



광운대학교
KwangWoon University

지표 데이터 비정상 원인분석, 시그널 패턴 분류 프로젝트

이상민

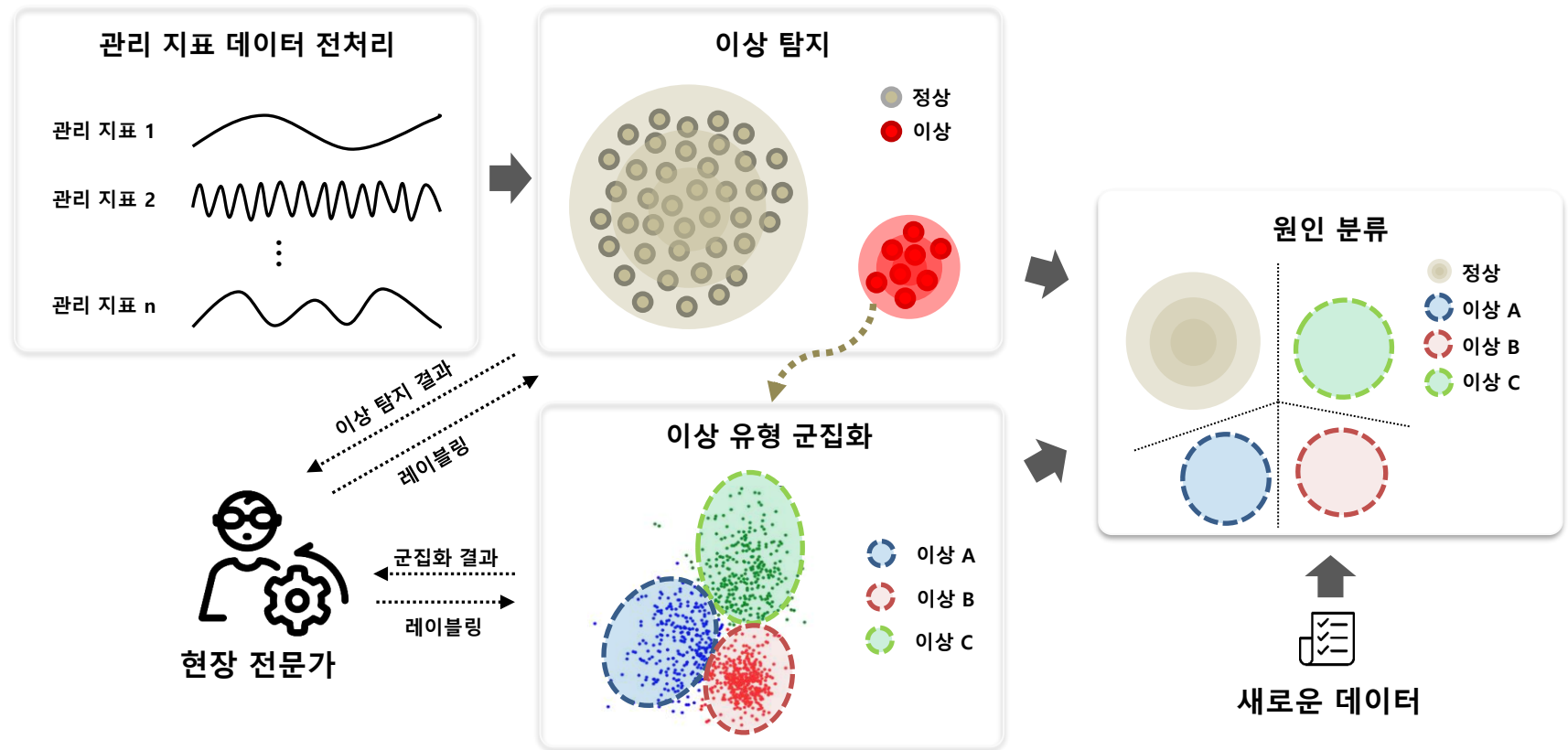
