팀 프로젝트

머신러닝을 활용한 미세먼지농도예측모델

홍승현

김진재

안세기

팀 프로젝트

기상환경에 따른 체감온도 예측

홍승현
김진재
안세기

목차

데이터 소개

기상자료포털

기상자료개방포털_지상기상환경

STN	지점
STN_NAME	지점명
TM	일시
AVG_TEMP	평균기온 (°C)
MIN_TEMP	최저기온 (°C)
MIN_TEMP_TIME	최저기온 시각 (hhmi)
MAX_TEMP	최고기온 (°C)
MAX_TEMP_TIME	최고기온 시각 (hhmi)
RAIN_DURATION	강수 계속시간 (hr)
MAX_RAIN_10M	10분 최다 강수량 (mm)
MAX_RAIN_10M_TIME	10분 최다강수량 시각 (hhmi)
MAX_RAIN_1H	1시간 최다강수량 (mm)
MAX_RAIN_1H_TIME	1시간 최다 강수량 시각 (hhmi)

DAILY_RAIN	일강수량 (mm)
MAX_GUST_SPEED	최대 순간 풍속 (m/s)
MAX_GUST_DIR	최대 순간 풍속 풍향 (16방위)
MAX_GUST_TIME	최대 순간풍속 시각 (hhmi)
MAX_WIND_SPEED	최대 풍속 (m/s)
MAX_WIND_DIR	최대 풍속 풍향 (16방위)
MAX_WIND_TIME	최대 풍속 시각 (hhmi)
AVG_WIND_SPEED	평균 풍속 (m/s)
WIND_RUN	풍정합 (100m)
DOMINANT_WIND_DIR	최다풍향 (16방위)
AVG_DEW_POINT	평균 이슬점온도 (°C)
MIN_HUMIDITY	최소 상대습도 (%)
MIN_HUMIDITY_TIME	최소 상대습도 시각 (hhmi)

데이터 소개

기상자료개방포털_지상기상환경

AVG_HUMIDITY	평균 상대습도 (%)
AVG_VAPOR_PRESSURE	평균 증기압 (hPa)
AVG_LOCAL_PRESSURE	평균 현지기압 (hPa)
MAX_SEA_PRESSURE	최고 해면기압 (hPa)
MAX_SEA_PRESSURE_TIME	최고 해면기압 시각 (hhmi)
MIN_SEA_PRESSURE	최저 해면기압 (hPa)
MIN_SEA_PRESSURE_TIME	최저 해면기압 시각 (hhmi)
AVG_SEA_PRESSURE	평균 해면기압 (hPa)
SUNSHINE_DURATION	가조시간 (hr)
TOTAL_SUNSHINE	합계 일조시간 (hr)
MAX_SOLAR_1H_TIME	1시간 최다일사 시각 (hhmi)
MAX_SOLAR_1H	1시간 최다일사량 (MJ/m²)
TOTAL_SOLAR	합계 일사량 (MJ/m²)
MAX_NEW_SNOW	일 최심신적설 (cm)
MAX_NEW_SNOW_TIME	일 최심신적설 시각 (hhmi)
MAX_SNOW_DEPTH	일 최심적설 (cm)
MAX_SNOW_DEPTH_TIME	일 최심적설 시각 (hhmi)
MIN_SEA_PRESSURE_TIME	최저 해면기압 시각 (hhmi)
AVG_SEA_PRESSURE	평균 해면기압 (hPa)
SUNSHINE_DURATION	가조시간 (hr)
TOTAL_SUNSHINE	합계 일조시간 (hr)
MAX_SOLAR_1H_TIME	1시간 최다일사 시각 (hhmi)
MAX_SOLAR_1H	1시간 최다일사량 (MJ/m²)
TOTAL_SOLAR	합계 일사량 (MJ/m²)
MAX_NEW_SNOW	일 최심신적설 (cm)
MAX_NEW_SNOW_TIME	일 최심신적설 시각 (hhmi)
MAX_SNOW_DEPTH	일 최심적설 (cm)
MAX_SNOW_DEPTH_TIME	일 최심적설 시각 (hhmi)
NEW_SNOW_3H	합계 3시간 신적설 (cm)
AVG_CLOUD_TOTAL	평균 전운량 (1/10)
AVG_CLOUD_LOW	평균 중하층운량 (1/10)
AVG_GROUND_TEMP	평균 지면온도 (°C)
MIN_SURFACE_TEMP	최저 초상온도 (°C)
SOIL_TEMP_5CM	평균 5cm 지중온도 (°C)
SOIL_TEMP_10CM	평균 10cm 지중온도 (°C)
SOIL_TEMP_20CM	평균 20cm 지중온도 (°C)

