

랜덤 포레스트 모형을 이용한 태양광 발전량 예측



통계적 참견 시점 x  한국동서발전|주|

오인광 김영선 이은지 임송현 오희준



모형 적합 02

- 랜덤 포레스트 알고리즘
- 단계적 예측 (2단계)

01 데이터 전처리

- 이상데이터 처리
- SOLAR 변수 생성

03 예보 시뮬레이션

- 스플라인을 통한 보간법
- 자동화된 발전량 예측





1

데이터 전처리

01 데이터 전처리 기상 데이터 전처리



◆ 15분 단위 데이터 추출

발전량 데이터의 15분 단위를 맞추기 위해 1분 단위의 데이터에서 15분 마다 순간 데이터 추출

◆ SOLAR 변수 생성

태양의 고도에 따라 변하는 이론상의 일사량 계산

◆ 전운량(CLOUD) 변수 변환

1시간 단위 데이터에 존재하는 전운량 정보를 Linear Interpolation을 통해 15분 단위로 바꾼 후 추가

◆ 강수량(RAIN) 변수 변환 & 강수여부(FRAIN) 변수 추가

누적강수량 변수를 차분한 뒤, Linear Interpolation을 사용해 15분간 강수량으로 변환

태양고도에 따라 바뀌는 이론 상의 일사량을 계산하는 식을 구하여 SOLAR 변수를 생성

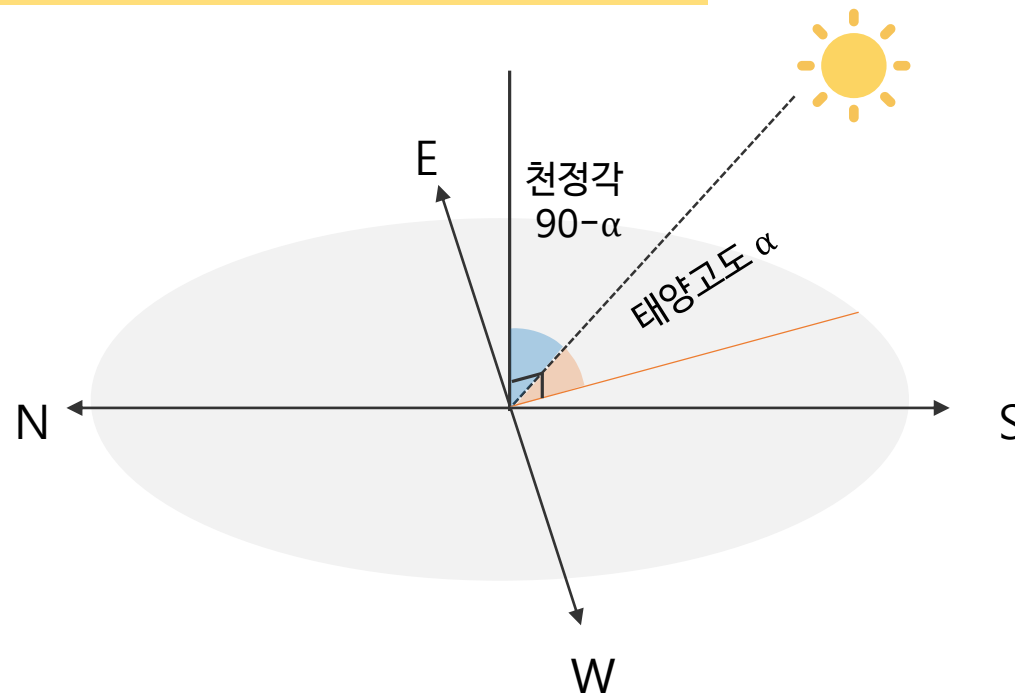
$$SOLAR = 1367 \times (0.7)^{AM}, \quad AM = \frac{1}{\cos(90 - \alpha)}$$

$$\sin(\alpha) = \cos(\Phi) \cos(\delta) \cos(w) + \sin(\Phi) \sin(\delta)$$

위도(Φ) 관측 위치 정보를 반영하는 위도 ex) 교로리: 37°

시간각(w) 관측 시간 정보(T)를 반영하는 시간각 $\rightarrow (12 - T)/96 \times 360^\circ$

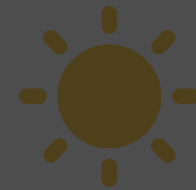
적위(δ) 관측 날짜 정보(N)를 반영하는 적위 ex) 하지: 23.5° , 동지: $-23.5^\circ \rightarrow 23.5 \times \sin[((N - 80)/365) \times 360]$



