

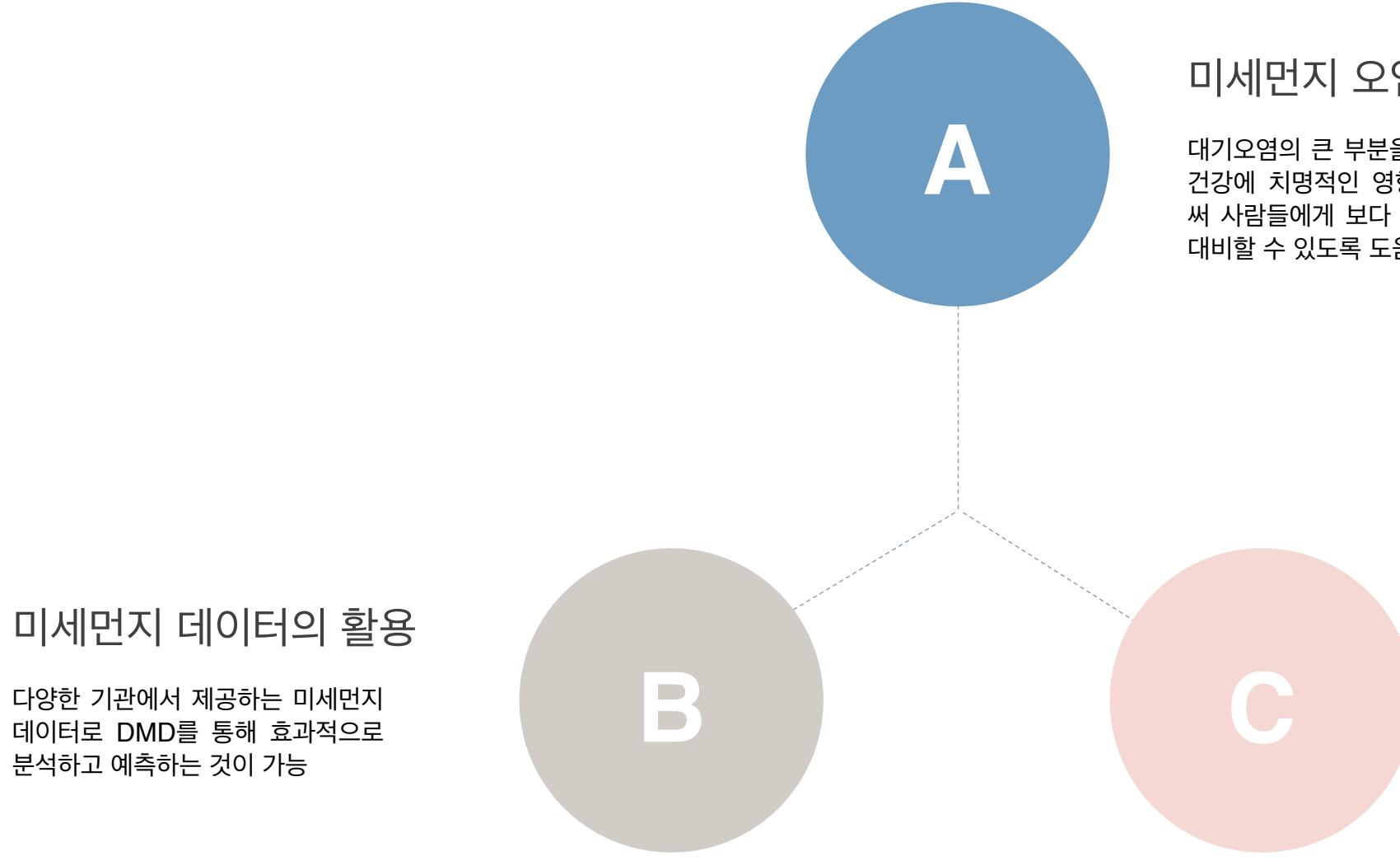
DMD를 이용한 미세먼지 오염도 예측



Part 1

주제 선정 이유

주제 선정 이유



미세먼지 오염도 예측의 중요성

대기오염의 큰 부분을 차지하는 미세먼지는 인간의 건강에 치명적인 영향을 미치며, 이를 예측함으로써 사람들에게 보다 정확한 정보를 제공하고, 미리 대비할 수 있도록 도움을 줄 수 있음

