

MAKINA PEBBLES

MACHINE INTELLIGENCE FOR MANUFACTURING

Creatively Designed and Beautifully Engineered

2019. 08

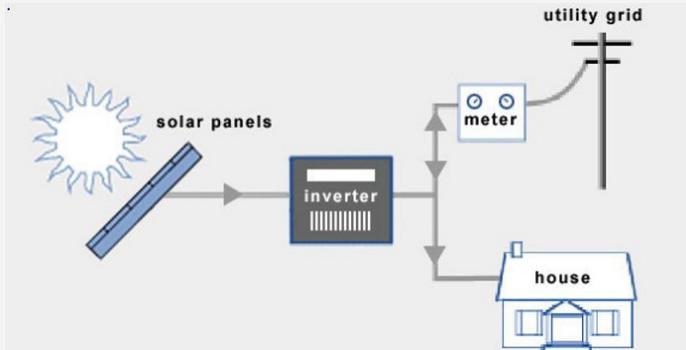
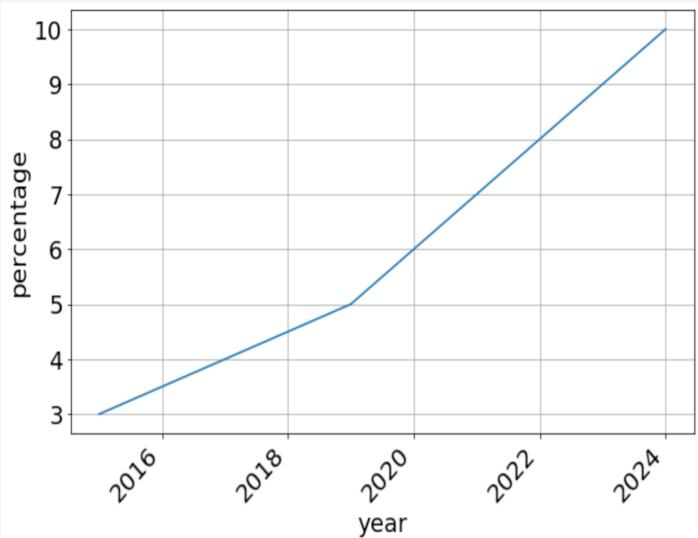
1. 태양광 발전량 예측의 필요성

Renewable Portfolio Standard (RPS)

On-Grid Issue

화력발전의 유연운전

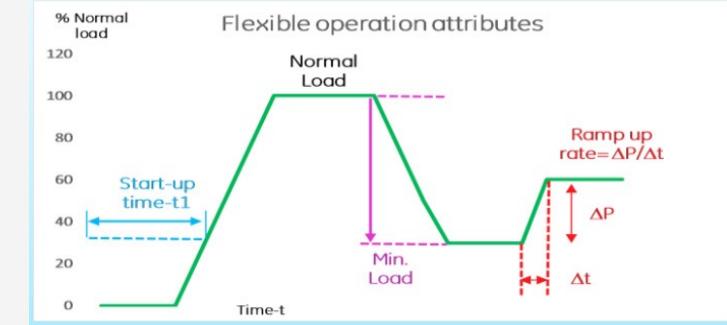
<2024년까지 공급의무자의 신재생에너지 공급의무비율>



신재생에너지의 불확실성 + 변동성
예비력 요구량 및 백업비용 증가
스케줄링이 가능한 발전원으로서의 가치 증대

RPS로 인한
태양광 발전 공급 확대 및
그에 따른 이슈 발생

전력구매자가 신재생에너지를
On-Grid System에 편입시키기 위해서는
발전량 예측이 필요



Source : Flexible Operation of Thermal Power Plant For Integration of Renewable Generation (2019)

화력발전의 유연운전을 통해
원가 절감, 오염배출 배출량 감소

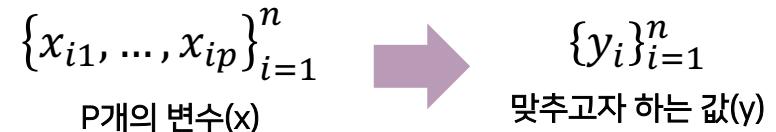
"유연운전 스케줄링, 설비 수명 보전 위해
신재생에너지 발전량 예측 필요"

Source : 'Machine Learning Can Boost the Value of Wind Energy', DeepMind

Source : '전력연구원, 2023년까지 '지능형 디지털 발전시스템' 개발', EPJ

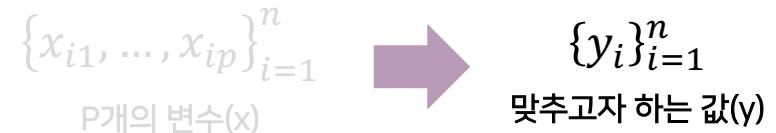
2. 문제정의 : Time Series Forecasting

Regression



- 예측을 위한 P개의 변수가 주어짐
 - 학습 구간과 예측 구간의 데이터 분포가 동일함

Time Series Forecasting



- Input으로 사용할 해당 시점의 변수가 존재하지 않거나
 - 학습 구간과 예측 구간의 데이터 분포가 다름

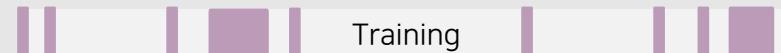
Data Leakage

Case 1) Interpolation



- 예측 시점보다 미래의 데이터가 Training Set에 포함되어 모델이 미래를 미리 보고 예측을 수행하는 경우

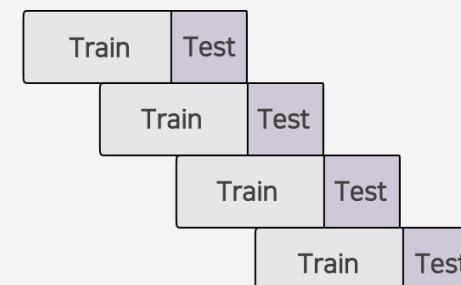
Case 2) Random Split Cross Validation



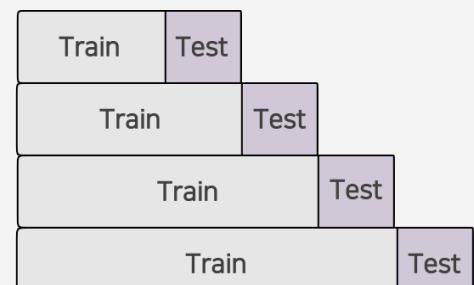
- 데이터를 섞은 다음 Cross Validation을 수행하여
마찬가지로 모델이 미래를 보고 예측을 수행하는 경우