01

데이터 전처리

기상 데이터 전처리 - SOLAR 변수

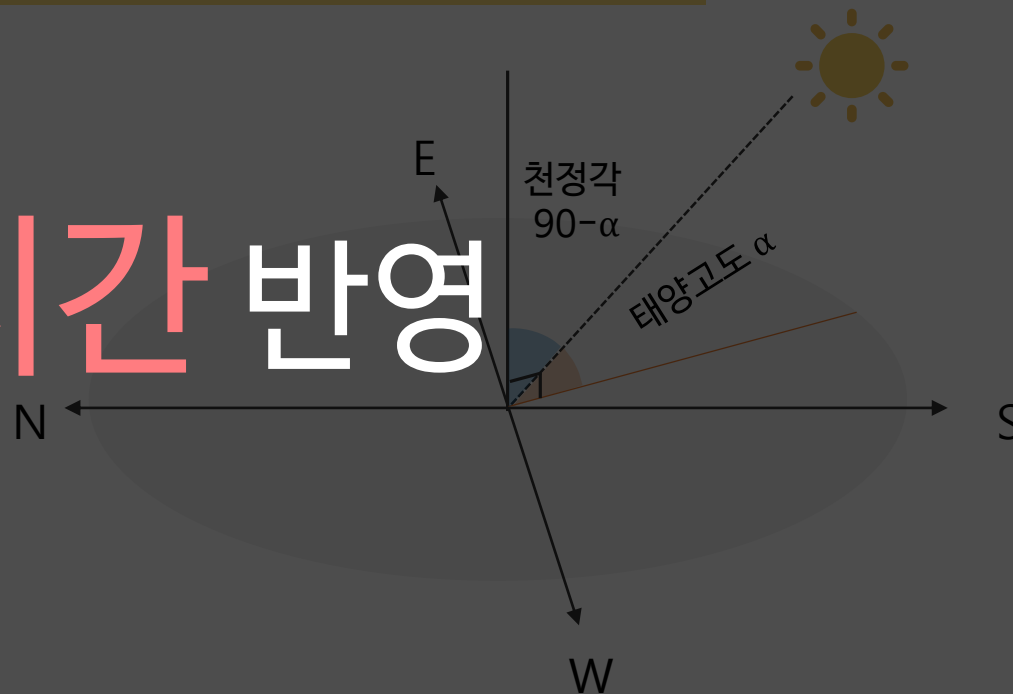


태양고도에 따라 바뀌는 이론 상의 일사량을 계산하는 식을 구하여 SOLAR 변수를 생성

$$SOLAR = 1367 \times (0.7)^{AM}, \quad AM = \frac{1}{\cos(90 - \alpha)}$$

SOLAR는 시간 반영

$$\sin(\alpha) = \cos(\Phi) \cos(\delta) \cos(w) + \sin(\Phi) \sin(\delta)$$



위도(Φ) 관측 위치 정보를 반영하는 위도 ex) 교로리: 37°

시간각(ω) 관측 시간 정보(T)를 반영하는 시간각 $\rightarrow (12 - T)/96 \times 360^\circ$

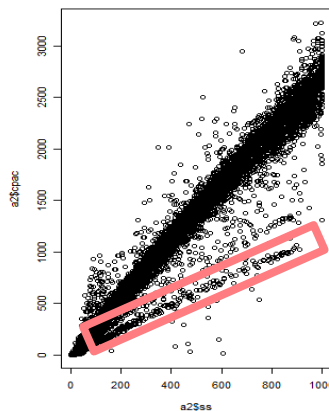
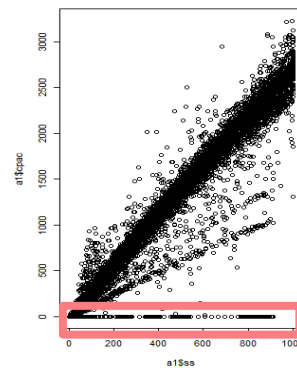
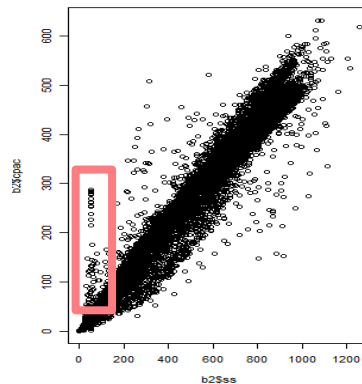
적위(δ) 관측 날짜 정보(N)를 반영하는 적위 ex) 하지: 23.5° , 동지: $-23.5^\circ \rightarrow 23.5 \times \sin[((N - 80)/365) \times 360]$

01 데이터 전처리

이상 데이터 처리



- ◆ 05 ~ 20시를 제외한 데이터 삭제
일사량이 있는 시점인 5시부터 20시까지의 데이터 사용
- ◆ 기기 오류로 간주되는 데이터 삭제
경사 일사량 or 현재 전력 둘 중 하나라도 0이면 삭제
인버터가 하나라도 작동하지 않는 경우
데이터가 하루 종일 변하지 않는 경우
- ◆ 추세선으로 떨어진 데이터 삭제
일반적인 추세를 예측하는 것이 주목적!