DMD의 장점

DMD는 데이터의 동적 패턴을 추출하여 시간적 변화를 분석하는 데 유용한 방법이기 때문에 시간에 따른 미세먼지 오염도를 효율적으로 예측할 수 있고, 단순하고 직관적인 해석이 가능

Part 2

데이터 전 처리

미세먼지 데이터

지역	망	측정소코드	측정소명	측정일시	SO2	CO	O3	NO2	PM10	PM25	주소
서울 중구	도시대기	111121	중구	2023010101	0.004	0.8	0.004	0.055	57	52	서울 중구 덕수궁길 15
서울 중구	도시대기	111121	중구	2023010102	0.004	0.8	0.004	0.051	65	60	서울 중구 덕수궁길 15
서울 중구	도시대기	111121	중구	2023010103	0.004	0.8	0.005	0.048	73	63	서울 중구 덕수궁길 15
서울 중구	도시대기	111121	중구	2023010104	0.004	0.7	0.014	0.032	76	71	서울 중구 덕수궁길 15
서울 중구	도시대기	111121	중구	2023010105	0.004	0.7	0.021	0.021	78	73	서울 중구 덕수궁길 15
서울 중구	도시대기	111121	중구	2023010106	0.004	0.7	0.022	0.021	68	61	서울 중구 덕수궁길 15
서울 중구	도시대기	111121	중구	2023010107	0.003	0.7	0.015	0.028	67	58	서울 중구 덕수궁길 15
서울 중구	도시대기	111121	중구	2023010108	0.003	0.8	0.009	0.034	73	65	서울 중구 덕수궁길 15
서울 중구	도시대기	111121	중구	2023010109	0.004	0.9	0.013	0.03	77	65	서울 중구 덕수궁길 15
서울 중구	도시대기	111121	중구	2023010110	0.004	0.8	0.019	0.025	81	68	서울 중구 덕수궁길 15
서울 중구	도시대기	111121	중구	2023010111	0.004	0.8	0.028	0.019	78	67	서울 중구 덕수궁길 15
서울 중구	도시대기	111121	중구	2023010112	0.005	0.8	0.032	0.019	75	64	서울 중구 덕수궁길 15
서울 중구	도시대기	111121	중구	2023010113	0.005	0.7	0.035	0.019	72	64	서울 중구 덕수궁길 15
서울 중구	도시대기	111121	중구	2023010114	0.005	0.6	0.04	0.015	64	54	서울 중구 덕수궁길 15
서울 중구	도시대기	111121	중구	2023010115	0.004	0.5	0.044	0.012	49	41	서울 중구 덕수궁길 15
서울 중구	도시대기	111121	중구	2023010116	0.004	0.5	0.044	0.011	43	34	서울 중구 덕수궁길 15
서울 중구	도시대기	111121	중구	2023010117	0.003	0.4	0.04	0.011	32	26	서울 중구 덕수궁길 15
서울 중구	도시대기	111121	중구	2023010118	0.004	0.4	0.037	0.014	23	19	서울 중구 덕수궁길 15
서울 중구	도시대기	111121	중구	2023010119	0.004	0.4	0.033	0.018	25	17	서울 중구 덕수궁길 15
서울 중구	도시대기	111121	중구	2023010120	0.004	0.4	0.032	0.019	26	22	서울 중구 덕수궁길 15
서울 중구	도시대기	111121	중구	2023010121	0.003	0.4	0.033	0.018	26	19	서울 중구 덕수궁길 15
서울 중구	도시대기	111121	중구	2023010122	0.003	0.4	0.036	0.013	23	17	서울 중구 덕수궁길 15
서울 중구	도시대기	111121	중구	2023010123	0.003	0.4	0.036	0.013	20	15	서울 중구 덕수궁길 15
서울 중구	도시대기	111121	중구	2023010124	0.003	0.4	0.033	0.014	19	11	서울 중구 덕수궁길 15
서울 중구	도시대기	111121	중구	2023010201	0.003	0.4	0.034	0.013	20	12	서울 중구 덕수궁길 15
서울 중구	도시대기	111121	중구	2023010202	0.003	0.4	0.034	0.014	15	14	서울 중구 덕수궁길 15
서울 중구	도시대기	111121	중구	2023010203	0.003	0.4	0.031	0.016	18	12	서울 중구 덕수궁길 15
서울 중구	도시대기	111121	중구	2023010204	0.003	0.4	0.03	0.015	17	15	서울 중구 덕수궁길 15
서울 중구	도시대기	111121	중구	2023010205	0.003	0.4	0.031	0.013	18	12	서울 중구 덕수궁길 15
서울 중구	도시대기	111121	중구	2023010206	0.003	0.4	0.03	0.012	19	13	서울 중구 덕수궁길 15
서울 중구	도시대기	111121	중구	2023010207	0.003	0.5	0.021	0.021	19	15	서울 중구 덕수궁길 15
서울 중구	도시대기	111121	중구	2023010208	0.003	0.5	0.017	0.023	19	13	서울 중구 덕수궁길 15