- 모델은 실제 운용상황과 같은 시나리오로 검증되어야 함
 - Train Set에 Overfitting된 모델로는 미래를 예측할 수 없음
 - Cross Validation Score로 발전량 예측 모델을 평가할 수 없음

- LightGBM Cross Validation 결과
 - Train R2 : 0.9619, Test R2 : 0.9600
 - 모델의 실제 Forecast 성능이라고 할 수 없음



1) Walk-Forward Optimization

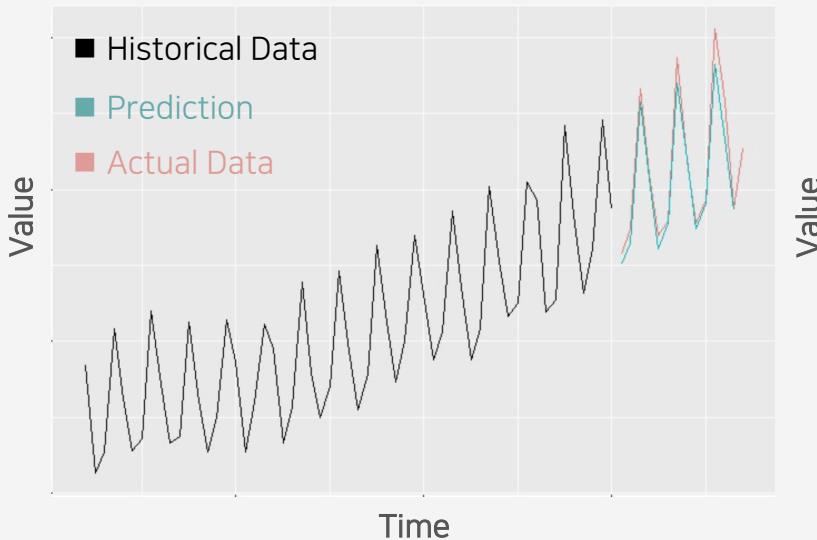


2) Nested Cross Validation

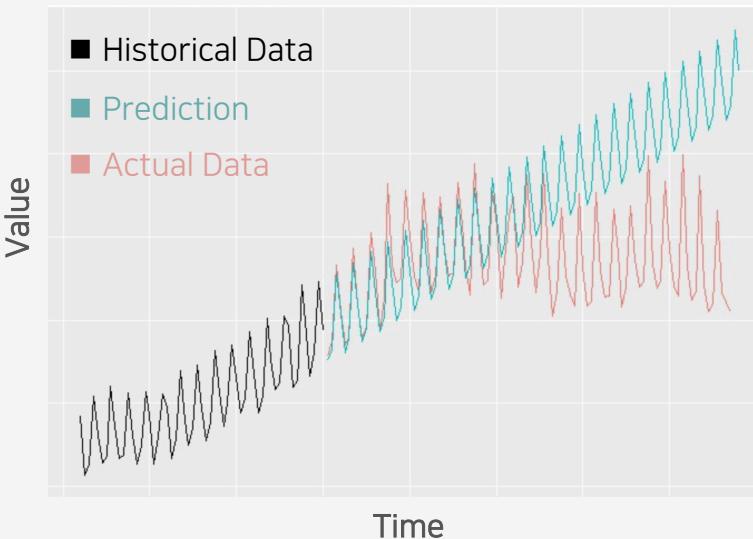
■ 2. 문제정의 : Non-Stationary Distribution

Time Series Data의 특성상 그 Distribution은 시간에 따라 변함

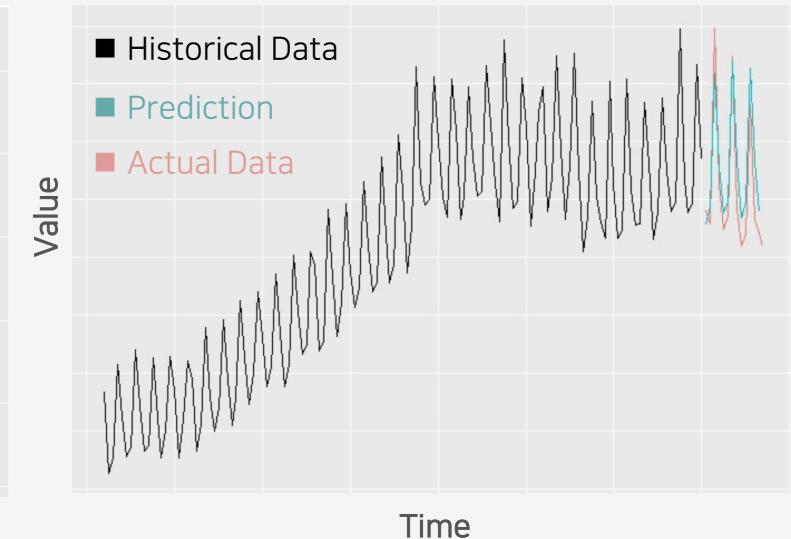
Train된 모델



운용 도중 Distribution이 바뀐 경우



새로운 데이터로 Training 필요



Model Management & Continual Learning

“시간에 따라 변하는 데이터의 Distribution을 적합한 방법을 통해 지속적으로 모델에게 학습시켜야 함”