기상자료개방포털 체감온도

일자	DATE
기온(°C)	TEMP
풍속(km/h)	WIND_SPEED
체감온도(°C)	FEELS_LIKE_TEMP

데이터 컬럼의 수는 약 60여개로 파악되며
이 데이터 셋을 활용하여 체감온도를
예측해볼 예정입니다.

데이터 전처리 과정

데이터 수집

수집한 데이터 자료가 많아 각 컬럼 값을 병합 데이터에 맞게 자료 수집

데이터 병합

수집된 2개의 DATA SET의 시간을 통하여 병합하여
NULL 값 을 가진 데이터를 제거 할 예정입니다.

학습 데이터 분리

단일값을 가진 데이터 및 불필요한 데이터 제거 후 train , test 데이터 파일 생성

EDA

각자 수집한 데이터를 합치는 과정에서 데이터 형식이 다르게 표기되어 데이터 형식을 맞추었습니다.

```
df_1["tm"] = pd.to_datetime(
    df_1["tm"],
    format="%Y%m%d%H%M"
)

df_1.columns = df_1.columns.str.upper()
```

Before

	tm
0	202301010000
1	202301010100
2	202301010200
3	202301010400
4	202301010600

	TM
0	2024-01-01 00:00:00
1	2024-01-01 02:00:00
2	2024-01-01 03:00:00
3	2024-01-01 04:00:00
