

-콩순이 분석놀이 -

곽홍재

박새롬

이한석

목차

1

공모 배경

2

분석 데이터 정의

3

활용 분석기법 및 모델링 결과

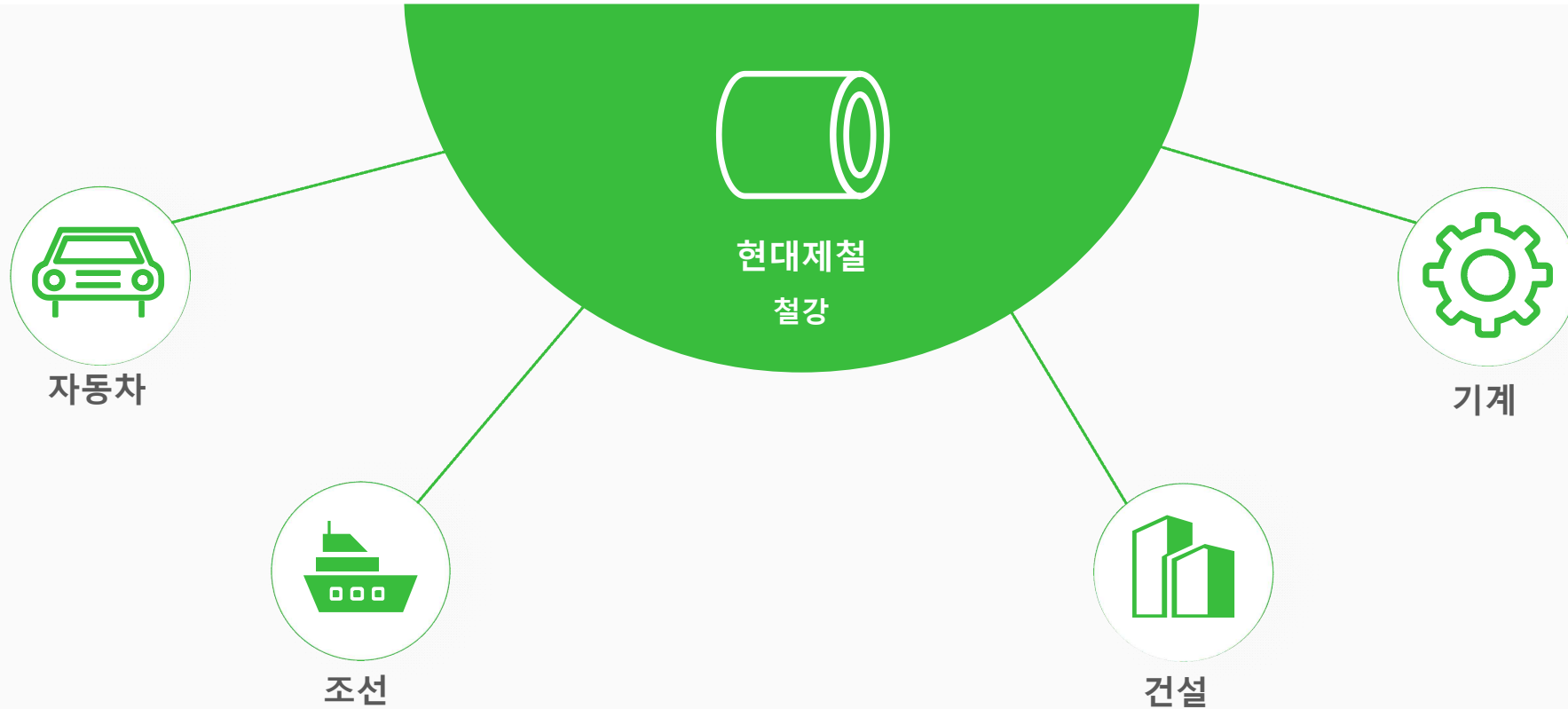
4

서비스 활용방안

5

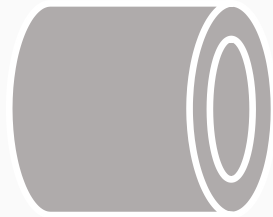
기대 효과

공모 배경

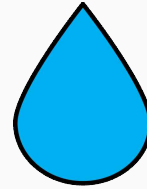
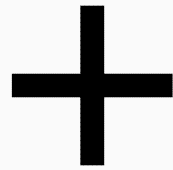