3. 사용 데이터

1

기상청 단기예보 데이터

날짜	오늘	내일(29일 목)									
시각	21	24	03	06	09	12	15	18	21	2	
날씨											
강수확률(%)	20	20	30	30	70	70	60	20	20		
강수량(mm)	-	-		10~19mm	10~19mm					-	
최저/최고(°C)	-/-				22/28						
기온(°C)	23	22	22	25	25	27	26	24	2		
풍향/풍속(km/h)											
습도(%)	90	95	95	85	80	70	75	80	8		

사용 변수	Feature Engineering
하늘상태	이론적 일사량/발전량
풍속	기대값(강수량)
습도	$\cos(\text{풍향})$
기온	$\sin(\text{풍향})$
풍향	파널 정면 방향 벡터로 분해된 풍속
강수상태	Polynomial Features
강수확률	Categorical Encoding
강수량	Location
적설량	
일최고기온	
일최저기온	

2

SK Weather Planet (웨더퐁) 단기예보



고해상도

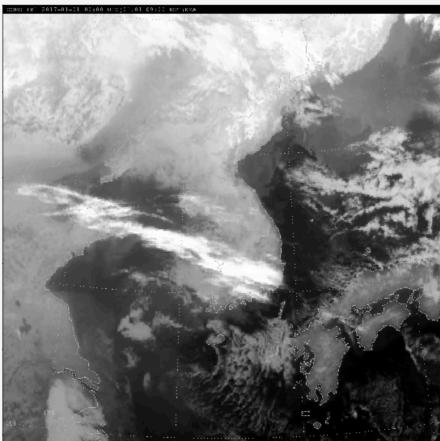
SK텔레콤 기지국 인프라에
기상 관측 센서를 설치, 전국 1~3km
해상도의 고해상도 기상 관측망

정확성

고해상도 기상 관측망 구축으로
기상청 데이터보다 정확한 관측정보
제공

3

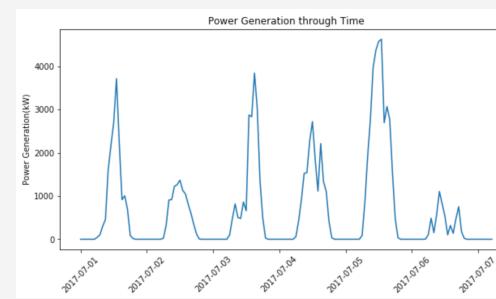
천리안 위성사진



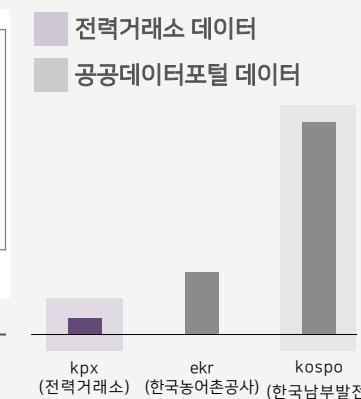
- 국가기상위성센터 제공
- 224 x 224 사이즈로 전처리
- 흑백으로 전처리
- 수집기간 : 2017.7 ~ 2019.7
- 구름의 흐름을 예측하고자 사용
- CNN-LSTM Model의 Input으로 사용되어,
52시간동안의 발전량을 예측