데이터 출처-에어코리아

데이터 전 처리 과정

에어코리아에서 2022.01~2024.01의 데이터를
행:날짜로부터 7일간의 평균치로
열: 측정소코드_오염물질명으로 정리

결측 값은 이전 값과 다음 값을 선형으로 채워 넣었음

데이터 값의 크기차이가 많이 나서 DMD로 예측할 때
작은 값은 반영이 거의 되지 않기 때문에 정규화
이때 미세먼지 오염도가 음수가 되지 않도록 Min-Max로 진행

최종 전 처리 데이터

	111121_CO	111121_NO2	111121_O3	111121_PM10	111121_PM25	111121_SO2	111123_CO	111123_NO2	111123_O3	111123_PM10	111123_PM25	111123_SO2	111124_CO	111124_NO2
2021-02-01	0.516407599	0.457142857	0.269230769	0.281464109	0.458960356	0.5	0.566610455	0.472222222	0.235294118	0.281466278	0.548749184	1	0.703642384	0.6
2021-02-08	0.678756477	0.8	0.211538462	0.369136713	0.760194175	1	0.792580101	0.777777778	0.196078431	0.355326571	0.844572547	1	0.84602649	0.933333333
2021-02-15	0.495682211	0.371428571	0.326923077	0.296820649	0.412014563	0.5	0.568296796	0.388888889	0.31372549	0.291917678	0.493212965	1	0.677152318	0.5
2021-02-22	0.507772021	0.6	0.230769231	0.242226204	0.300728155	0.5	0.568296796	0.583333333	0.215686275	0.231941214	0.36871873	1	0.754966887	0.7
2021-03-01	0.450777202	0.542857143	0.230769231	0.156964794	0.252002427	0.5	0.524451939	0.527777778	0.196078431	0.145114741	0.313226017	0.5	0.71192053	0.666666667
2021-03-08	0.687392055	0.8	0.326923077	0.468990461	0.847835761	0.5	0.741989882	0.722222222	0.294117647	0.44286755	1	0.5	0.943708609	1
2021-03-15	0.438687392	0.428571429	0.403846154	0.481257247	0.421804207	0.5	0.470489039	0.388888889	0.392156863	0.457241727	0.560408962	0.5	0.657284768	0.6
2021-03-22	0.552677029	0.714285714	0.346153846	0.39882866	0.62184466	1	0.569983137	0.611111111	0.333333333	0.3789543	0.598303241	1	0.771523179	0.966666667
2021-03-29	0.366148532	0.428571429	0.365384615	0.576947402	0.372390777	0.5	0.370994941	0.361111111	0.37254902	0.539837205	0.335240374	0.5	0.617549669	0.6
2021-04-05	0.274611399	0.285714286	0.596153846	0.209233688	0.344457929	0.5	0.284991568	0.25	0.607843137	0.196131861	0.356036546	1	0.455298013	0.466666667
2021-04-12	0.26597582	0.314285714	0.442307692	0.256970064	0.246480583	0	0.278246206	0.277777778	0.431372549	0.246098238	0.276571677	0.5	0.463576159	0.433333333
2021-04-19	0.388601036	0.457142857	0.711538462	0.300332033	0.435355987	1	0.387858347	0.388888889	0.725490196	0.293970131	0.462345008	1	0.607615894	0.7
2021-04-26	0.27806563	0.228571429	0.557692308	0.205452198	0.239906958	0.5	0.276559865	0.194444444	0.549019608	0.198268374	0.2458995	0.5	0.42218543	0.333333333
2021-05-03	0.322970639	0.314285714	0.442307692	1	0.60467233	0.5	0.310286678	0.305555556	0.431372549	1	0.518881879	0.5	0.44205298	0.4
2021-05-10	0.33851468	0.4	0.480769231	0.20815985	0.354631877	1	0.320404722	0.361111111	0.392156863	0.163853008	0.279856428	0.5	0.51986755	0.633333333
2021-05-17	0.298791019	0.257142857	0.519230769	0.111784547	0.183960356	1	0.291736931	0.25	0.490196078	0.116205273	0.199956493	1	0.400662252	0.366666667