4	2024-01-01 05:00:00

After

	TM
0	2023-01-01 00:00:00
1	2023-01-01 01:00:00
2	2023-01-01 02:00:00
3	2023-01-01 04:00:00
4	2023-01-01 06:00:00

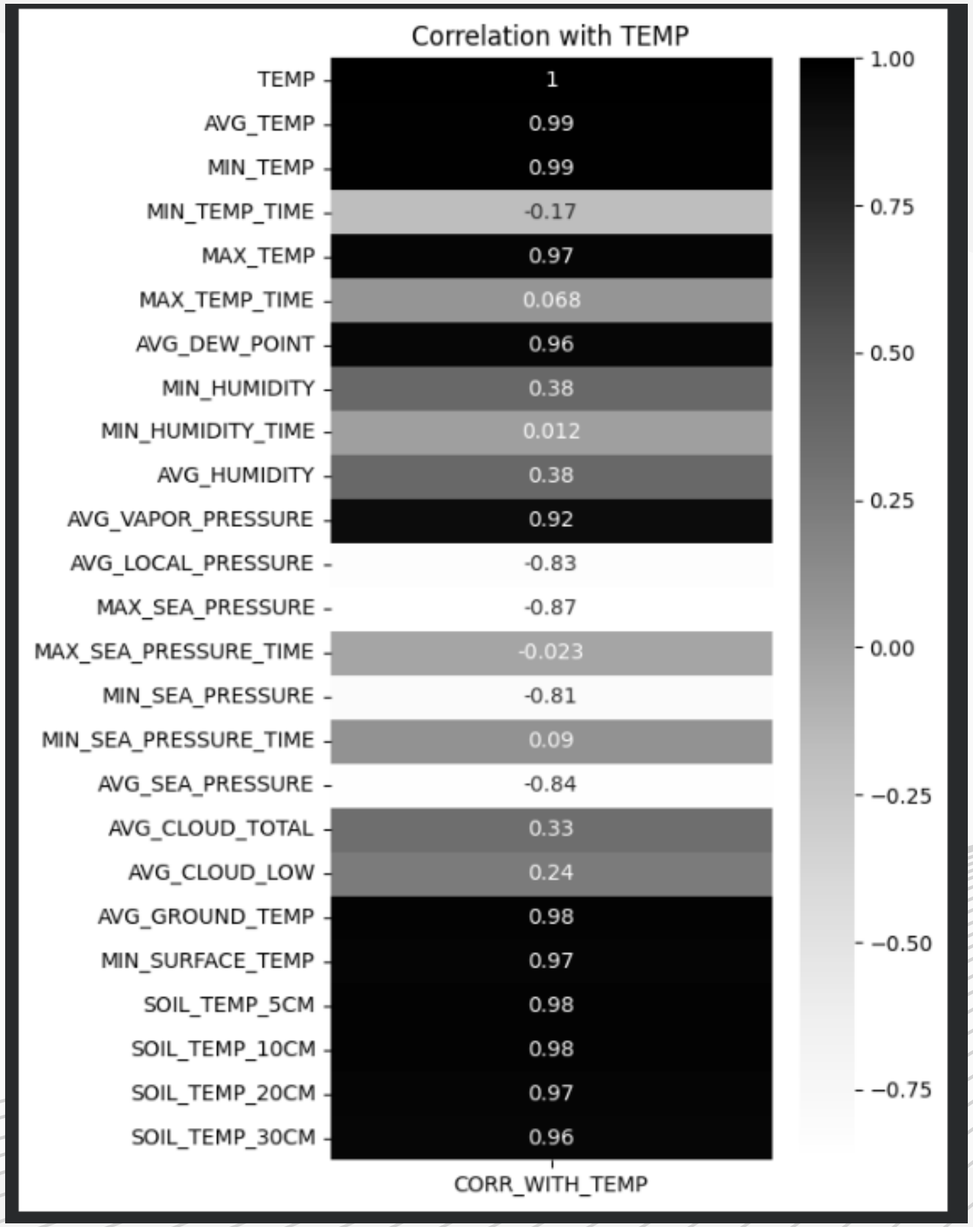
	TM
0	2024-01-01 00:00:00
1	2024-01-01 02:00:00
2	2024-01-01 03:00:00
3	2024-01-01 04:00:00
4	2024-01-01 05:00:00

EDA

TEMP(체감온도)컬럼을 제외한 나머지 컬럼의
NaN값을 제거한뒤 확인 한 결과 28개로
많은 컬럼수가 줄었으며, 0.2이하의 상관계수를 나타내는
데이터를 삭제하여 feature를 설정하였습니다.

```
[60]
✓ 0초
len(df.columns)
28
```

```
corr_with_temp = df.corr(numeric_only=True)['TEMP'].drop('TEMP')
corr_with_temp = corr_with_temp[corr_with_temp.abs() >= 0.2]
corr_with_temp
```



모델링

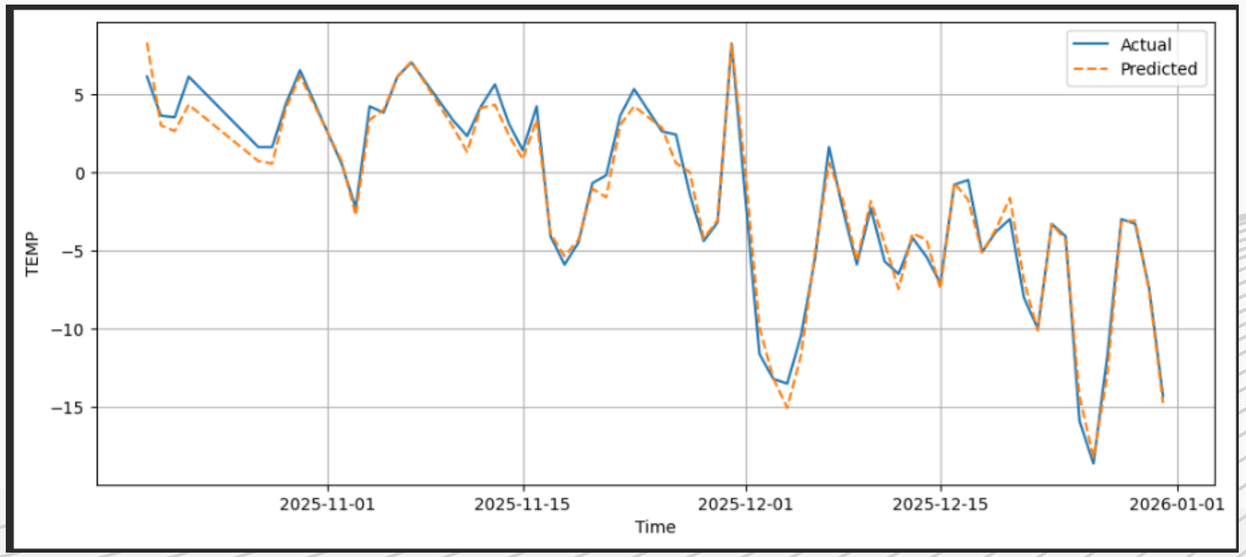
모델 : [RandomForest]