4

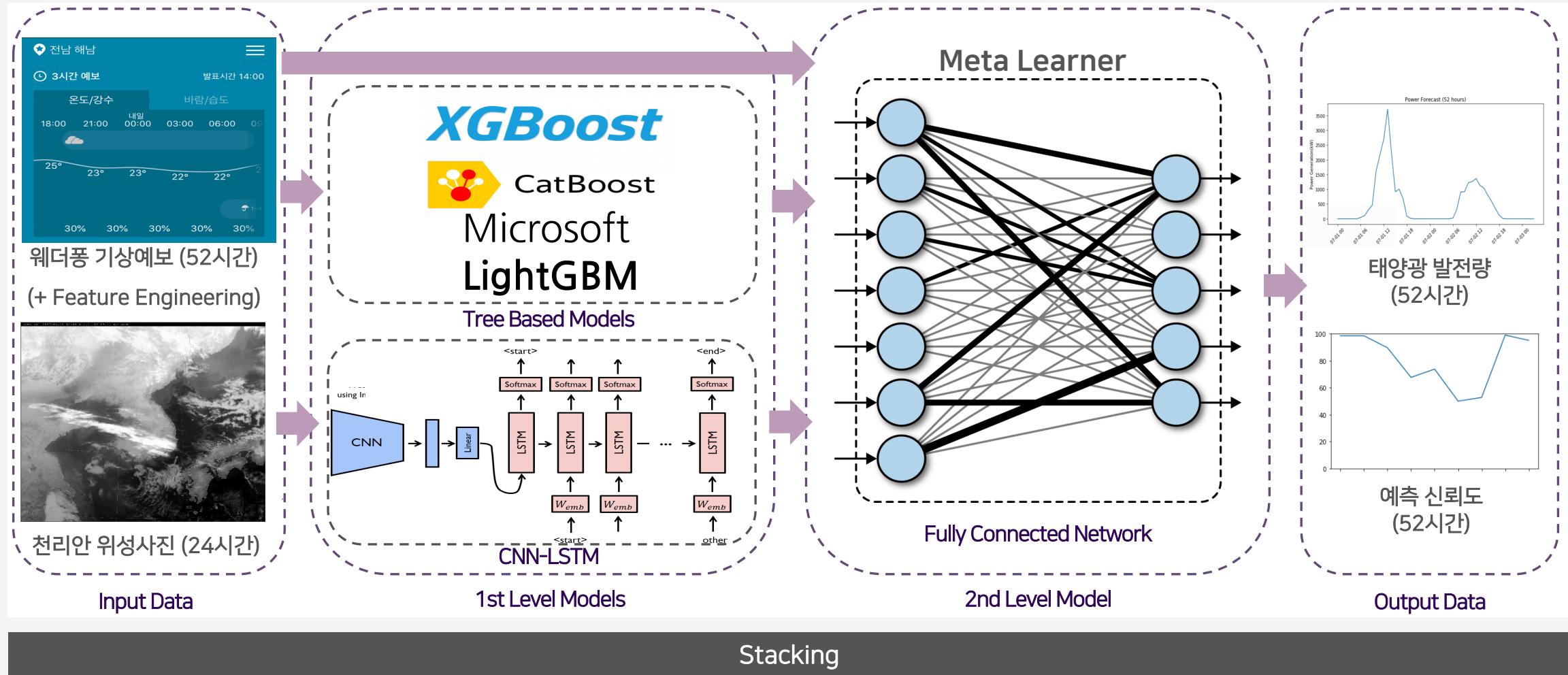
전력거래소 제공+ 타 발전소 태양광 발전량 데이터



각 발전소의 발전량 Location-invariant하게
인코딩하여 Data Aggregation으로 모델 성능 향상



4. 사용 모델 (Stacked Regressor)

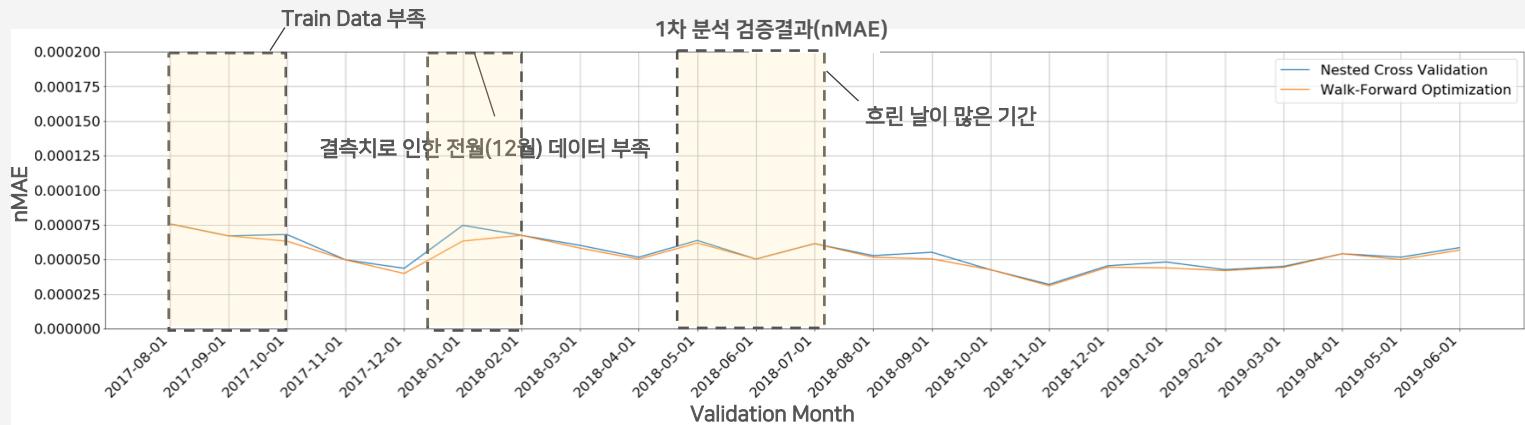


"Input Sample과 각 모델의 Prediction을 활용, 각 모델의 장점은 살리고 단점은 상쇄"

5. 초도결과

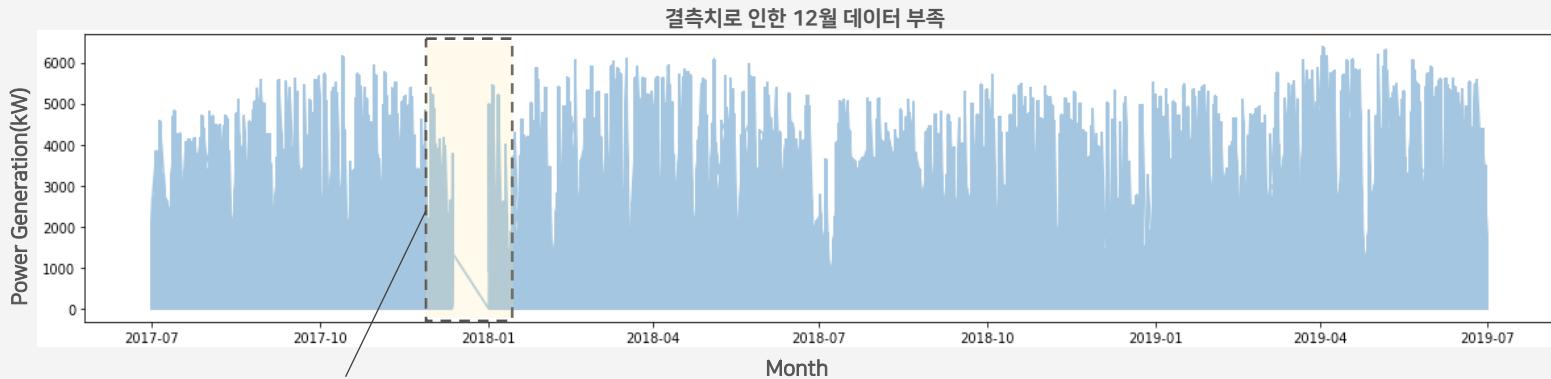
실험 세팅

- 1) 사용 데이터
 - 1) 웨더풀 기상예보
 - 2) 태양광 발전량(전력거래소 제공)
- 2) 사용 모델 (Stacked Regressor)
 - 1) Tree Based 1st Level Models
 - 2) FCN 2nd Level Model