2021-05-24	0.303972366	0.171428571	0.596153846	0.343621535	0.301719256	1	0.305227656	0.194444444	0.568627451	0.351235675	0.298368501	1	0.430463576	0.333333333
2021-05-31	0.343696028	0.257142857	0.711538462	0.22852983	0.420226537	1	0.333895447	0.222222222	0.705882353	0.231471882	0.381118121	1	0.516556291	0.433333333
2021-06-07	0.362694301	0.371428571	0.576923077	0.18746706	0.381998382	1	0.344013491	0.361111111	0.568627451	0.184258455	0.357385251	1	0.531456954	0.6
2021-06-14	0.354058722	0.257142857	0.557692308	0.158862127	0.329025081	1	0.313659359	0.194444444	0.588235294	0.12977388	0.30432891	1	0.476821192	0.333333333
2021-06-21	0.30224525	0.171428571	0.615384615	0.131844893	0.2677589	0.5	0.271500843	0.111111111	0.68627451	0.111308807	0.280313248	1	0.42218543	0.3
2021-06-28	0.303972366	0.228571429	0.403846154	0.113550121	0.246945793	0.5	0.254637437	0.138888889	0.450980392	0.091190563	0.254056994	0.5	0.438741722	0.333333333
2021-07-05	0.303972366	0.257142857	0.557692308	0.137273374	0.274514563	0.5	0.173693086	0.222222222	0.588235294	0.117375102	0.276071351	1	0.610927152	0.6
2021-07-12	0.184801382	0.228571429	0.403846154	0.103081849	0.19848301	0.5	0.087689713	0.166666667	0.450980392	0.081495699	0.224559495	1	0.529801325	0.533333333
2021-07-19	0.224525043	0.342857143	0.461538462	0.110717297	0.233616505	1	0.15851602	0.25	0.470588235	0.073264872	0.251294322	0.5	0.448675497	0.466666667
2021-07-26	0.248704663	0.142857143	0.653846154	0.149559924	0.337398867	1	0.217537943	0.166666667	0.725490196	0.12026815	0.368479443	1	0.369205298	0.4
2021-08-02	0.150259067	0.114285714	0.326923077	0.079048962	0.141868932	0.5	0.107925801	0.111111111	0.352941176	0.055962622	0.179595388	0.5	0.17218543	0.166666667
2021-08-09	0.2417962	0.142857143	0.557692308	0.096651997	0.188187702	0.5	0.202360877	0.083333333	0.607843137	0.073096753	0.225821188	0.5	0.302980132	0.2
2021-08-16	0.227979275	0.114285714	0.5	0.049231844	0.109263754	0.5	0.15682968	0.027777778	0.529411765	0.044250329	0.144463781	0.5	0.32615894	0.2
2021-08-23	0.455958549	0.342857143	0.326923077	0.10246258	0.231027508	0.5	0.36762226	0.277777778	0.333333333	0.12107372	0.233935175	0.5	0.587748344	0.466666667
2021-08-30	0.288428325	0.142857143	0.423076923	0.036543428	0.091039644	0.5	0.225969646	0.138888889	0.392156863	0.041742554	0.106808788	0.5	0.408940397	0.266666667
2021-09-06	0.331606218	0.285714286	0.423076923	0.057710551	0.137864078	0.5	0.286677909	0.277777778	0.431372549	0.071961949	0.182662606	0.5	0.483443709	0.433333333
2021-09-13	0.25388601	0.142857143	0.442307692	0.020771846	0.041302589	0.5	0.165261383	0.111111111	0.411764706	0.029441845	0.070437242	0.5	0.337748344	0.233333333

