

Duck Curve

이형권, 장성호, 김태연, 김현진
한국공학대학교

POSTECH X DERShare
태양광 발전량 예측 경진대회

CONTENTS

01 서론

02 EDA

03 태양광 발전 모델링

04 결론 및 향후 계획

서론

01 서론

Data set

Data Table	항목	설명
Gens	ID Time Amount	해당 발전소 ID 한시간 단위 해당 발전소의 해당 시간대 발전량
Envs	ID Time Nins Mtemp	해당 발전소 ID 기상센서의 계측시간 일사량계 계측 값 태양전지 모듈 표면 온도 계측 값
Weather1	ID Time Temperature Humidity Dew_point Wind_dir Wind_spd Uv_idx Visibility Cloudiness Celling Pressure Precip_1h	해당 발전소 ID 관측시간 기온, °C 습도, % 이슬점 온도, °C 풍향, ° 풍속, km/h 자외선 강도, 지수(0~12) 시정, km 운량, % 최저운고, m 기압, mb 강수량, mm

01 서론

Data set

Data Table	항목	설명
Weather2	ID Time Temperature Humidity Wind_dir Wind_spd Cloudiness Pressure Precip_1h	해당 발전소 ID 관측시간 기온, °C 습도, % 풍향, ° 풍속, km/h 운량, % 기압, mb 강수량, mm
Weather3	ID Time Temperature Humidity Wind_dir Wind_spd Precip_1h	해당 발전소 ID 관측시간 기온, °C 습도, % 풍향, ° 풍속, km/h 강수량, mm
Forecasts1,2,3	기존 weather1,2,3 항목에 Fcst_time Precip_prob Precip_1,6h 추가	예측 시점 해당시간에 예보된 강수확률 예측된 시간 이전 1, 6시간 동안의 예측된 강수량