01

데이터 전처리 데이터 완성



A_site_total.csv - Excel

File Home Insert Page Layout Formulas Data Review View Help Team Tell me what you want to do

Clipboard Font Alignment Number Styles Cells Editing

Y45

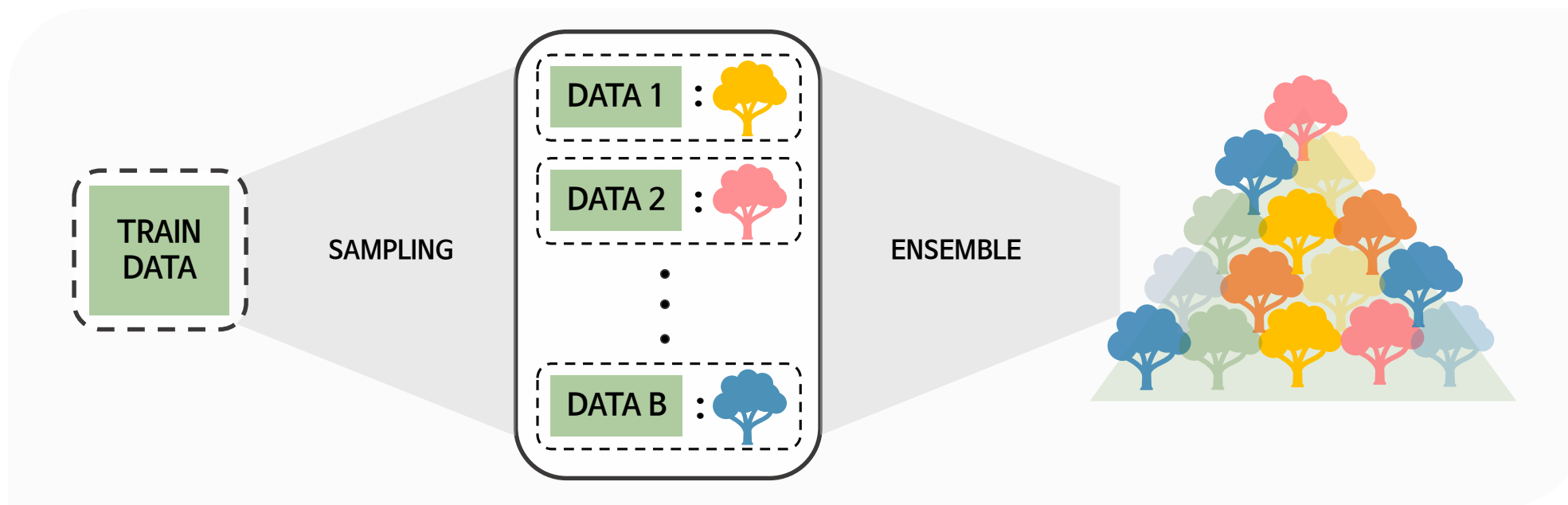
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	date	time	ss	hs	otemp	mtemp	cpac	tpg	temp	dwind	swind	rhumid	ap	apsea	cdsol	cdsun	rain	frain	cloud	solar
2	2017-07-07	10:00	213.1236	143.6855	27.99145	34.21479	257.0816	528.0693	27	165.1	2.4	98.1	1005.3	1008.6	1.09	0	0	0	10	906.9996
3	2017-07-07	10:15	89.67624	152.7103	27.28698	31.9435	84.19556	555.8715	27	167.2	2.4	98.1	1005.4	1008.7	1.16	0	0	0	10	916.5393
4	2017-07-07	10:30	193.2502	184.0253	27.62563	32.89985	244.3689	604.2584	27.4	164.7	2.4	98.1	1005.3	1008.6	1.28	0	0	0	10	924.5772
5	2017-07-07	10:45	180.3124	218.6728	27.51349	32.43766	210.9268	656.4768	27.7	184.3	4.1	95.6	1005.4	1008.7	1.4	0	0	0	10	931.2176
6	2017-07-07	11:00	269.7566	293.6901	28.78304	34.89423	351.6949	727.3232	27.9	180	4.6	94.1	1005.2	1008.5	1.54	0	0	0	10	936.544
7	2017-07-07	11:15	238.4022	279.8726	30.00469	37.05759	275.4471	808.5172	28	173.9	3	95.7	1005	1008.3	1.69	0	0	0	10	940.622
8	2017-07-07	11:30	192.4094	224.2523	30.51205	37.05902	219.1637	874.5689	27.8	170.3	4.1	96.1	1004.8	1008.1	1.83	0	0	0	10	943.5008
9	2017-07-07	11:45	193.6022	217.0072	30.78888	36.7715	251.2215	935.853	27.7	167.7	4.6	95.6	1004.8	1008.1	1.98	0	0	0	10	945.2146
10	2017-07-07	12:00	236.3673	193.8621	30.76977	34.88449	294.0252	992.3468	27.4	170.1	3.7	96.5	1005	1008.3	2.07	0	0	0	10	945.7836
11	2017-07-07	12:15	222.3531	271.1865	30.00063	33.35459	283.7921	1057.389	27.4	168.7	4.3	96.5	1005.1	1008.4	2.17	0	0	0	10	945.2146
12	2017-07-07	12:30	151.9719	164.8163	30.89427	34.80914	152.4456	1113.074	28	174.1	3.9	95	1004.9	1008.2	2.34	0	0	0	10	943.5008
13	2017-07-07	12:45	144.481	247.5607	30.05995	33.57908	146.3649	1167.11	28.1	171.8	3.7	93.9	1004.8	1008.1	2.52	0	0	0	10	940.622
14	2017-07-07	13:00	202.3219	222.9949	31.06124	36.03617	269.3799	1235.728	28.1	171.1	4.7	92.8	1004.6	1007.9	2.69	0	0	0	10	936.544
15	2017-07-07	13:15	208.2563	194.8307	30.82006	34.4612	265.725	1296.418	28.2	177.9	5.1	91.2	1004.6	1007.9	2.86	0	0	0	10	931.2176
16	2017-07-07	13:30	233.0928	258.7377	30.74467	33.93483	271.5951	1356.355	28.1	179.3	4.7	91.1	1004.5	1007.8	2.97	0	0	0	10	924.5772
17	2017-07-07	13:45	270.0161	214.1975	31.16604	35.0672	328.868	1431.268	28.2	184.4	5.4	90.7	1004.6	1007.9	3.12	0	0	0	10	916.5393
18	2017-07-07	14:00	129.3485	171.306	30.19108	33.33153	146.3478	1510.971	28.2	180.2	6.4	89.8	1004.6	1007.9	3.25	0	0	0	10	906.9996
19	2017-07-07	14:15	188.0636	284.1786	29.42328	32.35777	592.0112	1670.543	28.1	185.3	6.3	90.8	1004.7	1008	3.39	0	0	0	10	895.8304
20	2017-07-07	14:30	282.5864	304.0284	30.23046	33.52894	887.2602	1893.258	27.9	182.9	5.8	91.9	1004.6	1007.9	3.53	0	0	0	10	882.8766
21	2017-07-07	14:45	208.3427	183.7027	30.01229	32.91615	644.6152	2065.92	27.6	183.6	5.3	93.4	1004.6	1007.9	3.64	0	0	0	10	867.9507