검증결과

- Validation nMAE : 0.000053
 - 검증기간 : 2017.08.01 ~ 2019.06.01
- Test nMAE : 0.000056
 - 검증기간 : 2019.08.12 ~ 2019.08.14
- Generalization이 잘 된 모델
(Validation Score과 Test Score의 차이가 적음)



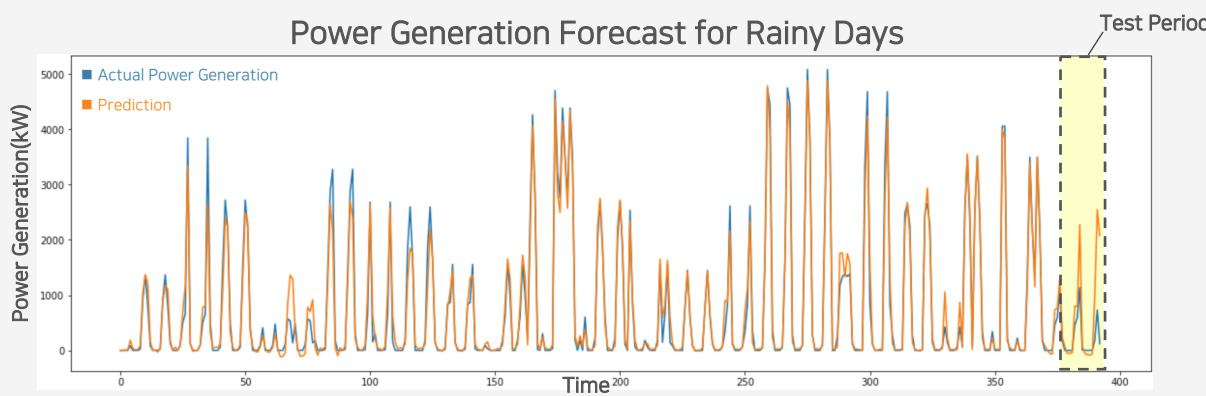
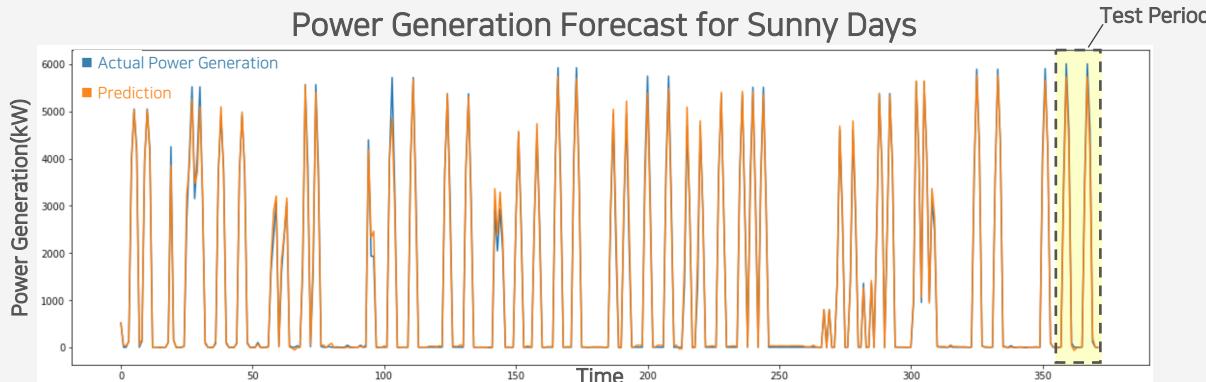
“기상예보 결측값(강수상태)로 인해 12월 데이터 누락”

초도 분석 결과

“Train Data가 부족할 때, 직전 데이터가 업데이트되지 않았을 때, 흐린 날이 많을 때 예측 정확도 저하”

6. 흐린 날 예측의 어려움

맑은 날은 적은 데이터로도 높은 정확도로 예측 가능, 흐린 날 예측 성능이 중요



맑은 날은 적은 양의 데이터로도 높은 정확도로 예측 가능

- Training Set : 372시간
- Test Set : 52시간
- Train nMAE : 0.000006
- Test nMAE : 0.000048
- Train R2 : 0.9945
- Test R2 : 0.9921

전체 기간 맑은날, 흐린날 nMAE 비교



1. Unseen Pattern

흐린날의 기상 패턴은 맑은날보다 다양함

2. Forecast Error

기상예보 오차로 인한 발전량 예측의 오차

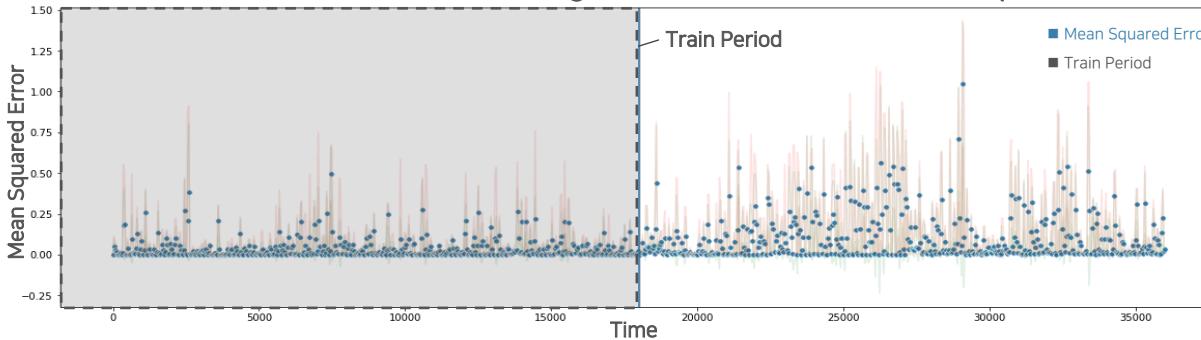
모델의 정확도 향상을 위해서 흐린 날 예측 성능 향상이 필수적

Unseen Pattern, Forecast Error로 인한 흐린 날 예측 성능 저하를 해결할 방법?

