

제3회 K-인공지능 제조데이터 분석 경진대회 보고서

프로젝트명	시계열 분석 기반 고장시기 예측 및 공정별 이상탐지 모델
팀명	Hustar
내용요약	<p>열처리 공정은 금속 재료를 사용 목적에 따라 소재를 가열하고, 냉각하여 금속의 구조와 성질을 변화시키는 작업으로 금속의 경도, 내마모성, 가공성, 자성등의 특성을 얻기위하여 수행하는 공정이다. 본 경진대회에서 제공된 데이터의 열처리 공정은 Austempering 을 목적으로 설계된 열처리 공정이다.</p> <p>열처리 내부는 800℃의 고온으로 내부의 상태를 확인하고 파악하기 어렵다. 소모품 교체시 내부의 온도가 200℃ 이하로 내려갔을때에만 가능하며 내부의 온도를 낮추기 위해서는 하루이상의 가동이 정지되어야한다. 제조업 공정에서 설비를 멈춘다는 것은 큰손일이다. 그렇기때문에 설비를 자주 멈출 수 없을 뿐더러 설비의 중단에 신중할 필요가 있다.</p> <p>설비의 고장시기 예측 모델의 개발로 갑작스러운 고장사태에 대한 불안감을 해소하고 대략적인 고장시점의 인지로 생산스케줄의 유동적인 조절로 공정의 중단으로 인한 손실을 최소화하고, 공정별 이상탐지 모델의 개발로 예의주시 해야하는 공정을 알려줌으로써 예지보전 및 수리에 소모되는 인력 및 에너지 등의 비용을 감소하고자 한다.</p> <p>고장시기 예측 모델은 ARIMA(Autoregressive Intergrated Moving Average) 방식을 이용한 시계열 데이터 분석과 예측을 위한 통계적 방법중 하나를 이용하였다. ARIMA 모델은 이렇게 AR(AutoRegressive), MA(MoingAverage), Intergrated을 결합하여 시계열 데이터를 모델링하고 예측하는 모델이다. ARIMA(p, d, q)와 같이 표현되며, 각 부분의 차수는 데이터에 적합하게 선택되어야 한다. 모델 학습 후에는 미래 값을 예측할 수 있으며, 시계열 데이터의 패턴을 분석하고 예측하는데 사용되는 모델을 사용하였다.</p> <p>공정별 이상탐지 모델은 USAD(UnSpervised Anomaly Detection on Multivariate TimeSeries) 라는 정통적인 시계열 이상탐지 방법인 AutoEncoder 방식과 생성적 모델 방법인 GANs 방법을 융합하여 두 모델의 장점을 결합하여 조금 더 정밀한 데이터의 이상탐지를 수행할 수 있는 모델이다. Adversarial Train을 적용하여 모델이 Real Data와 Fake Data를 좀더 정밀하게 구분할 수 있도록 서로가 서로의 AutoEncoder를 훈련시키는 모델을 사용하였다.</p>
<p>상기 본인(팀)은 위의 내용과 같이 제3회 K-인공지능 제조데이터 분석 경진대회 결과 보고서를 제출합니다.</p> <p>2023 년 11 월 3 일</p> <p>팀장 : 이 병 화 (서명)</p> <p>팀원 : 김 재 희 (서명)</p> <p>팀원 : (서명)</p> <p>한국과학기술원장 귀중</p>	

□ 문제정의

○ 분석 배경

- 공정 개요

* 열처리 공정은 금속 재료를 사용 목적에 따라 소재를 가열하고, 냉각하여 금속의 구조와 성질을 변화시키는 작업으로 금속의 경도, 내마모성, 가공성, 자성등의 특성을 얻기위하여 수행하는 공정이다.

* 본 경진대회에서 제공하는 데이터의 열처리 공정은 Austempering을 목적으로 설계된 열처리 공정으로 소재에 기본적으로 주어진 경도에 대하여 더 높은 강도 및 인성, 연성을 얻고, 충격에 대한 저항과 얇은 소재의 경우 왜곡을 방지하기 위하여 수행하는 공정이다.

* Austempering 공정은 Austenizing - Quenching - Cooling - Tempering 으로 이루어진다. 기존 열처리 공정과의 차이점은 담금질 및 템퍼링 사이에 장시간 동안 담금질 온도에서 공작물을 유지하는 작업이 포함된다.

* 기존 열처리 공정에 비해 제조 및 성능상 이점이 있고, 비용적으로도 많은 절감이 이루어졌다.

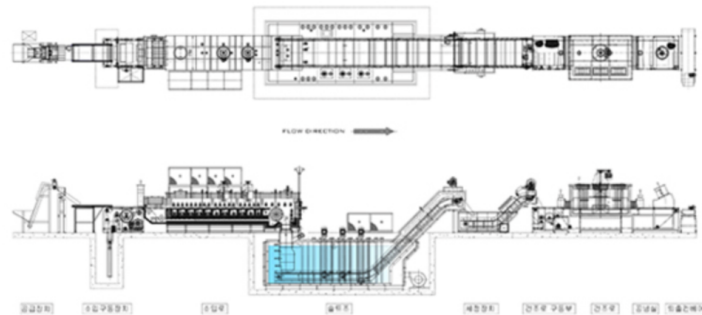


Figure 3 열처리 공정도

- 이슈 사항

* 문제현황

- 열처리 내부는 800℃의 고온으로 내부의 상태를 확인하고 파악하기 어렵다. 소모품 교체시 내부의 온도가 200℃ 이하로 내려갔을때에만 가능하며 내부의 온도를 낮추기 위해서는 하루이상의 가동이 정지되어야한다.

