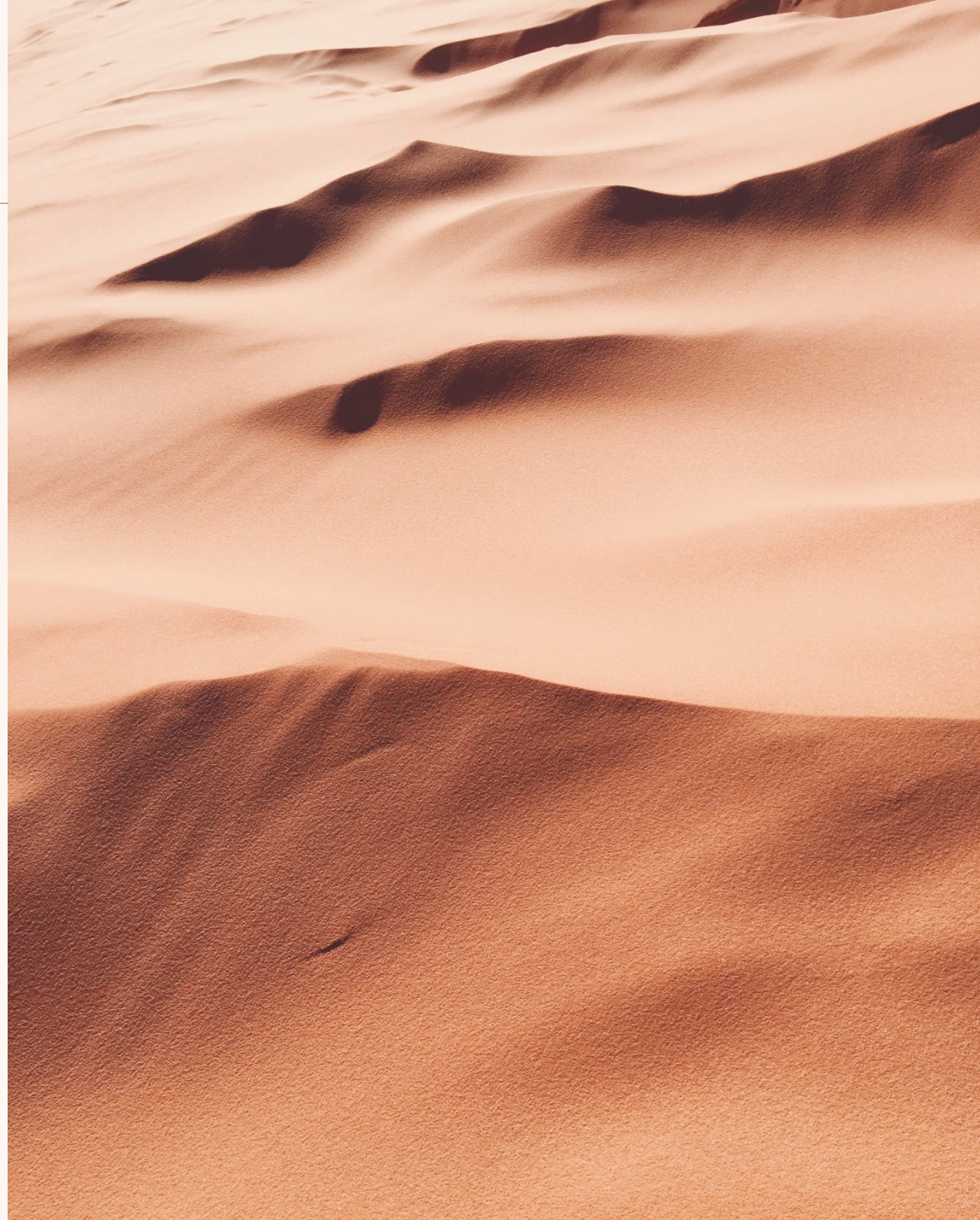


시계열 분석 기반 고장시기 예측 및 공정별 이상탐지 모델

목차 A table of contents.

- 1 문제 정의
- 2 해결 방안
- 3 분석 모델 소개
- 4 분석 모델 적용
- 5 기대효과 및 확장



[1. 문제 정의]

1-1. 이슈 사항

문제 현황

- 열처리 내부는 약 800°C로 내부의 상태를 확인하기 어렵다.
- 200°C이하로 내려 갔을때에만 소모품 교체를 할 수 있다.
- 온도를 낮추기 위해서는 하루 이상의 가동을 멈춰야한다.

열처리 설비는 빈도수가 낮지만
고장 발생 시 손실이 높은 설비이기 때문에
특히 중소기업에서의 예지보전은 중요하다.