현대 제철의 철강 제품은 다양한 산업에 이용되고 있다

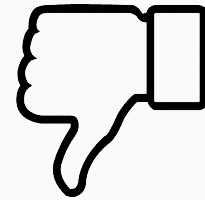
공모 배경



철강



결로



품질 저하

현대 제철의 철강 제품은 일정기간 창고에 보관되고 있으며
이 시기에 제품에 결로가 발생하면 철강에 녹이 슬어 상품의 질이 저하된다

공모 배경



품질 저하는 연계된 다양한 산업에 피해를 입히고, 현대 제철은 경제적 손실을 받게 되며, 큰 규모의 산업들이 많기 때문에 안전사고의 우려도 높아진다

공모 배경

결로 예측은 제철 산업 입장에서 매우 중요하고
결로는 온도 습도와 같은 기상 상황에 크게 연관이 되어 있기 때문에
날씨 데이터와 융합 데이터 분석이 중요하다.

분석 데이터 정의

현대 제철 데이터

- 1공장, 2공장으로 구분
- 내부센서(온도, 습도, 코일온도)
- 외부센서(온도, 습도)
- 결로 여부

기상 데이터

- 당진 AWS
- 기온, 풍향, 풍속, 강수량, 습도

서산 ASOS, 신평 AWS의 데이터도 고려했으나
가장 가까운 당진 AWS 데이터를 사용했다.

활용 분석기법 및 모델링 결과

데이터 전처리

- Plant1_train과 Plant2_train을 하나의 데이터 셋으로 병합
- 검증 셋과 형식을 맞추기 위해 melt를 통해 Long form으로 변환
 - > 결측 값 모두 지움
- 당진 AWS 데이터와 병합(시간 기준)
 - > 이 과정에서 생긴 결측 값은 이전 시점의 데이터로 대체
- Standard Scaler를 사용해 데이터 정규화