x_train = 상관관계 0.2 이상데이터
x_test,y_test = 25년도 겨울 데이터

회귀분석으로 테스트 한결과
R2의 점수가 91%로 잘 가져오는것을
볼수 있으나 train 컬럼이 너무 많아
새로운 데이터 수집에 어려움을 겪을수도
있다 판단하여 컬럼을 수정,
풍속(-0.15)의 상관계수는 낮지만 겨울에는
체감온도에 영향을 준다는 자료를
바탕으로 추가하여 진행해보았습니다.

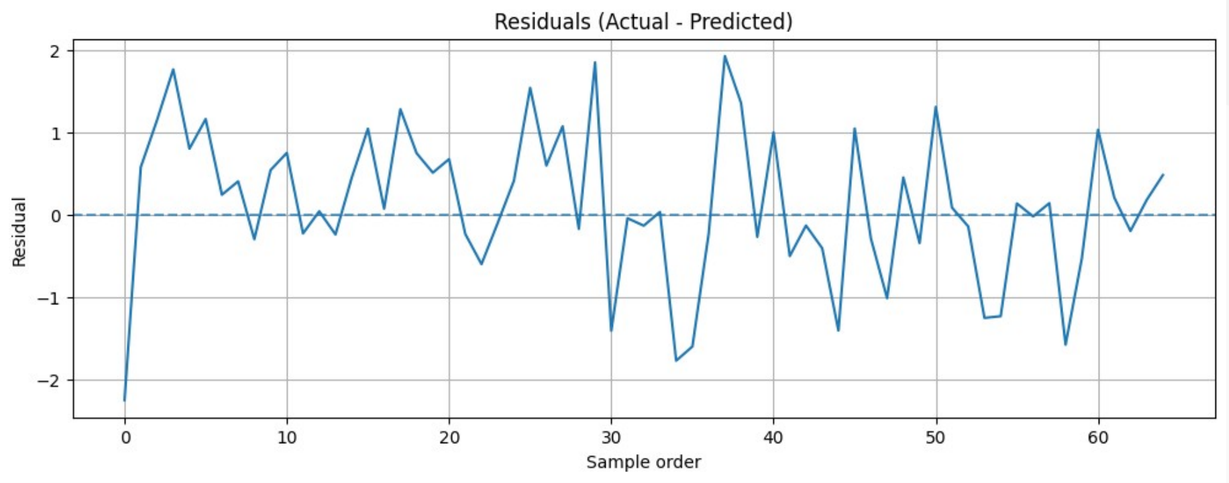
```
R2   : 0.9170
MAE  : 0.8704
RMSE : 1.7684
```

MAX_WIND_SPEED	-0.066385
MAX_WIND_DIR	-0.077862
MAX_WIND_TIME	0.128774
AVG_WIND_SPEED	-0.155309



모델링

컬럼을 정리한 뒤 25년도 겨울데이터를 모델을 적용한 결과,
약 98%에 가까운 예측 성능이 나왔습니다.
잔차 그래프를 통해 실제값과 예측값의 차이를
확인한 결과, 최대 오차는 1.8°C 정도이며
전반적으로 0을 중심으로 분포해 높은 정확도를 보였습니다.