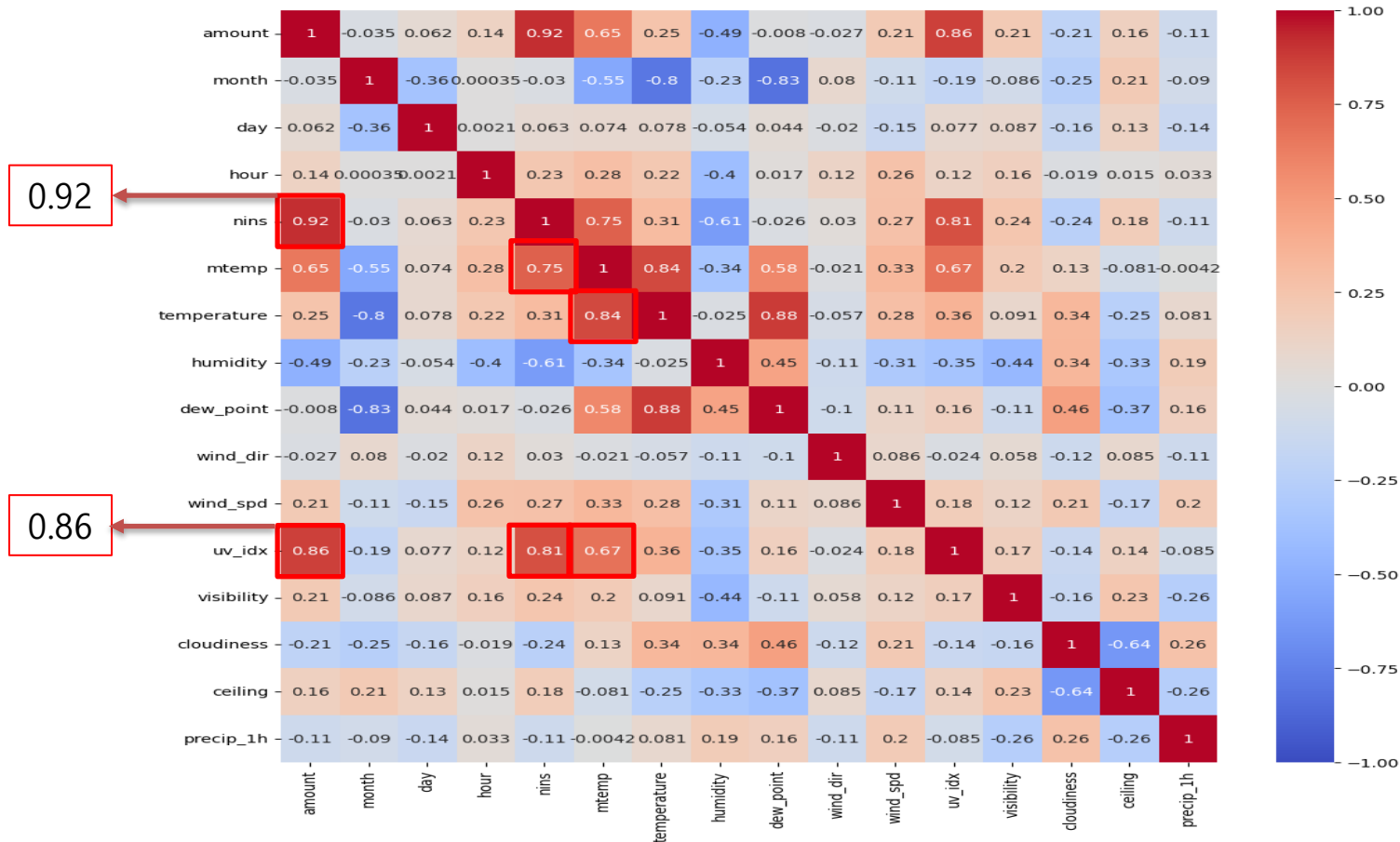
01 서론 - 발전소와 기상관측소

	id	capacity	wth1_id	wth1_dist	wth2_id	wth2_dist	wth3_id	wth3_dist	asos_station	asos_dist
0	0	472.39	1	4539	1	5486	1	2497	108	16196
1	1	30.00	2	1140	2	1745	2	867	119	14969
2	2	99.90	3	429	3	7929	3	2054	279	8111
3	3	97.90	4	355	4	5562	4	2391	279	17267
4	4	99.60	5	870	4	6749	5	1735	137	16766
5	5	50.00	6	1061	5	2351	6	2198	156	6842
6	6	70.00	7	2228	6	5921	7	2380	108	11351
7	7	99.00	8	633	7	2297	8	2382	108	10633
8	8	49.50	9	574	1	2026	9	2014	108	13931
9	9	36.00	10	377	8	4254	10	2481	108	12626
10	10	99.00	11	190	9	22494	11	1525	174	16968
11	11	100.00	12	752	10	10761	12	2761	254	11498
12	12	99.00	13	510	9	14308	13	1876	174	18284
13	13	99.00	14	1169	11	11039	14	1253	156	19369
14	14	100.00	15	521	12	3137	15	2061	252	5552
15	15	97.90	16	844	13	10285	16	2586	260	26012
16	16	99.00	17	829	14	12989	17	3272	165	14802
17	17	365.00	18	78	15	11304	18	1221	156	9336
18	18	97.20	19	551	14	19213	19	2215	268	18020
19	19	300.00	20	178	16	2497	20	461	244	2219
20	20	98.60	21	1415	17	4934	21	2059	129	2607