2. 해결 방안

2-1. 해결 방안



시계열 분석을 이용하여 설비의 고장시기를 예측

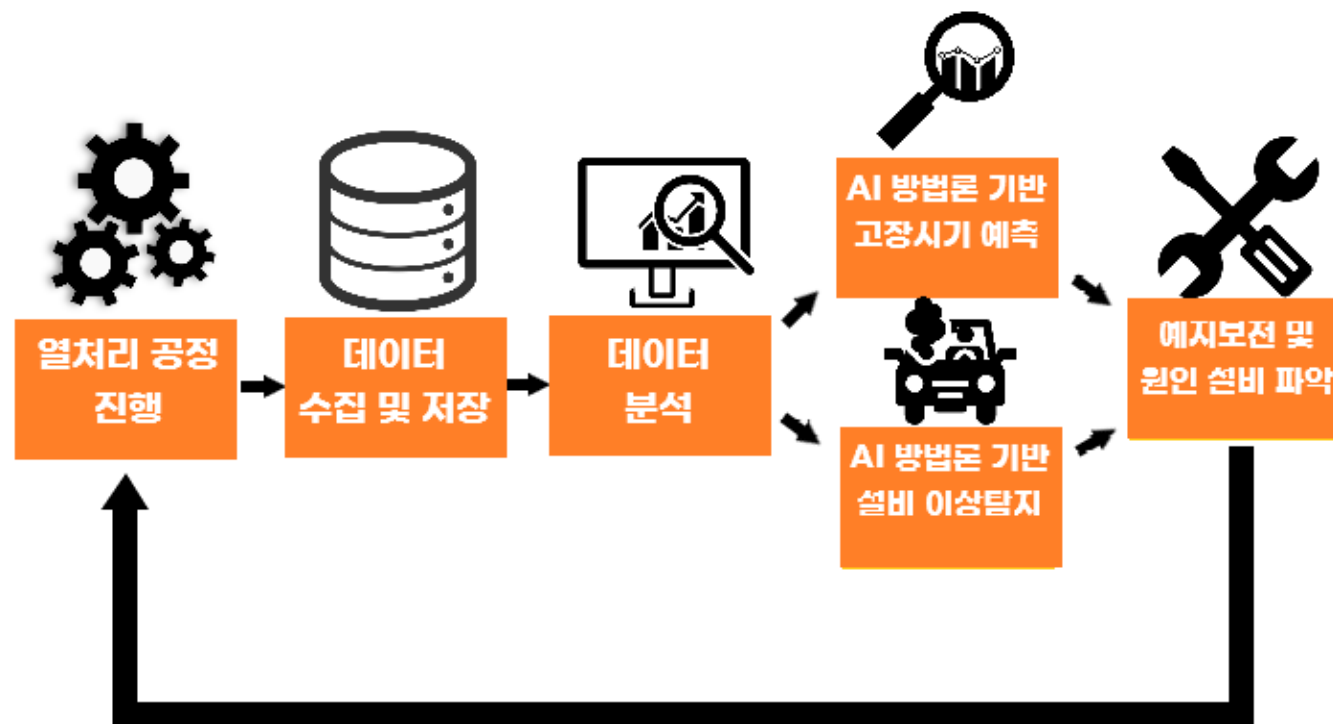


각 공정별 이상을 탐지



예지보전 및 원인 설비 파악에 기여

2-2. 데이터 흐름도



[3. 분석 모델 소개]

3-1 ARIMA

What's ARIMA?

ARIMA(Autoregressive Integrated Moving Average)모델은 시계열 데이터 분석과 예측을 위한 통계적 방법 중 하나이다. 시계열 데이터의 패턴과 경향을 이해하고 미래 값을 예측하는데 사용된다.

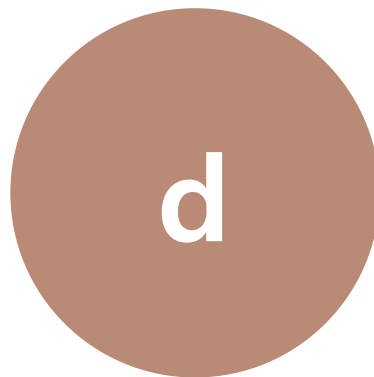
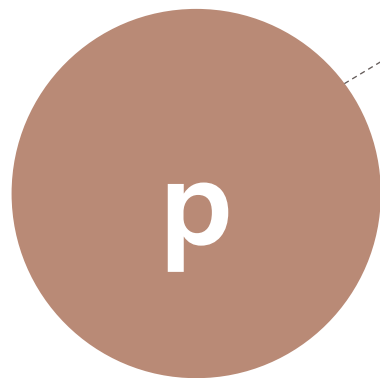


3-1 ARIMA(p, d, q)

AR(Autoregressive)

AR은 현재 값이 이전 값들의 선형 조합으로 설명될 수 있다는 것을 기반으로 한다.

AR 차수(p)는 몇 개의 이전 예측 오차를 사용할지를 나타낸다.



차분(Integrated)

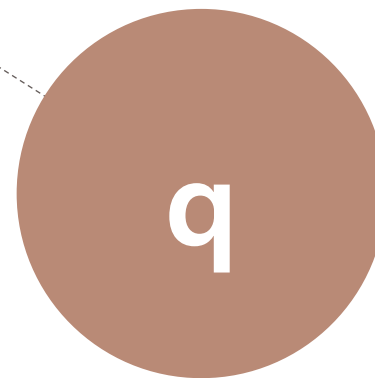
차분은 시계열 데이터의 불규칙성, 계절성, 추세 등을 제거하기 위한 것이다.

이를 통해 정상성을 갖는 시계열 데이터로 변환된다. I(d)는 몇 번의 차분을 수행할지를 나타낸다.

MA(Moving Average)

MA는 현재 값이 이전 예측 오차들의 선형 조합으로 설명될 수 있다는 것을 기반으로 한다.

MA차수(q)는 몇 개의 이전 예측 오차를 사용할지를 나타낸다.

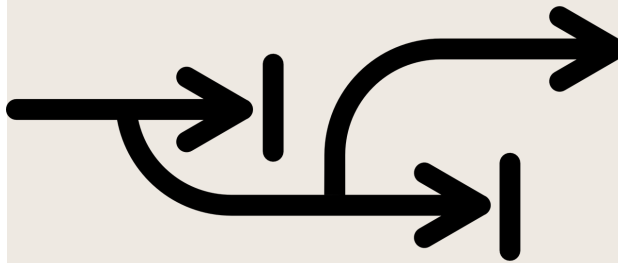


3-1 Why ARIMA?