Before

```
R2   : 0.9170
MAE  : 0.8704
RMSE : 1.7684
```

After

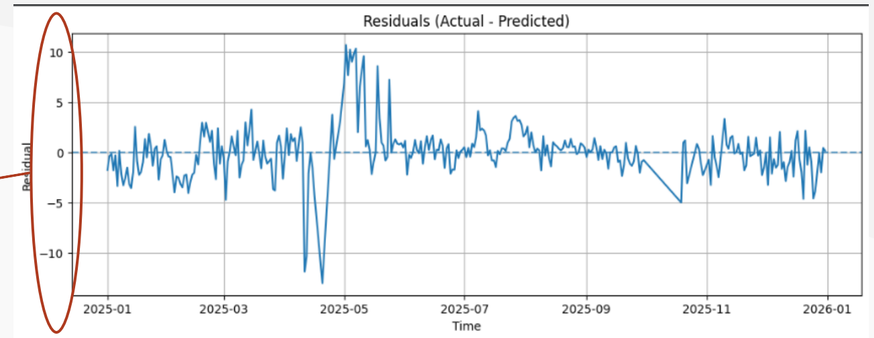
```
R2   : 0.9781
MAE  : 0.7038
RMSE : 0.9090
```

모델성능비교

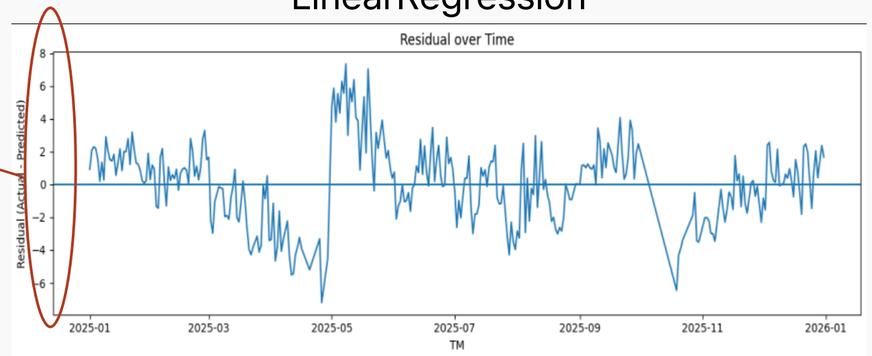
각 모델 성능비교를 확인하고자
25년을 예측해보았으며 각 모델은
예측성능이 좋게 나오는 것을
확인할 수 있었습니다.

전차 그래프를 통해 전차를 확인해본 결과에
따라서 Linear Regression 모델을
최종 예측 모델로 선정하였습니다.