정보융합학부, 소프트웨어융합대학

광운대학교

프로젝트 개요

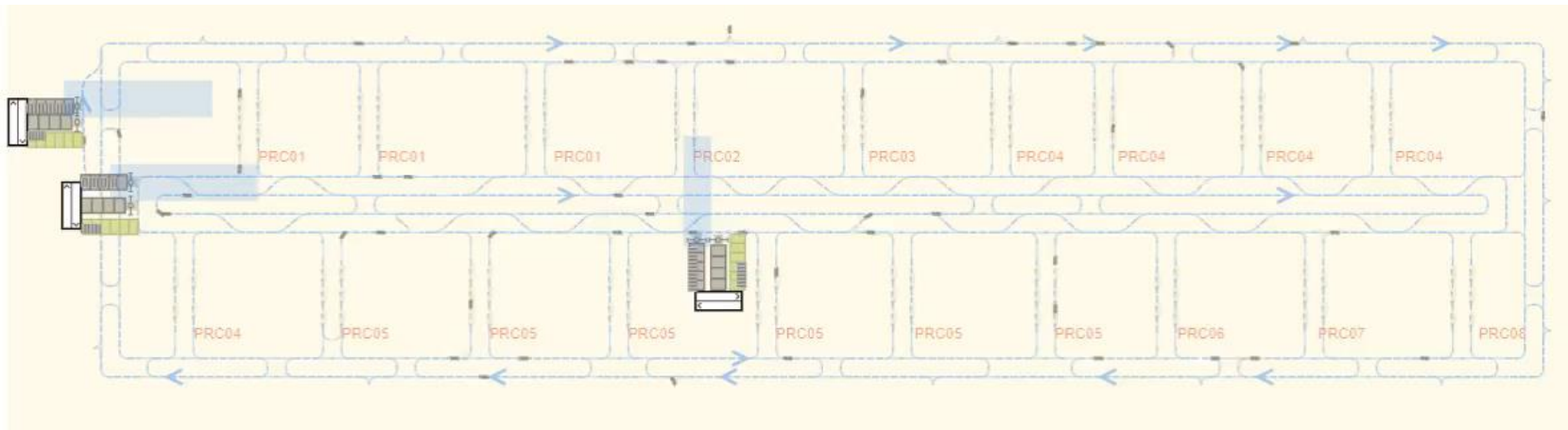
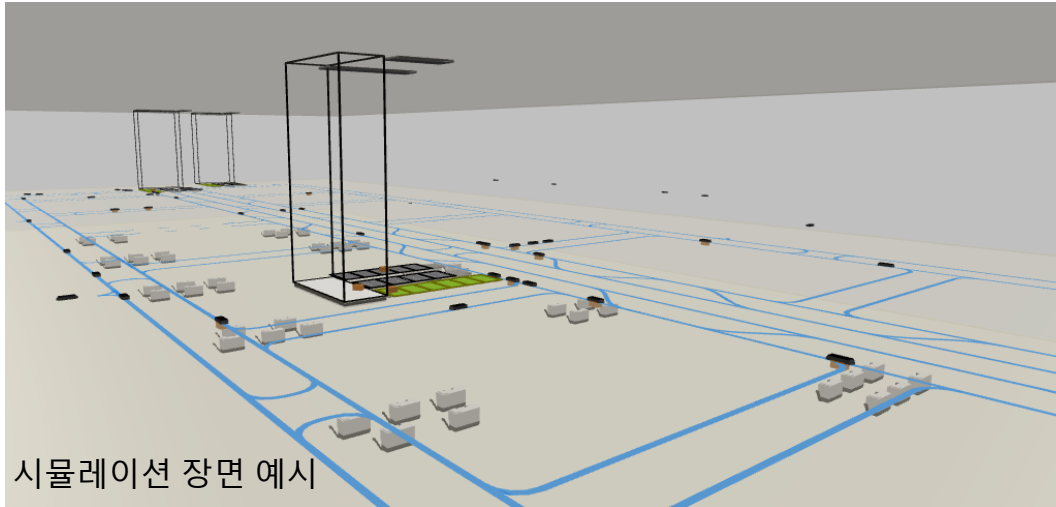
전체 프로세스