EDA

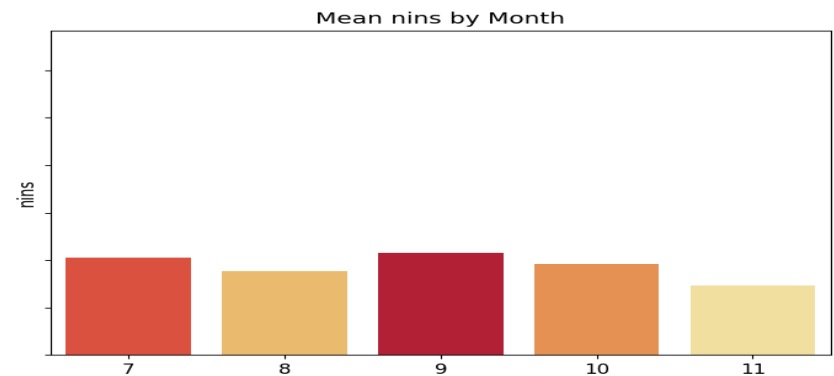
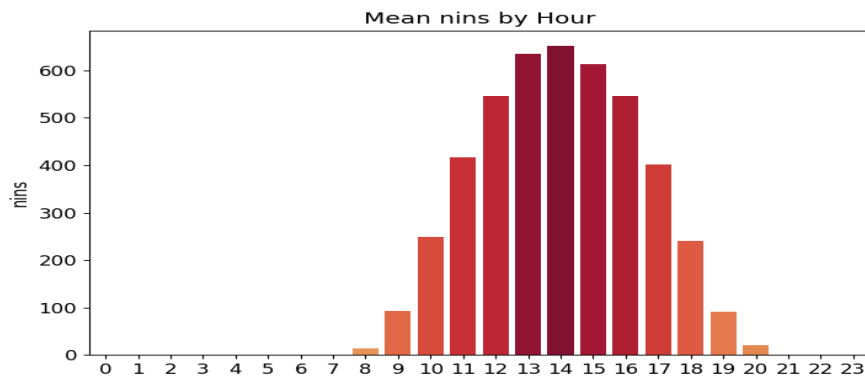
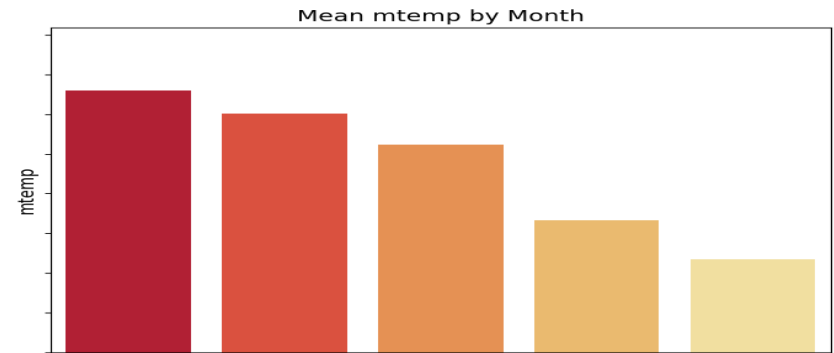
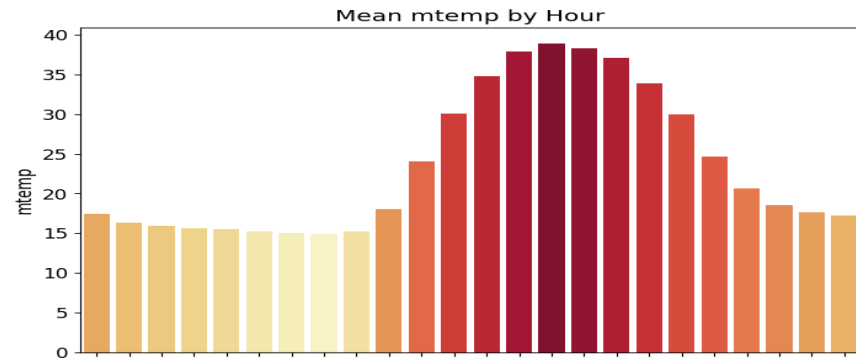
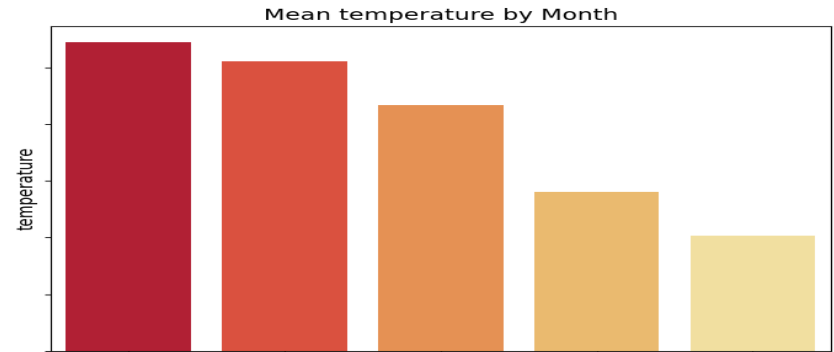
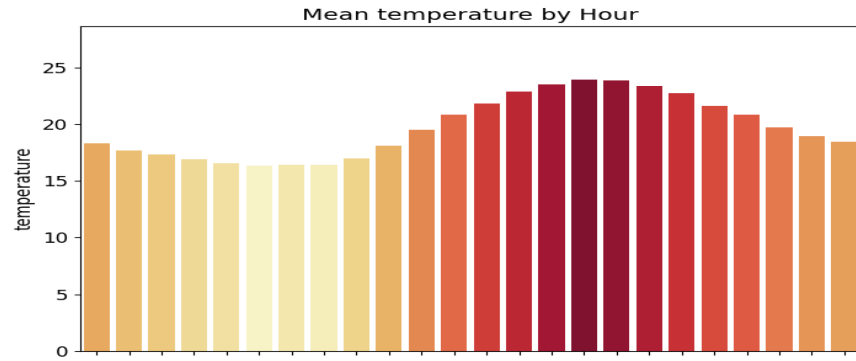
02 EDA - 변수간의 관계

