

기상 데이터 기반 결로 예측 모델링

CONTENTS

목차

- 1. 공모배경
 - 2. 데이터 구성
 - 3. 데이터 분석
 - 4. 모델링
 - 5. 활용방안
-

1. 공모배경

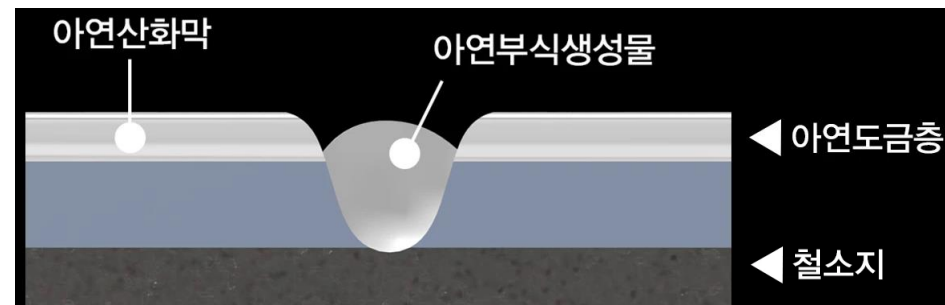
• 현대제철 냉연 제품



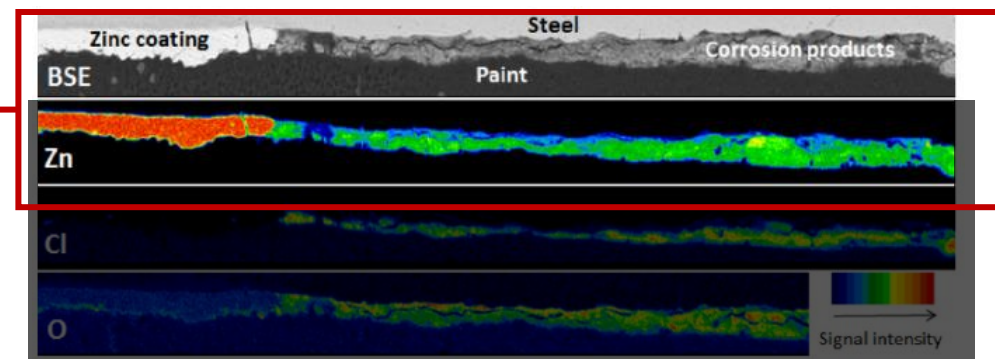
제품구분	제품특성
산세도유강판	열연강판 불순물 제거 후 표면에 오일처리
열연도금강판	냉연강판 표면에 아연도금하여 내식성이 우수
전기아연도금강판	냉연 및 열연강판 표면을 전기도금처리
용융아연도금강판	냉간압연된 강판 표면을 용융아연도금처리

목표 공장의 보관제품인 현대제철 냉연강판은
도유 및 아연도금 처리하여 내식성이 우수

• 아연도금 처리 시 효과



<아연도금 희생피막 원리>

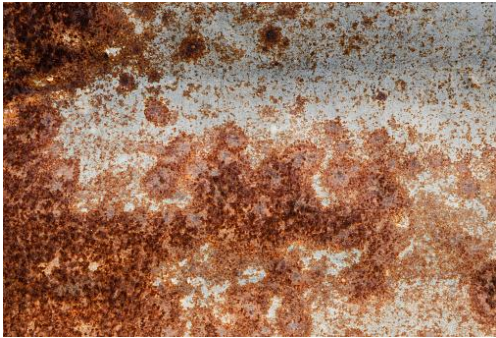


<아연도금 부식 방지효과>

도금처리된 아연이 더 빠른 속도로 산화하면서
희생효과를 일으켜 철소지의 녹 발생을 지연

현대제철은 내식성이 강한 도유 및 아연 도금 냉연 제품을 생산하고 있음

• 제품 부식 발생 종류



< 적청 : Red rust >



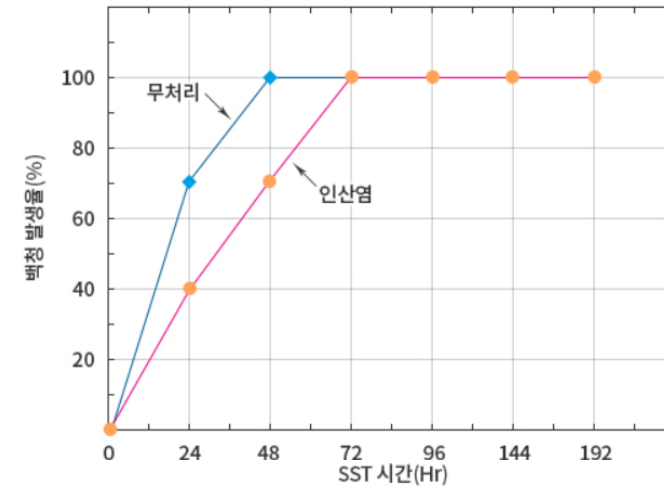
< 백청 : White rust >

제품구분	제품특성
적청 현상	철소지가 수분과 산소와 전기화학적으로 결합, 산화하면서 붉은 물질인 산화철로 변환
백청(백화) 현상	아연 도금된 구조물이 물이나 습기와 같은 수분과 만날 경우 백색의 물질을 형성하여 표면에 침착

