



2023

기상에 따른 계절별 지면온도 산출 기술 개발

날씨 빅데이터 콘테스트

Contents

1. 분석 배경 및 목표
2. 데이터 정의 및 EDA
3. 데이터 전처리
4. 모델링
5. 서비스 활용방안 및 기대효과



1. 분석배경 및 목표

1

2

3

4

5

지면온도는 수많은 분야에서 유용하게 활용되고 있으며, 기후변화, 농작물의 생육 조건 파악, 지열 발전의 잠재력 평가, 지구 온난화 현상 추적, 가뭄 감시 등 연구적 측면에서까지 중요한 기상인자로 사용되고 있음

이러한 실질적 수요에 입각하여 지면온도를 정확히 파악하고 데이터를 체계적으로 구축하는 것은 사회적, 경제적 측면에서 대단히 중요한 성과로 이어질 수 있음

따라서 본 콘테스트에서는 시공간적으로 상세한 계절별 지면온도 산출 기술을 개발하여 양질의 지면온도 데이터를 확보하는 것이 주된 목적임



1. 분석배경 및 목표

1

2

3

4

5

본 콘테스트의 목적에 따라 경희의 온도 팀은
기상 빅데이터를 통해
계절별 지면온도 예측과 밀접한 관련이 있는 특징들을 탐색하고
특정 상황에 맞는 파생변수를 만들어
모델의 정확도와 설명력을 높이는 것을 목표로 하고자 함

2.1. 데이터 정의

1

2

3

4

5

〈표 1〉 Train set 변수 개요

변수명	정의	변수명	정의	변수명	정의
STN	지점번호(1~10)	HM	1시간 평균 상대습도(%)	SI	1시간 누적 일사량(MJ)
YEAR	년도(A, B, C, D, E)	WS	1시간 평균 풍속(m/s)	SS	1시간 누적 일조량(초)
MMDDHH	월/일/시간	RN	1시간 누적 강수량(mm)	SN	00분에 측정된 적설 깊이(cm)
TA	1시간 평균 기온(°C)	RE	1시간 누적 강수유무(분)	TS(*종속변수)	1시간 평균 지면온도(°C)
TD	1시간 평균 이슬점 온도(°C)	WW	현천계 현천(S:눈/R:비/F:안개/H:박무/G:연무/C:맑음,X:모름)		
봄(106,999 rows x 14 columns) 여름(110,342 rows x 14 columns) 가을(110,328 rows x 14 columns) 겨울(110,374 rows x 14 columns)					
*결측값: -99, -99.9, -999					