❖ 1. 이상 탐지 → 2. 이상 유형 군집화 → 3. 이상 원인 분류



분석 관심 대상, 반도체 공장 물류자동화

데이터 분석 대상



H社 반도체 공장 인접성 예시



네이버 맵 캡처 화면, 2021-10-06

S社 반도체 공장 인접성 예시



네이버 맵 캡처 화면, 2021-10-06

관심 대상, 반도체 공장 물류자동화

데이터 분석 대상

시뮬레이션 장면 예시 - 3D simulations



프로젝트 개요

전체 프로세스

❖ 단계별 프로젝트 수행 내역

1. 이상치 탐지

- 물류 시스템 데이터 수집 및 전처리
- 이상 탐지 모델 구축

2. 이상 유형 군집화

- 이상 관측치를 유사한 유형별로 군집화
- 현업 관리자에 의한 군집별 이상 유형 검증
- 미검증 관측치에 대한 이상 유형 레이블 할당

3. 분류 모델 구축

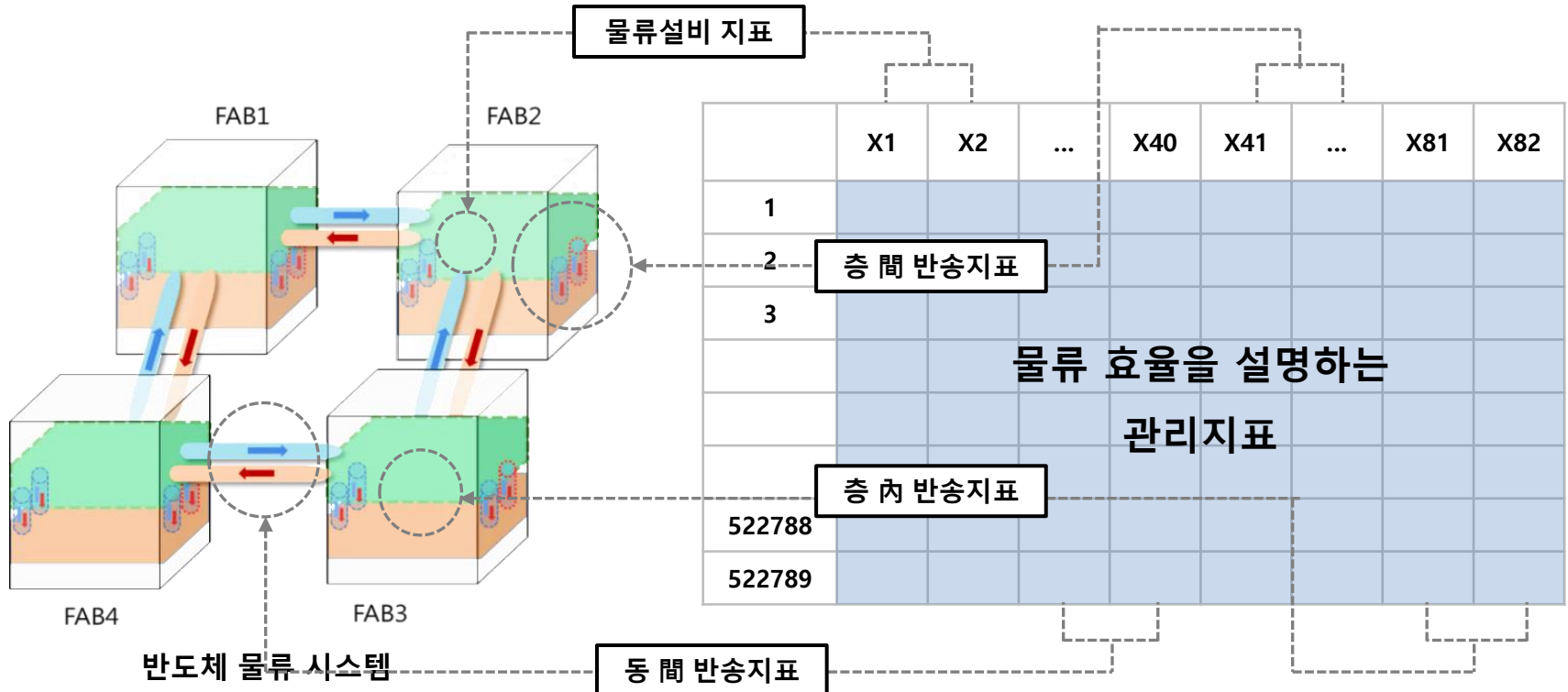
- 정상 및 이상 유형을 자동으로 분류하는 모델 구축

물류 관리지표 데이터 전처리

물류 관리지표

❖ 데이터 수집

- 반도체 물류 시스템에서 발생한 82개 관리지표
- 20XX년 1년 동안의 데이터를 1분 단위로 수집

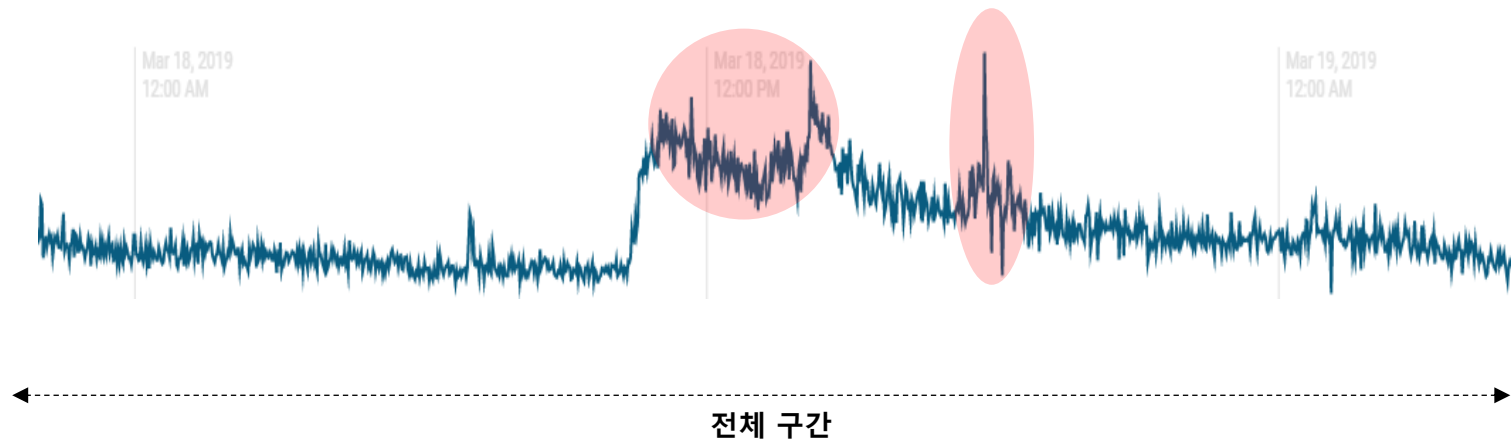


물류 관리지표 데이터 전처리

구간 분할

❖ 구간 분할의 필요성

- 물류시스템 지표는 물량생산계획, 설비 증설 등 생산환경에 따라 변함
→ **시간에 따라 정상 데이터의 특성이 변함**
- 전 구간을 하나의 모델로 분석할 때, 정상 관측치가 이상으로 잘못 판단될 수 있음