제품 표면 물질에 따라
부식 발생의 종류는 적청, 백청으로 구분됨

• 냉연강판 백청 발생 위험

< POSCO 전기아연도금강판 내식성 >



SST(Salt Spray TEST)에서
48시간 노출될 경우

백청 발생률이 100%에 달함

보관기간 중 누적 결로 노출 시간

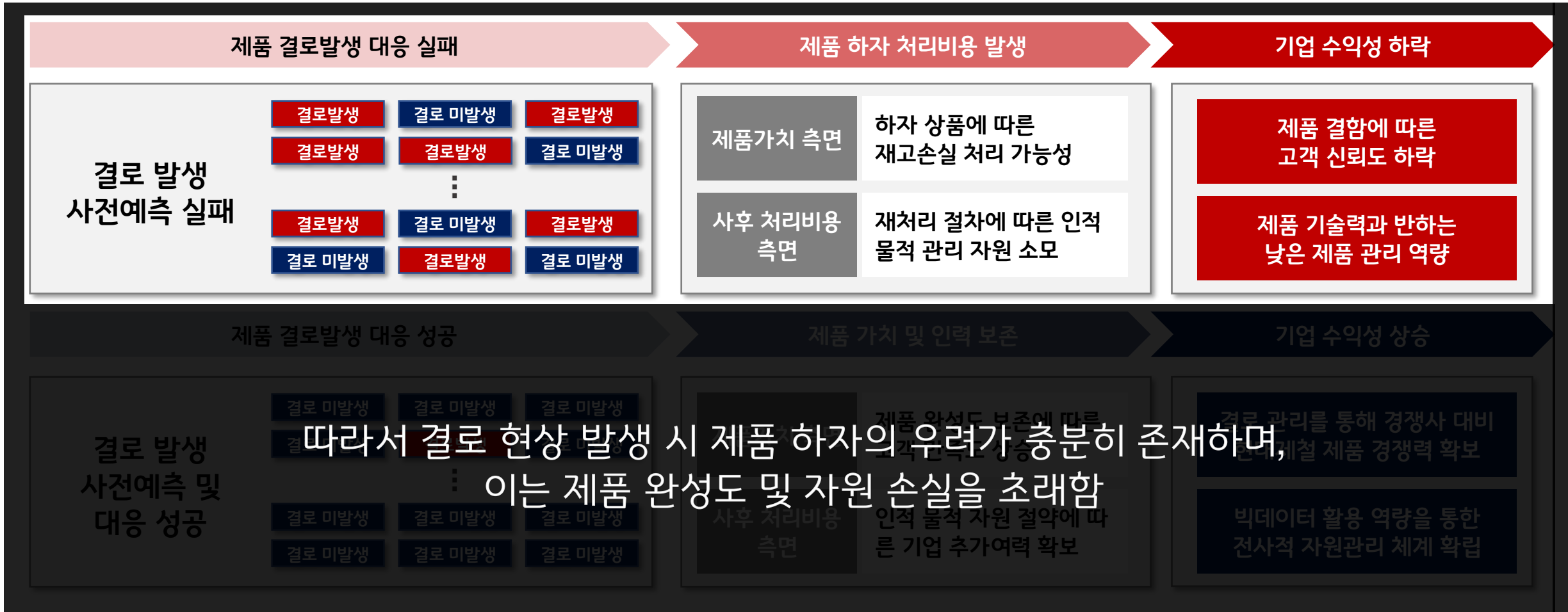
1d	2d	3d	...	N-1	N
3h	3h	7h	...	35h	38h

보관 기간이 장기화될 경우,
백청 발생의 위험이 증가

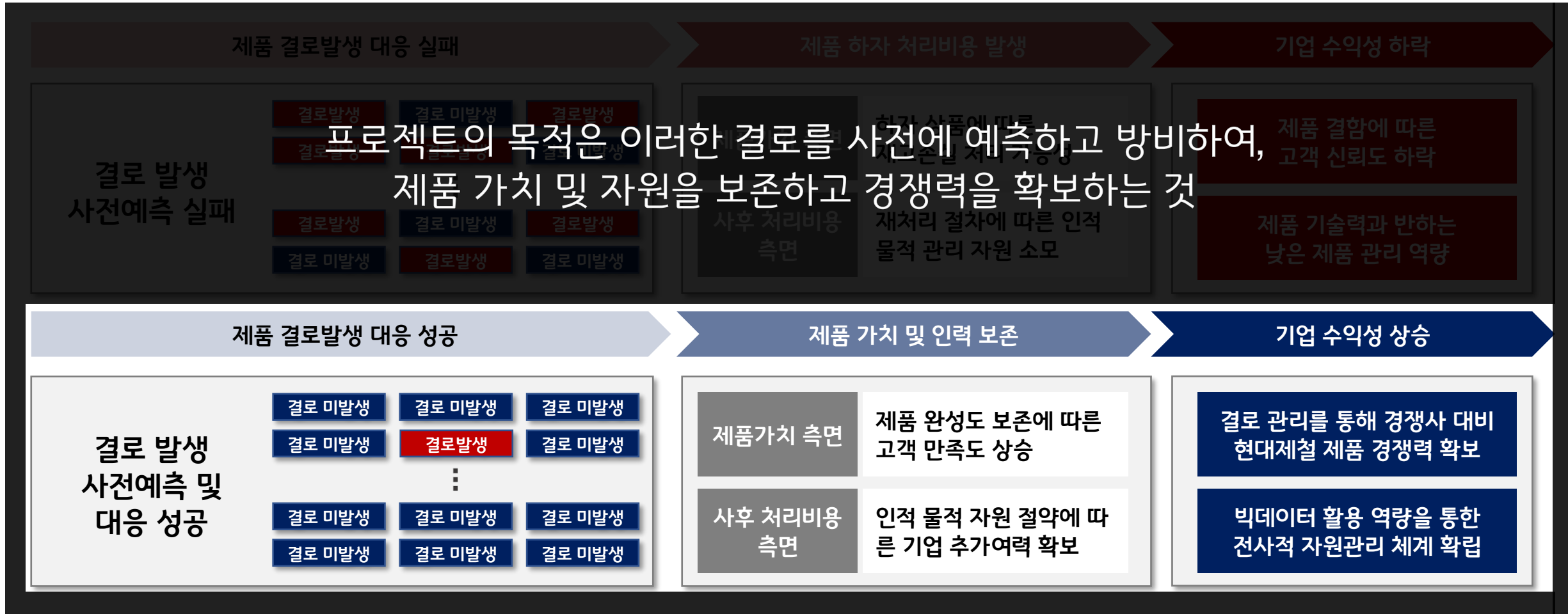
냉연 강판의 경우 제품 자체의 결함인
적청의 우려는 적으나, 백청 발생의 위험이 상존함

철소지의 내식성은 강하나 아연도금 표면은 습기에 노출 시 백청의 우려가 있음

• 결로 발생에 따른 수익 효과 저하



• 결로 발생에 따른 수익 효과 저하

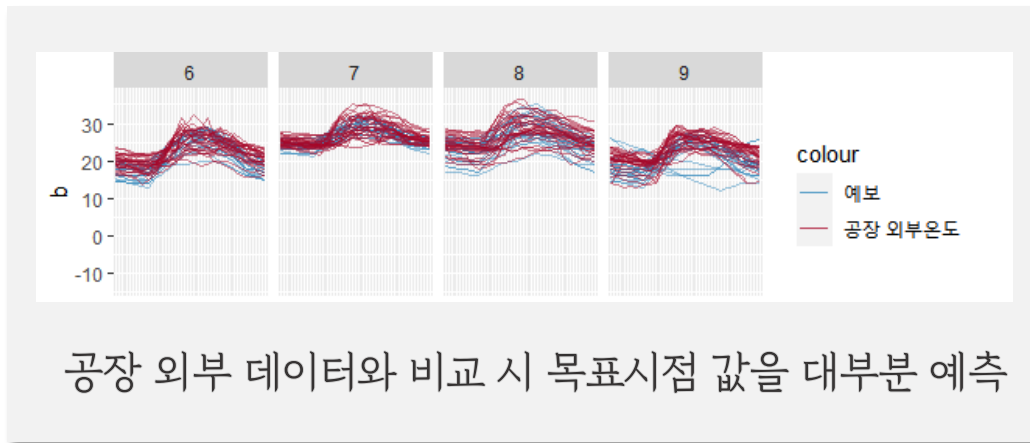


2. 데이터 구성

• 송악읍 동네예보 데이터 활용

변수 종류	데이터 형식 및 처리
24, 48시간 후 기온	°C 기준의 기온으로 수치 데이터
24, 48시간 후 습도	0~100까지의 상대습도를 나타내는 수치 데이터
24, 48시간 후 강수확률	0~100의 확률 값을 나타내는 수치 데이터
24, 48시간 후 풍향	0~360 수치를 4분할하여 동서남북으로 구분
24, 48시간 후 풍속	풍속에 해당하는 수치 데이터