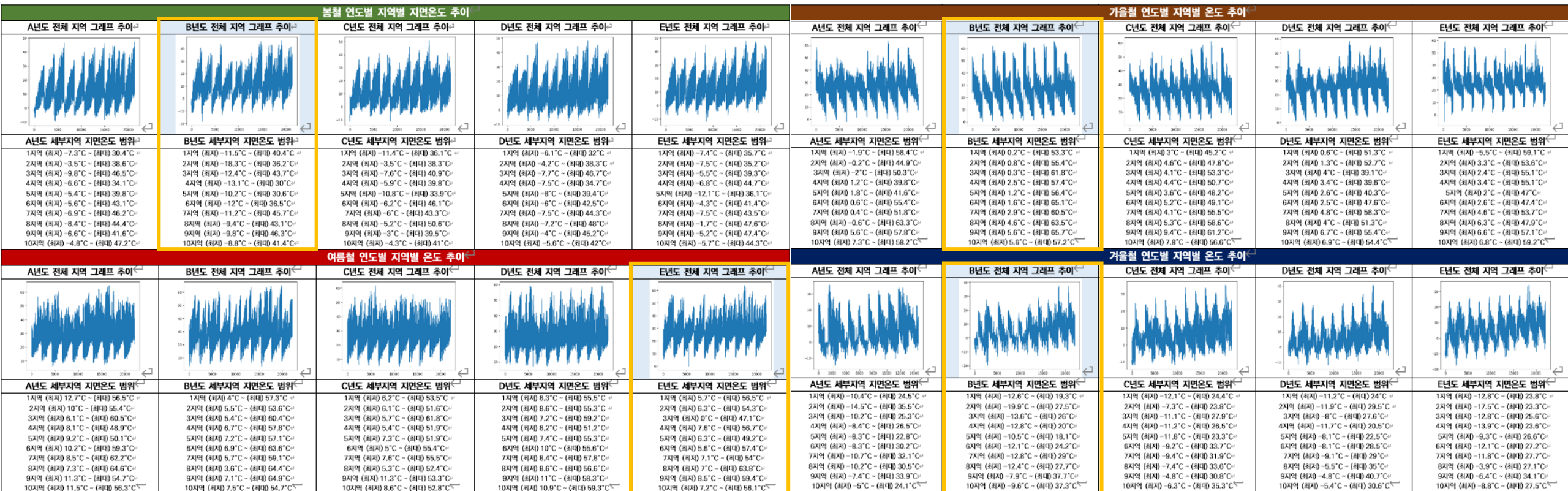
〈표 2〉 Test set 변수 개요

변수명	정의	변수명	정의	변수명	정의
STN	지점번호(a, b, c)	HM	1시간 평균 상대습도(%)	SI	1시간 누적 일사량(MJ)
YEAR	년도(F, G)	WS	1시간 평균 풍속(m/s)	SS	1시간 누적 일조량(초)
MMDDHH	월/일/시간	RN	1시간 누적 강수량(mm)	SN	00분에 측정된 적설 깊이(cm)
TA	1시간 평균 기온(°C)	RE	1시간 누적 강수유무(분)	현천계 현천(S:눈/R:비/F:안개/H:박무/G:연무/C:맑음,X:모름)	
TD	1시간 평균 이슬점 온도(°C)	WW			
봄(6,408 rows x 13 columns) 여름(6,624 rows x 13 columns) 가을(6,624 rows x 13 columns) 겨울(6,624 rows x 13 columns)					
*결측값: -99, -99.9, -999					

- 본 콘테스트에서 제공한 데이터는 **A년 2월 ~ F년 1월까지의 국내 10개 지점의 11개의 기상 데이터를 가진 학습 데이터 (Trainset)**와 **F년 2월 ~ G년 1월까지의 특정 3지점의 11개의 기상 데이터를 가진 검증 데이터 (Test set)**이며 〈표 1〉, 〈표 2〉에서 해당 변수들을 확인할 수 있음
- 본 콘테스트에서 제시한 계절별 기간은 **봄(2월~4월), 여름(5월~7월), 가을(8월~10월), 겨울(11월~1월)**이며 해당 기준을 바탕으로 학습 데이터와 검증 데이터를 계절별로 분리하여 사용함



〈표 3〉 계절별, 지역별, 연도별 지면온도 파악



- 위 <표 3>과 같이 계절별, 연도별로 전체 지면온도 수치를 계산 및 시각화 했을 때, 전반적인 지면온도 범위의 폭이 넓은 것을 확인할 수 있었음
 - 특히, **봄의 지면온도 범위**는 B년도에 $-18.3^{\circ}\text{C} \sim 46.3^{\circ}\text{C}$ 로 폭이 가장 넓었으며, **여름**은 E년도에 $0^{\circ}\text{C} \sim 63.8^{\circ}\text{C}$ 의 넓은 지면온도 범위를 보였으며, **가을 및 겨울**은 B년도에 각각 $0.2^{\circ}\text{C} \sim 65.7^{\circ}\text{C}$, $-19.9^{\circ}\text{C} \sim 37.7^{\circ}\text{C}$ 로 가장 넓은 지면온도 범위폭을 나타냄
- ➔ 이러한 결과는 봄/여름/가을/겨울의 범위가 통상적인 구간과는 다를 뿐만 아니라, 10개 지역별 기상 정보가 큰 차이를 보이기 때문으로 추정할 수 있음

3.1. 데이터 전처리 - 결측치 분포 파악

1

2

3

4

5

〈표 4〉 계절별 데이터 결측치 개수 및 비율

봄				여름				가을				겨울			
변수명	관측된 결측값	개수(개)	비율(%)	변수명	관측된 결측값	개수(개)	비율(%)	변수명	관측된 결측값	개수(개)	비율(%)	변수명	관측된 결측값	개수(개)	비율(%)
RE	[-99]	301	0.0028	RE	[-99]	736	0.0067	RE	[-99]	481	0.0044	RE	[-99]	424	0.0038
TA	[-99.9]	111	0.001	TA	[-99.9]	99	0.0009	TA	[-99.9]	117	0.0011	TA	[-99.9]	106	0.001
TD	[-99.9]	112	0.001	TD	[-99.9]	106	0.001	TD	[-99.9]	120	0.0011	TD	[-99.9]	111	0.001
HM	[-99.9]	103	0.001	HM	[-99.9]	92	0.0008	HM	[-99.9]	95	0.0009	HM	[-99.9]	91	0.0008
WS	[-99.9]	167	0.0016	WS	[-99.9]	162	0.0015	WS	[-99.9]	256	0.0023	WS	[-99.9]	225	0.002
RN	[-99.9]	2,277	0.0213	RN	[-99.9]	1,955	0.0177	RN	[-99.9]	1,291	0.0117	RN	[-99.9]	2,275	0.0206
TS	[-99.9]	114	0.0011	TS	[-99.9]	111	0.001	TS	[-99.9]	111	0.001	TS	[-99.9]	100	0.0009
SI	[-99.9]	48,956	0.4575	SI	[-99.9]	41,873	0.3795	SI	[-99.9]	49,451	0.4482	SI	[-99.9]	59,549	0.5395
SS	[-99.9]	49,045	0.4584	SS	[-99.9]	41,520	0.3763	SS	[-99.9]	48,775	0.4421	SS	[-99.9]	59,526	0.5393
SN	[-99.9]	104,778	0.9792	SN	[-99.9]	110,342	1.0	SN	[-99.9]	110,328	1.0	SN	[-99.9]	105,401	0.9549