Heatmap - gen0

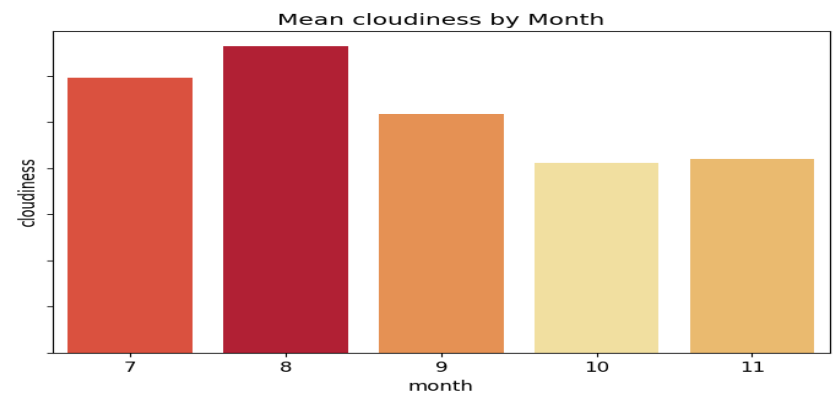
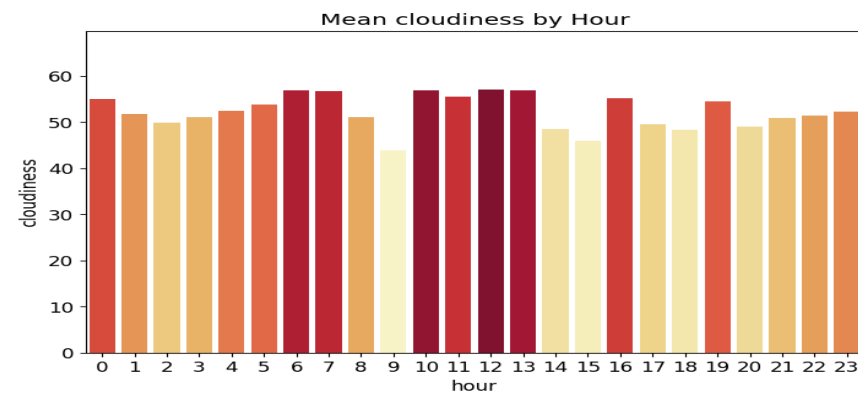
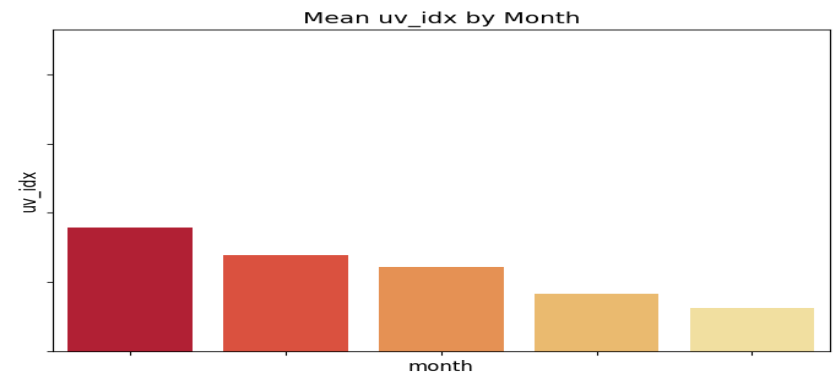
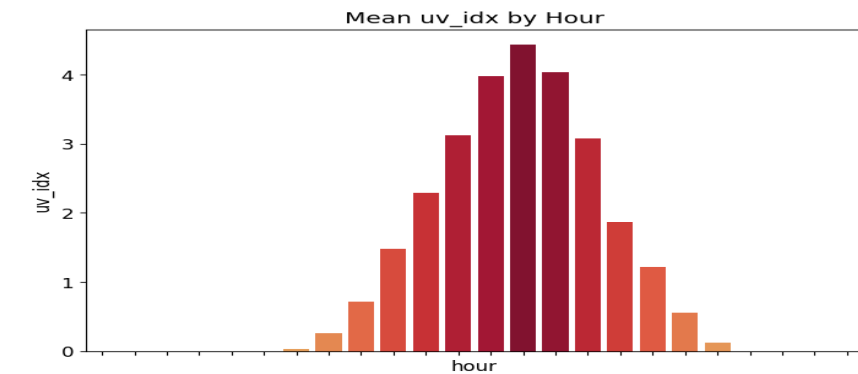
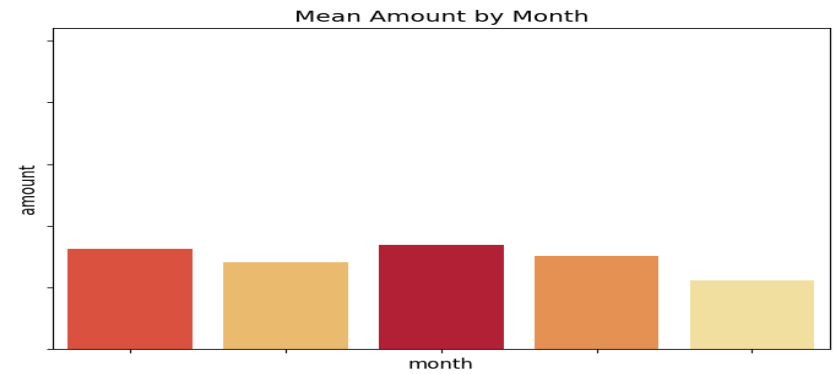
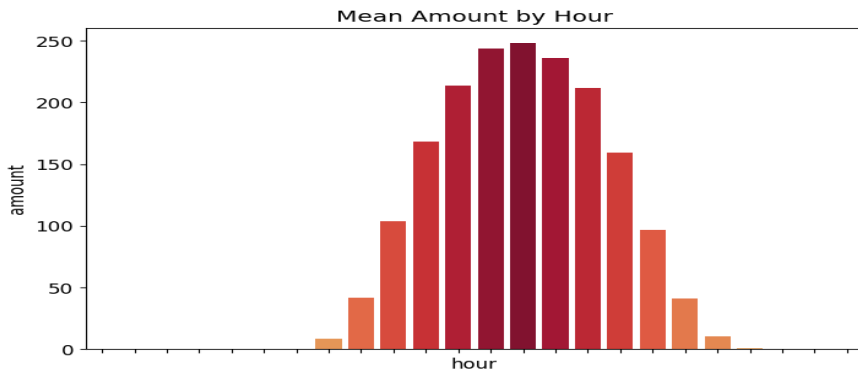