Ready



2

모형 적합



- ◆ 다수의 의사결정나무(Decision Tree)를 학습하는 앙상블 방법
 - B개의 decorrelated된 트리 생성
 - 각 트리는 부트스트랩(Bootstrapped) 데이터에서 생성
 - 각 트리의 매 split마다 적합에 사용되는 변수를 랜덤하게 선택 (4개)

- ◆ 앙상블 기법으로 예측값을 종합
 - 분류(Classification) : 예측값들의 투표(Voting)
 - 회귀(Regression) : 예측값들의 평균

랜덤포레스트

- ◆ 훈련 데이터의 변화에 크게 민감하지 않아 모형이 안정적
 - 실시간 데이터 업데이트 시
 - 새로운 발전 단지(D site) 적용 시
- ◆ VIF로 파악한 중요 변수를 통한 인사이트를 도출
- ◆ 다른 모형에 비해 간단하고 예측력도 좋음

서포트벡터머신 & 인공신경망

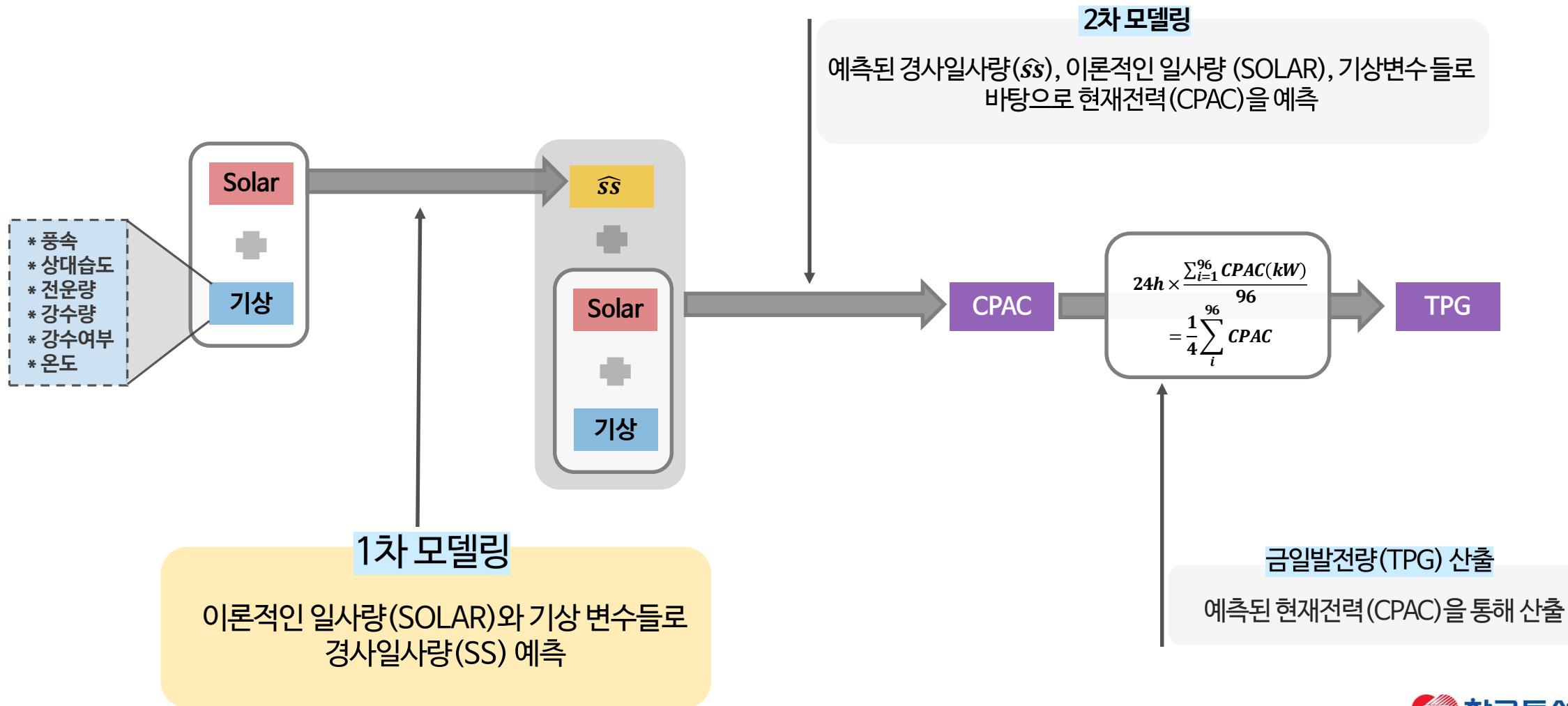
- ◆ Grid Search를 통해 오차를 최소화할 수 있는 hyperparameter tuning 필요
- ◆ 훈련 데이터에 바뀔에 따라 최적의 hyperparameter가 변동
- ◆ hyperparameter tuning에 오랜 시간 소요
- ◆ 해석이 어려움