- 본 데이터의 계절별 각 변수의 결측치 개수 및 비율은 〈표 4〉와 같음
 - 전체적인 결측값의 분포는 각 계절별로 유사한 양상을 보였고, 특히 **SI, SS, SN의 결측치 비율**이 높게 나온 것으로 확인되었음
 - **타 기상 변수 정보와 맞지 않는** 현천(VW) 데이터가 존재하는 것을 파악할 수 있었음
- ➔ 기상청에서 명시한 기간별로 전체 데이터셋을 나누어 각 계절별 데이터셋을 생성한 다음, 세부적인 분석을 진행하여 결측치를 처리하는 과정을 진행하였음



3.2. 데이터 전처리 - 결측치 처리 과정

1

2

3

4

5

1. 결측치 행 제거

- 전체 훈련 데이터 기준으로 종속변수를 포함한 모든 변수 값이 결측치인 197개 행 제거
- 계절별 데이터셋으로 분할 후 지면온도(TS)의 값이 결측치인 행 제거(*봄 114개, 여름 111개, 가을 111개, 겨울 100개 총 436개 행 제거)

2. 일사량(SI), 일조량(SS) 변수 결측치 처리

- 일사량(SI), 일조량(SS) 변수는 각각 1시간 누적 일사량 및 일조량하므로 **해가 떠 있는 시간만 관측이 가능함**
- 각 월 별로 일사량(SI), 일조량(SS) 데이터를 확인했을 때, 특정 시간대 결측치가 몰려있는 양상을 확인함(*해당 구간의 결측치를 0으로 대체)

3. 적설(SN) 변수 결측치 처리

- 봄: **2월 17일 이후** 적설(SN) 변수가 모두 결측치로 되어 있었고, 실제로 적설이 관측되지 않아 해당 결측치를 **모두 0으로 처리**하였음
- 여름 및 가을: 적설(SN) 변수가 모두 결측치로 관측되었고, 해당 결측치를 **모두 0으로 처리**하였음
- 겨울: **최초 적설 관측일이 11월 24일**이었고, **24일 이전**의 적설(SN)의 결측치는 **모두 0으로 처리**하였음





3.2.1. 일사량(SI) 결측치 처리 예시

1

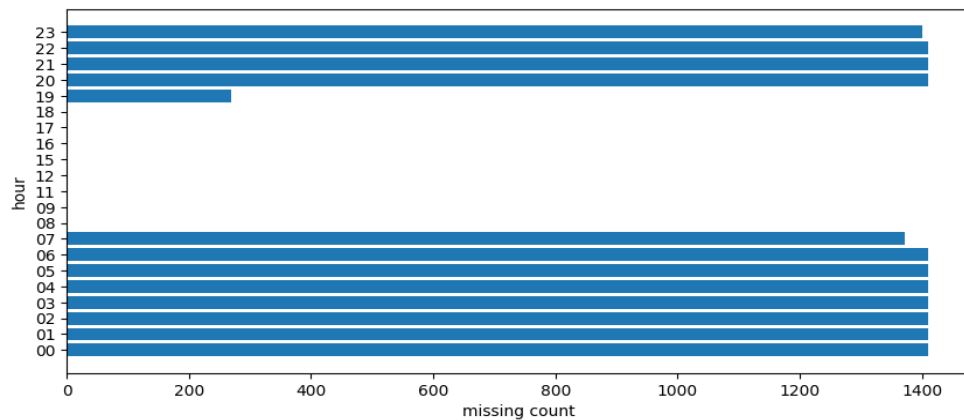
2

3

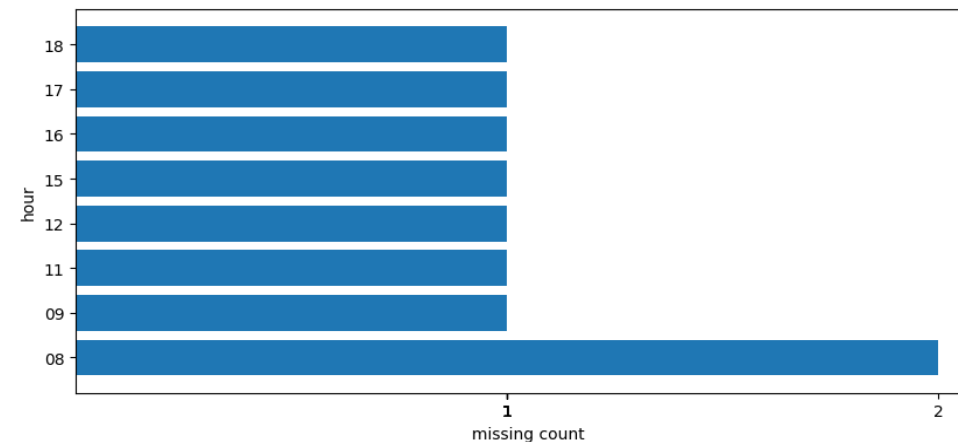
4

5

Feb's missing value counts by hour



Feb's missing value counts by hour



2월 SI 결측치 처리 전

month	hour	si
2	07	-99.9
2	19	-99.9
2	20	-99.9
2	21	-99.9
2	22	-99.9



2월 SI 결측치 처리 후

month	hour	si
2	07	0
2	19	0
2	20	0
2	21	0
2	22	0

(*일조량(SS) 결측치 처리도 이와 같은 절차로 진행)