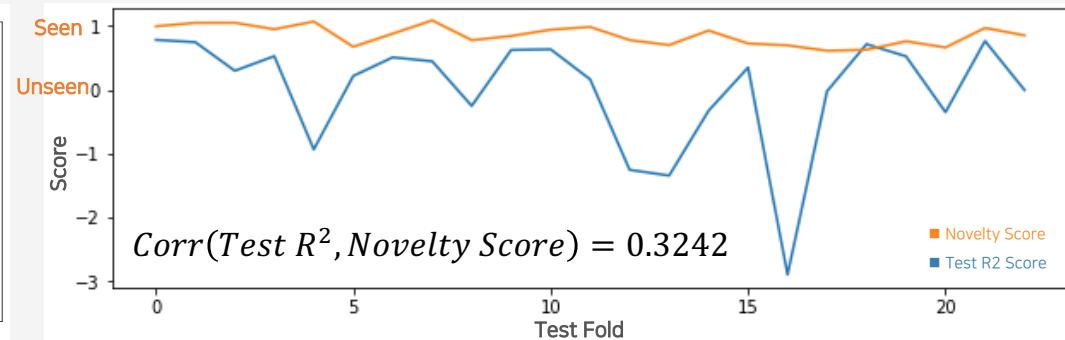
7. Novelty (Unseen Pattern)

Stationary한 맑은 날의 기상에 비해 흐린 날은 기상은 훨씬 다양한 패턴을 가짐

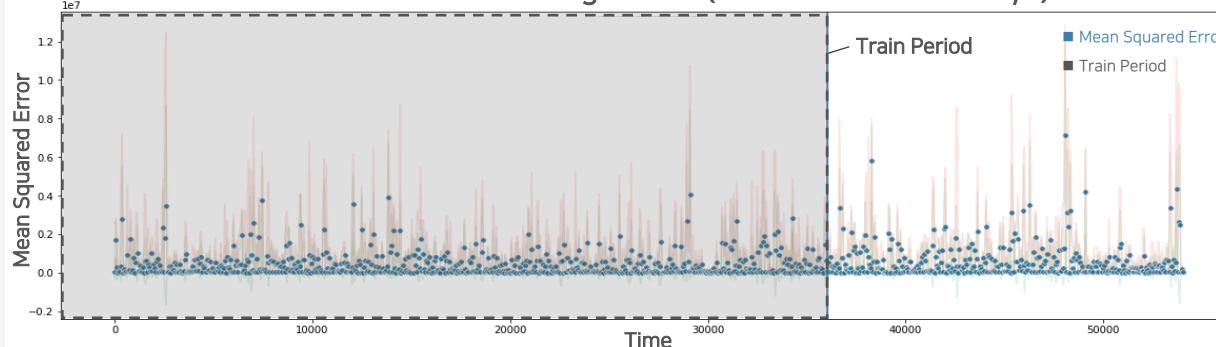
Prediction Error through Time (Train Period = 30 days)



Correlation between Novelty Score, Test R2 Score



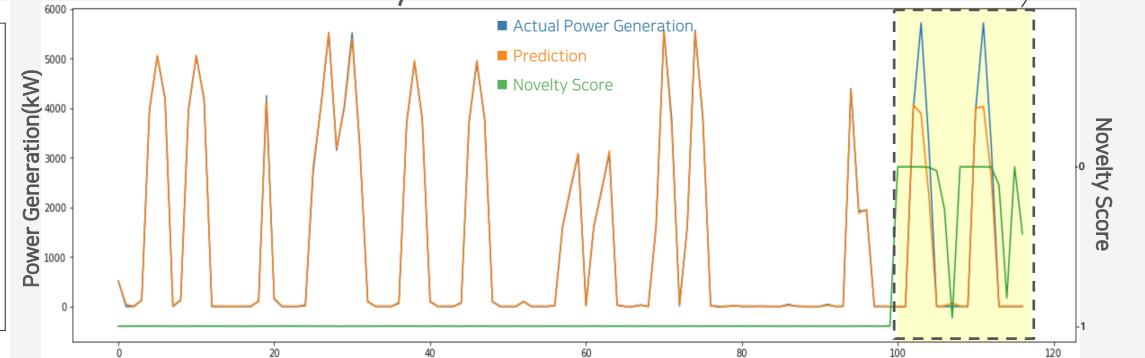
Prediction Error through Time (Train Period = 60 days)



Transfer Learning

"타 발전소 데이터 Location-Invariant Encoding하여 학습에 활용"