단순성

비교적 간단한 구조로 복잡한 모델보다 이해하고 구현하기 쉬우며, 초기 모델링 단계에서 시계열 데이터를 빠르게 분석할 수 있다.



유연성

다양한 시계열 데이터 패턴에 대응할 수 있다.
이전의 데이터 분석 경험을 필요로 하지 않고,
데이터 경향, 계절성, 불규칙성 등을 잘 다룰 수 있다.



과거 데이터 활용

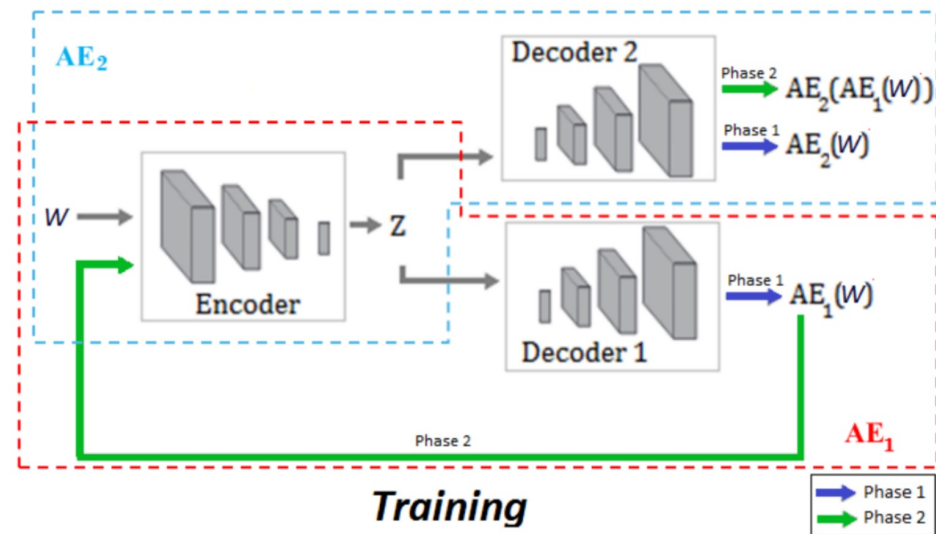
과거 시계열 데이터를 기반으로 미래 값을 예측한다.
데이터의 정보를 최대한 활용하며, 데이터 기반의 패턴 및 동향 파악에 도움을 준다.

3-2 USAD

What's USAD?

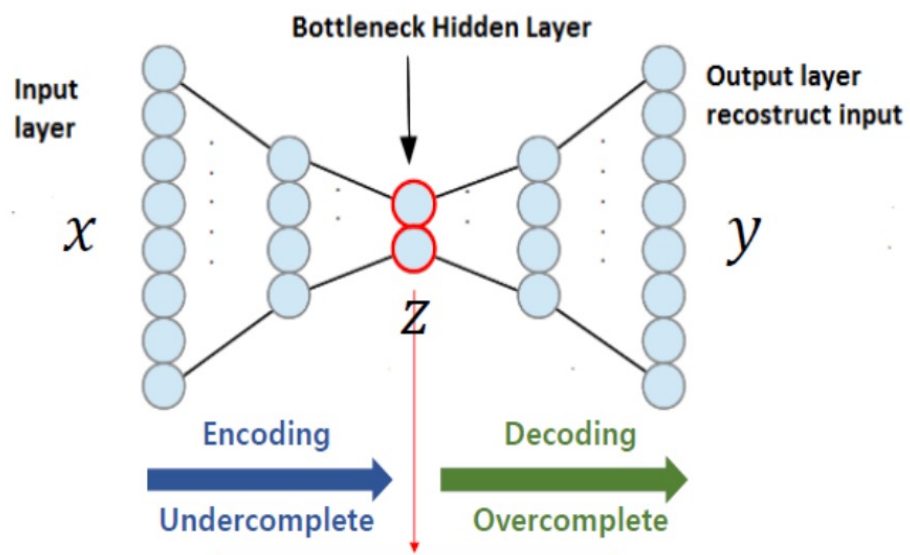
UnSupervised Anomaly Detection on Multivariate Time Series
AE 알고리즘과 GAN기반의 알고리즘을 융합하여 서로의 장단점을 보완한 알고리즘이다.

GAN모델로 비정상 데이터를 발생시켜 AE모델은 점점 정교한 비정상 데이터를 구분할 수 있게되고 정확도 높은 Anomaly Detection을 수행할 수 있다.



3-2 USAD Auto Encoder

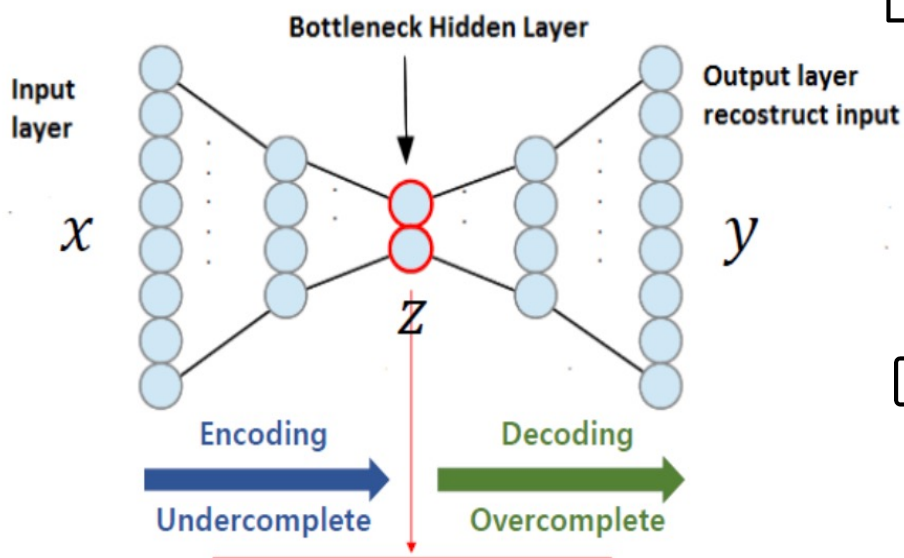
What's Auto Encoder?