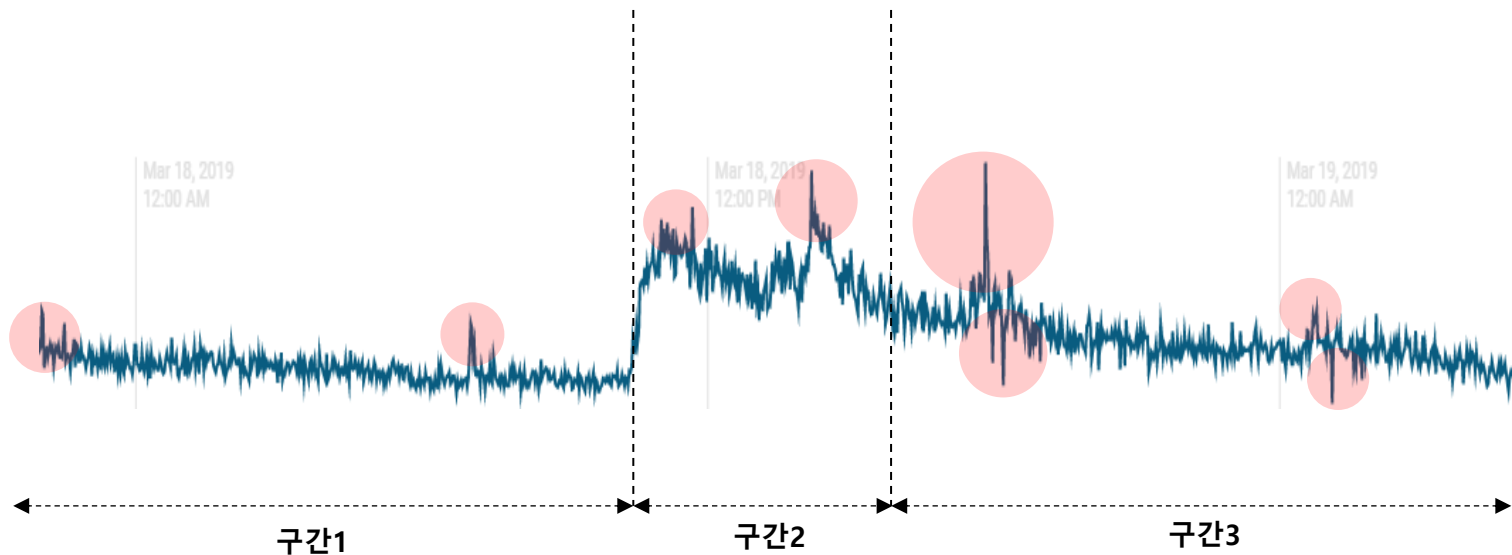
[구간 분할 **前** 이상 탐지 시점]

물류 관리지표 데이터 전처리

구간 분할

❖ 구간 분할의 필요성

- 물류시스템 지표는 물량생산계획, 설비 증설 등 생산환경에 따라 변함
→ **시간에 따라 정상 데이터의 특성이 변함**
- 전 구간을 하나의 모델로 분석할 때, 정상 관측치가 이상으로 잘못 판단될 수 있음



[구간 분할 後 이상 탐지 시점]

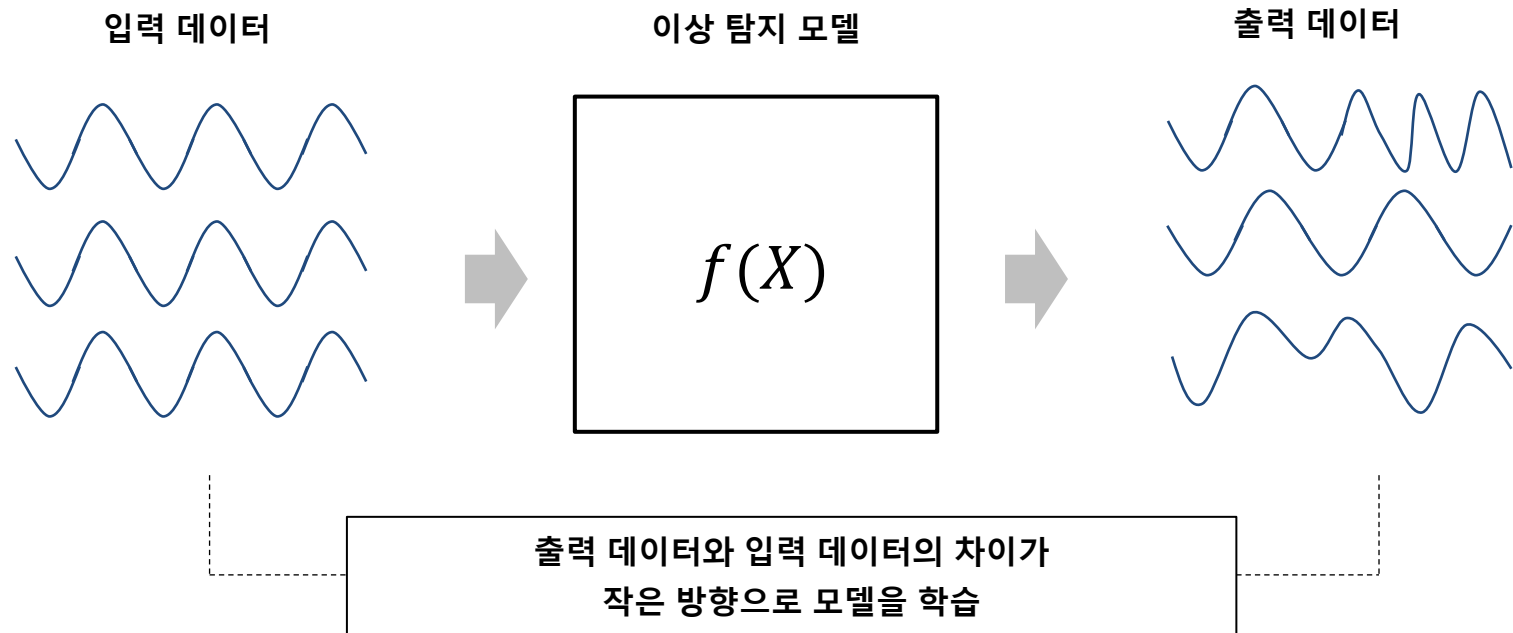
이상 탐지

재구축 기반 모델

❖ 재구축 기반 이상 탐지 모델

- 정상 데이터만을 사용하여 모델 구축 → 정상 패턴을 학습한 모델 구축
- 재구축 오차 = 입력 데이터 - 모델로 부터 나온 데이터
→ 정상 데이터는 자기 자신과 유사하게 생성하도록 (재구축 오차가 작게) 학습