< 예보 데이터와 실제 데이터 일치 수준 >



< 보관 위치와 관측소 사이 거리 >



냉연 1공장에서 거리 : 7.49km

냉연 2공장에서 거리 : 8.51km

미래 시점의 결로를 예측하기 위해 공장 환경과 가장 유사한 예보데이터 활용

• 공장 데이터 결측값 파악

1. 결측값 주변에 누락이 많아 보간이 어려운 경우

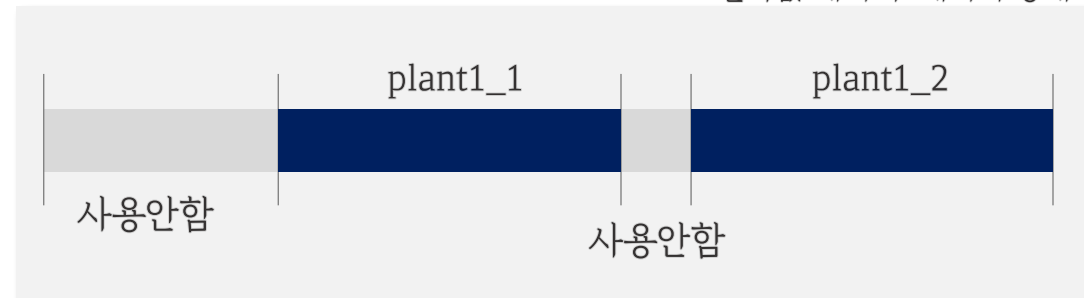
날짜	일별 데이터 수	데이터 활용 여부
2016-12-27	22	보간하여 활용
2016-12-28	16	제외
2016-12-29	9	제외
2016-12-30	24	

2. 결측값 주변에 값이 충분하여 보간이 용이한 경우

mea_ddhr(시간)	결측 데이터	데이터 활용 여부
2018-07-10 13:00:00	NA	보간하여 활용
2018-07-11 23:30:00	NA	
2018-07-12 23:30:00	NA	
2018-07-13 23:30:00	NA	

• 결측값 처리 방안

< 결측값 제거 후 데이터 상태 >



날짜	기존 데이터 수	보간 후 데이터 수	데이터 사용 비율
제 1공장 데이터	58749	46957	79.93%
제 2공장 데이터	57914	46963	81.09%

보간이 어려운 결측 데이터 구간을 제외하고 추린 결과
기존 두 공장 데이터의 80% 수준 확보

보간 가능 여부에 따라 결측값을 처리하여 적합 데이터를 구성

• 공장 데이터의 다중대치

< 대치 적용 전 데이터 >

mea_ddhr(시간)	tem_in_loc1	hum_in_loc1	...
2018-07-06 23:30:00	28.56	44.17	...
2018-07-06 23:40:00	28.5	44.14	
2018-07-06 23:50:00	NA	NA	
2018-07-06 00:00:00	28.25	44.23	

< 대치 적용 후 데이터 >

mea_ddhr(시간)	tem_in_loc1	hum_in_loc1	...
2018-07-06 23:30:00	28.56	44.17	...
2018-07-06 23:40:00	28.5	44.14	
2018-07-06 23:50:00	25.9	32.57	
2018-07-06 00:00:00	28.25	44.23	

• 예보 데이터의 선형보간

< 선형보간 적용 전 데이터 >

mea_ddhr(시간)	공장 데이터(온도)	예보 데이터(온도)	...
2018-05-01 02:00:00	17.02	12.00	...
2018-05-01 03:00:00	16.38	NA	
2018-05-01 04:00:00	16.23	NA	
2018-05-01 05:00:00	16.53	11.00	

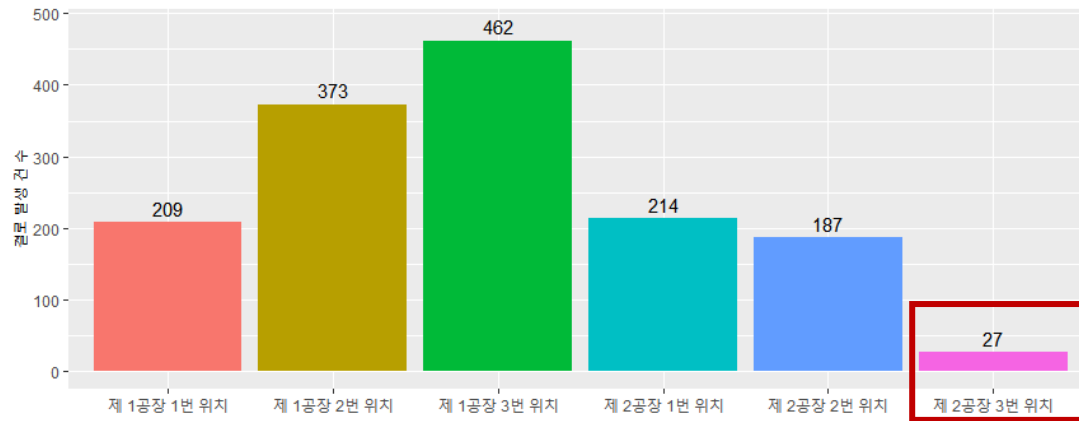
< 선형보간 적용 후 데이터 >

mea_ddhr(시간)	공장 데이터(온도)	예보 데이터(온도)	...
2018-05-01 02:00:00	17.02	12.00	...
2018-05-01 03:00:00	16.38	11.67	
2018-05-01 04:00:00	16.23	11.33	
2018-05-01 05:00:00	16.53	11.00	