- 제조업 공정에서 설비를 멈춘다는 것은 큰손일이다. 그렇기때문에 설비를 자주 멈출 수 없을 뿐더러 설비의 중단에 신중할 필요가 있다.

* 극복방안

- 시계열 데이터의 트렌드와 자기회귀를 이용하여 설비의 고장시기를 예측하고, 이상탐지에 활용되는 전통적인 모델인 Auto Encoder 기반의 모델을 사용하여 각 공정별 이상을 탐지하여 비단 열처리 공정만의 문제점이 아닌 모든 제조업에서 공통적으로 가지고 있는 고질적인 문제를 해결하고 솔루션을 제공하고자 한다.

○ 분석 목표

- 분석목표

* 시계열 데이터 분석으로 고장시기를 예측하는 모델과 공정별 이상탐지 모델의 개발을 목표를 한다.

- 기대효과

- * 설비의 고장시기의 예측으로 갑작스러운 고장사태에 대한 불안감 해소 및 안정성 증감.
- * 대략적인 고장시점 인지로 생산스케줄의 유동적인 조절로 기대되는 생산성 향상
- * 공정의 이상탐지 모델로 예의주시해야하는 공정을 알려줌으로써 예지보전 및 수리에 소모되는 인력 및 에너지 등의 비용감소.

□ 제조데이터 정의 및 처리과정

○ 제조 데이터 소개

- 데이터 유형 / 구조

* 데이터셋에는 일자, 시간 배치번호 및 다양한 공정 변수 정보가 포함되며, 주요 변수는 <Figure 2>과 같다. 종속변수인 ‘이상치개수’ 이상치가 발생한 변수의 개수를 더하여 생성한 변수다.

속성	설명	비고
TAG_MIN	데이터가 수집된 시간	데이터형: datetime
배정번호	공정의 작업 지시 번호-배정번호별로 생산	데이터형: int
건조 1~2존 OP	각 건조 온도 유지를 위한 출력 량(%)	데이터형: float
건조로 온도 1~2 Zone	각 건조로 Zone의 온도 값	데이터형: float
세정기	세정기 온도 값	데이터형: float
소입1~4존 OP	각 소입존 온도 유지를 위한 출력 량(%)	데이터형: float
소입로 CP 값	침탄 가스의 침탄 능력의 량(%)	데이터형: float
소입로 온도1~4 Zone	각 소입로 Zone의 온도 값	데이터형: float
솔트 컨베이어 온도 1~2 Zone	각 솔트 컨베이어 Zone의 온도 값	데이터형: float
솔트조 온도 1~2 Zone	각 솔트조 Zone의 온도 값	데이터형: float
소입로 CP 모니터 값	모니터의 소입로 CP 값	데이터형: float
건조_error	건조 설비 이상치 발생 개수	데이터형: float
세정_error	세정 설비 이상치 발생 개수	데이터형: float
소입_error	소입 설비 이상치 발생개수	데이터형: float
솔트_error	솔트 설비 이상치 발생 개수	데이터형: float
이상치개수	이상치가 발생한 변수의 개수의 합	데이터형: int

Figure 2 데이터 속성

- 주요변수 기술통계

변수	개수	평균	표준 편차	최솟값	중앙값	최댓값
건조 1존 OP	2935040	69.89	4.01	47.25	70.51	87.29
건조 2존 OP	2935040	20.44	5.21	0.00	21.28	47.53
건조로 온도 1 Zone	2935040	100.00	0.43	97.34	100.00	102.46
건조로 온도 2 Zone	2935040	100.02	0.36	97.87	100.01	101.84
세정기	2935040	67.71	1.63	60.62	67.69	71.49
소입1존 OP	2935040	75.64	25.16	0.00	82.21	100.00
소입2존 OP	2935040	54.86	4.42	8.62	55.66	77.27
소입3존 OP	2935040	53.85	2.66	0.04	53.88	66.01
소입4존 OP	2935040	71.09	2.55	0.00	71.04	87.39
소입로 CP 값	2935040	0.44	0.01	0.00	0.45	0.90
소입로 CP 모니터 값	2935040	0.00	0.00	0.00	0.00	0.00
소입로 온도 1 Zone	2935040	859.21	3.63	840.29	859.57	877.22
소입로 온도 2 Zone	2935040	860.00	0.55	855.92	860.02	866.03
소입로 온도 3 Zone	2935040	860.00	0.35	858.28	860.00	870.11
소입로 온도 4 Zone	2935040	860.00	0.45	857.99	860.00	882.14
솔트 컨베이어 온도 1 Zone	2935040	283.99	9.51	266.23	284.59	298.53
솔트 컨베이어 온도 2 Zone	2935040	279.92	6.61	266.42	280.01	291.69
솔트조 온도 1 Zone	2935040	331.80	0.78	328.16	332.01	332.71
솔트조 온도 2 Zone	2935040	332.17	0.87	328.07	332.42	333.17

Figure 3 주요변수 기술통계

- 독립변수 / 종속변수 정의

- * 독립변수란, 다른 변수에 영향을 받지 않는 변수로, 입력 값이나 원인을 나타낸다.
- * 종속변수란 독립변수의 변화에 따라 어떻게 변하는지를 알고 싶어 하는 변수를 말하며, 결과물이나 효과를 나타낸다.