- Encoding 과 Decoding으로 데이터의 압축과 복원을 수행함.
- 압축과 복원 과정을 통하여 정상데이터를 복원할 수 있도록 학습함.
- 이렇게 학습된 모델은 정상데이터일 경우 정상적으로 복원할 수 있음.
- 비정상 데이터일 경우 올바르게 복원되지 않아 Reconstruction Error가 커짐.

3-2 USAD Auto Encoder

Auto Encoder Advantage & Disadvantage



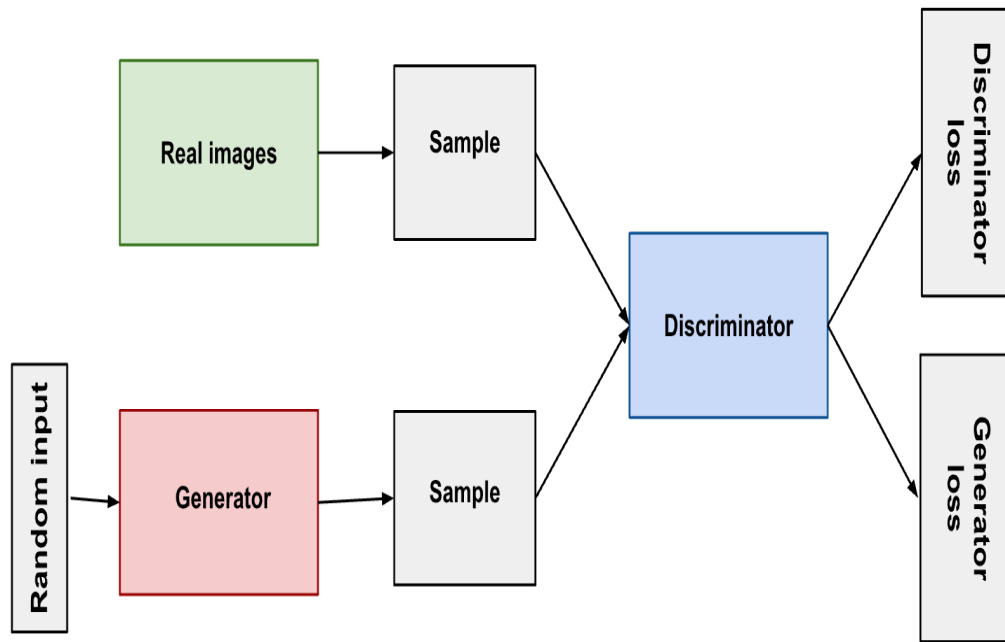
- Unsupervised Learning 방식으로 라벨링이 되어있지 않은 데이터로 학습가능.
- Curse of Dimensionality 문제를 예방 가능함.
- Manifold learning으로 중요한 Feature를 찾을 수 있음.



- 정상 데이터와 비슷한 배열의 이상 데이터 판별에 어려움

3-2 USAD GANs

What's GANs?

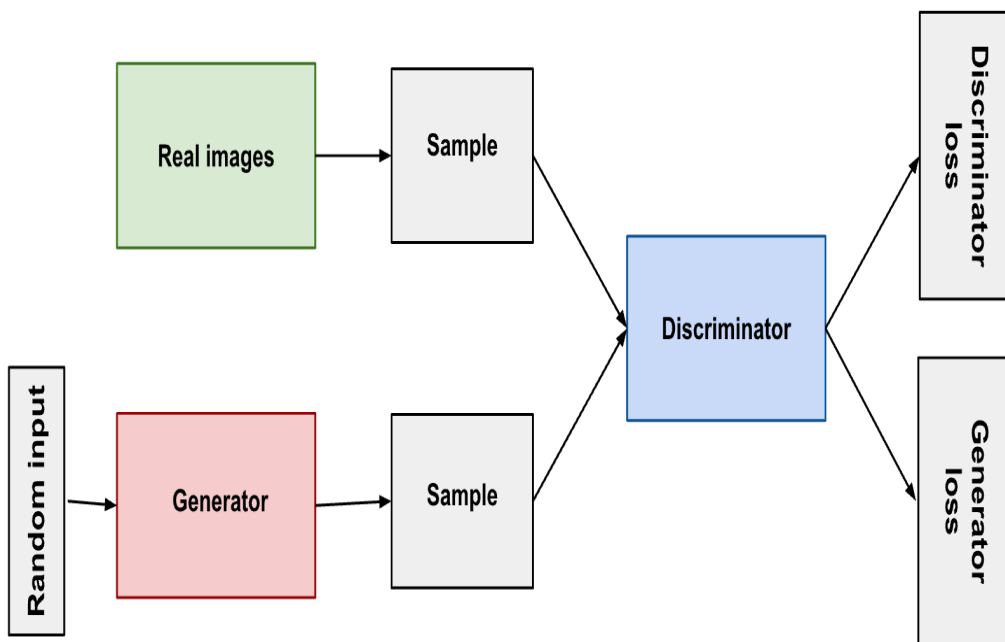


- Generative Adversarial Networks 의 약자로 적대적 생성 모델이다.
- Generator 와 Discriminator로 구성되어 있다.
- Generator는 Fake Data를 생성.
- Discriminator 는 Real Data 와 Fake Data를 구분한다.
- Generator는 Real Data와 같은 정교한 Fake Data를 생성하는 방향으로 학습.
- Discriminator는 Real Data와 정교한 Fake Data를 구분하도록 학습.