데이터의 변형 정도가 최소화 되도록 대치 및 보간하여 결측 데이터 처리

3. 데이터 분석

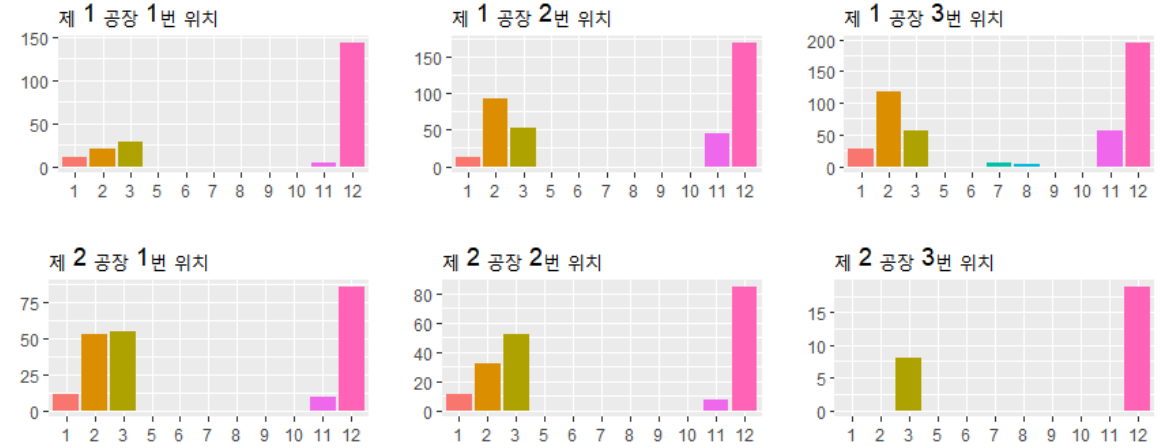
• 공장 위치별 결로발생 현황



	1번 위치	2번 위치	3번 위치
제 1 공장	209	373	462
제 2 공장	214	187	27

제 2 공장 3번위치는 결로 데이터 수가 극단적으로 적으므로 추가적인 방안 필요

• 계절별 결로발생 현황

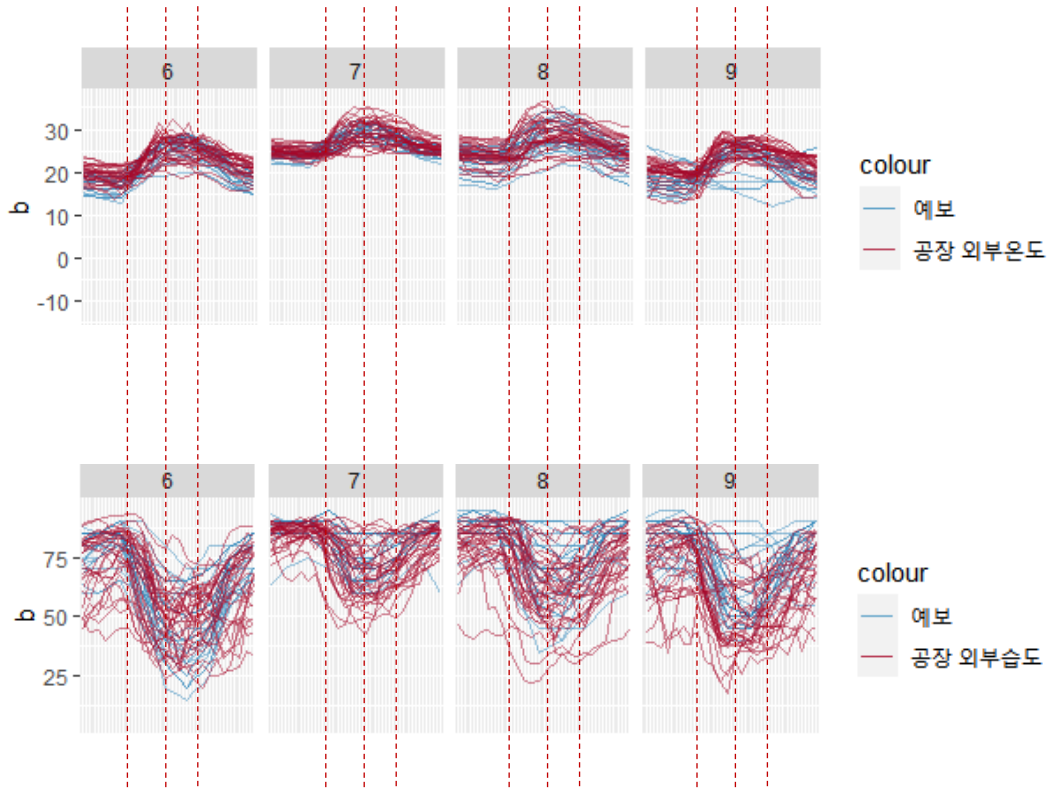


수치형 계절변수	1	2	3	4	5	6	7	8	9	10	11	12
명목형 여름변수	비-여름(0)								여름(1)			

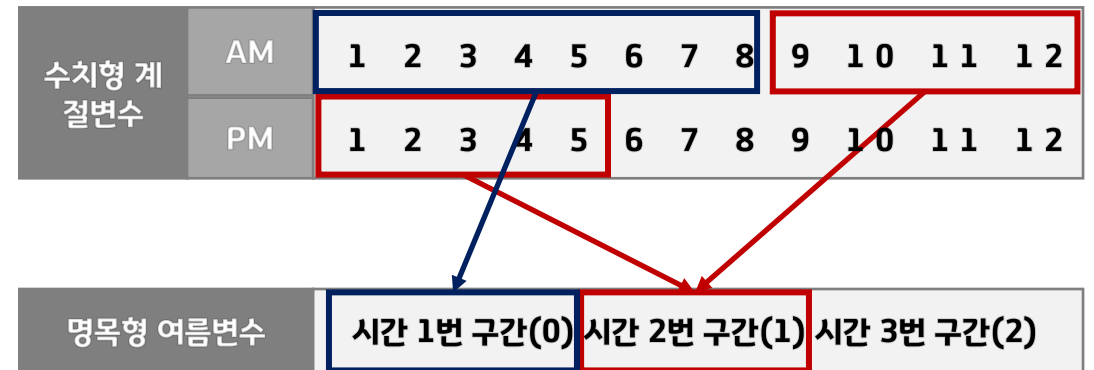
여름을 제외한 계절에 결로발생이 많으므로 여름과 비 여름으로 명목변수로 추가