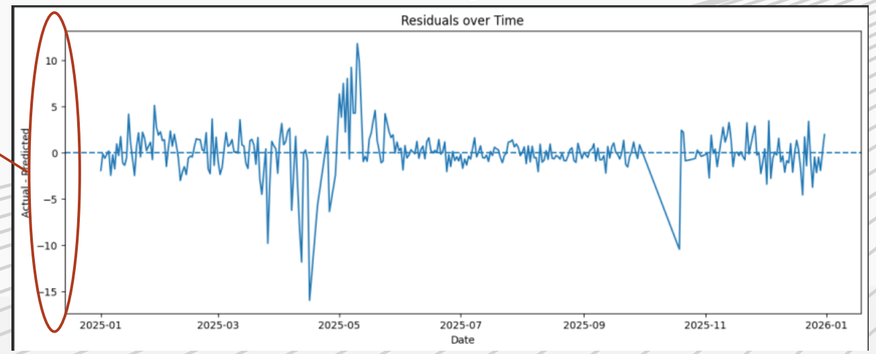
RandomForestRegressor



LinearRegression



XGBRegressor



RANDOM FOREST

R2 : 0.9807
MAE : 1.3895
RMSE : 2.3671

Linear_model

R2 Score: 0.9811145008318106
RMSE: 2.3390357132411537
MAE: 1.7983426746827942

XGboost

MAE : 0.9688483511560534
RMSE: 1.8642999697968927
R2 : 0.9880267288760919

시스템 개발 순서

학습 데이터 저장 — 일기예보 API — 예측 매핑 — 대시보드

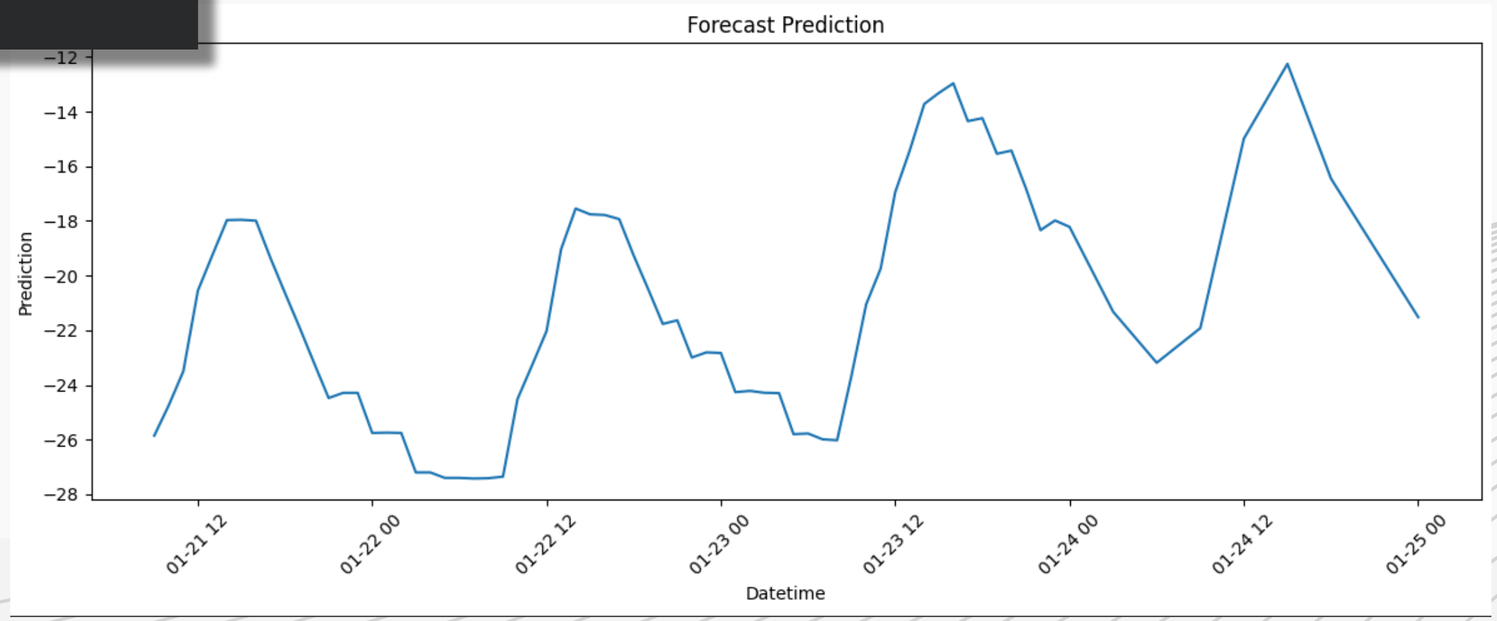
일기예보 api

```
df_fcst.columns

Index(['fcstDate', 'fcstTime', 'PCP', 'POP', 'PTY', 'REH',
'SKY', 'SNO', 'TMN',
       'TMP', 'TMX', 'UUU', 'VEC', 'VVV', 'WAV', 'WSD',
'datetime'],
```

학습데이터 DF

```
#컬럼명 변경
merge_df = merge_df.rename(columns={
    'AVG_TEMP': 'TMP',
    'AVG_WIND_SPEED': 'WSD',
    'AVG_DEW_POINT': 'AVG_TD',
    'AVG_HUMIDITY': 'REH',
})
```