3.2.2. 적설(SN) 결측치 처리 예시

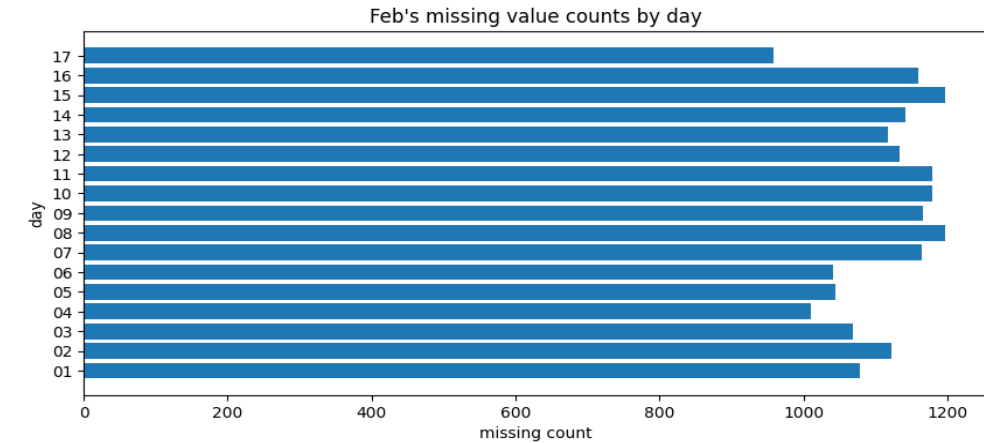
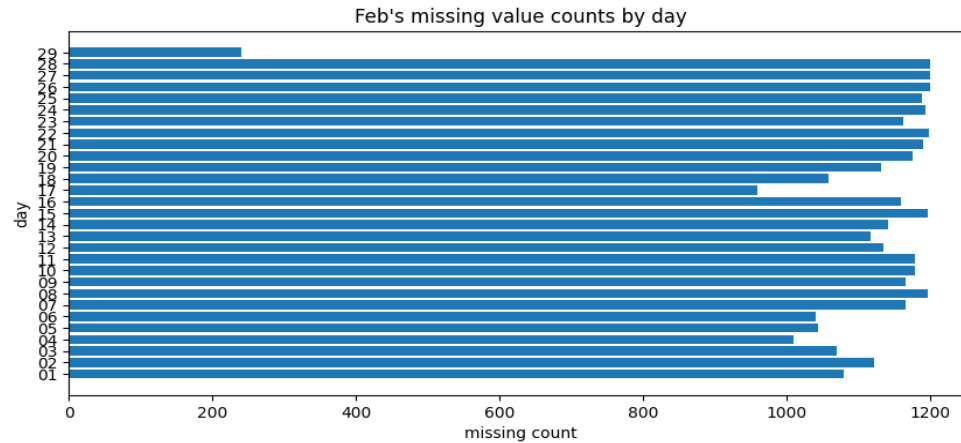
1

2

3

4

5



2월 SN 결측치 처리 전

month	day	sn
2	18	-99.9
2	20	-99.9
2	21	-99.9
2	23	-99.9
2	29	-99.9



2월 SN 결측치 처리 후

month	day	sn
2	18	0
2	20	0
2	21	0
2	23	0
2	29	0

3.3. 데이터 전처리 - 현천(WW) 이상치 데이터 제거



〈표 5〉 현천(WW) 이상치 데이터 제거 방식

봄철 E년도 눈(S) 이상치 사례



- 앞선 데이터 EDA 과정에서 **현천(WW)이 눈(S)이지만 기온(TA) 및 지면온도(TS)가 10°C를 넘어서는 이상치 데이터**를 발견하여 〈표 5〉와 같은 과정을 거쳤음(*봄 14개, 여름 11개, 가을 11개의 총 36개 이상치 데이터 제거)
- 추가적으로 **기온이 10°C 내**에서는 이슬점 온도 및 상대습도, 풍속의 영향에 따라 **눈이 내릴 수 있으며**, 본 분석에서 10°C 내의 눈(S)으로 기록된 데이터의 이슬점 온도, 상대습도, 풍속 수치 확인 후 사용함



3.4. 다항식 보간법(Polynomial Interpolation)