결로발생 현황 분석 결과, 특정 위치 결로 데이터 보정 및 변수 추가의 필요성 발견

하루 시간별 온도 추세

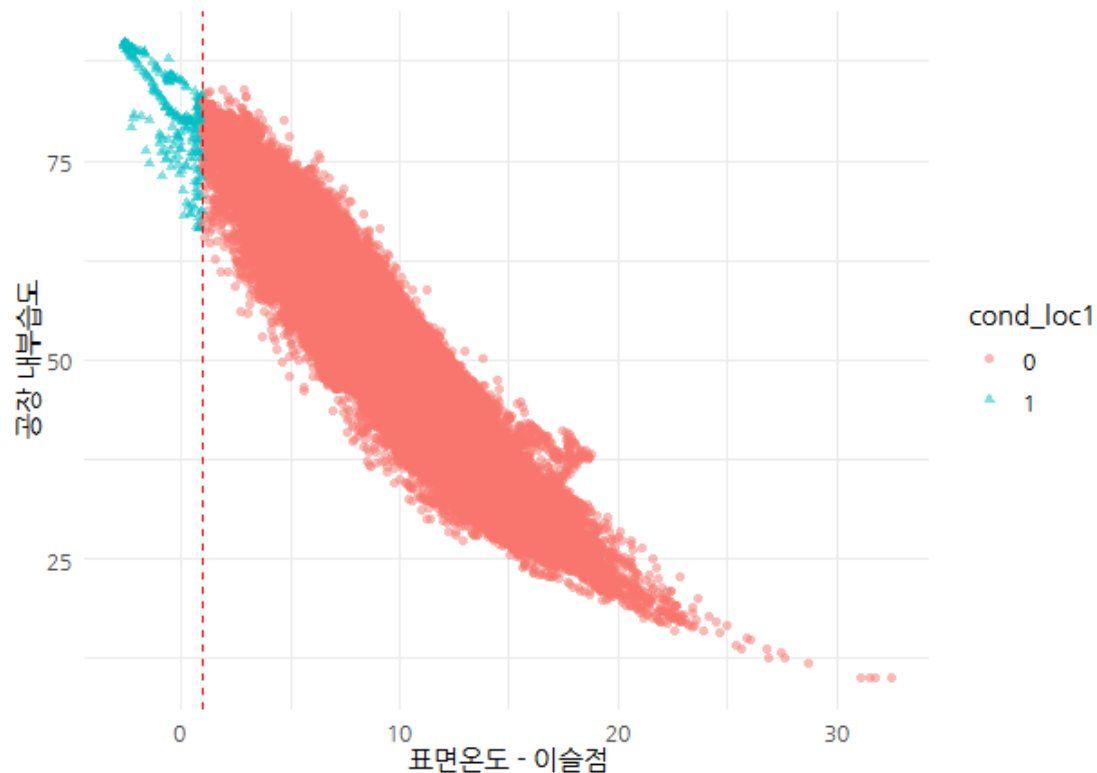


온도, 습도, 코일온도는 오전 8시, 오후 3시를
시점으로 규칙적으로 변화



습도, 온도, 코일온도 데이터의 변화지점을 파악하여 계절 명목 변수로 활용

- (코일온도 - 이슬점)과 결로여부 상관관계



(표면온도-이슬점) 값은
결로 여부를 결정짓는 가장 중요한 변수

- (코일온도 - 이슬점)의 결과가 약 0.99 보다 낮을 경우 결로가 발생
- 이러한 (코일온도 - 이슬점) 변수는 습도와 강한 음의 상관관계를 보임
- 따라서 목적 변수인 결로 발생 여부를 연결하는 중요한 변수로 파악

(코일온도 - 이슬점) 값은 결로 발생 모델에 직결되는 핵심 변수이므로 활용 변수에 포함

• 사용변수

1. 해당 위치 공장 내부 습도
2. 해당 위치 공장 내부 온도
3. 해당 위치 공장 코일 온도
4. 여름 더미변수
5. 시간 더미변수
6. 24시간, 48시간 후 습도 예보
7. 24시간, 48시간 후 온도 예보
8. 24시간, 48시간 후 강수확률 예보
9. 24시간, 48시간 후 풍속
10. 24시간, 48시간 후 Heat Index
11. 24시간, 48시간 후 풍향 예보
12. 기타 ,변수들의 상호작용 또는 가중치를 둔 변수들