구분	명칭
독립변수	TAG_MIN
	배정번호
	건조 1~2존 OP
	건조로 온도 1~2 Zone
	세정기
	소입1~4존 OP
	소입로 CP 값
	소입로 온도1~4 Zone
	솔트 컨베이어 온도 1~2 Zone
	솔트조 온도 1~2 Zone
	소입로 CP 모니터 값
파생변수	건조_error
	세정_error
	소입_error
	솔트_error
	이상치개수

Figure 4 독립변수 종속변수

○ 제조 데이터 전처리 과정

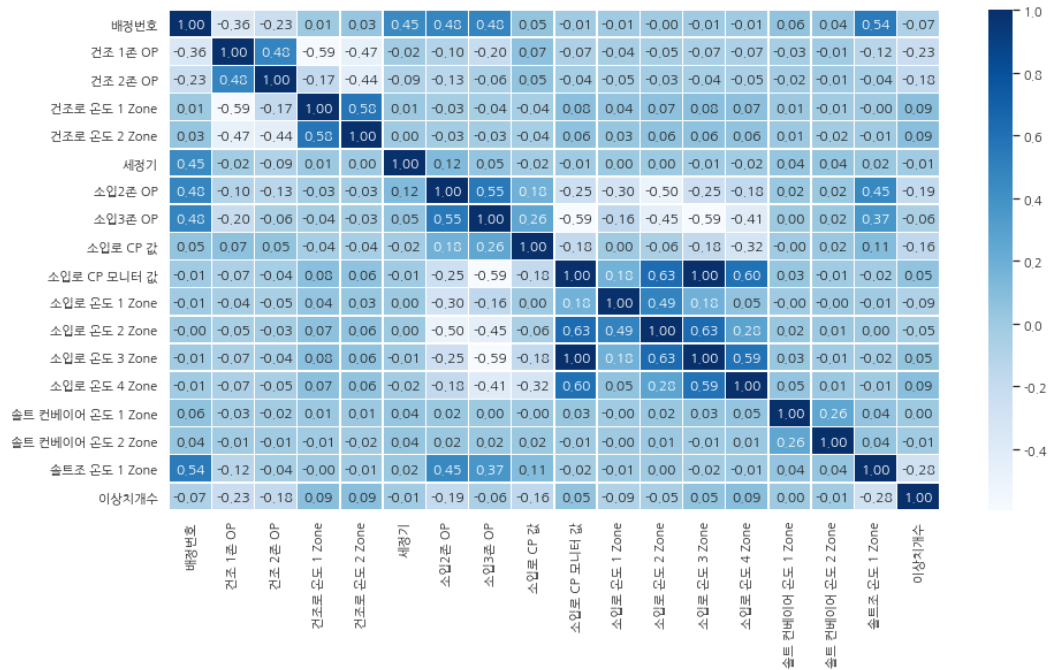


Figure 5 상관관계 계수 플롯

- 전처리 및 파생변수 생성

- * 고장시기 예측 모델
- 결측치 해당 변수 평균값 대치
- 이상치 개수 파생변수 생성.

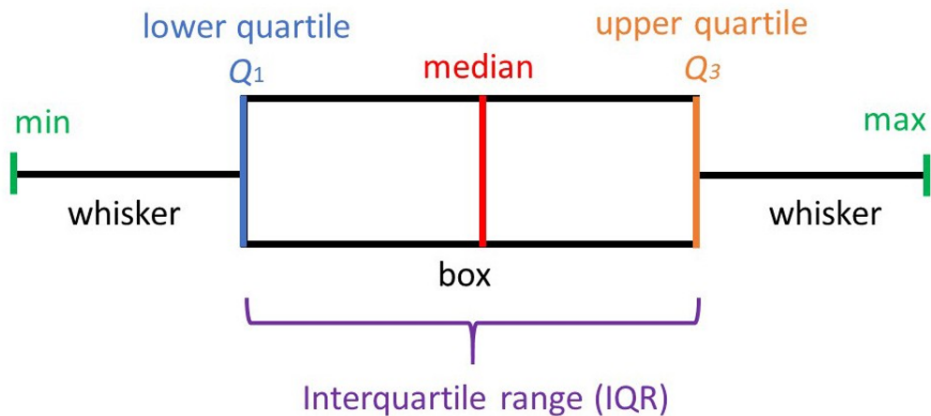


Figure 6 IQR

- IQR(InterQuartile Range) 사분위수 이상탐지, IQR 범위 밖의 값을 이상치로 판단하여 이상치 측정.
- $IQR = Q3 - Q1$
- Upper Outlier = $Q3 + 1.5 * IQR$
- Lower Outlier = $Q1 - 1.5 * IQR$
- IQR 방식으로 각 변수의 이상치 밖에 해당하는 변수의 개수를 구함.