Amount와 상관계수가 높은 변수는 nins, uv_idx
Nins는 mtemp와 uv_idx
Mtemp는 temperature와 uv_idx

02 EDA - 시간 당 평균 및 월별 평균



02 EDA - 시간 당 평균 및 월별 평균



태양광 발전 모델링

Hyper Parameter

XGBoost, AdaBoost Regression은 대표적인 앙상블 ML 기법으로 ML의 과적합 문제에 강건한 특성을 가지며, 안정적인 예측이 중요한 태양광 발전량 예측 문제에 적합함

XGBoost Regression

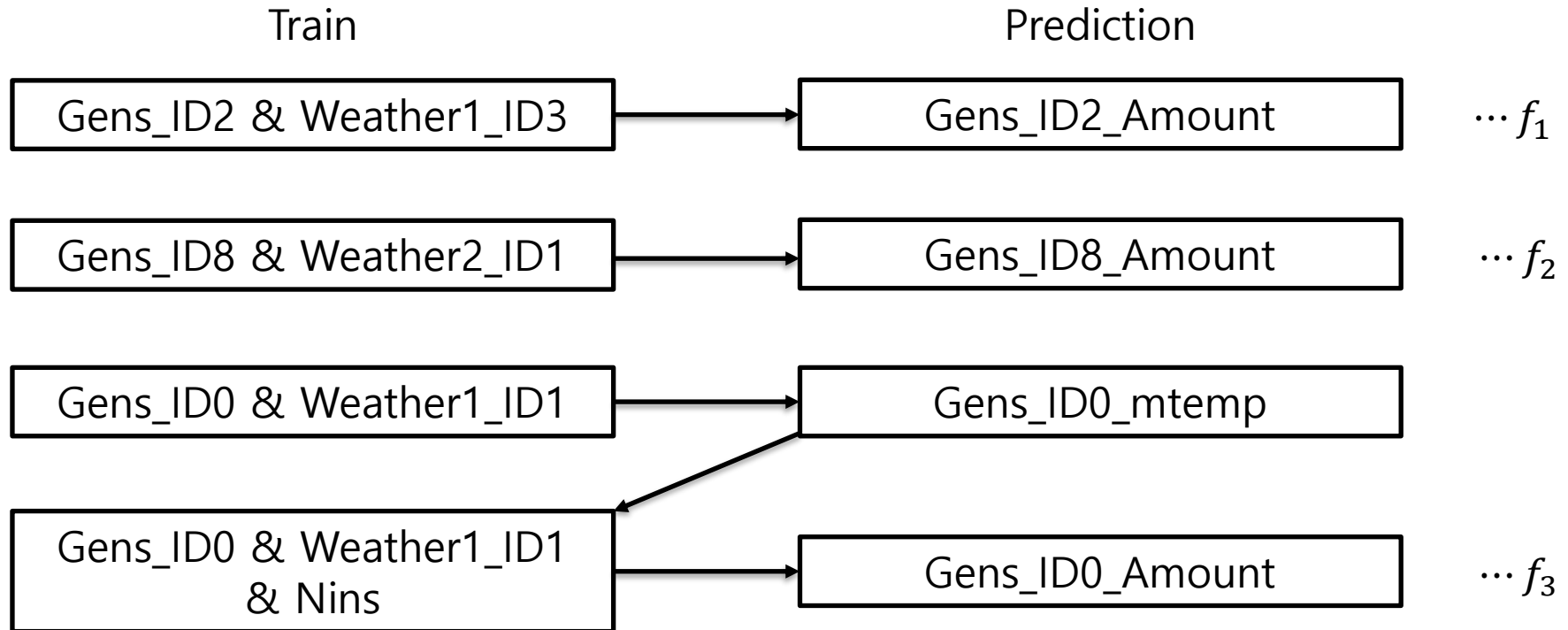
	Max_Depth	N_Estimators	Learning_Rate	Loss	R2
Mtemp	17	170	0.05	Linear	0.91

AdaBoost Regression

	Max_Depth	N_Estimators	Learning_Rate	Loss	R2
Nins	14	190	0.1	Linear	0.943
Gen0	12	175	0.05	Square	0.95
Gen2	12	135	0.1	Square	0.96
Gen8	12	170	0.1	Linear	0.935