제철 공장 데이터

+

기상청 예보 데이터

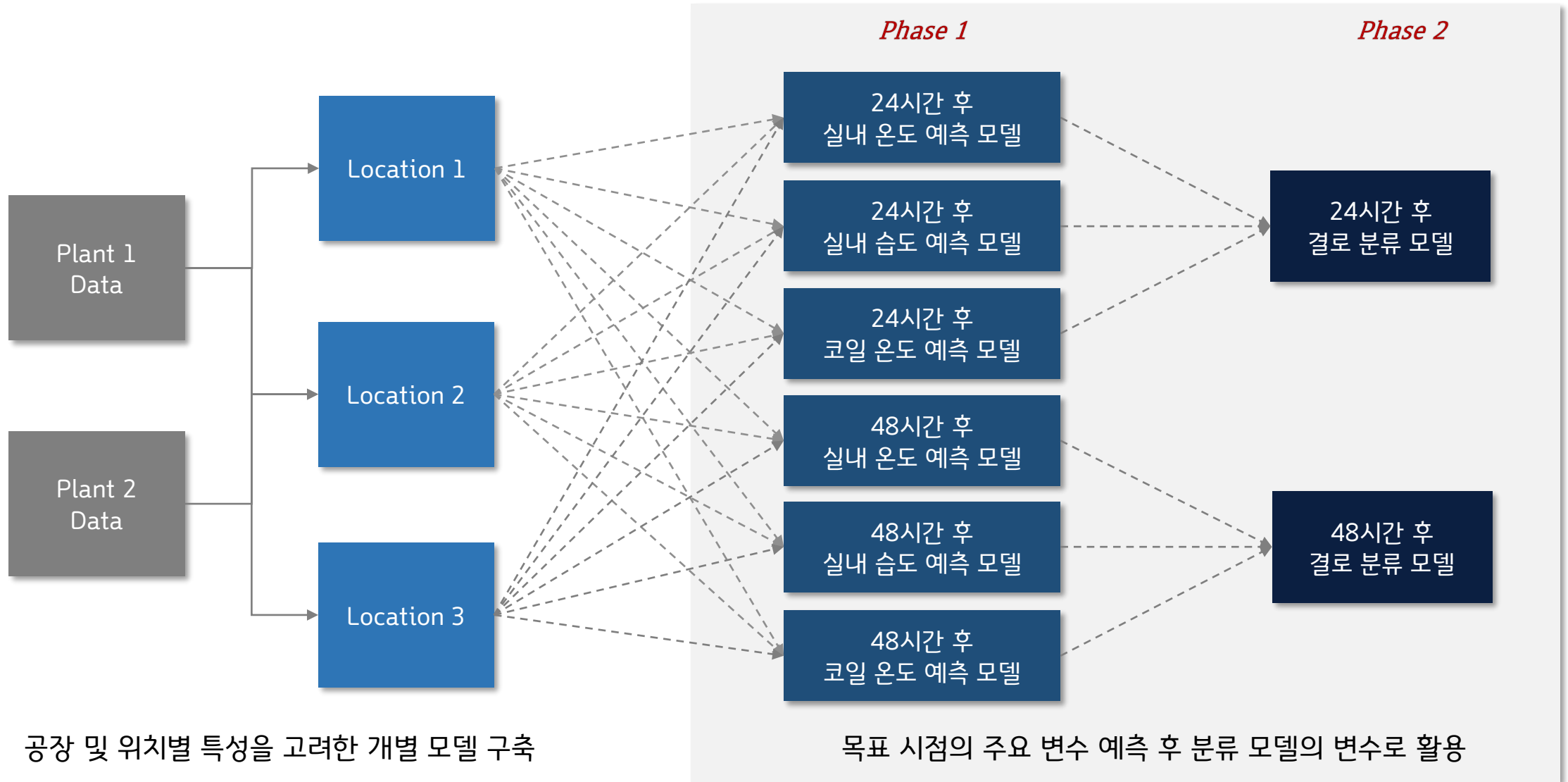
+

분석 후 추가 변수

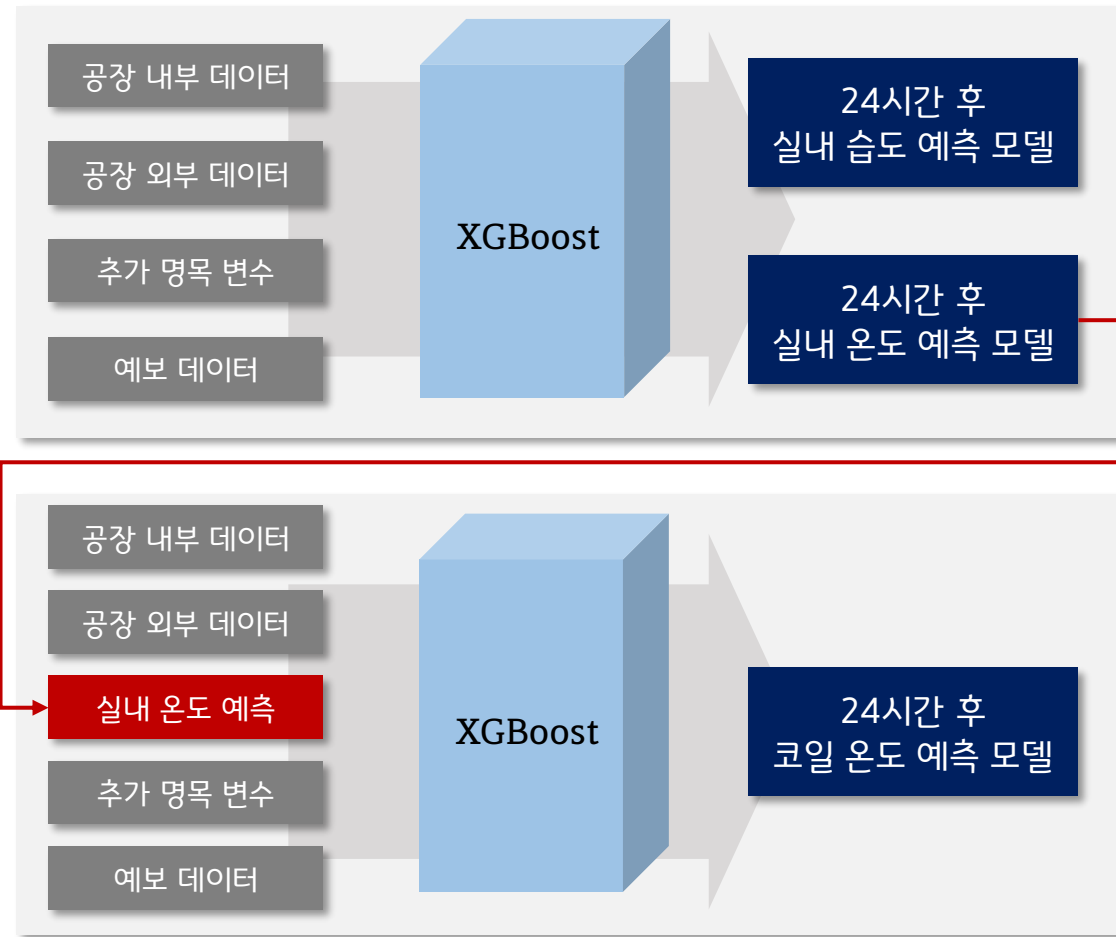
=

예측변수 설정

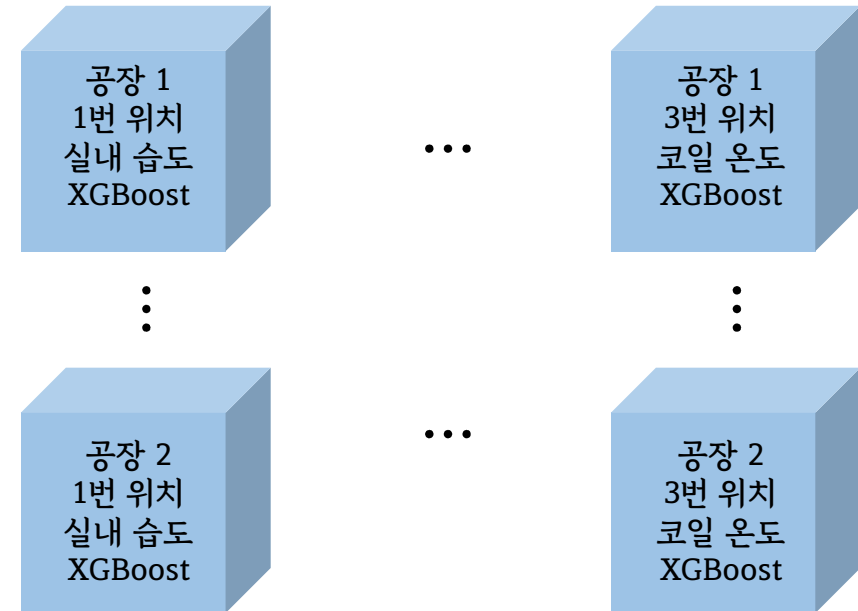
4. 모델링



예측 모델링 구성요소 *Phase 1*



GridSearch 결과 최적 모델 모수 적용

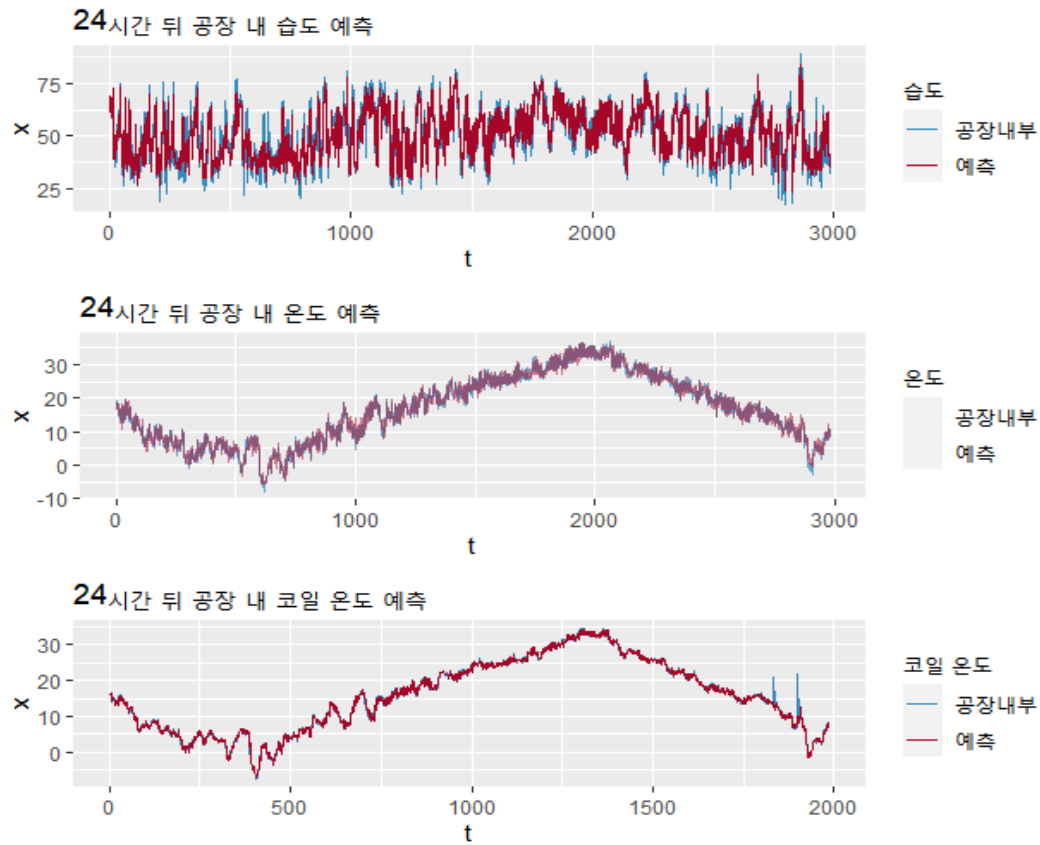


$$2 \times 3 \times 3 \times 2 = 36 \text{개 모델}$$

공장·위치·변수·시간별 XGBoost 예측모델 구축

XGBoost 기법으로 학습시켜 내부 습도, 내부 온도 그리고 코일 온도 예측 모델을 구축

• 실제 데이터-예측값 비교 예시



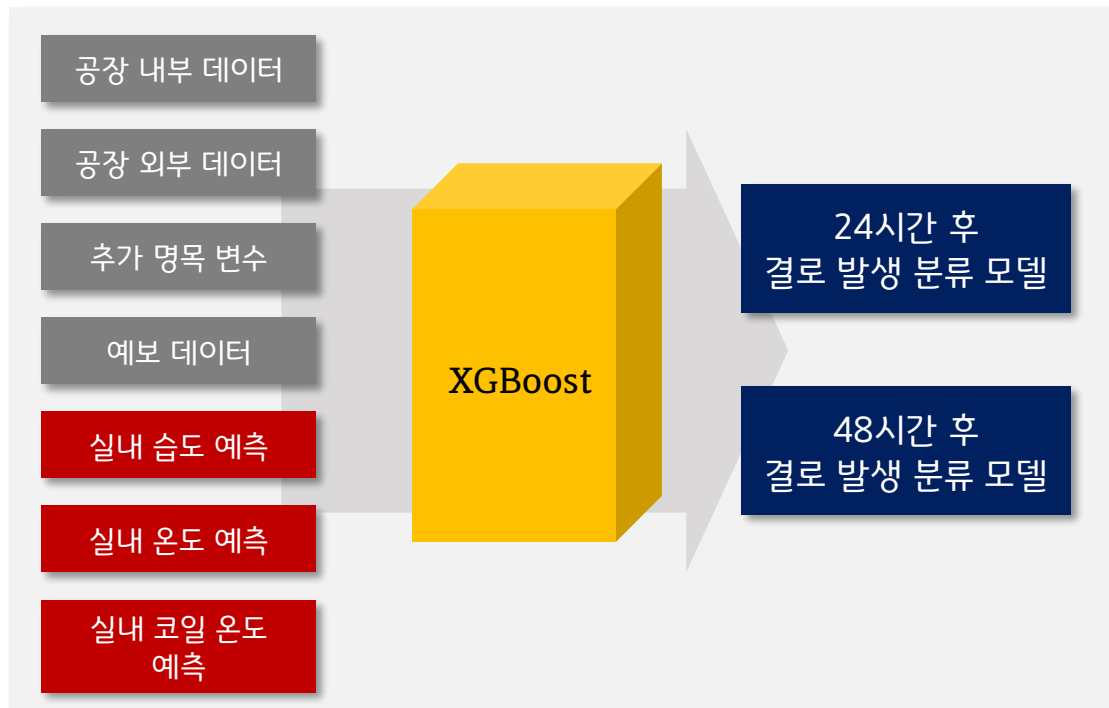
• 예측모델 RMSE 결과

	RMSE			RMSE	
제 1 공장	24시간 후	48시간 후	제 2 공장	24시간 후	48시간 후
1번 위치 습도	3.149	3.432	1번 위치 습도	3.543	3.871
1번 위치 온도	0.575	0.710	1번 위치 온도	0.652	0.762
1번 위치 코일온도	0.419	0.581	1번 위치 코일온도	0.579	0.633
2번 위치 습도	3.525	3.814	2번 위치 습도	3.381	3.749
2번 위치 온도	0.608	0.766	2번 위치 온도	0.920	0.899
2번 위치 코일온도	0.318	0.369	2번 위치 코일온도	0.814	0.526
3번 위치 습도	3.733	3.957	3번 위치 습도	3.325	3.590
3번 위치 온도	0.736	0.829	3번 위치 온도	0.616	0.799