Novelty Score and Prediction Error



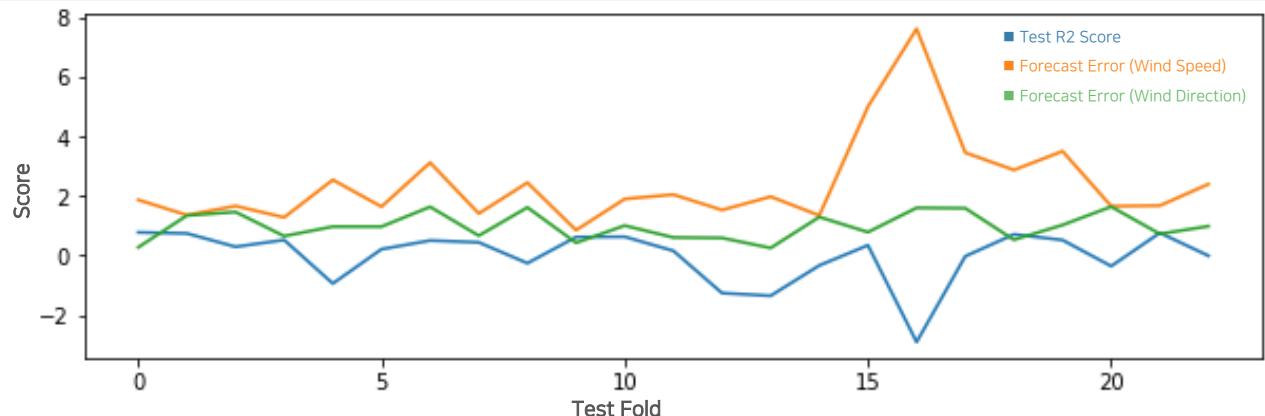
Forecast Confidence

"Novelty Score를 통해 예측 신뢰도 제공"

7. Forecast Error

단기예보가 구름의 흐름을 제대로 예측하지 못하는 경우 예측 정확도 저하

Correlation between Forecast Error, Test R2 Score



Train Set Error

Train Set에서 Threshold 이상의 Forecast Error를 가진 Sample을 제외, 모델이 잘못된 관계를 학습하는 것을 방지

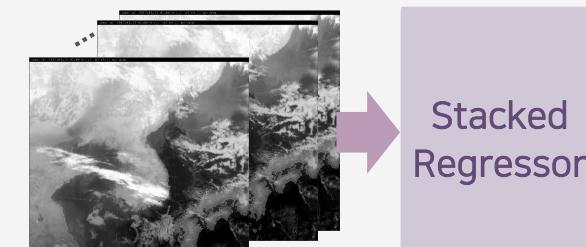
Test Set Error

천리안 위성사진 활용하여 모델의 예측이 틀리는 원인인 구름의 흐름을 직접 예측

천리안 위성사진 활용

$$\text{Corr}(\text{Forecast Error(Wind Speed)}, \text{Test } R^2) = -0.5336$$
$$\text{Corr}(\text{Forecast Error(Wind Direction)}, \text{Test } R^2) = -0.2129$$

- 풍속, 풍향 예측오차가 커질수록 Test Score가 감소
- 흐린 날 태양광 발전을 가장 방해하는 요인은 구름의 흐름
- 구름의 흐름을 나타내는 풍속, 풍향을 틀렸을 때 예측오차가 증가



위성사진

모델

태양광 발전량

“CNN-LSTM으로 구름의 흐름을 예측하여 흐린 날 예측 정확도 향상”

8. 최종결과

실험 세팅

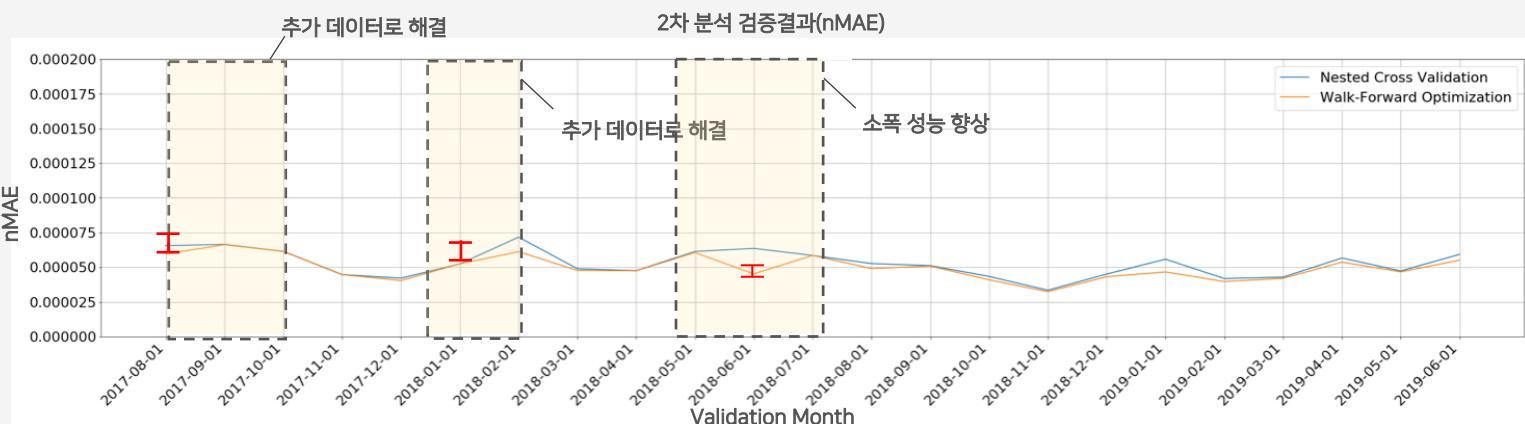
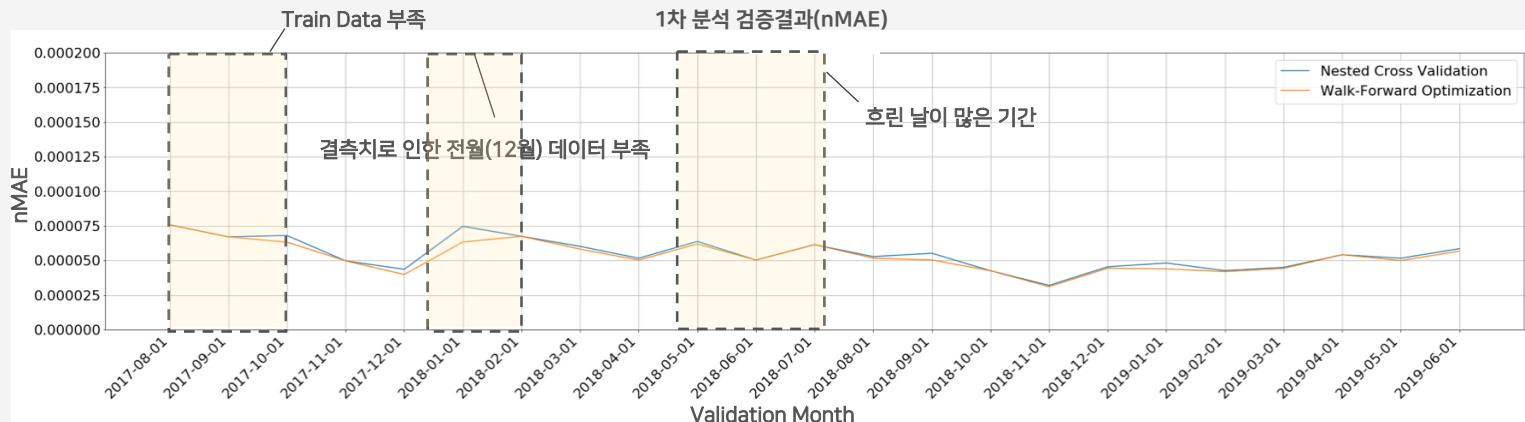
- 1) 사용 데이터
 - 1) 웨더퐁 기상예보
 - 2) 천리안 위성사진
 - 3) 태양광 발전량(전력거래소, **한국농어촌공사, 한국남부발전**)
- 2) 사용 모델 (Stacked Regressor)
 - 1) Tree Based 1st Level Models
 - 2) CNN-LSTM 1st Level Model
 - 3) FCN 2nd Level Model