03 모델링

Ensemble Model



ML의 근본적인 문제점인 학습되지 않은 데이터에 대해서 모델링의 제약이 존재
다양한 데이터 학습을 위해 다른 지역의 데이터 사용 최종 식은 <수식 1>과 같음

$$\hat{y} = \frac{f_1 + f_2 + f_3}{3}$$

< 1 >

Moving Average

태양광 발전 공식은 <수식2>으로 표현할 수 있으며

$$\text{Power Generation(kW)} = \text{Generator Capacity(kW/h)} * \text{Insolation(h)} \quad < 2 >$$

<수식 2>을 통해 발전량이 Insolation에 비례하는 것 알 수 있음

EDA를 통해 연관성이 높은 변수들을 파악할 수 있었고, 이를 통해 다변량 ML 분석을 사용

또한 선행연구를 통해 태양광 발전에 Time Series Analysis가 효과적인 것을 파악

다변량 ML과 시계열 분석을 같이할 수 있는 방안에 대한 모색

ML의 Output에 Moving Average를 사용해 노이즈를 필터링해 보다 안정적인 모델 제작

현재 ML모델은 $f(x_1, x_2, x_3 \cdots x_n)$ 으로 구성되어 있고 외삽 문제 등으로 추후 오차가 발생할 확률이 있고, 이를 $f(h)$ 를 이용한 시계열 분석을 사용해 Hybrid하는 것이 최종 목표