활용 분석기법 및 모델링 결과

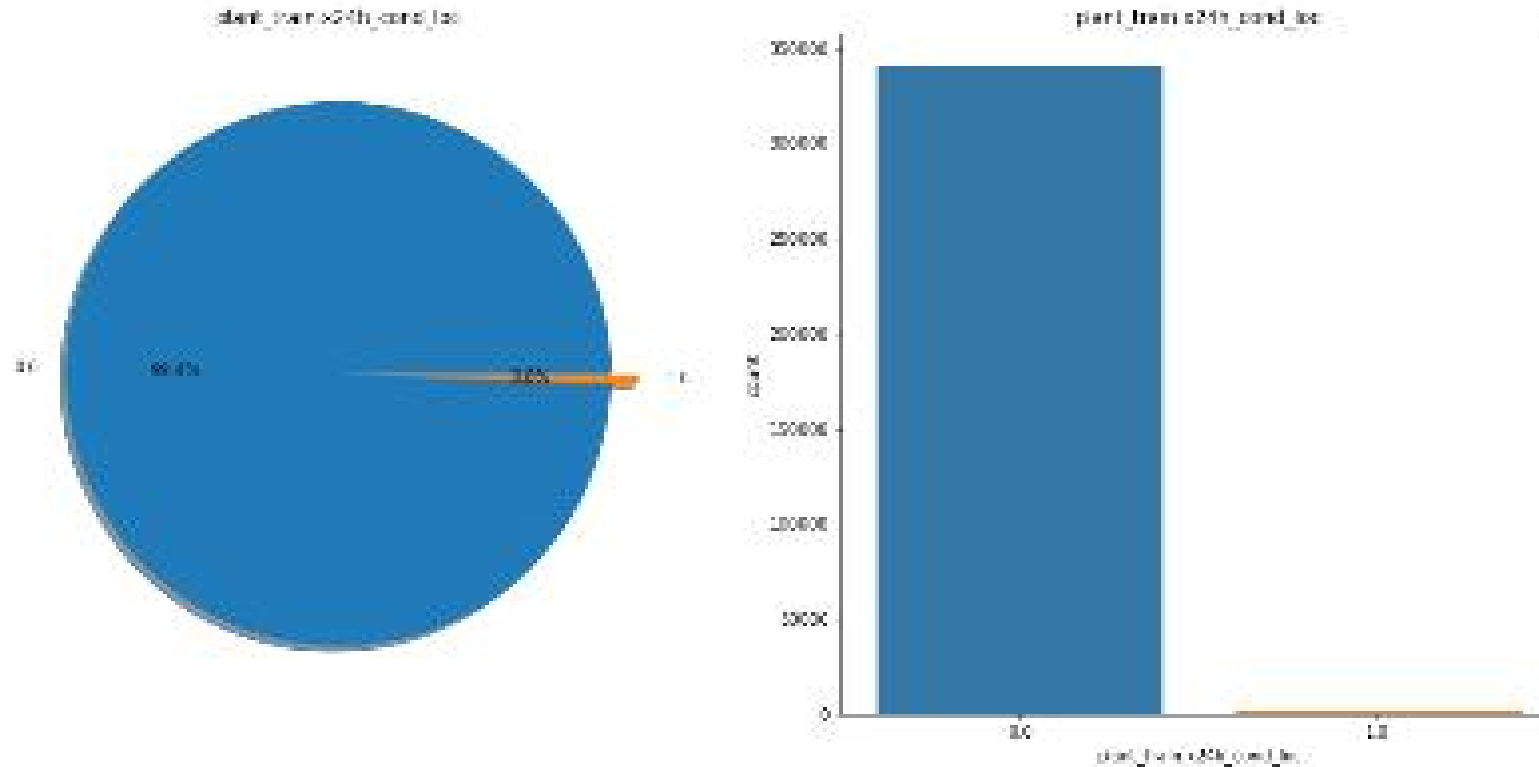
데이터 전처리

변환(melt)한 학습 데이터 형식

	plant_train.mea_ddhr	plant_train.plant	plant_train.loc	plant_train.tem_in	plant_train.hum_in	plant_train.tem_coil_in	plant_train.tem_out_loc1	plant_train.cond_loc	plant_test.x24h_tma	plant_train.x24h_cond_loc	plant_test.x48h_tma	plant_train.x48h_cond_loc
0	2016-04-01 00:00:00	1	1	16.00	24.00	11.00	9.00	0.0	2016-04-02 00:00:00	0.0	2016-04-03 00:00:00	0.0
1	2016-04-01 00:00:00	1	3	13.00	32.00	10.00	9.00	0.0	2016-04-02 00:00:00	0.0	2016-04-03 00:00:00	0.0
2	2016-04-01 00:00:00	1	2	14.00	23.00	11.00	9.00	0.0	2016-04-02 00:00:00	0.0	2016-04-03 00:00:00	0.0
3	2016-04-01 03:00:00	1	1	14.00	28.00	10.00	7.00	0.0	2016-04-02 03:00:00	0.0	2016-04-03 03:00:00	0.0
4	2016-04-01 03:00:00	1	2	12.00	32.00	9.00	7.00	0.0	2016-04-02 03:00:00	0.0	2016-04-03 03:00:00	0.0
...
348202	2019-03-31 23:40:00	1	3	9.82	38.95	9.83	6.31	0.0	2019-04-01 23:40:00	NaN	2019-04-02 23:40:00	NaN
348203	2019-03-31 23:40:00	2	3	8.26	39.62	9.57	4.47	0.0	2019-04-01 23:40:00	NaN	2019-04-02 23:40:00	NaN
348204	2019-03-31 23:50:00	1	2	10.55	36.69	11.01	6.34	0.0	2019-04-01 23:50:00	NaN	2019-04-02 23:50:00	NaN
348205	2019-03-31 23:50:00	1	3	9.82	39.35	9.83	6.34	0.0	2019-04-01 23:50:00	NaN	2019-04-02 23:50:00	NaN
348206	2019-03-31 23:50:00	1	1	10.49	38.74	10.52	6.34	0.0	2019-04-01 23:50:00	NaN	2019-04-02 23:50:00	NaN

활용 분석기법 및 모델링 결과

Imbalance Learning



결로가 아닌 경우가 99.4%로 아주 높은 비중을 차지하는 것을 알 수 있다.