3-2 USAD GANs

GANs

Advantage & Disadvantage



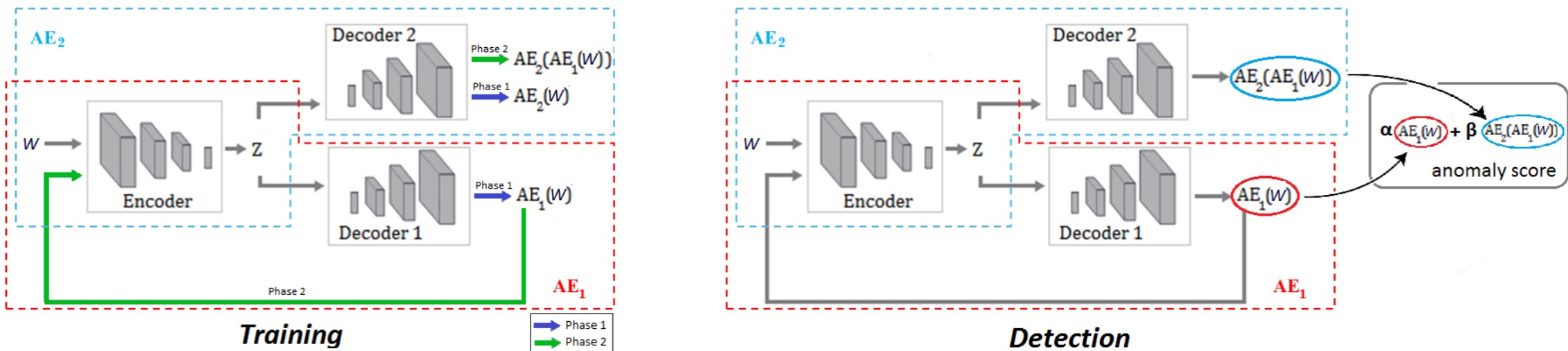
- Unsupervised Learning 방식으로 데이터를 학습하기 위한 라벨링이 필요 없음.
- 분류 모델의 효율적인 학습데이터가 적어도, 생성 모델을 활용하여 보충할 수 있음.



- 학습이 불안정하다.

3-2 USAD

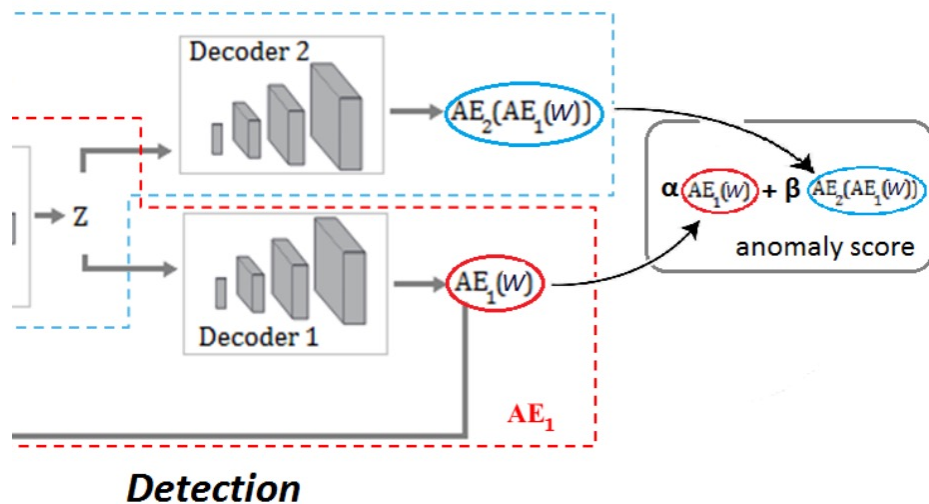
What's USAD?



- Auto Encoder 모델과 GANs 모델의 융합으로 단점을 보완한 모델.
- Adversarial Train 과정으로 Generator AE1와 Discriminator AE2 모델 학습.
- 이 과정으로 정상 데이터 배열의 이상 데이터 탐지 가능.

3-2 USAD

Why USAD?



- 두개의 AE 모델을 거쳐 나온 데이터를 비교하여 Anomaly Score 측정.
- Alpha 와 Beta 값의 세팅으로 Anomaly Data에 대한 민감도 설정가능.
- 기존 AE 보다 우수한 판별성능.