- Train, Test Data split Train 90%, Test 10%
- * 공정별 이상탐지 모델
- 결측치 제거 (row).
- 각 건조, 세정, 소입, 숏트 공정별 IQR 밖의 이상치의 변수의 개수 파생변수 생성.
- Normal Data 와 Abnormal Data로 분류하고, Normal 은 Train data, Abnormal은 test Data
- Train, Validation Data split Train 80%, Validation 20%
- Data Normalization
 - MinMax Scaling
 - $(x - \min(x)) / (\max(x) - \min(x))$
- Scale 된 데이터를 12개의 Row로 묶은 Rolling Window로 구성.

○ 품질지수 측정

- 완전성 품질지수

$$\text{완전성 품질지수} = \left(1 - \frac{\text{결측데이터의개수}}{\text{전체데이터의개수}}\right) \times 100$$

결측 데이터의 개수 = 4677

전체 데이터의 개수 = 2939722

$$\text{완전성 품질지수} = \left(1 - \frac{4677}{2939722}\right) \times 100 = 99.84\%$$

- 유일성 품질지수

$$\text{유일성 품질지수} = \left(1 - \frac{\text{결측데이터의개수}}{\text{전체데이터의개수}}\right) \times 100$$

결측 데이터의 개수 = 0

전체 데이터의 개수 = 2939722

$$\text{유일성 품질지수} = \left(1 - \frac{0}{2939722}\right) \times 100 = 100\%$$

□ 분석모델 개발

○ 고장시기 예측 모델

- 분석 모델 소개

* ARIMA(Autoregressive Integrated Moving Average)

ARIMA모델은 시계열 데이터 분석과 예측을 위한 통계적 방법 중 하나이다. ARIMA는 시계열 데이터의 패턴과 경향을 이해하고 미래 값을 예측하는 데 사용된다. ARIMA모델은 세 가지 구성요로 이루어져 있다.

첫번째로는 AR(Autoregressive)이다. AR은 현재 값이 이전 값들의 선형 조합으로 설명될 수 있다는 것을 기반으로 한다. AR차수(p)는 몇 개의 이전 관측치를 사용할지를 나타낸다.

두 번째로는 MA(Moving Average)이다. MA는 현재 값이 이전 예측 오차들의 선형 조합으로 설명될 수 있다는 것을 기반으로 한다. MA차수(q)는 몇 개의 이전 예측 오차를 사용할지를 나타낸다.

마지막으로는 차분(Integrated)이다. 차분은 시계열 데이터의 불규칙성, 계절성, 추세 등을 제거하

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

ARIMA모델은 이렇게 AR, MA, 차분을 결합하여 시계열 데이터를 모델링하고 예측하는 모델이다. $ARIMA(p, d, q)$ 와 같이 표현되며, 각 부분의 차수는 데이터에 적합하게 선택되어야 한다. 모델 학습 후에는 미래 값을 예측할 수 있으며, 시계열 데이터의 패턴을 분석하고 예측하는데 사용된다.

* 해당 AI 방법론(알고리즘) 선정 이유

ARIMA모델은 비교적 간단한 구조로, 몇 가지 파라미터만 설정하면 모델을 학습하고 예측할 수 있다. 복잡한 모델보다 이해하고 구현하기 쉬우며 초기 모델링 단계에서 시계열 데이터를 빠르게 분석할 수 있다.

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

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

○ 공정별 이상탐지 모델

- 분석 모델 소개

* USAD(UnSupervised Anomaly Detection on Multivariate TimeSeries)

먼저 Anomaly detection은 학습 단계(training phase)와 탐지 단계(detection phase)로 구분할 수 있다. AE(Auto Encoder) 기반 모델의 학습 단계에서는 정상 데이터를 압축&복원 과정을 거치는데, 이 때 복원된 시계열과 원본 시계열 간의 차이인 Reconstruction error를 Minimize 하여 정상 데이터를 잘 복원하는 모델을 구축한다.

정상 데이터만을 학습한 AE 모델은 비정상 데이터를 입력하였을 경우 큰 값의 reconstruction error를 갖는다. 탐지 단계에서는 위 성질을 활용하여 정상과 비정상의 데이터가 혼합되어 있는 데이터를 AE에 Input으로 입력하여 복원 후 발생한 Reconstruction Error 를 Anomaly Score 라고 명명한다.

AE모델은 레이블이 없는 정상데이터만을 학습데이터로 사용하기 때문에 비지도 학습으로써 학습에 용이하지만 정상 데이터와 유사한 비정상 데이터가 들어올 경우 잘 구별하지 못한다는 어려움이 있다. 이는 AE의 Encoding 과정에서 복원에 불필요한 정보를 제거하기 때문이다. 학습단계에서 정상데이터만 사용하는 특성과 비정상을 탐지하는 Abnormal Information이 소거되는 특징이 있다. 정리하면, AE의 학습과정에서 데이터를 최대한 정상에 가깝게 복원하려는 성질이 있기 때문에 미세한 차이의 Abnormal Data를 검출하지 못한다.