활용 분석기법 및 모델링 결과

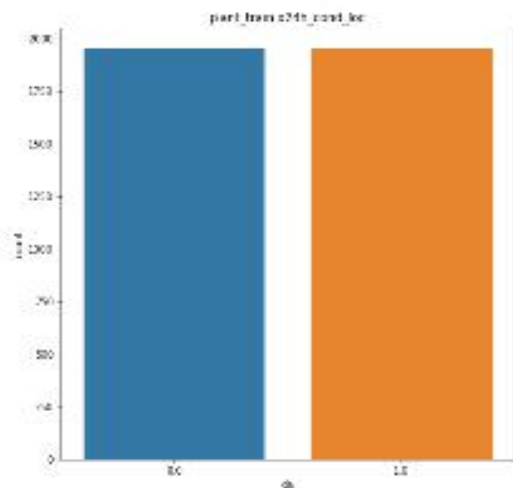
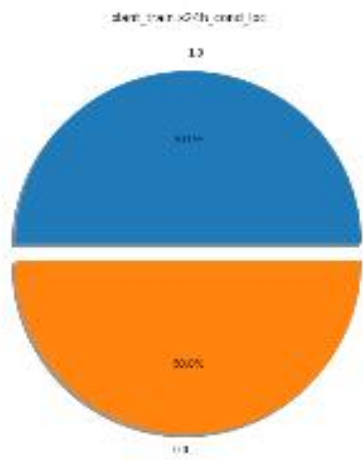
Imbalance Learning

- Target이 심한 Imbalance 상황이다.
- 그대로 모형을 학습시킨다면 모델은 일단 '결로가 아니다'라고 예측해도 정확도가 99%가 나온다.
- 결로 예측 문제에서 중요한 것은 정확도가 아니라 결로인 상황을 얼마나 정확하게 예측하는 것이다. (검증 지표 CSI에서도 나타난다.)
- 결로와 결로가 아닌 상황의 비율을 맞춰줘야 한다.