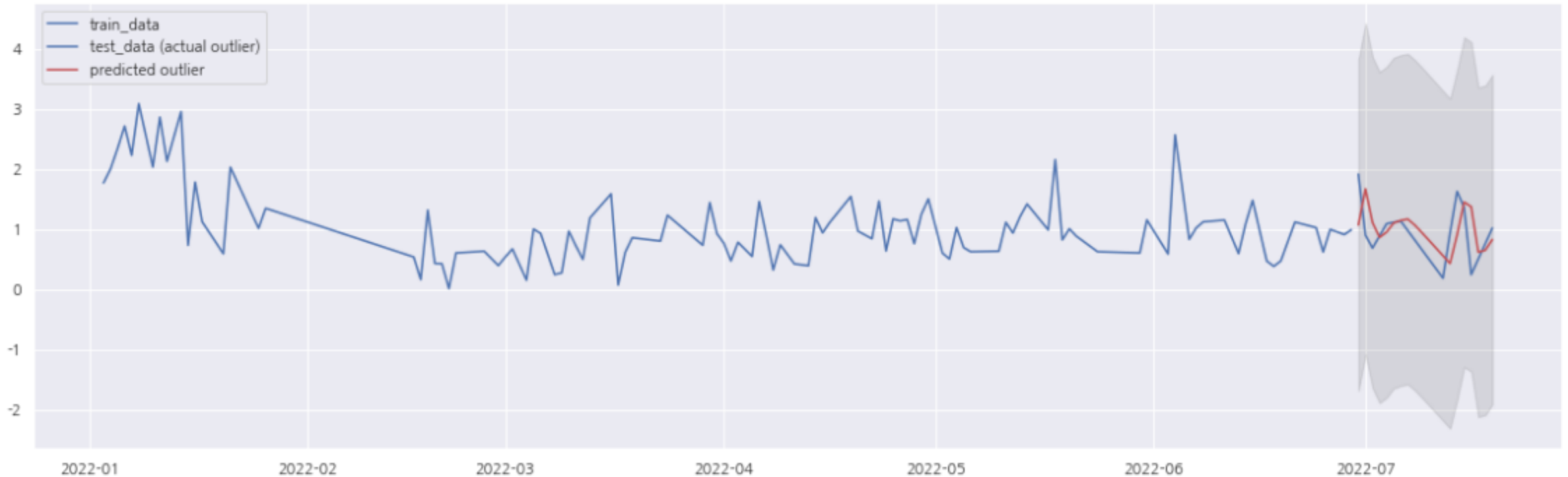
Parameter setting	# of detection	Detection sensitivity
$\alpha > \beta$	Reduce	Low
$\alpha < \beta$	Increase	High

[4. 분석 모델 적용]

4-1. ARIMA 모델 평가

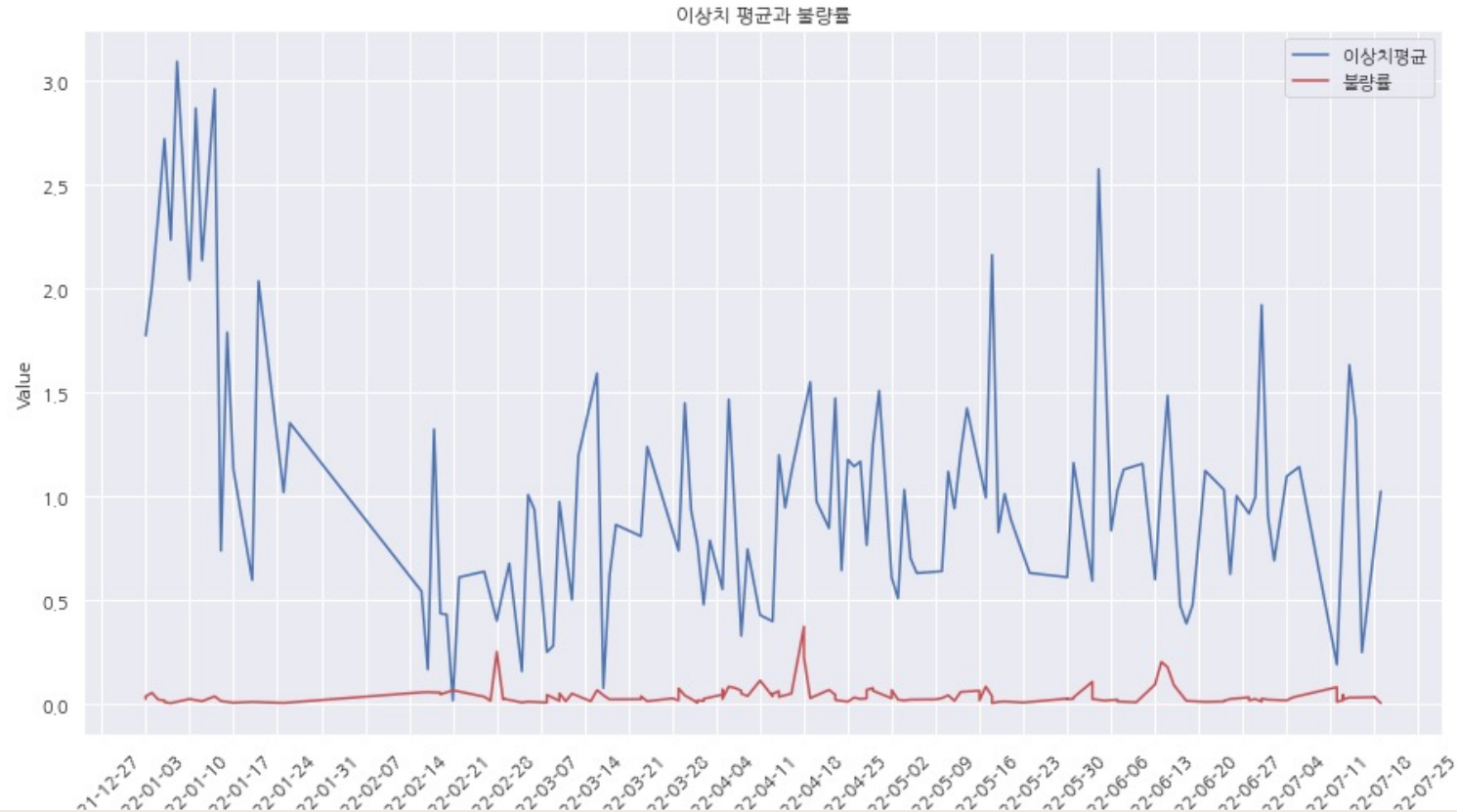
평가 지표	MSE	MAE	RMSE
의미	오차의 제곱을 모두 더한 후 평균을 내는 값	절대값을 모두 더한 후 평균을 내는 값	MSE의 제곱근으로 오차의 표준편차를 나타내는 값
식	$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$	$MAE = \frac{1}{n} \sum_{i=1}^n x_i - x $	$RMSE = \sqrt{MSE} = \sqrt{\frac{\sum(\hat{y}-y)^2}{n}}$
결과	1.9672484874046297	0.9913108344012007	1.4025863564874106

4-2. ARIMA 모델 시사점



파란색이 **실제데이터** 이고, 빨간색이 **예측**한 결과
->자기회귀를 통해 이상치 평균을 예상할 수 있다.