1

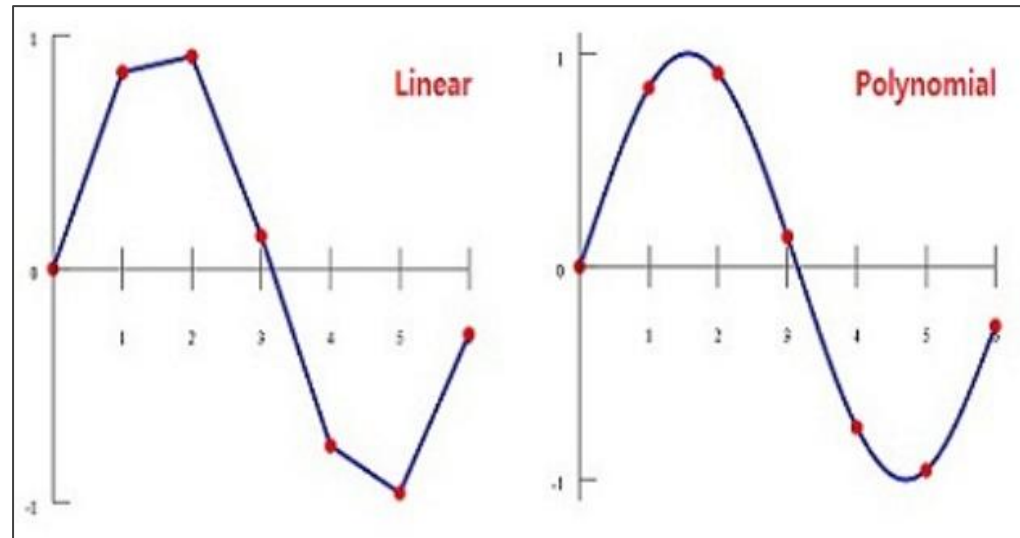
2

3

4

5

〈그림 1〉 선형 및 다항 보간 그래프



- 선형 보간법은 관측치를 직선으로 연결하여 그 사이의 결측치를 해당 직선 위의 값으로 추측하는 방법
- 하지만 단순히 선형으로 보간할 경우 다차원의 변수로 이루어진 본 데이터 셋에서는 정확도가 현저히 낮아질 수 있음
- 이에 대한 대안으로 실제 연구 중 기상 데이터에, **직선이 아닌 다항식을 활용한 보간법이 많이 사용되는 것을** 확인함(Chin et al., 2023; Antal et al., 2021).
- 본 실험에서는 **2차 다항식 보간법**을 사용하였음
- 변수에 따라 데이터의 맨 앞 또는 뒤 값이 결측치에 해당하여 다항식 보간법이 이뤄지지 않은 경우는 해당 결측치의 바로 앞 또는 뒤의 관측치로 대체함
- 최종적으로, 종속변수인 **지면온도(TS)외의 결측치가 존재하던 10개 변수에서 결측치가 모두 대체**되었음

3.5. 파생변수 소개

1

2

3

4

5

〈표 6〉 파생변수 소개

생성변수	생성 방법	생성 과정 및 설명	활용된 계절별 모델
월(month), 일(day), 시간(hour)	월일시간(mmddhh) 변수로부터 월(month), 일(day), 시간(hour) 변수 생성	- 앞선 결측치 처리 과정에서 월, 일, 시간 단위로 세부적인 판단이 필요한 상황이 있었음	봄, 여름, 가을, 겨울
강수여부(rs_yn)	현천(ww) 변수가 눈(S)이거나 비(R)인 경우는 1, 그렇지 않은 경우 0으로 처리	- 현천(WW) 변수의 활용을 위해 사용	봄, 여름, 가을, 겨울
체감온도(sense_ta)	$13.12 + 0.6215T - 11.37W^{0.16} + 0.3965W^{0.16}T$ *(T: 기온(ta), W: 풍속(ws))	- 체감온도란 인간이 느끼는 더위나 추위를 수량으로 나타낸 것 - 여름철(5~9월)과 겨울철(10~익년 4월)을 구분하여 계산 가능 (*본 공모전 데이터의 경우 여름철 체감온도 산출에 필요한 습구온도가 존재하지 않으므로, 겨울철 체감온도만 활용하여 봄/겨울 데이터 셋에서만 체감온도를 산출)	봄, 겨울
화씨온도(f_ta)	$(T \times 9/5) + 32$ *(T: 기온(ta))	- 열지수 산출에 사용하기 위해 생성	봄, 여름, 가을, 겨울
열지수(h_idx)	<p>(1) <i>value</i> $= -42.379 + 2.04901523F + 10.14333127H - 0.22475541FH - 0.00683783F^2 - 0.05481717H^2 + 0.00122874F^2H + 0.00085282FH^2 - 0.00000199F^2H^2$</p> <p>(2) <i>value2</i> = (<i>value</i> - 32) × 5/9</p> <p>(3) <i>value3</i> = round(<i>value2</i>)/10.0</p> <p>*(F: 화씨온도(f_ta), H: 상대습도(hm))</p>	<p>- 열지수란 기온과 습도에 따라 사람이 실제로 느끼는 더위를 지수화한 것</p> <p>- Rothfus의 회귀방정식을 적용했고, 해당 식은 화씨 80도 미만에서는 유효하지 않음</p> <p>- 따라서, 화씨온도가 80도 이상인 경우만 식을 적용하고, 아닌 경우는 기온값을 열지수로 적용</p> <p>- 지면 온도(TS)가 30~50°C의 값을 전반적으로 나타내는 여름/가을 데이터 셋에서 열지수 변수 생성</p>	여름, 가을
불쾌지수(u_idx)	$(T - 0.55 \times (1 - (0.01 \times H))) \times (T - 14.5)$ *(T: 기온(ta), H: 상대습도(hm))	<p>- 불쾌지수란 사람이 느끼는 불쾌감의 정도</p> <p>- 기온과 습도의 조합으로 산출 가능</p> <p>- 열지수와 마찬가지로 30~50°C의 값을 나타내는 여름/가을 데이터 셋에서 생성</p>	여름, 가을