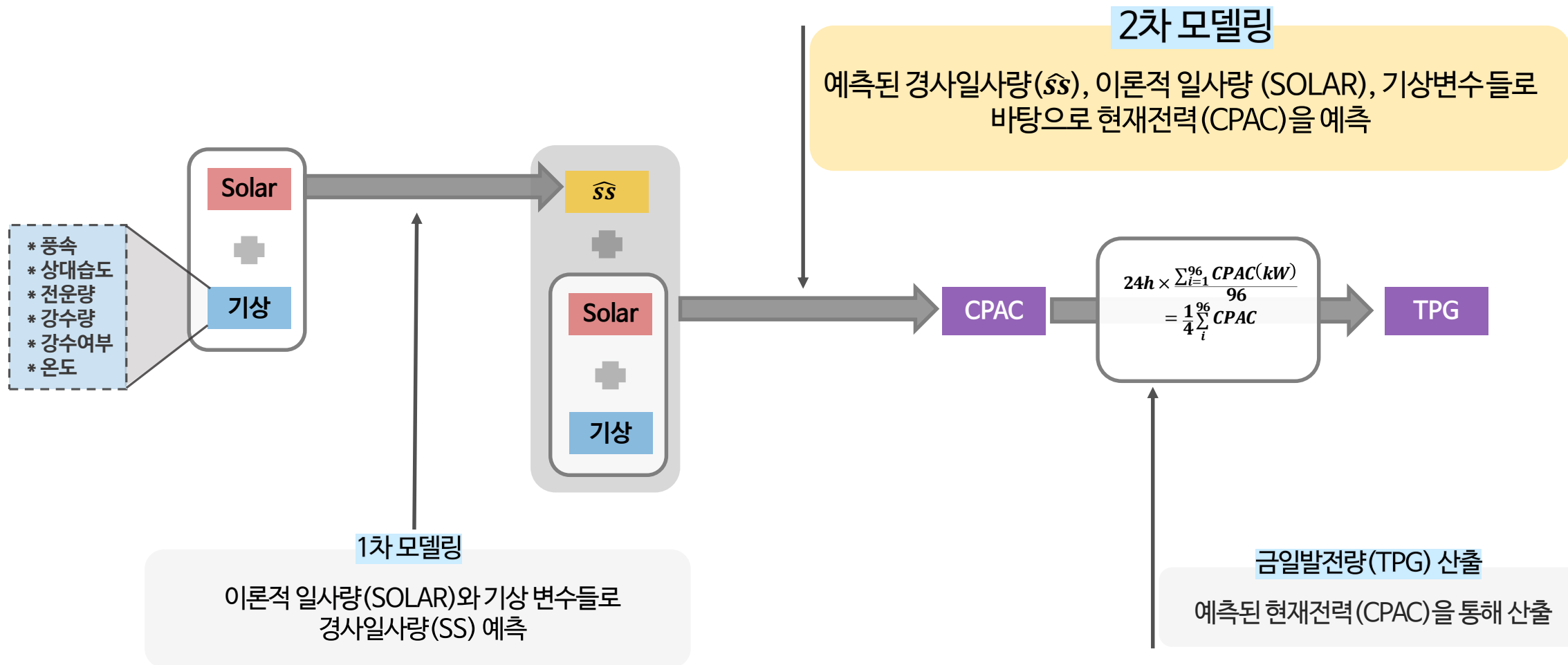
랜덤포레스트

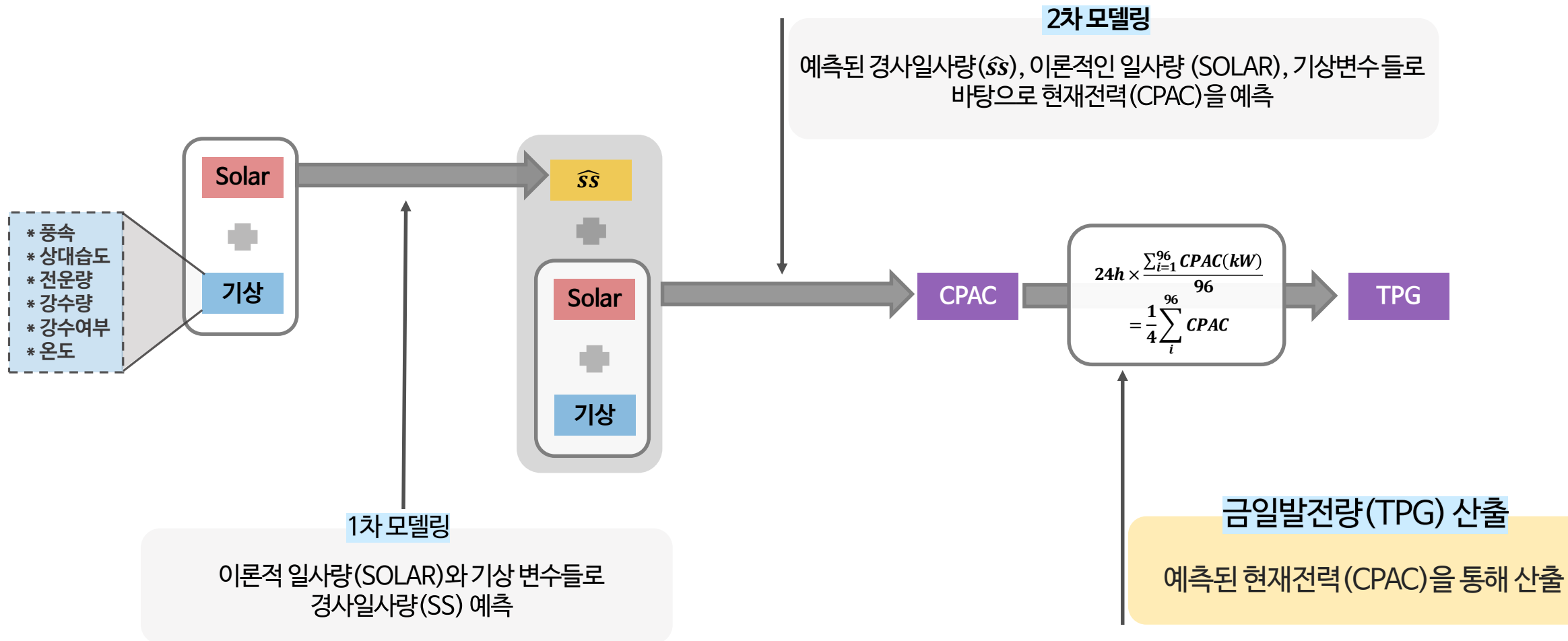
- ◆ 훈련 데이터의 변화에 크게 민감하지 않아 모형이 안정적
 - 실시간 데이터 업데이트 시
 - 새로운 발전 단지(D site) 적용 시
- ◆ VIF로 파악한 중요 변수를 통한 인사이트를 도출
- ◆ 다른 모형에 비해 간단하고 예측력도 좋음

서포트벡터머신 & 인공신경망

- ◆ Grid Search를 통해 오차를 최소화할 수 있는 hyperparameter tuning 필요
- ◆ 상대적으로 parameter tuning이 쉽고 최적의 hyperparameter가 변동
- ◆ 성능이 좋은 **랜덤 포레스트** 선택
- ◆ hyperparameter tuning에 오랜 시간 소요
- ◆ 해석이 어려움











































3

예보 시뮬레이션

03

예보 시뮬레이션 예보 데이터 입력



날짜	내일(13일 수)								모레(14일 목)								
시각	24	03	06	09	12	15	18	21	24	03	06	09	12	15	18	21	00
날씨																	
강수확률(%)	10	20	20	0	0	0	0	0	0	0	0	0	20	30	60	60	
강수량(mm)	-		-		-		-		-		-		-		1~4mm		
최저/최고(℃)	1/8								0/11								
기온(℃)	2	2	3	4	6	8	6	3	2	2	0	3	7	11	10	8	6
풍향/풍속(m/s)	 4	 3	 4	 5	 7	 7	 6	 3	 2	 2	 0	 1	 2	 4	 2	 1	 2
습도(%)	35	45	60	45	30	20	20	30	45	55	70	60	40	30	30	60	70



