6. User의 Needs를 반영한 시각화 평가지표 구성 여부 비교



End of Document