4.1. 모델링 - 기본 모델 구축

1

2

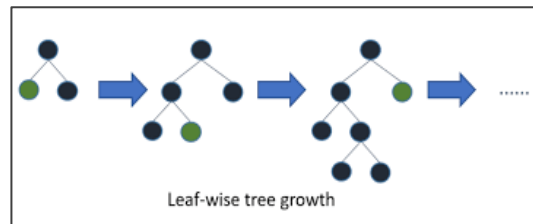
3

4

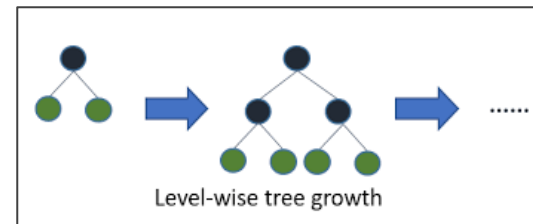
5

$$\frac{X_{it} - Q_1(X_i)}{Q_3(X_i) - Q_1(X_i)}$$

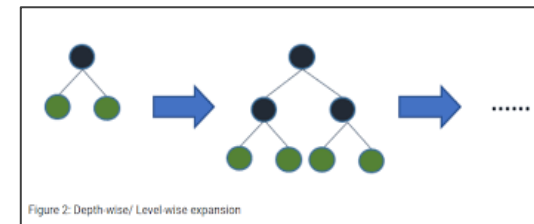
〈그림 2〉 RobustScaler 산출식



〈그림 3〉 LGBM 구현 방식



〈그림 4〉 XGBoost 구현 방식



〈그림 5〉 Catboost 구현 방식

- 본 실험에서는 이상치를 모두 반영하여 모델을 구축하므로 중앙값과 IQR을 사용하여 **이상치의 영향을 최소화하는 RobustScaler를 사용**하여 데이터를 표준화함(Raju et al., 2020).
- 이후 기상 예측 관련 선행연구들을 조사했을 때, **LGBM**(Light Gradient Boosting Model), **CatBoost**, **XGBoost**(Extreme Gradient Boosting)와 같은 boosting 계열 모델들이 좋은 성능을 보인 것을 확인할 수 있었음(Han et al., 2023; Niu et al., 2021; Ma & Ji, 2020).
- 따라서 기본 모델을 **LGBM, CatBoost, XGBoost**로 선정함
- 기본 모델의 **random_state 값은 42로 고정**하였으며, **Parameter 값은 기본으로 설정**하여 각 모델에 데이터를 학습시키고 MAE를 산출함
- MAE 산출 결과, **봄, 여름 및 겨울 데이터에서는 LGBM 모델이, 가을 데이터에서는 CatBoost 모델이** 각각 기본 모델 중에서 가장 좋은 성능을 보임

4.2. 모델링 - 변수 중요도 파악

1

2

3

4

5

〈표 7〉 각 계절별 모델 변수 중요도 파악

변수명 / 계절	봄(LGBM)	여름(LGBM)	가을(CatBoost)	겨울(LGBM)
월일시간(mmddhh)	0.1466	0.1736	1.841757	0.1286
기온(ta)	0.1326	0.1173	20.172713	0.131
이슬점 온도(td)	0.0923	0.0873	1.493564	0.0666
상대습도(hm)	0.079	0.0833	2.121682	0.0656
풍속(ws)	0.0666	0.07	1.106124	0.056
강수(rn)	0.01	0.0043	0.055587(제거)	0.0053
강수유무(re)	0.0113	0.0083	0.174311	0.0123
현천(ww)	0.0256	0.0356	0.43827	0.025
일사량(si)	0.1536	0.136	27.72064	0.142
일조시간(ss)	0.0323	0.036	2.354042	0.026
적설(sn)	0.0443	0(제거)	0(제거)	0.1836
월(month)	0(제거)	0(제거)	0.053896(제거)	*0.0036(제거)
일(day)	0.0456	0.0553	1.226451	0.038
시간(hour)	0.1113	0.112	5.25577	0.0766
강수여부(rs_yn)	0.0046	0.003	0.105852	0.0006(제거)
체감온도(sense_ta)	0.0436	해당 없음	해당 없음	0.0386
화씨온도(f_ta)	0(제거)	0(제거)	19.337002	0(제거)
열지수(h_idx)	해당 없음	0.023	10.849456	해당 없음
불쾌지수(u_idx)	해당 없음	0.0543	4.35758	해당 없음