USAD 알고리즘에서는 AE 알고리즘과 GAN 기반의 알고리즘을 융합하여 두 기반의 서로가 각 알고리즘의 장단점을 보완한다. GAN모델로 비정상 데이터를 발생시켜 AE모델은 점점 더 데이터를 구분하기 어려워지고 정확도 높은 Anomaly Detection을 수행할 수 있게되었다.

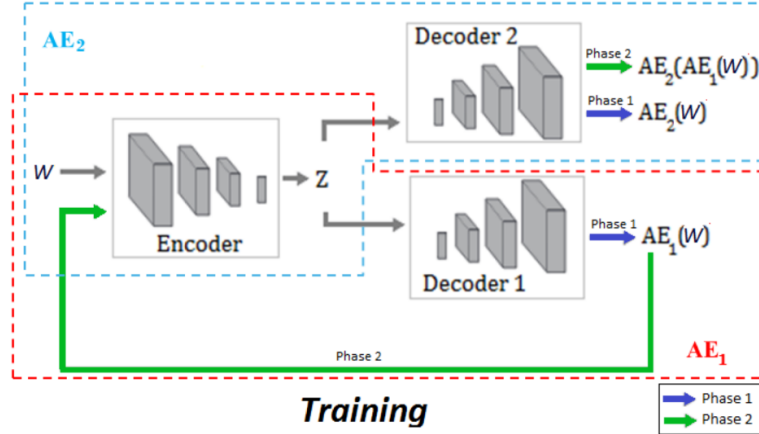


Figure 6 USAD Training Architecture

USAD AutoEncoder Architecture에서 기존 AE모델과 다른구조는 Adversarial Training을 적용하기 위해 두개의 Decoder을 사용하였다. USAD는 모델의 두 단계를 거쳐 학습을 진행한다.

Phase 1에서는 Input W (Real, Normal)를 Decoder 1 과 Decoder 2 를 거쳐 잘 복원하도록 학습을 진행한다. Phase 2에서는 Adversarial Traing 과정을 진행한다.

Adversarial Traing과정에서는 AE_2 는 AE_1 로부터 복원된 Fake data $AE_1(W)$ 를 구분하도록 학습한다. 반면에 AE_1 은 AE_2 의 판별 능력을 떨어트리는 것을 목적으로 학습을 진행한다.

즉, AE_1 은 GANs의 Generator역할, AE_2 는 GANs의 Discriminator역할을 담당한다.

$$\mathcal{L}_{AE_1} = + \|W - AE_2(AE_1(W))\|_2$$

$$\mathcal{L}_{AE_2} = - \|W - AE_2(AE_1(W))\|_2$$

Figure 7 Loss Function

Generator 역할을 하는 AE_1 은 AE_2 를 속이기위한 Fake 데이터를 생성하고, Discriminator 역할을 하는 AE_2 Real data W 와 Fake data $AE_1(W)$ 를 잘 구분해야 한다.

따라서, AE_1 은 Fake data에 대한 Reconstruction Error를 줄이도록 학습하고, AE_2 는 Fake data에 대한 Reconstruction Error를 키우도록 학습해야한다. 이를 학습하기위한 Loss Fuction은 Figure 7을 참고한다.

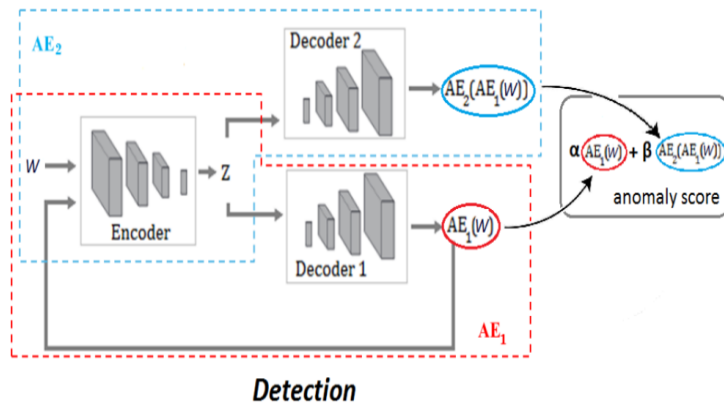


Figure 8 cnomaly score Aalculate method

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

Figure 9 Parameter setting Chart

Anomaly Detection시 Anomlay score을 산정하여 이상여부를 탐지한다. Anomaly score 산정 공식은 $\alpha(AE_1(W)) + \beta(AE_2(AE_1(W)))$ 를 이용하여 산정한다. $\alpha + \beta$ 의 값은 1로 산정되며, Parameter setting에 따라 아래의 Figure 9 와 같은 특성을 얻을 수 있다.

* 분석모델의 적설정 및 타당성

대부분의 공정에서 사용되는 설비는 시계열성을 띤다. 또 공정의 설비데이터의 특성상 고해상도 스케일로 데이터가 수집이 되기때문에 대다수의 정상적인 데이터가 들어오고 그리고 Anomaly 데이터 라고 하더라도 정상 데이터와 비슷한 배열을 가진 경우 탐지에 큰 어려움이 있다. 그래서 우리는 전통적인 시계열 데이터의 이상탐지를 위한 Auto Encoder를 사용함과 동시에 GANs 모델을 융합하여 조금 더 정확한 이상탐지를 위해 USD 모델을 사용한다.