3번 위치 코일온도	0.5900	0.704	3번 위치 코일온도	0.824	0.822

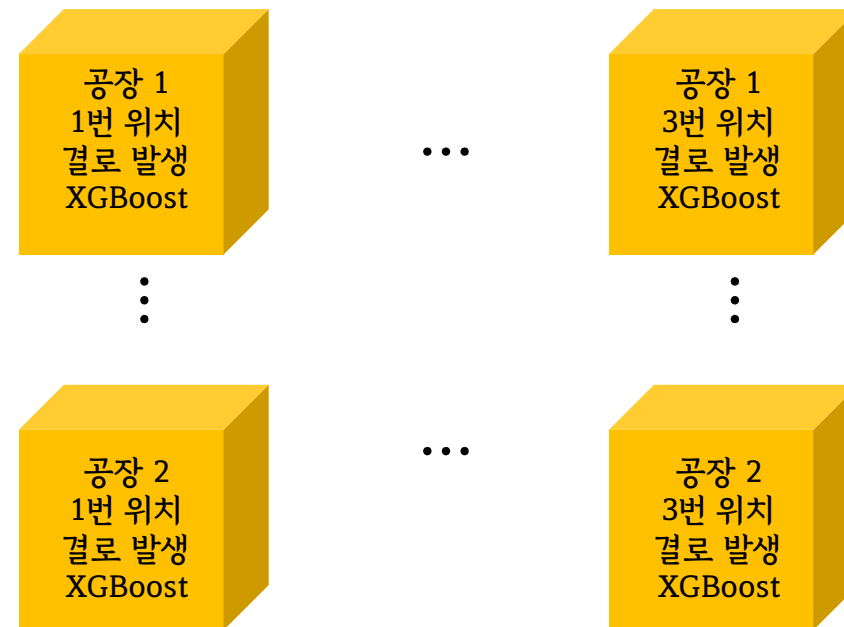
	습도 예측	온도 예측	코일온도 예측
RMSE 평균	3.589	0.739	0.598

각 변수 모델의 RMSE 수준이 3.589, 0.739, 0.598 정도로 예측력이 높은 것으로 판단됨

- 분류 모델링 구성요소 *Phase 2*



GridSearch 결과 최적 모델 모수 적용
데이터 불균형 문제를 위한 SMOTE 기법 적용



$$2 \times 3 \times 2 = 18 \text{개 모델}$$

공장·위치·시간별 XGBoost 예측모델 구축

XGBoost 기법에 데이터 및 내부 예측 변수를 학습시켜 각 위치별 결로발생 분류 모델을 구축

• 혼동행렬 결과 예시

Confusion Matrix and Statistics		
Prediction	Reference	
	0	1
	0 9326	1
	1 2	30
Accuracy : 0.9997		
95% CI : (0.9991, 0.9999)		
No Information Rate : 0.9967		
P-Value [Acc > NIR] : 1.808e-10		
Kappa : 0.9522		
McNemar's Test P-Value : 1		
Sensitivity : 0.9998		
Specificity : 0.9677		
Pos Pred Value : 0.9999		
Neg Pred Value : 0.9375		
Prevalence : 0.9967		
Detection Rate : 0.9965		
Detection Prevalence : 0.9966		
Balanced Accuracy : 0.9838		
'Positive' Class : 0		