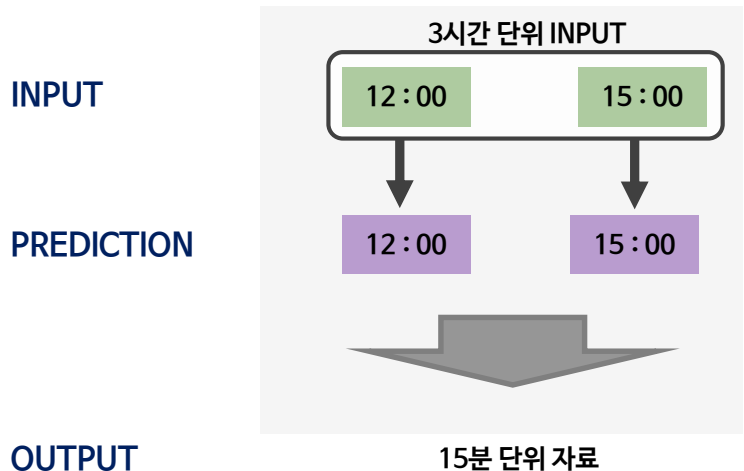
date	time	temp	swind	cloud	rhumid	probrain	rain	frain
2019-03-05	0:00	4	5	1.25	90	0	0	0
2019-03-05	3:00	3	3	1.25	95	0	0	0
2019-03-05	6:00	3	4	1.25	95	0	0	0
2019-03-05	9:00	6	3	7	90	10	0	0
2019-03-05	12:00	13	2	4	60	10	0	0
2019-03-05	15:00	13	5	7	50	20	0	0
2019-03-05	18:00	10	4	7	65	20	0	0
2019-03-05	21:00	6	3	7	85	20	0	0
2019-03-06	0:00	6	3	7	80	20	0	0