*LGBM 모델 = 변수 중요도 총 합 1 / CatBoost 모델 = 변수 중요도 총 합 100

- 기본 모델 구축 결과, 계절별로 가장 좋은 성능을 보인 모델별로 내장 함수(feature_importances_)를 적용해 변수 중요도가 0.1%(LGBM = 0.001, CatBoost = 0.1) 미만인 변수를 각 계절별 모델에서 제거하였음

(*단, 겨울의 변수 중 월(month)의 경우 타 모델의 변수 중요도를 종합적으로 파악해 본 결과, 다른 모델에서 모두 0의 값을 나타내어 제거하기로 결정하였음)

4.3. 모델링 - 튜닝 과정 (1)

1

2

3

4

5

〈표 8〉 Optuna Hyper Parameter 적용 후 모델 성능 결과

봄		여름		가을		겨울	
Model	MAE	Model	MAE	Model	MAE	Model	MAE
LGBM(Best)	2.724	LGBM(Best)	3.232	LGBM(Best)	2.061	LGBM(Best)	1.954
CatBoost	3.003	CatBoost	3.774	CatBoost	2.275	CatBoost	2.232
XGBoost	3.656	XGBoost	3.857	XGBoost	2.302	XGBoost	2.235
Best Model Hyper Parameter 값		Best Model Hyper Parameter 값		Best Model Hyper Parameter 값		Best Model Hyper Parameter 값	
N_estimators = 265 Max_depth = 6 Learning_rate = 0.03848576560913784 Min_child_samples = 24 Colsample_bytree = 0.9275432756814064 Subsample = 0.4818277531049522		N_estimators = 1380 Max_depth = 7 Learning_rate = 0.00379864792169426 Min_child_samples = 19 Colsample_bytree = 0.47828626887767617 Subsample = 0.4474447500029784		N_estimators = 1968 Max_depth = 15 Learning_rate = 0.14199190766877956 Min_child_samples = 26 Colsample_bytree = 0.9471423809959226 Subsample = 0.8804923171057708		N_estimators = 2000 Max_depth = 15 Learning_rate = 0.19481757590060347 Min_child_samples = 7 Colsample_bytree = 0.9898081343609673 Subsample = 0.4668480569567963	

- 변수 중요도가 Cut-Off 기준 0.1% 미만인 변수들을 제거한 Baseline 모델의 성능을 높이기 위해 최적의 Hyper Parameter를 탐색하는 **Optuna 소프트웨어**를 적용하여 결과를 산출함
- A, B, C, D년도를 학습 데이터, 나머지 E년도를 검증 데이터로 지정하여 Optuna를 적용함
- 그 결과, 하이퍼 파라미터 튜닝된 LGBM이 모든 계절에서 가장 우수한 성능을 보임

4.3. 모델링 - 튜닝 과정 (2)

1

2

3

4

5

〈표 9〉 Ensemble 모델 성능 결과

봄 Ensemble	여름 Ensemble	가을 Ensemble	겨울 Ensemble
[기본 Ensemble] MAE = 1.789	[기본 Ensemble] MAE = 2.016	[기본 Ensemble] MAE = 1.926	[기본 Ensemble] MAE = 1.982
LGBM*0.4	LGBM*0.3	LGBM*0.7	LGBM*0.6
CatBoost*0.4	CatBoost*0.6	CatBoost*0.2	CatBoost*0.3
XGBoost*0.2	XGBoost*0.1	XGBoost*0.1	XGBoost*0.1
[가중치 적용 Ensemble] MAE = 1.784	[가중치 적용 Ensemble] MAE = 2.013	[가중치 적용 Ensemble] MAE = 1.71	[가중치 적용 Ensemble] MAE = 1.879
최종 MAE 평균 값 = 1.846			

*기본 Ensemble(기본 LGBM+기본 CatBoost+기본 XGBoost)