• 분류모델 혼동행렬 결과

제 1 공장	Specificity	Kappa	제 2 공장	Specificity	Kappa
1번 위치 24시간	0.9911	0.9940	1번 위치 24시간	0.9764	0.9865
1번 위치 48시간	0.9929	0.9964	1번 위치 48시간	0.9847	0.9915
2번 위치 24시간	0.9797	0.9896	2번 위치 24시간	0.9927	0.9948
2번 위치 48시간	0.9888	0.9943	2번 위치 48시간	0.9762	0.9878
3번 위치 24시간	0.9818	0.9873	3번 위치 24시간	1.0000	0.9844
3번 위치 48시간	0.9839	0.9918	3번 위치 48시간	0.9770	0.9882

각 분류 모델의 정확도 수준은 Kappa값 기준 0.99정도로 예측력이 높은 것으로 판단됨

- 교차검증-GridSearch 기반 Cut-off 설정

	0	0.05	...	0.45	...	
Confusion Matrix 결과값 비교				오분류 감소 가장 높음		

- 검증데이터 검증 결과

참가번호 203266 의 모델 종합 정확도(CSI)는 36.11%, 평균 AUC는 0.8372 입니다.

Cut-off 설정 후 검증 결과, CSI 점수 36.11 , 평균 AUC 0.8372로 평가됨

5. 활용방안

최적의 습도를 찾고



결로 취약 위치를 찍는



DEW **POINTER**

• 조건부확률 기반 최적 습도 도출

< 결로 발생 예상 시점 발견 >

습도	온도	코일온도	...	결로확률	결로 발생
78	22	10		0.7	1

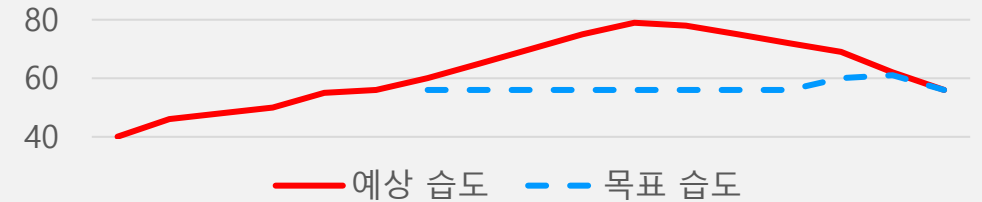
< 습도 시나리오별 최적 결로확률 도출 >

습도	온도	코일온도	...	결로 확률	결로 발생
1	22	10		0.0	0
2	22	10		0.0	0
...					
56	22	10		0.3	0
...					
67	22	10		0.5	1
...					
100	22	10		0.7	1

24시간, 48시간 후 결로 발생 시
습도 1~100의 시나리오 중 최적 습도 검색

• 최적 습도 기반 대응방안 제안

< 결로 발생 차단 습도 유지 >



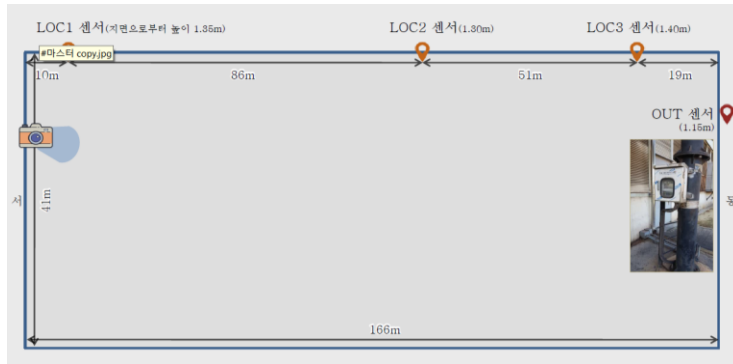
결로가 발생하지 않는 시점까지 최적 습도 유지

< 4단계 결로 경보 시스템 >

안전 단계	결로 발생 확률이 충분히 낮아 안전한 수준
환기 단계	결로발생이 예상되나 외부 습도가 충분히 낮아 환기로 결로가 예방이 가능한 단계
제습 단계	결로 발생이 예상되고 외부습도가 높아 외부공기 차단 및 제습기 가동이 필요한 단계
야간 제습 단계	새벽시간에 결로발생이 예상되어 제습기의 자동 가동을 위해 퇴근 전 외부공기 차단이 필요한 단계

• 역거리법 기반 결로 취약위치 세분화

< 공장 센서 위치 기반 격자별 결로 확률 >



0.0	0.3	0.3	0.4	0.6	0.8
0.0	0.2	0.3	0.4	0.7	0.8
0.1	0.2	0.2	0.5	0.6	0.7
0.1	0.1	0.2	0.4	0.5	0.6
0.0	0.1	0.1	0.3	0.4	0.5
0.0	0.0	0.0	0.1	0.3	0.4

< 공장 센서 위치 기반 격자별 결로 확률 >

1. 공장 위치 격자를 설정하고 3곳 센서의 습도, 온도, 코일온도 평균값 적용
2. 각 격자의 3곳 센서의 거리의 반비례하는 가중을 두어 습도, 코일온도, 습도 조정
3. 격자별 결로 발생확률 예측

0.5	0.3	0.5	0.5	0.5	0.8
0.5	0.5	0.5	0.5	0.5	0.5
0.5	0.5	0.5	0.5	0.5	0.5
0.5	0.5	0.5	0.5	0.5	0.5
0.5	0.5	0.5	0.5	0.5	0.5
0.5	0.5	0.5	0.5	0.5	0.4

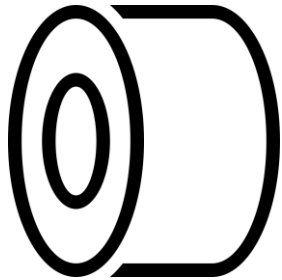
4. 세분화된 위치별 결로 확률에 따른 대처방법 적용

0.0	0.3	0.3	0.4	0.6	0.8
0.0	0.2	0.3	0.4	0.7	0.8
0.1	0.2	0.2	0.5	0.6	0.7
0.1	0.1	0.2	0.4	0.5	0.6
0.0	0.1	0.1	0.3	0.4	0.5
0.0	0.0	0.0	0.1	0.3	0.4

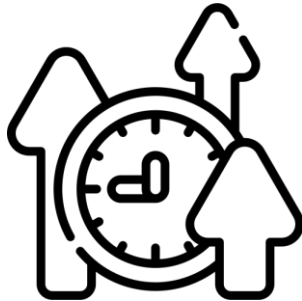
DEW POINTER

 최적 습도 제안

 결로 위치 세분화



제품 완성도



자원 효율성



공정 개선 기회

현대제철 제품 경쟁력 및 생산역량 향상

감사합니다!

Q&A



콘테스트 공모 문제해결 과정별 팀원 참여도

구분	김정의	송지수	정현준
문제이해 및 자료조사	40	30	30
데이터 전처리	30	25	45
데이터 모델링	25	45	30
분석결과 정리 및 보고서 작성	30	30	40
활용 방안 아이디어 제시	30	30	40