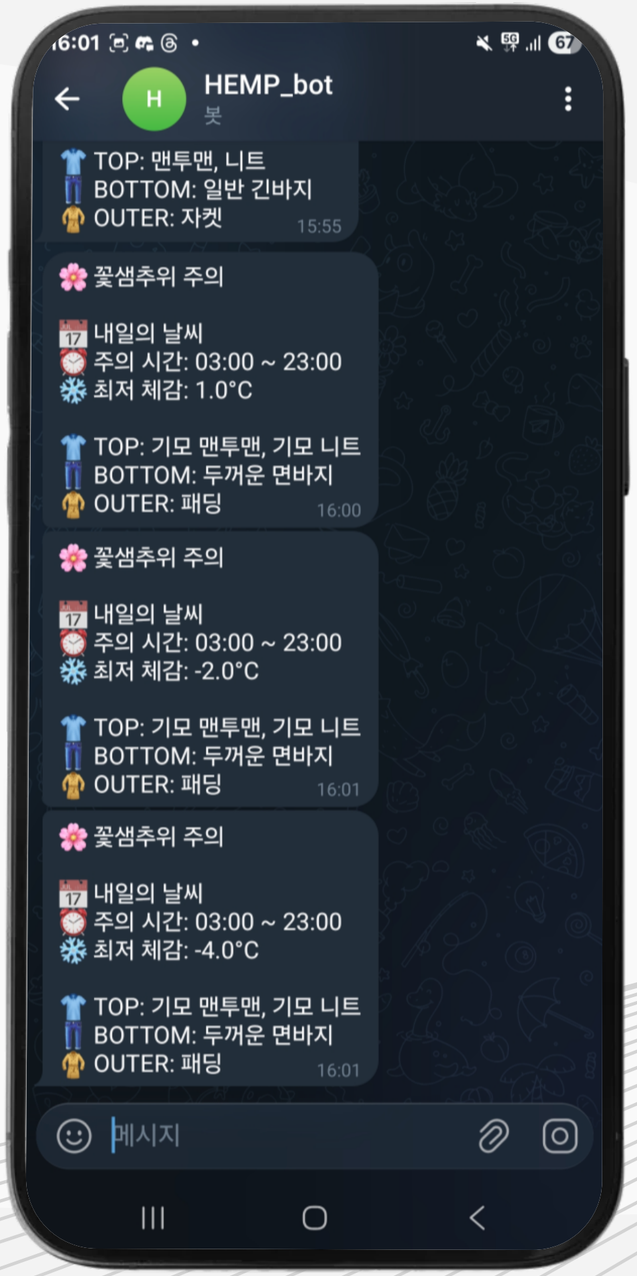
꽃샘추위,가을추위 감지 알림 서비스

내일의 체감온도를 예측하여 봄,가을 기간
10도 이하일때 알림을 통하여 복장을 추천해주고
내일의 날씨 정보를 채팅 봇을 통하여
전달해주며 실시간으로 확인이 가능합니다.

```
내 chat_id = 5690128314
🌸 꽃샘추위 주의
🗓️ 날짜: 2026-04-24
⏰ 체감 10° C 이하: 00:00 ~ 23:00
❄️ 최저 체감: -26.0° C

👕 TOP: 히트텍, 두꺼운 니트, 목도리(필수)
👖 BOTTOM: 방한 바지, 기모 레깅스(내복)
🧥 OUTER: 롱패딩, 두꺼운 코트(대체)
```

```
이번 달은 (3~5월, 8~10월) 시즌이 아니라 알림을 보내지 않았어요.
```



대시보드 시연

최근날씨 조회 및 복장 추천 서비스

체질에 따른 프리셋 설정

향후계획

대시보드 보완

대시보드에서 프리셋 로직을 짠 뒤 프리셋이 가능하게 하고, api 호출('/refresh')를 스케쥴러에 적용

복장 세분화

더욱 다양한 의상을 추가하여 패션 쪽으로도 옷차림을 정할수 있게끔 설계

콘텐츠 확장

복장 세분화에 따라 의류 쇼핑몰을 통하여 날씨에 맞는 복장을 모델 사진을 통하여 추천 받을수 있도록 만들 예정