❖ 모델 학습 과정

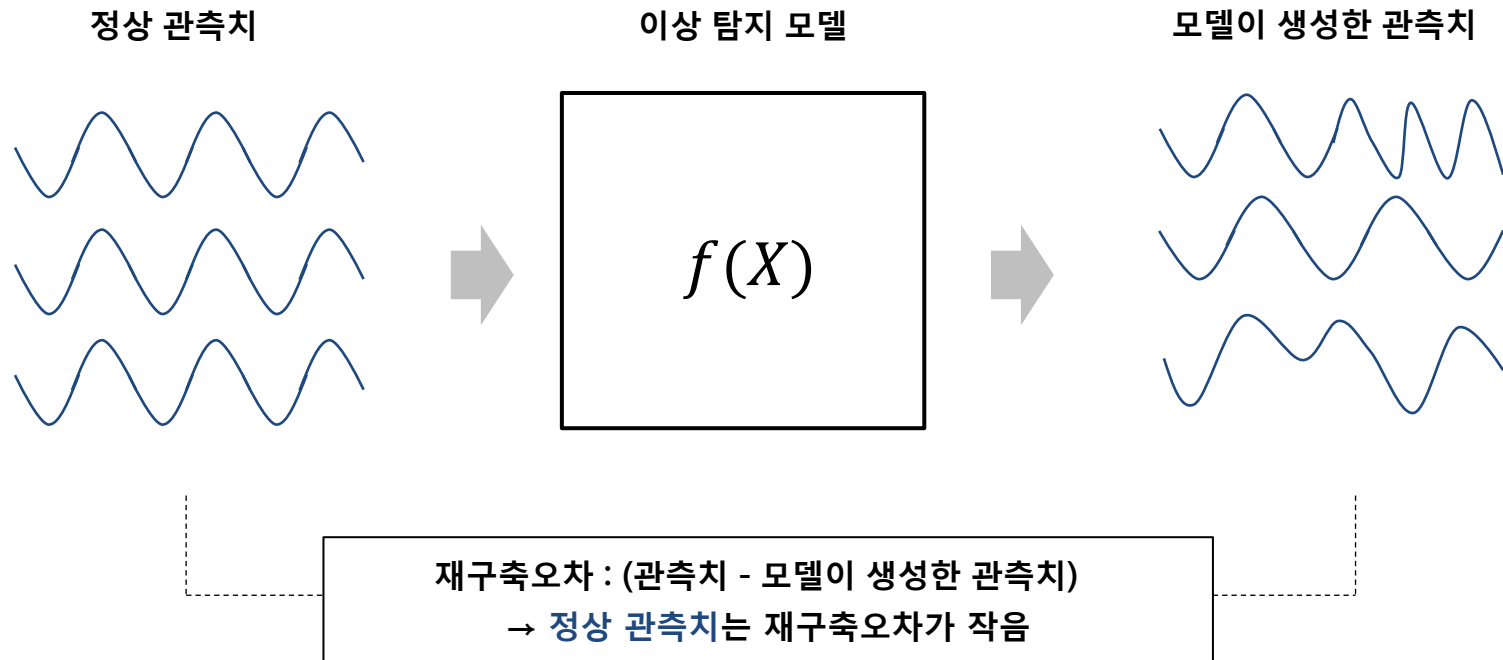


이상 탐지

재구축 기반 모델

❖ 이상 탐지 과정

- 정상 데이터 입력시 → 재구축 오차가 작기 때문에 정상으로 분류

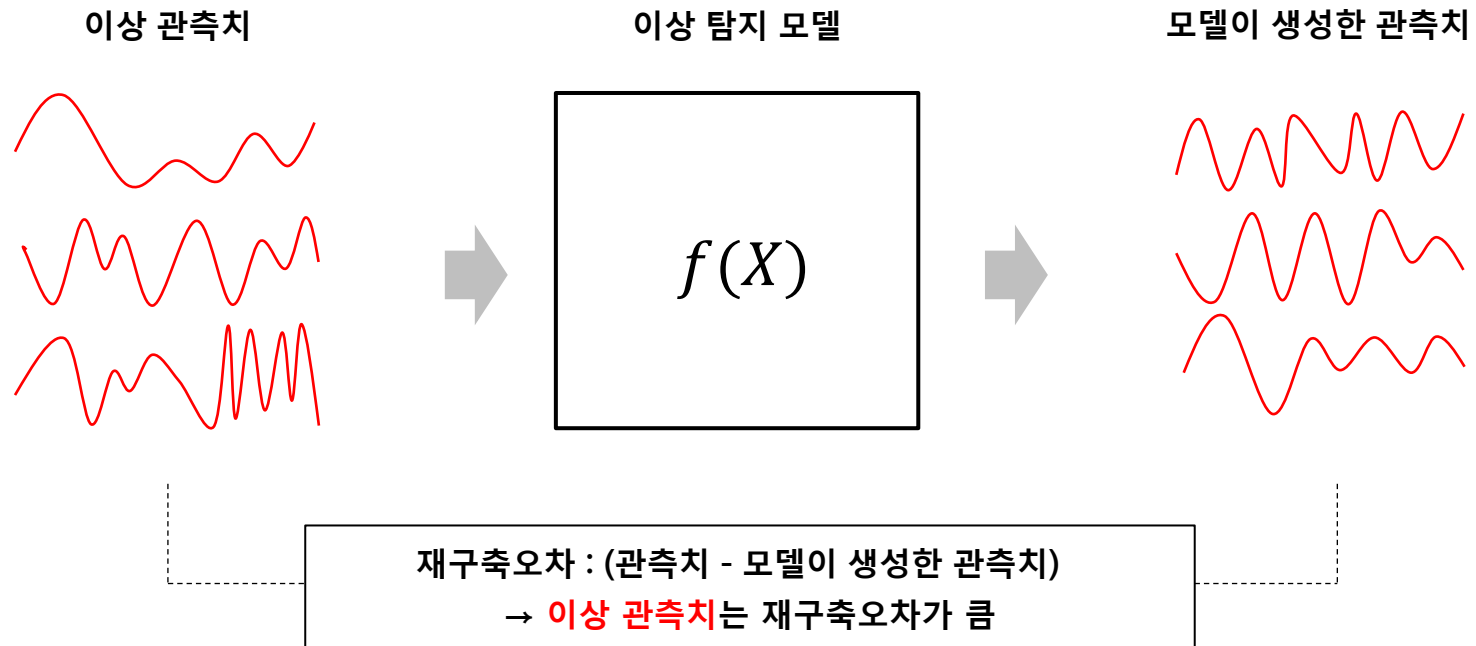


이상 탐지

재구축 기반 모델

❖ 이상 탐지 과정

- 이상 데이터 입력시 → 재구축 오차가 크기 때문에 이상으로 분류



이상 탐지

이상 탐지 결과

❖ 이상 탐지 모델 성능 확인

- 이상 검출 정확도를 기반으로 모델의 성능을 평가
- **CAE 모델이 이상 검출 정확도 90%**로 우수한 성능을 보임

CAE		예측값		이상 검출 정확도
		정상	이상	
실제값	정상	437	55	90.00%
	이상	12	108	

$$\text{※ 이상 검출 정확도} = \frac{D}{C+D}$$

실제 이상치를 모델이 이상으로 예측한 비율

PCA: 81%

CAE: 90% ✓

PCA: Principal Component Analysis