4-2. ARIMA 모델 시사점



이상치의 평균이 높으면 불량률 발생 확률이 높다.
->모델로 이상치평균을 확인하여 예지보전 가능

4-2. ARIMA 모델 시사점



향후 이상치의 평균을 예측하게 되면, 증가하게 되는 것을 알 수 있다.
이후에도 계속해서 실제의 데이터를 수집하여 사용한다면, 더 확실하게
이상치의 평균을 예측할 수 있을 것이다.

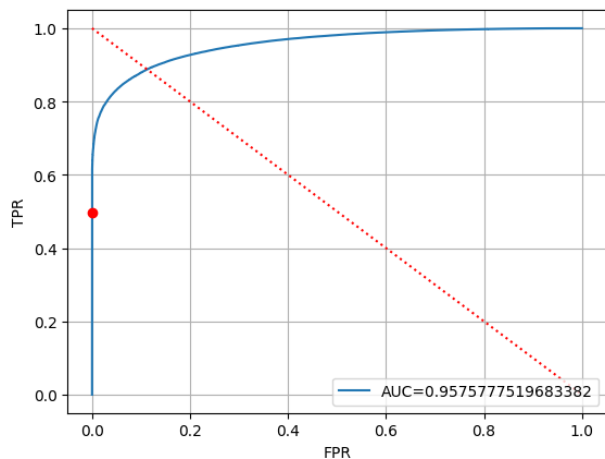
4-1. USAD 모델 평가

공정	건조	세정	소입	솔트
F1-Score	0.6643	0.6654	0.6660	0.6660
AUC	95.75%	99.99%	93.32%	90.85%

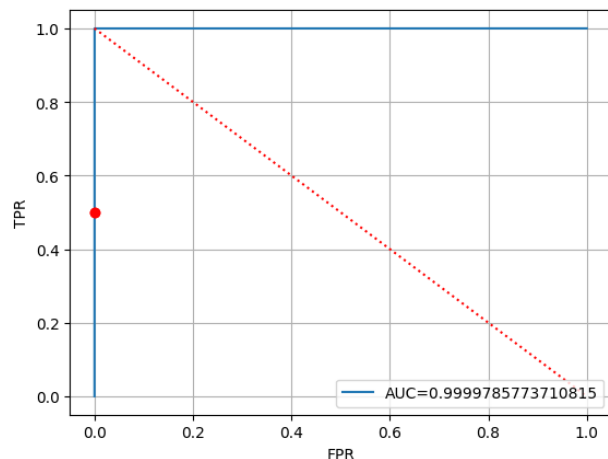
F1-Score : 정밀도와 재현율의 조화 평균.
Area Under Curve(AUC): 모델의 정확도 판단지표.