□ 분석결과 및 시사점

○ 고장시점 예측 모델 성능 결과(ARIMA)

- 모델 학습 평가

MSE: 1.9672484874046297
MAE: 0.9913108344012007
RMSE: 1.4025863564874106

Figure 10 Error Score

* MSE는 오차의 제곱을 모두 더한 후 평균을 내는 값으로, 낮은 값일 수록 더 나은 모델임을 의미한다.

* MAE는 오차의 절대값을 모두 더한 후 평균을 내는 값으로, 낮은 값일 수록 더 나은 모델임을 의미한다.

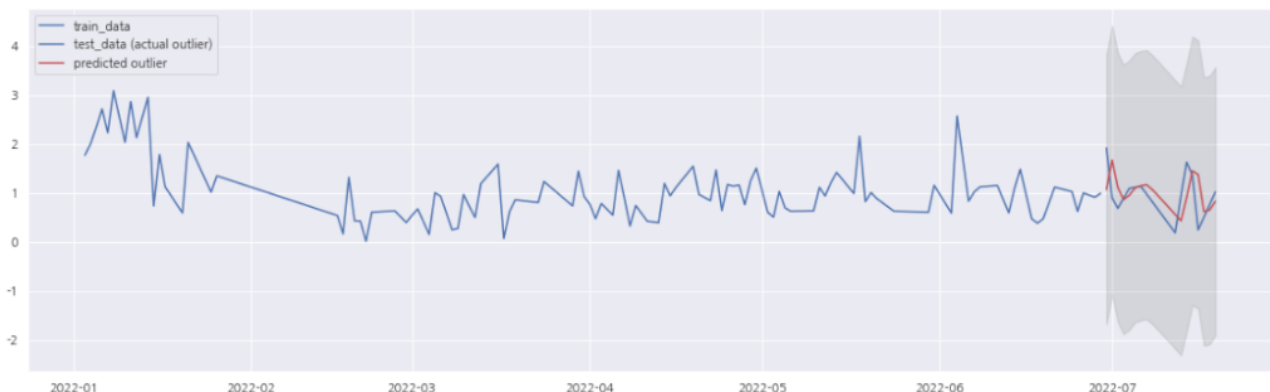


Figure 11 Predict chart using Predice model

* RMSE는 MSE의 제곱근으로 오차의 표준편차를 나타내는 값으로, 낮은 값일 수록 더 나은 모델임을 의미한다. 세 지표의 값이 낮은 편이므로 모델 성능이 좋은 편이라고 판단할 수 있다.

- 시사점

* 파란색이 실제데이터이고, 빨간색이 예측한 결과이다. 자기회귀를 통해 이상치를 예상할 수 있다.



Figure 13 Predict Chart using Model

*이상치평균과 불량률 그래프인데, 이상치평균이 높으면 불량률이 발생할 확률이 높다. 그러므로 ARIMA로 이상치평균을 확인하여 설비를 예지보전하며, 비용을 줄일 수 있다.