CAE: Convolutional Autoencoder

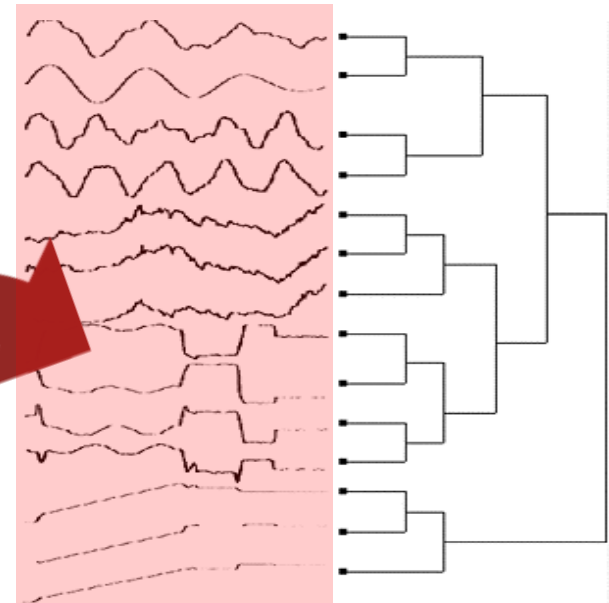
이상 탐지

구간별 이상 탐지 결과

- ❖ 구간별로 이상 관측치를 정의
- ❖ 필터링된 이상 관측치를 군집화하여 이상 패턴을 탐색하고자 함

	구간별 정상 / 이상 관측치 수	
구간	정상	이상
1	63,646	11,232
2	26,878	4,743
3	37,760	6,664
4	19,831	3,500

이상 탐지 결과



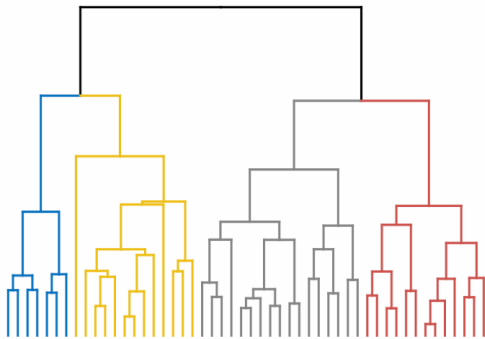
이상 관측치 군집화

이상 유형 군집화

군집 알고리즘 비교 분석

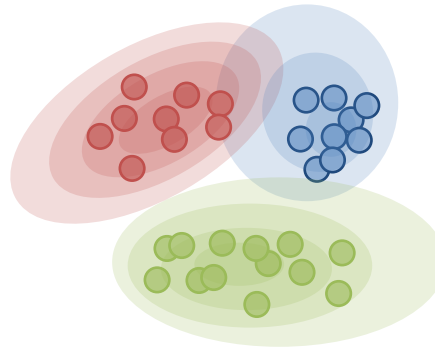
❖ 군집 알고리즘

- 패턴이 유사한 관측치들을 군집으로 묶는 과정
- Hierarchical, Gaussian Mixture, K-means Clustering 비교 실험
- 모델 검증 결과 **Hierarchical Clustering**이 가장 우수한 성능을 보임