03 모델링

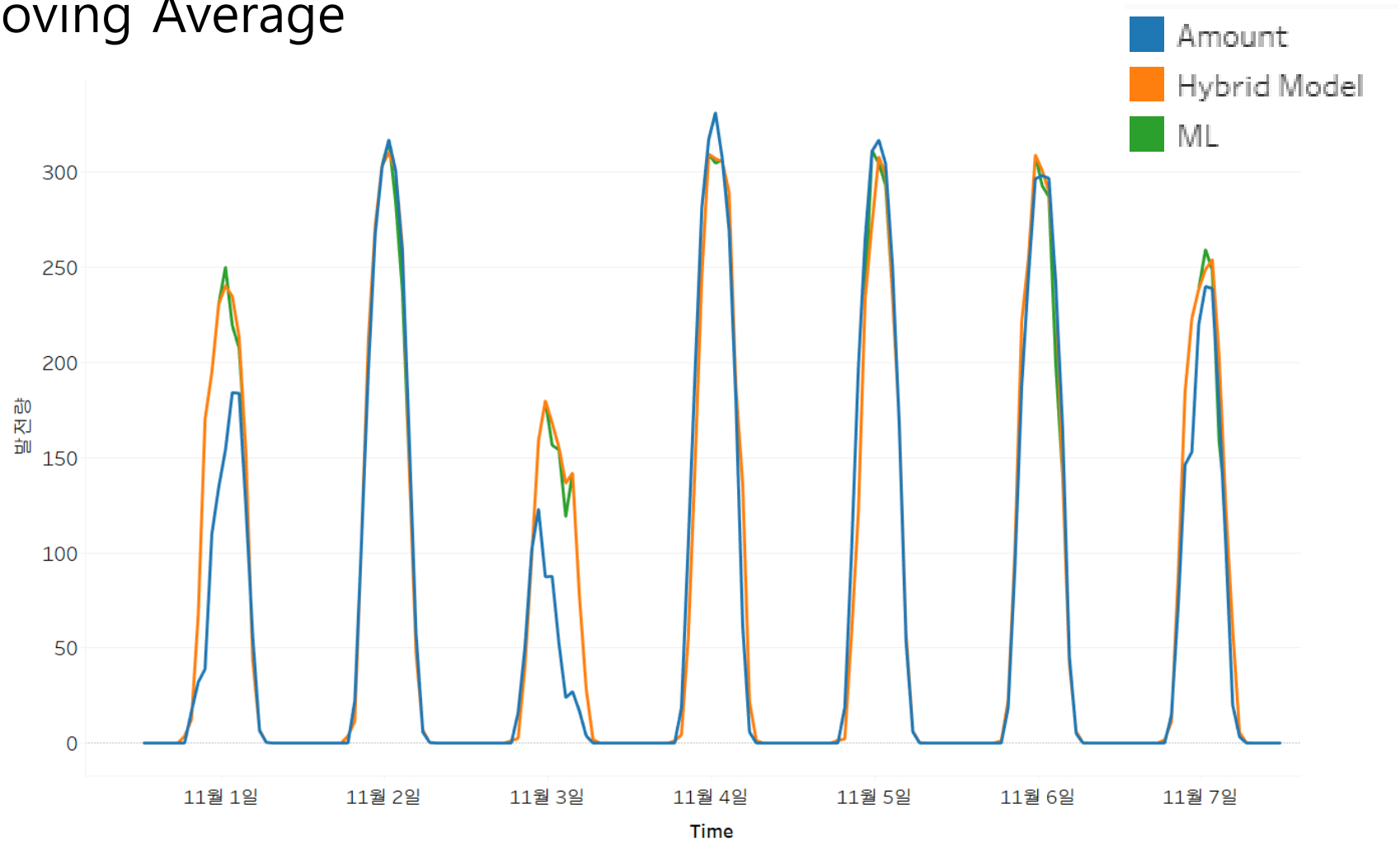
Moving Average

Time	ML Model	Hybrid Model	MA(2)	MA(3)	MA(4)
22-11-01	1.516554	1.494666	1.461888	1.488164	1.519675
22-11-02	0.985425	1.033768	1.311841	1.895737	2.472662
22-11-03	1.499788	1.441852	1.626103	1.829362	1.979579
22-11-04	1.453834	1.424473	2.006493	2.587715	3.173254
22-11-05	1.232199	1.254815	1.74895	2.321157	2.904118
22-11-06	1.104115	1.10076	1.125376	1.668569	2.233995
22-11-07	1.089628	1.080891	1.152877	1.398703	1.795282
Average	1.268792	1.261604	1.490504	1.884201	2.296938

발전량 Peak Time인 13시~15시까지 Moving Average 적용을 통해 Noise제거
Hybrid Model은 기존 AdaBoost Regression Model에 5.7% 향상된 성능 보유

03 모델링

Moving Average



결론 및 향후 계획

04 결론 및 향후계획

결론

- ML 중 Boosting 모델로 태양광 예측 진행 시 Peak Time를 제외한 시간은 대회 Error 기준 0.05 미만의 값을 보이고 있음
- ML과 Time Series 분석을 Hybrid하면 ML의 정확도가 내려갈 시 Moving Average의 보정으로 인해 Noise제거로 Error값 감소