검증결과(초기 모델)

- Validation nMAE : 0.000053
- Test nMAE : 0.000056

검증결과(최종 모델)

- Validation nMAE : 0.000050
- Test nMAE : 0.000024

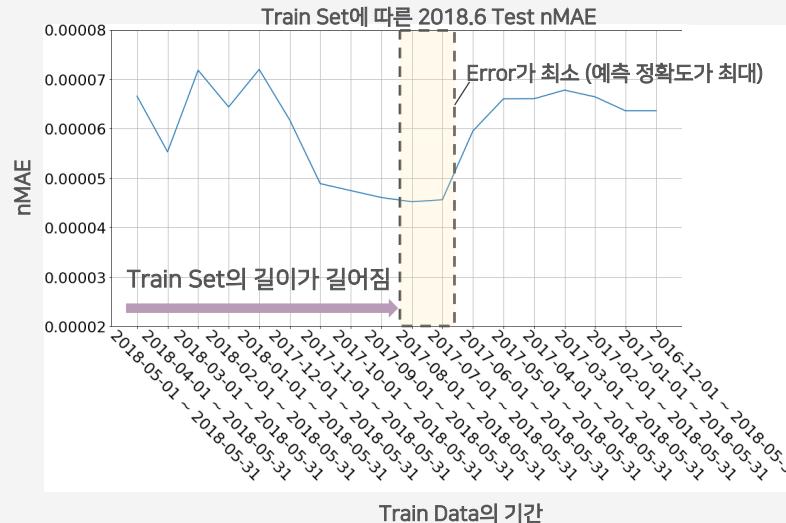


최종 분석 결과

"Location-invariant 인코딩 기술, 위성사진 활용 모델로 흐린 날 예측 성능 향상"

9. Future Work : Model Management

모델 산업 적용의 어려움 : 모델의 온라인 적용은 배포를 포함한 운영 측면의 요소들을 고려하여야 함



태양광 발전량 예측에서 발견된 문제점

Train Data가 추가됨에 따라 계속 성능이 올라가는 것이 아니라,
어느정도까지 성능이 증가하다가 최고성능을 찍고 난 뒤에는 추가로
들어오는 과거데이터가 모델의 성능을 저하시킴

-> Model Management, Continual Learning 기술 필요

Model이 가지는
Conflict Objectives

Continual Learning
의 필요성

Triggering Policy

1. Continual Learning

모델은 가장 최신의 Distribution과 $y = f(x)$ 관계를 학습해야 함

2. Forgetting

모델은 오래된 Distribution, 현재에 유효하지 않은 $y = f(x)$ 를 잊어버려야 함

3. Training Strategy

Train Set의 길이는 모델의 학습시킬 수 있을 만큼 충분히 길어야 함

1. 예측할 수 없는 환경변화에 능동적으로 대응할 수 없음

2. 모델의 초기 성능을 유지하는 것이 불가능

1. Time window-based Policy : 일정한 주기마다 업데이트
2. On-demand Policy : Event-base로 업데이트

"Continual Learning을 통해 지속적으로 변화하는 환경에 대응하는 시스템 구축"

■ 9. Future Work : Forecast Confidence

Novelty Score과 과거 Error를 기반으로 Forecast Confidence 제공

1 Out-of-Distribution Detection

- 목적 : 새로운 Sample이 In-Distribution (Training Set에서 본 Sample)인지, Out-of-Distribution (Training Set에서 보지 못한 Novelty)인지 탐지하여 이를 반영한 예측 신뢰도 도출

2 관련 논문

- 'A Baseline for Detecting Misclassified and Out-of-Distribution Examples in Neural Networks (2017)', Dan Hendrycks et al.
- 'Learning Confidence for Out-of-Distribution Detection in Neural Networks (2018)', Terrance DeVries et al.

3 Sub-Model for Providing Forecast Confidence

- Input Sample, Novelty Score를 통해 Prediction Error를 예측하는 Sub-Model을 개발
- Post-Processing 하여 Forecast Confidence로 활용

“Forecast Confidence 제공하여 모델의 활용성 개선”