03

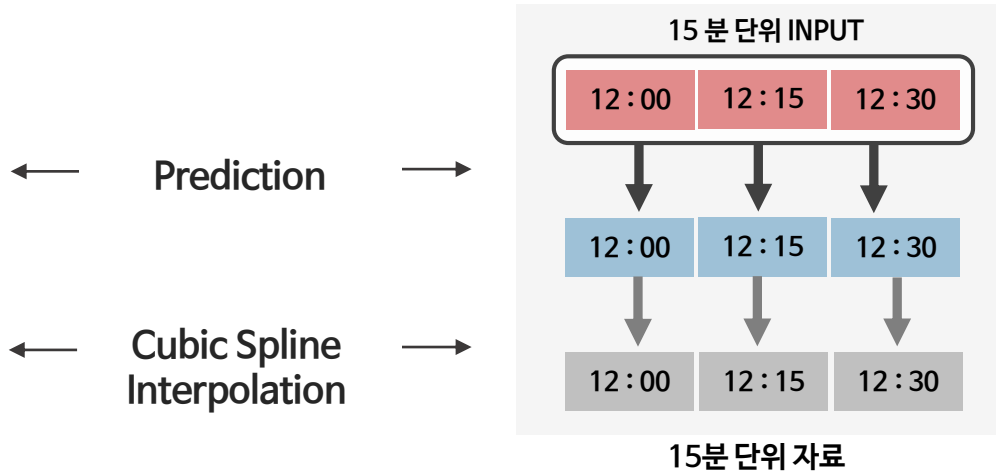
예보 시뮬레이션 보간법 선택



보간방법 A : 3시간



보간방법 B : 15분



INPUT

3시간 단위로 주어진 예보 데이터 그대로 사용

PREDICTION

적합한 모형을 통해 예측된 3시간 단위 자료

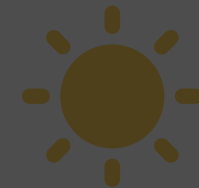
OUTPUT

예측된 자료를 Cubic Spline Interpolation을 통해 15분 단위로 변환

3시간 단위로 주어진 예보 데이터를 Cubic Spline Interpolation을 통해 15분 단위 입력자료 생성

적합한 모형을 통해 예측된 15분 단위 자료

예측된 15분 단위 그대로 사용



보간방법 A : 예측되는 시점 매우 적음

보간방법 B : 일출, 일몰 시점에 과대예측

INPUT 3시간 단위로 주어진 예보 데이터 그대로 사용

PREDICTION 적합한 모델을 통해 예측된 3시간 단위 자료

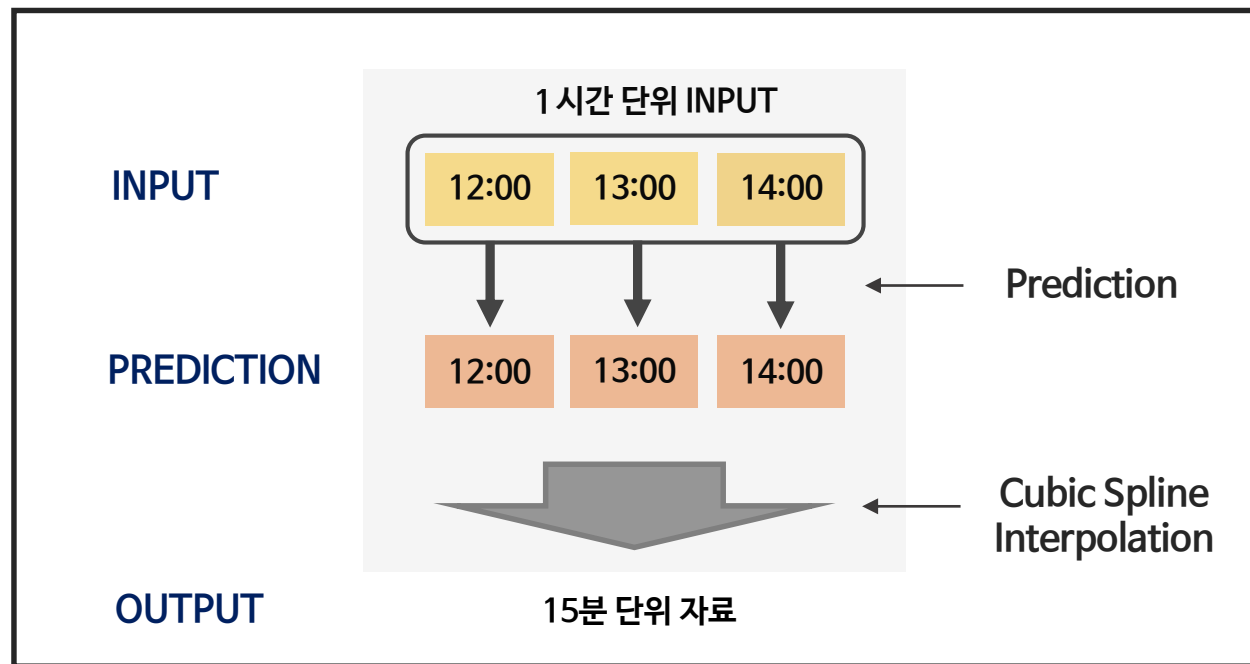
OUTPUT 예측된 자료를 Cubic Spline Interpolation을 통해 15분 단위로 변환

3시간 단위로 주어진 예보 데이터를 Cubic Spline Interpolation을 통해 15분 단위 입력자료 생성

적합한 모델을 통해 예측된 15분 단위 자료

예측된 15분 단위 그대로 사용

보간방법 C : 1시간

**INPUT**

3시간 단위로 주어진 예보 데이터를 Cubic Spline Interpolation을 통해 1시간 단위 입력자료 생성

PREDICTION

적합한 모형을 통해 예측된 1시간 단위 자료

OUTPUT

예측된 자료를 Cubic Spline Interpolation을 통해 15분 단위로 변환

현재 전력량 예측

- ◆ 일출, 일몰 시간 사이는 랜덤 포레스트 모형으로 예측, 그 외는 0으로 고정

$$\widehat{CPAC} = \begin{cases} \widehat{CPAC}, & (\text{time} = [\text{일출}, \text{일몰}]) \\ 0, & \text{otherwise} \end{cases}$$

금일 발전량 산출

- ◆ 금일 발전량은 현재 전력량 예측값에서 다음의 식으로 추정

$$24h \times \frac{\sum_{i=1}^{96} \widehat{CPAC} (kW)}{96} = \frac{1}{4} \sum_i^{96} \widehat{CPAC}$$

$$\widehat{TPG}(kWh)$$

엄지영, 최형진, 조수환. (2015).
기상정보를 활용한 도시규모-EMS용 태양광 발전량 예측모델. 전기학회논문지, 64(3), 393-398.

김민수, 지상민, 오수영, 정재학. (2016).
그림자 효과를 고려한 태양전지 모듈의 발전량 예측 연구. Current Photovoltaic Research, 4(2), 80-86.

차왕철, 박정호, 조욱래, 김재철. (2015).
온도와 풍속에 따른 태양광발전 효율 실증분석 연구. 전기학회논문지 P, 64P(1), 1-6.

권오현, 이경수. (2018).
기상 환경 모니터링 데이터를 이용한 태양광발전시스템 발전량 성능 분석. 한국태양에너지학회 논문집, 38(4), 11-31.

양동현, 여나영, 마평수. (2017).
효율적인 태양광 발전량 예측을 위한 Dynamic Piecewise 일사량 예측 모델. 정보과학회 컴퓨팅의 실제 논문지, 23(11), 632-640.



감사합니다

동서발전과 함께 서투른 솜씨였지만 발전된 모습으로 전념했습니다