- 모델 성능의 추가적인 개선을 위해 Ensemble 기법을 적용함(*Ensemble이란 여러 개의 예측 모델을 생성한 다음, 각 모델의 예측 값을 설정된 가중치를 통해 결합함으로써 보다 정확한 예측을 도출하는 기법)
- Ensemble에는 기본 LGBM, CatBoost, XGBoost를 투입하여 모델을 구축하였음
- 〈표 9〉의 결과를 확인했을 때, Hyper Parameter를 적용한 LGBM 모델에 비해 모든 계절에서 Ensemble 모델이 향상된 결과를 보였음
- 추가로 Ensemble 내 각 모델에 Optuna를 적용해 파라미터 튜닝된 Ensemble 모델을 생성하였으나, 기본 Ensemble 모델에 비해 성능이 떨어지는 것으로 나타냄
- 따라서 **기본 LGBM, CatBoost, XGBoost 모델을 포함하는 Ensemble을 최종 모델로 선정**하고, 추가적으로 Ensemble 내 각 모델의 예측 **가중치를 지속적으로 조정**하여 최종 MAE 결과를 산출함



5. 서비스 활용방안 및 기대효과

1

2

3

4

5

1) 서비스 활용방안

- 지면온도는 자연재해 예방 및 도시계획 수립에 있어 대단히 중요한 역할을 할 수 있음
- 구체적으로, 지면온도를 정확히 예측하여 **미래에 발생할 수 있는 자연재해로 인한 피해를 사전에 예방**할 수 있으며, 도시계획 정책 수립 및 건축물 건설 등에 지면온도 정보를 활용하여 **기후 친화적인 도시를 개발**하여 시민들에게 편의를 제공할 수 있음
- 본 공모전에서 구축한 계절별 통합 모델은 효과적으로 지면온도를 예측하여 해당 분야에서 활용이 가능할 것으로 기대됨

2) 실제 서비스 구현 예시

◆ 공공 서비스 측면

- 예측된 지면온도를 토양온도/토양수분 등의 데이터와 연계하여 **가뭄 예상지역 및 기간을 예측**한 뒤, 해당 지역에 농업 용수 등의 자원을 추가 분배하는 등 **피해를 최소화하기 위한 효율적인 자원 분배**를 달성할 수 있음
- **도심 열섬 현상의 원인을 파악하기 위해** 지역/계절별로 지면온도를 예측한 뒤, 지면온도가 지나치게 높은 지역을 추가 분석하여 해당 지역의 문제점을 효과적으로 포착하고, 도심 열섬 현상을 완화하기 위한 가로수 확충, 시민공원 조성 등의 **친환경 개발안을 미래 도시개발 계획에 반영**할 수 있음

◆ 기업 서비스 측면

- 건설업체는 예측된 지면온도 데이터를 반영하여 특정 **건축물 설립에 알맞은 부지를 선정**하고, 해당 부지의 **계절별 최저/최고 지면온도를 견딜 수 있는 적절한 자재**를 사용해 **사고 확률을 줄이고 추가 비용을 절감**할 수 있음

3) 향후 서비스 구현을 위한 제언

- 제안 모델의 EDA 분석과정에서 계절/지역/연도별로 지면온도 세부 추세 확인 시, 지역별로 지면온도 분포가 광범위한 것을 확인함
- 제안 모델은 모든 지역을 고려한 계절별 통합 모델이지만, 실무적으로 더욱 정확한 지면온도 예측을 위해서는 **각 지역의 특성이 반영된 보다 세분화된 모델**의 구축이 필요함
- 추후, **각 지역의 계절별로 데이터를 세부적으로 분할**하여 EDA 및 전처리를 거친 **계절-지역별 모델을 구축하면 더욱 실효적인 서비스가 될 것**으로 기대함





- Antal, A., Guerreiro, P. M., & Cheval, S. (2021). Comparison of spatial interpolation methods for estimating the precipitation distribution in Portugal. *Theoretical and Applied Climatology*, 145(3-4), 1193-1206.
- Chin, R. J., Lai, S. H., Loh, W. S., Ling, L., & Soo, E. Z. X. (2023). Assessment of Inverse Distance Weighting and Local Polynomial Interpolation for Annual Rainfall: A Case Study in Peninsular Malaysia. *Engineering Proceedings*, 38(1), 61.
- Han, W., Duan, S. B., Tian, H., & Lian, Y. (2023). Estimation of land surface temperature from AMSR2 microwave brightness temperature using machine learning methods. *International Journal of Remote Sensing*, 1-22.
- Ma, X., Fang, C., & Ji, J. (2020). Prediction of outdoor air temperature and humidity using Xgboost. In *IOP conference series: earth and environmental science* (Vol. 427, No. 1, p. 012013). IOP Publishing.
- Niu, D., Diao, L., Zang, Z., Che, H., Zhang, T., & Chen, X. (2021). A machine-learning approach combining wavelet packet denoising with Catboost for weather forecasting. *Atmosphere*, 12(12), 1618.
- Raju, V. G., Lakshmi, K. P., Jain, V. M., Kalidindi, A., & Padma, V. (2020, August). Study the influence of normalization/transformation process on the accuracy of supervised classification. In *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)* (pp. 729-735). IEEE.

Thank you for Listening to the Presentation



감사합니다

2023 날씨 빅데이터 콘테스트