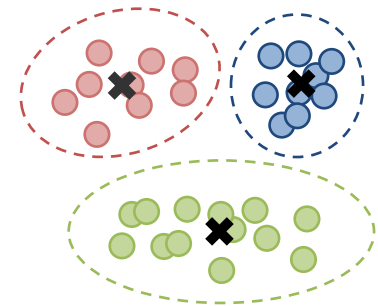
Hierarchical Clustering

- 거리가 가까운 관측치들을 묶어가면서 군집 형성



Gaussian Mixture Clustering

- 각 군집이 정규분포를 따르도록 관측치를 군집에 할당



K-means Clustering

- 군집의 중심과 거리를 계산하여 가까운 관측치들을 각 군집으로 할당

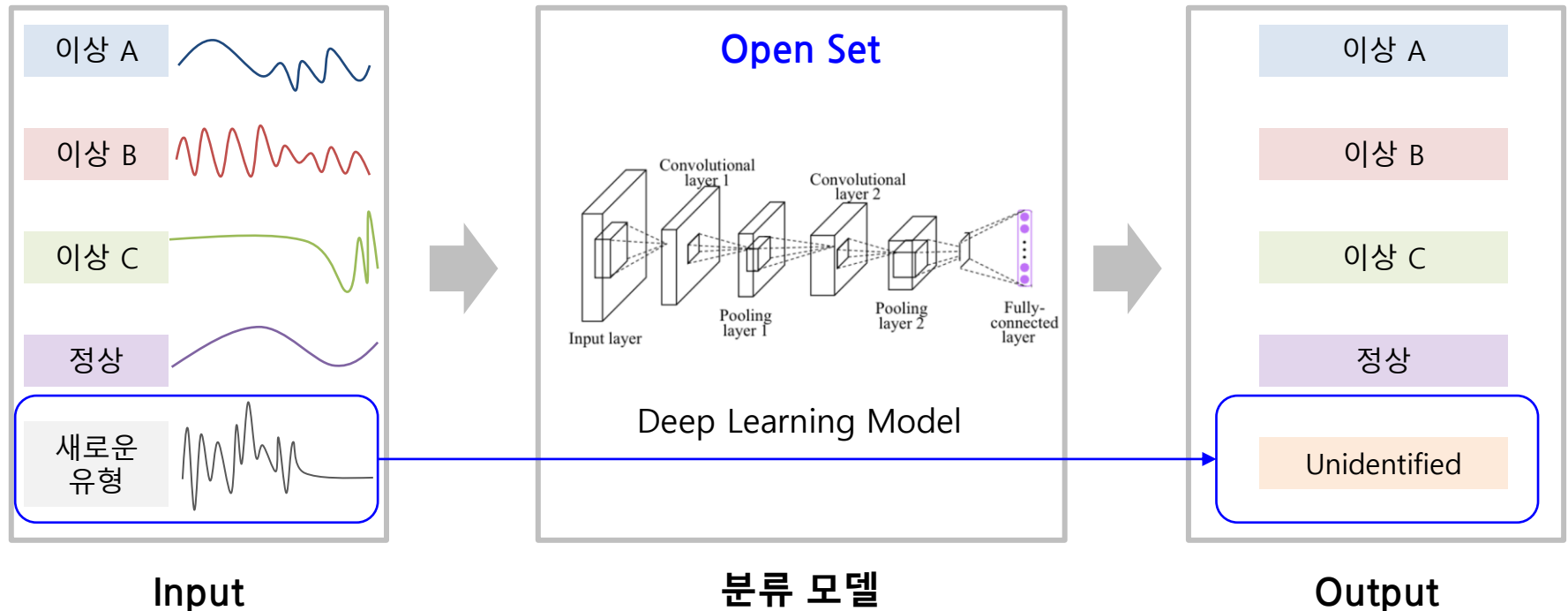
분류 모델 구축

Open Set Classification

❖ Open Set 분류 모델 적용

- Unidentified는 너무 많은 패턴이 섞여있어 하나의 클래스로 지정할 수 없음
- 테스트 단계에서 A B, C, 정상 이외의 유형이 발생하였을 때 **Unidentified**로 분류