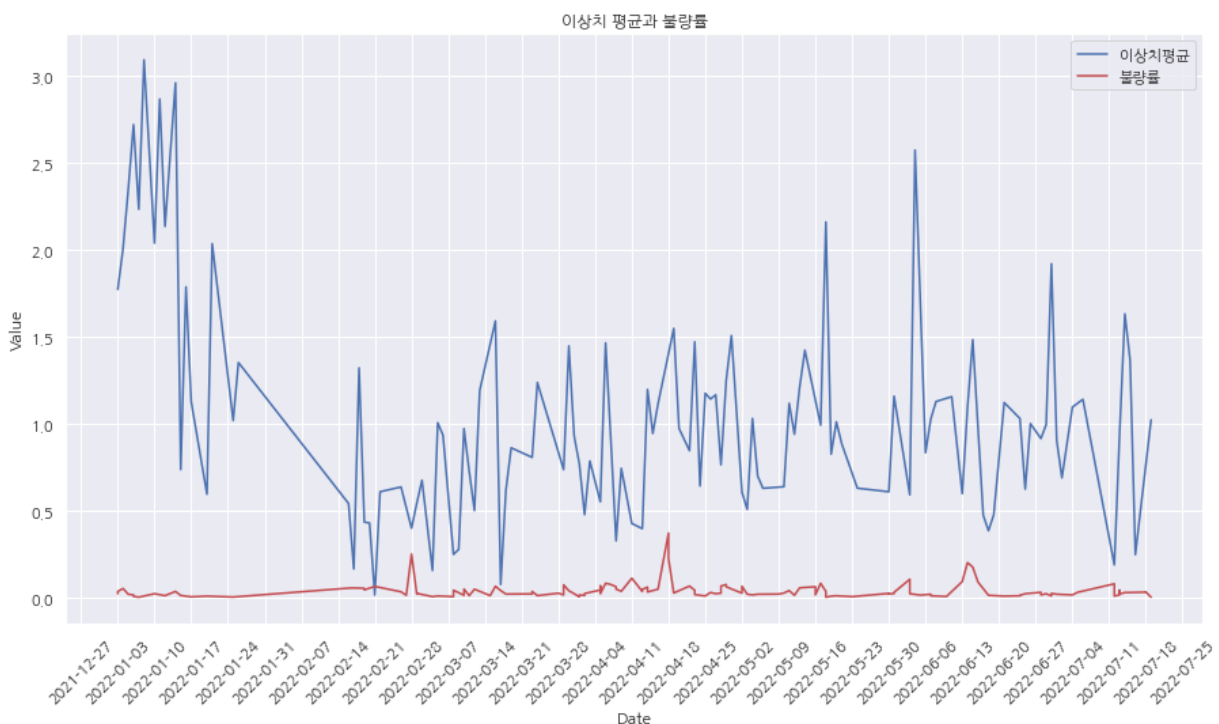


Figure 12 이상치 평균과 불량률 차트

* 향후 이상치를 예측하게되면, 이상치가 점점 증가하게 되는 것을 알 수 있다. 이후에도 계속해서 실제의 데이터를 수집하여 사용한다면, 더 확실하게 이상치의 평균을 예상할 수 있을 것이다.

○ 이상탐지 모델 성능 결과

- 분류 모델 평가지표

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Figure 13 Confusion Matrix

* 정확도(Accuracy) : 입력된 데이터에 대한 예측 정도.

$$\text{정확도} = \frac{TP + TN}{TP + TN + FN + FP}$$

* 정밀도(Precision) : 예측 값의 정확한 예측을 나타내는 지표.

$$\text{정밀도} = \frac{TP}{TP + FP}$$

* 재현율(Recall) : 실제 값 중에서 모델이 검출한 실제 값의 비율.

$$\text{재현율} = \frac{TP}{TP + FN}$$

* F1-Score : 정밀도와 재현율의 조화 평균.

$$F1Score = 2 \times \frac{\text{재현율} \times \text{정밀도}}{\text{재현율} + \text{정밀도}}$$

- 모델 학습 평가

```
drying F1 Score : 0.6643982608524707
clenaing F1 Score : 0.6654174474286904
quenching F1 Score : 0.6660700789067933
salt F1 Score : 0.6666071361822131
```