4-1. USAD 모델 평가

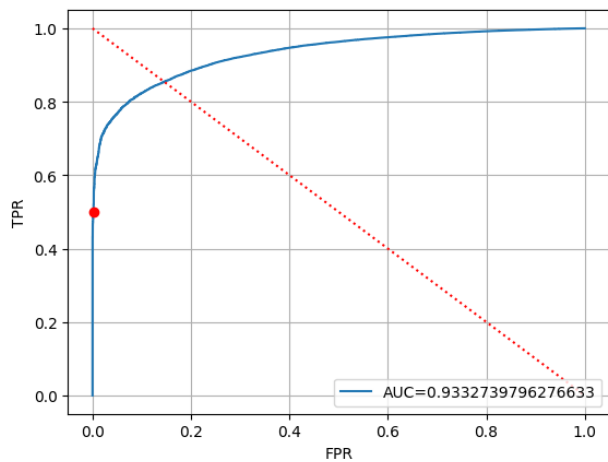
Drying



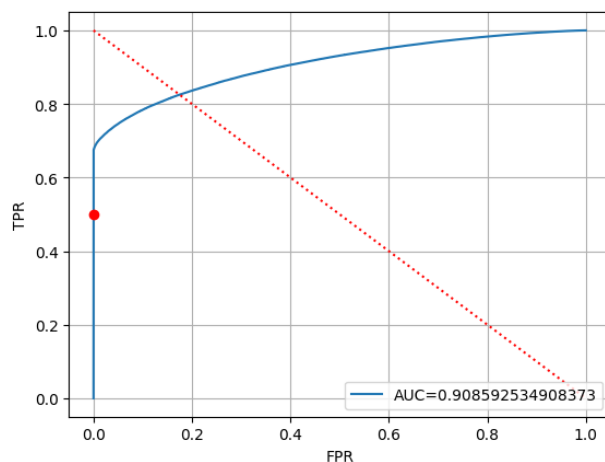
Cleanin



Quenching



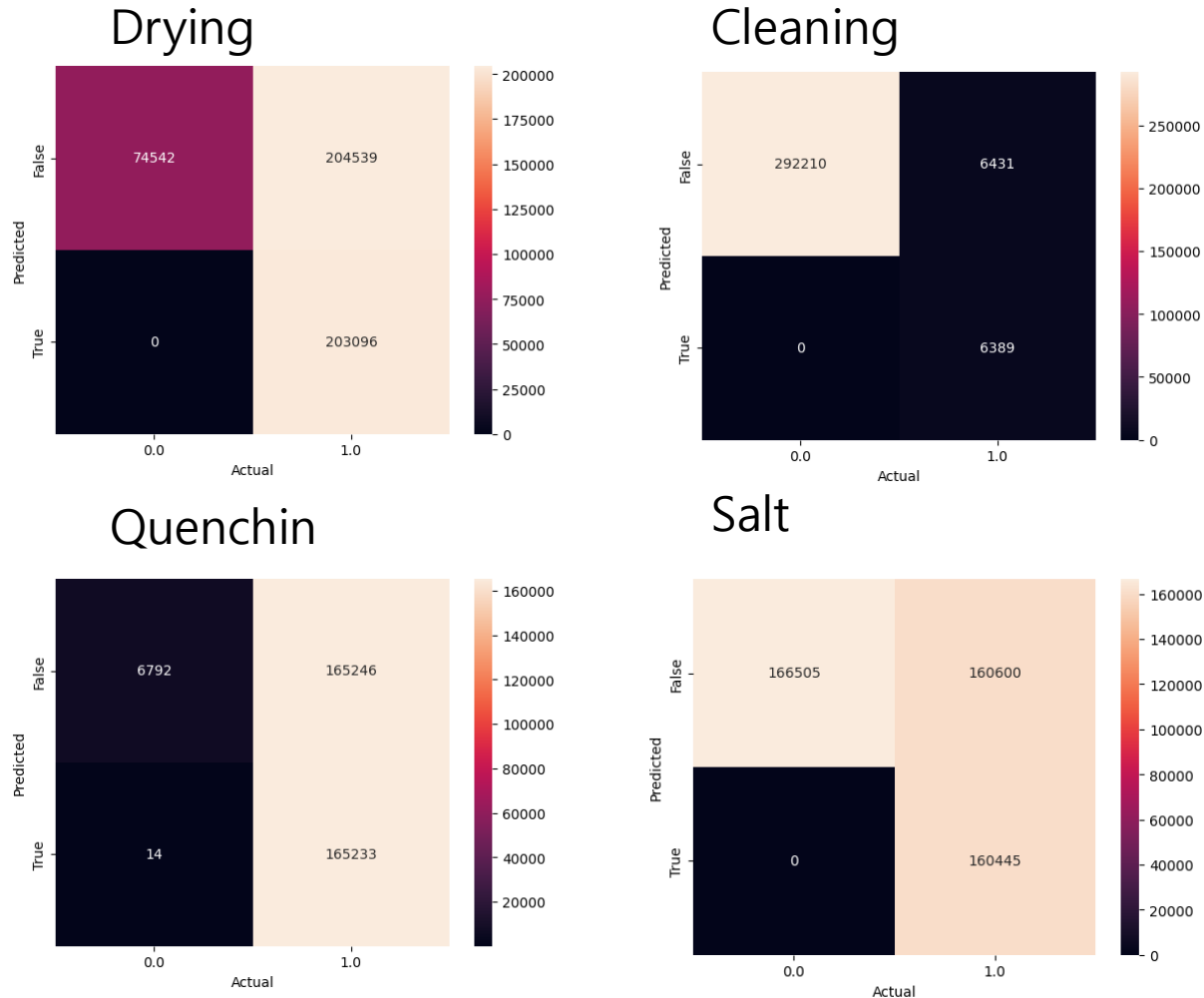
Salt



- ROC Curve
- Anomaly Score의 Threshold 측정.
- Threshold에 대한 정확도 측정.

Area Under Curve(AUC)	Evaluation
$AUC \geq 0.9$	Excellent
$0.9 > AUC \geq 0.8$	Good
$0.8 > AUC \geq 0.7$	Fair
$AUC < 0.7$	Poor

4-1. USAD 모델 평가



- 각 공정에 대한 Confusion Matrix
- Alpha 와 Beta 값의 조절
- > 더 높은 정밀도가 예상됨.

[5. 기대효과 및 확장]

5-1. 기대효과



불안감 해소

설비의 고장시기 예측으로 갑작스러운 고장에 대한 불안감 해소



생산성 향상

고장시점 인지로 생산스케줄의
유동적인 조절로 기대되는 생산
성 향상



예지보전 및 비용감소

공정의 이상탐지 모델로 예의주시
해야하는 설비를 알려줌으로써
예지보전 및 비용감소

5-2. 확장: 제조업 공정의 공통적인 문제점

1

공정의 중단으로 인한 손실

2

예기치 못한 설비의 고장

3

제조 제품의 불량률

5-2. 확장: 타 제조현장으로의 확장

해결 방안

- 3가지 공통적인 문제점을 해결하기 위해 타 제조현장에도 본 분석을 현장에 맞게 적용한다.

- 공정 운영 데이터를 수집하여 전산화하여 관리하고 있다면 본 분석은 적용 가능하다.
- 열처리 공정 외에도 시계열로 데이터 수집이 이루어지는 경우 적용할 수 있다.
- 본 분석에서는 설비 데이터를 바탕으로 이상치를 찾아, 고장시기의 예측과 이상탐질지를 하는 모델을 구축하였다. 계속해서 데이터를 수집하여 적용한다면, 더 좋은 모델로 개선할 수 있을 것으로 기대된다.

[

QnA

]

[감사합니다]