향후 계획

- Spatio Temporal 분석을 통해 모든 발전소와 기상관측소에 대한 정보를 거리에 따른 Weight를 가진 Input으로 받아 베이지안 업데이트를 통해 다른 신설 지역과 같이 양질의 데이터가 없는 지역에 적용 가능한 모델 제작
- Average Sunshine, Isolation, Solar Cell Array Output 등의 정보를 사용해 ML과 Physics Model을 결합하면 해당 지역에 맞는 모델 제작으로 외삽 문제에 대한 해결책 제시

감사합니다

Physics Hybrid Model

$$P_{PV}(\gamma) = \frac{\eta_C}{K_C} (G_t)^2, \quad 0 < G_t < K_C$$

$$P_{PV}(\gamma) = \eta_C G_t, \quad G_t < K_C$$

$$P_{PV.ave} = P_{PV}(\gamma) A Y_P (kWh)$$

$$= \text{Average Array Efficiency}(h) * \text{Capacity}(kW/h) * \text{Average Sunshine}(h)$$

$$= \frac{\text{Power generation of Sunshine}(kW)}{\text{Capacity}(kW/h)} * \text{Capacity}(kW/h) * \text{Average Sunshine}(h)$$

$$= \text{Power generation of Sunshine}(kW) * \text{Average Sunshine}(h)$$

$P_{PV.ave}$: 발전전력량(kWh)
 $P_{PV}(\gamma)$: 태양전지 어레이 출력(kW)
 A : 총 어레이 면적(m^2)
 Y_P : 시스템 가동시간($hour$)