Figure 14 F1 Score

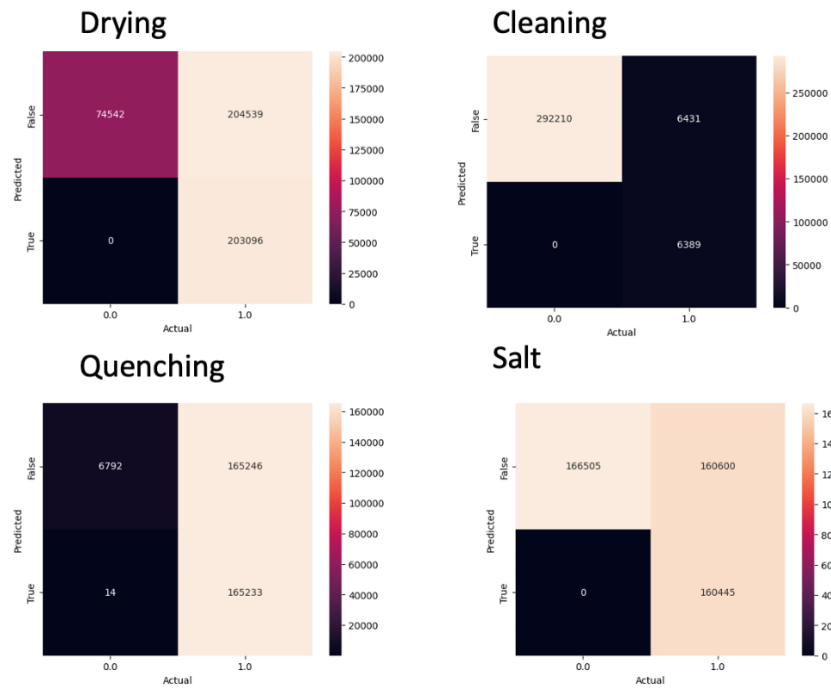


Figure 15 공정별 Confusion Matrix

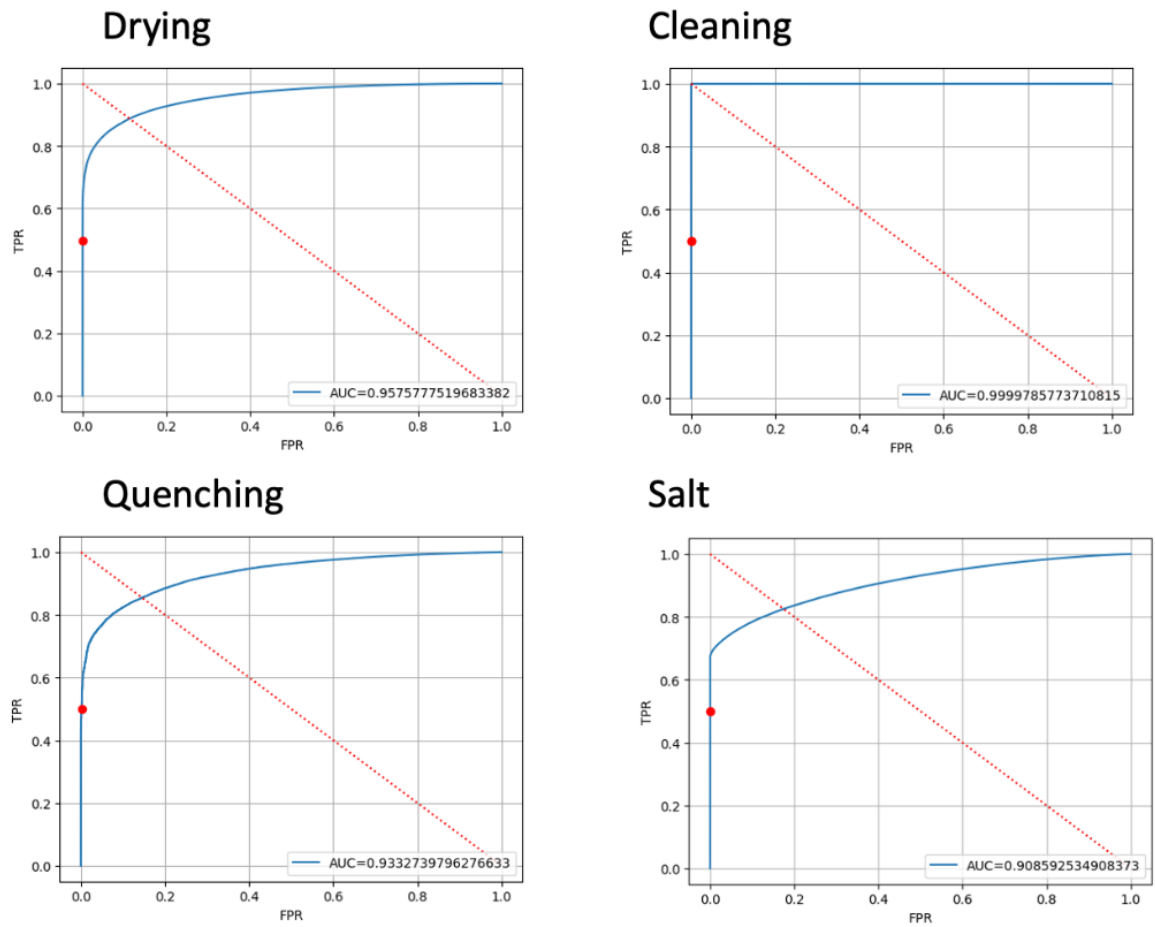


Figure 16 ROC Curve & AUC

- * 각 공정별 학습한 모델의 Confusion Matrix 와 F1 Score.
- * 각 공정별 AUC.

- 시사점

* 각 공정별 USAD 이상탐지 모델로 각공정의 이상을 탐지 할 수 있고 공정별 문제점을 조기에 판단 할 수 있다.

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

Figure 17 AUC Evaluation

* Area Under Curve(AUC) 의 면적이 90% 이상으로 높은 정확도 수용률을 보인다.

Drying AUC: 95.75%

Cleaning AUC: 99.99%

Quenching AUC: 93.32%

Salt AUC: 90.85%

* F1 Score는 모두 0.66이다.

* α 와 β 값을 조절하여 이상탐지 데이터에 대한 민감도를 조절할 수 있으며, Hyperparameter의 조절로 더 높은 F1-Score을 얻을 수 있을 것으로 기대된다.

□ 중소제조기업에 미치는 파급효과

○ 본 분석이 잠재하고 있는 확장성

- 제조업 공정의 공통적인 문제점

* 제조업 공정의 특성상 지속적으로 설비가 가동되며 물품을 생산한다. 열처리 공정에서 뿐만아니라 모든 제조업 공정이 가진 문제점은 비슷하다.

첫번째, 공정의 중단으로 인한 손실 방지

두번째, 예상치 못한 공정의 고장

세번째, 생산품의 불량률

이 외에 다른 문제 점도 있겠지만 공통적인 큰 문제점을 위 3가지로 축소한다.

- 문제점 해결방안

* 위 3가지 공통적인 문제점을 해결하고자 예지보전을 위한 고장시점 예측과 이상탐지 모델을 개발하여 문제를 해결한다.

- 본 분석의 적용가능 범위

* 공정 운영 데이터를 수집하여 관리하고 있는 체계가있는 대부분의 제조업 현장에 본 분석이 적용이 가능하다.

* 열처리 공정 외에도 데이터의 특성이 시계열성을 띄는 경우 적용할 수 있다.

* 본 분석에서 설비 데이터의 이상치를 바탕으로 고장시점의 예측과 이상탐지

를 수행하였다. 데이터가 계속해서 수집된다면 더 좋은 모델로 개선될 수 있을 것으로 기대된다.

- 타 공정 적용시 주의사항

* 열처리 공정의 문제를 해결하기 위해 적용된 상기모델은 열처리 공정의 데이터에 맞게 학습 되어있다. 타 공정에서 적용시 모델의 구조는 변경할 필요가 없으나, 해당 공정 설비 데이터로 모델을 재학습 할 필요가 있다.