활용 분석기법 및 모델링 결과

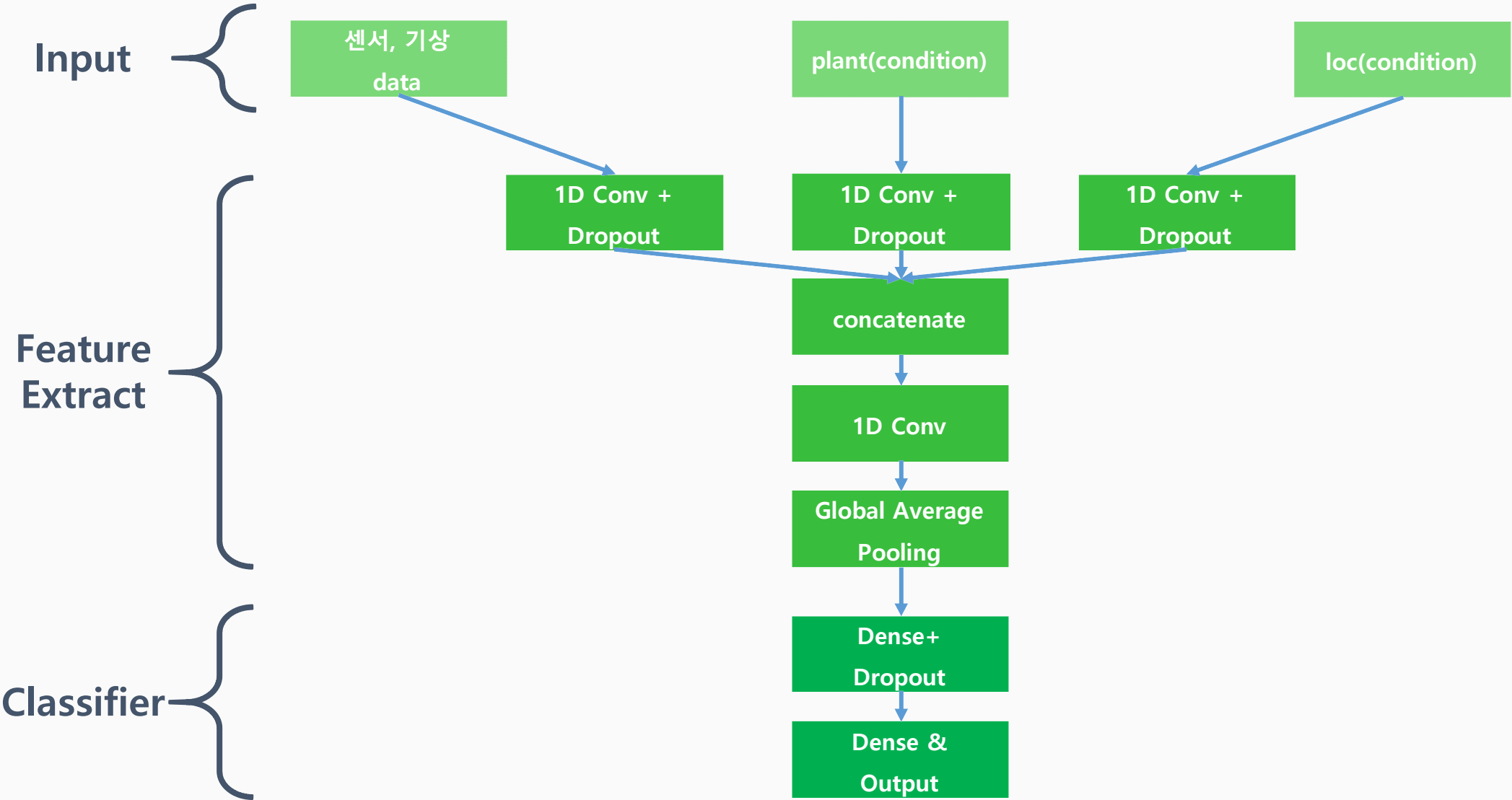
Imbalance Learning

- 비율을 맞추는 적은 쪽을 늘리거나 많은 쪽을 줄이거나 방법은 두가지가 있다.
- 적은 쪽을 늘리는 방법은 데이터가 시계열 특징을 가지고 있기도 하고 워낙 수가 적어 한계가 뚜렷했기 때문에 채택하지 않았다.
- 많은 쪽을 줄이는 방법을 채택했는데 imblearn의 Random Under Sampler를 사용했다.



활용 분석기법 및 모델링 결과

Modeling



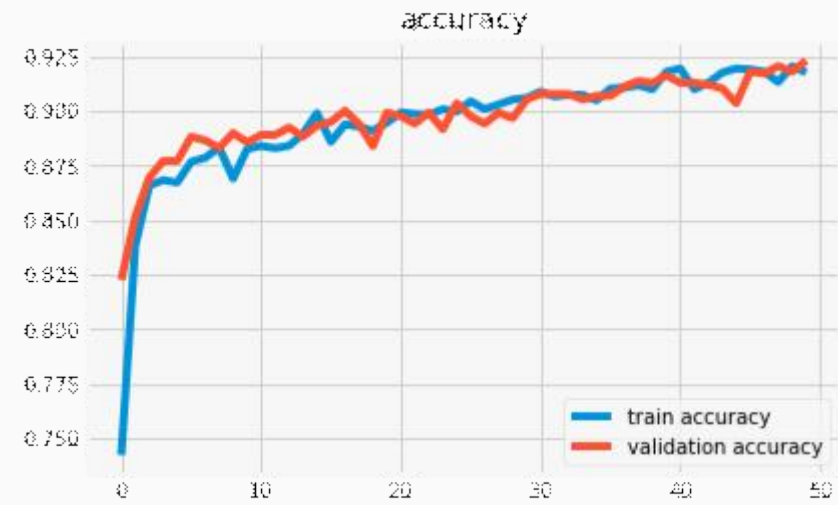
활용 분석기법 및 모델링 결과

Modeling

- 1, 2공장, 센서1,2,3을 한번에 고려하는 모형을 만들기위해 Generative 모형에서 많이 쓰는 Condition 개념을 사용했다.
 - > 학습 할 때 condition별로 특징 공간을 나눠서 학습할 수 있다.
- Sequence 데이터의 특징을 가지고 있다고 판단 해 1D Cnn을 통해 특징을 추출했다.
- Dense layer로 마지막 Classifier를 구성했다.
- End to End 모형이지만 24시간 후 모형과 48시간 후 모형은 구분했다.
 - > multi output도 생각했지만 특징을 공유하지 않는 것이 성능면에서 더 좋을 것이라고 생각했다.