*모델 테스트 단계



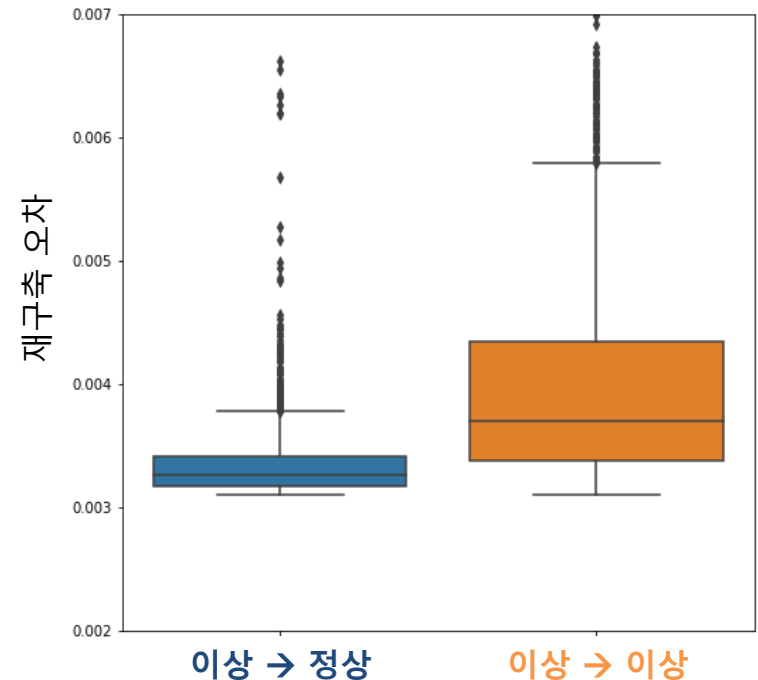
분류 모델 구축

사후분석

❖ 이상 관측치의 분류 결과에 대한 사후분석 수행

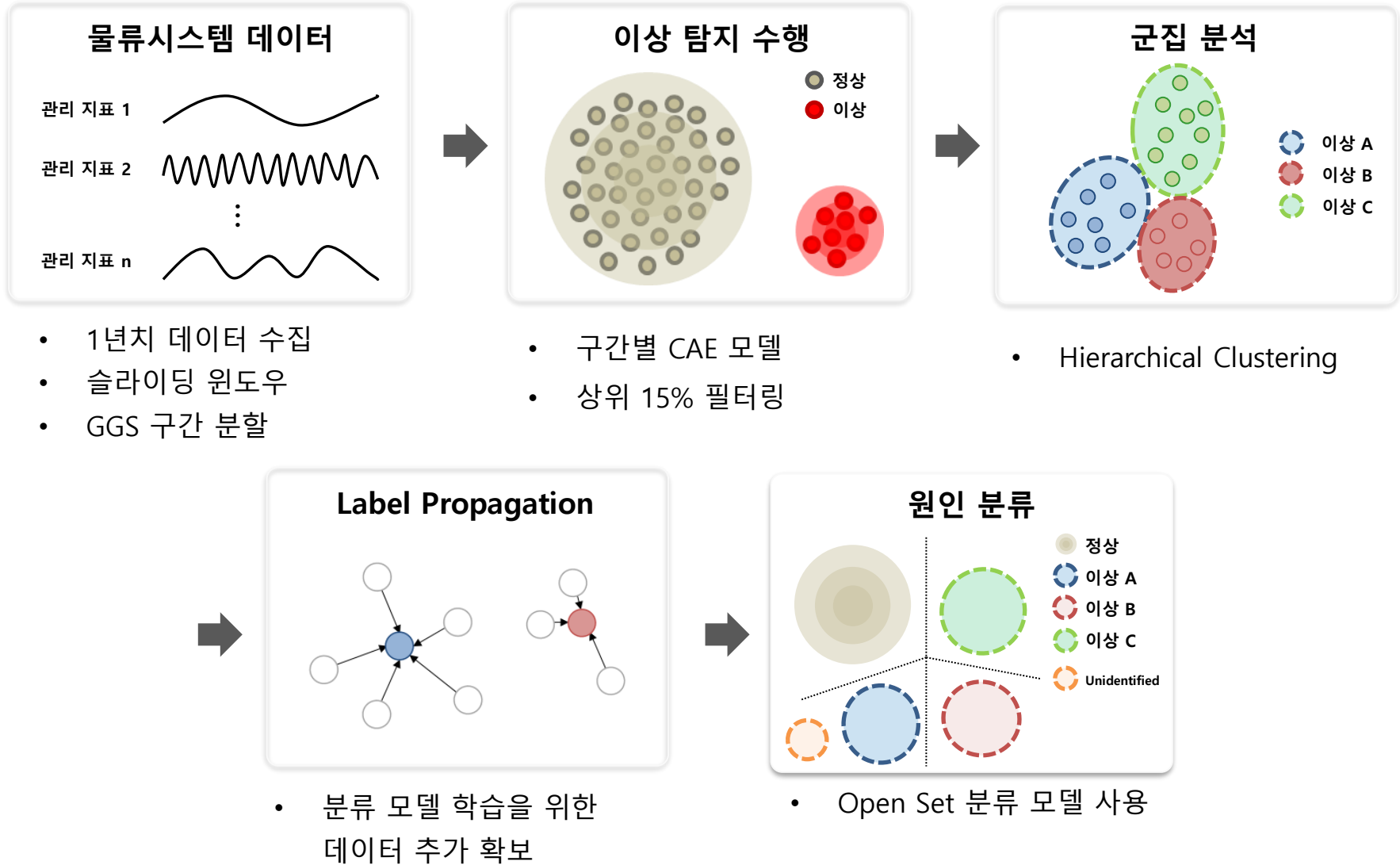
- 이상임에도 불구하고 정상으로 분류된 관측치들이 다수 존재
- 이상 → 정상으로 분류된 관측치의 재구축 오차는 상대적으로 작은 것을 Boxplot을 통해 확인 (이상 → 정상 vs 이상 → 이상)

		예측				
		A	B	C	정상	Unidentified
실제	A	89	0	0	57	32
	B	5	34	0	8	2
	C	0	0	14	3	11
	정상	23	1	1	7832	204
	Unidentified	645	146	50	2142	899



결론

전체 요약



결론

Contribution 및 기술 활용 방안

❖ Contribution

- 부족한 학습 데이터 확보를 위한 레이블링 기술 확보
 - ✓ 소수 관측치에 대한 전문가의 확인만으로도 **모델을 학습시키기에 충분한 이상 유형 레이블 확보**할 수 있는 방법론 개발
- 다변량 시계열 지표의 변화를 종합적으로 고려할 수 있는 딥러닝 기반 분류 모델 구축
 - ✓ GGS 구간 분리 알고리즘을 활용하여 **물류 생산환경 변화를 고려한 이상 탐지** 수행
 - ✓ 시스템 이상 발생 시 **이상 유형별 분류를 수행**하여 현장 전문가의 신속한 대응을 가능하게 함
 - ✓ 전문가가 경험하지 못한 **Unidentified 유형까지 분류**하여 문제상황 인지에 도움

❖ 기술 활용 방안

- Unidentified 이상 유형에 대한 전문가 분석 후 Label Propagation 기반 **새로운 이상 유형에 대한 레이블 추가 확보**
- 물류 관리지표에 대한 통합 모니터링 시스템 구축

결론

향후 계획

❖ 향후 연구 계획

- Unidentified 데이터의 특성 분석
 - Unidentified 데이터 특성을 분석하여 전문가에 제공
 - 전문가가 불량 유형을 파악하는데 도움
- 물류 클러스터 통합 모니터링 시스템
 - 분석 대상을 확장하여 물류 클러스터 내 모든 관리지표의 변화를 분석할 수 있는 통합관리체계 구축