활용 분석기법 및 모델링 결과

Modeling



안정적으로 학습이 이루어지는 모습을 볼 수 있다.

활용 분석기법 및 모델링 결과

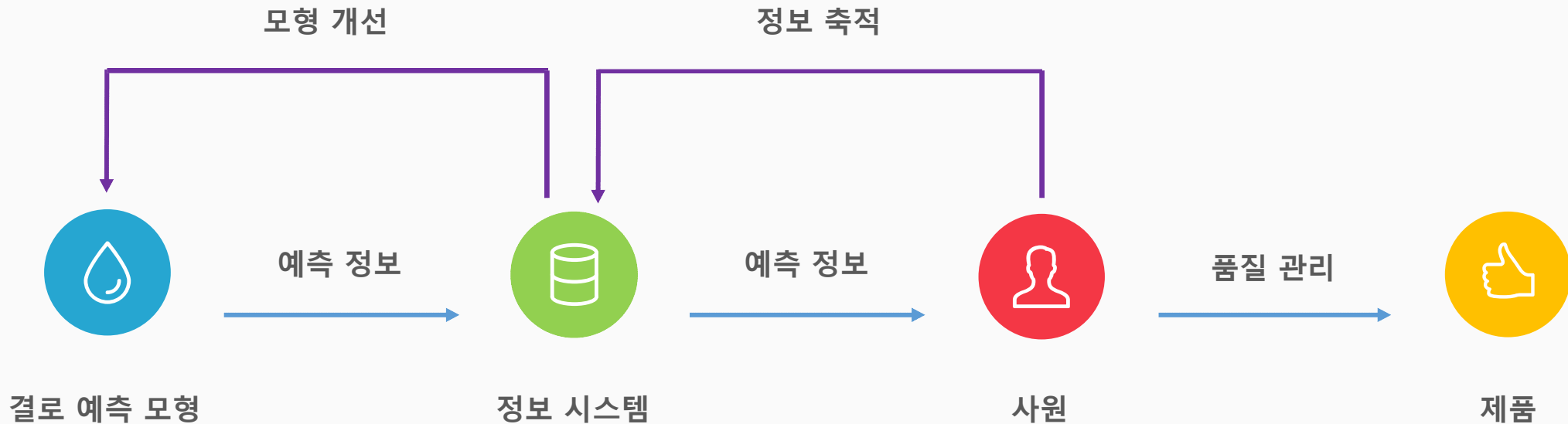
Modeling & Validation performance

- seed를 바꿔가며 Validation AUC가 가장 높은 모형의 Validation AUC에서 표준편차를 뺀 값보다 좋은 성능을 낸 모형을 저장했고 기하평균을 통해 Ensemble을 수행했다.

최종 성능

참가번호 203608 의 모델 종합 정확도(CSI)는 23.28%, 평균 AUC는 0.8765 입니다.

서비스 활용방안



결로 예측 모형을 사내 정보시스템과 연계해서 결로 예측 정보를 사원들에게 제공한다.
사원들은 이 정보를 통해 품질 관리를 하고 품질 관리에 대한 정보를 시스템에 축적한다.
이 정보를 바탕으로 모형을 개선하면 품질관리를 더 잘할 수 있게 되는 선순환 구조를 갖게 된다.

기대 효과

1. 철강 제품의 품질 저하로 인한 경제적 손실과 안전문제 최소화
2. 현대 제철 제품에 대한 신뢰도 상승을 통한 브랜드 이미지 재고
3. 다른 산업에서도 기상 데이터를 이용한 위험 감지 모형을 고려할 계기 마련