Part 3

DMD로 미세먼지 오염도 예측

DMD란 ?

행렬 A의 정의. $Ax(t_i) = x(t_{i+1})$.

$$X = \begin{bmatrix} x(t_0) & x(t_1) & \cdots & x(t_{m-1}) \end{bmatrix}$$

$$Y = \begin{bmatrix} x(t_1) & x(t_2) & \cdots & x(t_m) \end{bmatrix}$$

$\times A$ $\times A$ $\times A$

1. 행렬 X에 대한 SVD : $X \approx \tilde{U} \tilde{\Sigma} \tilde{V}^T$ (rank r)
2. 행렬 \tilde{A} ($r \times r$) 계산 : $\tilde{A} = \tilde{U}^T Y \tilde{V} \tilde{\Sigma}^{-1}$
3. 행렬 \tilde{A} 의 eigenvalue, eigenvector 찾기 $\tilde{A} W = W \Lambda$
4. 행렬 A의 DMD-mode (eigenvector), eigenvalue :
 $(DMD\text{-}mode) \Phi = Y \tilde{V} \tilde{\Sigma}^{-1} W \in \mathbb{C}^{n \times r}$. $\Phi = [\phi_1, \phi_2, \dots, \phi_r]$
 $(DMD\text{-}eigenvalue) \Lambda = [\lambda_1 \dots \lambda_r]$

행렬 A의 근사

STEP 1

```
import pandas as pd
import numpy as np
from pydmd import DMD
from sklearn.metrics import mean_squared_error, mean_absolute_error

csv_file_path = r"C:\Users\dispo\OneDrive\문서\카카오톡 받은 파일\normalized_data1.csv"
pm_data = pd.read_csv(csv_file_path, index_col=0)
pm_data.index = pd.to_datetime(pm_data.index)
```

- Python에서 DMD를 사용하기 위해 pydmd 라이브러리를 불러오기
- 전 처리한 데이터를 읽어오기

STEP 2

```
train_start, train_end = '2021-02-01', '2023-01-31'  
val_start, val_end = '2023-02-01', '2024-01-31'  
  
train_data = pm_data.loc[train_start:train_end].to_numpy()  
val_data = pm_data.loc[val_start:val_end].to_numpy()  
  
dmd = DMD()  
dmd.fit(train_data.T)
```

- 훈련 데이터를 2021년 02월 부터 2023년 01월까지 설정
- 검증 데이터를 2023년 02월 부터 2024년 01월까지 설정
- 훈련 데이터로 DMD 모델을 학습

STEP 3

```
reconstructed_data = np.real(dmd.reconstructed_data.T)
val_predictions = reconstructed_data[-val_data.shape[0]:]
rmse = np.sqrt(mean_squared_error(val_data, val_predictions))
mae = mean_absolute_error(val_data, val_predictions)

print("RMSE:", rmse)
print("MAE:", mae)
```

```
RMSE: 0.204499148612537
MAE: 0.15510920828586225
```

- 검증 데이터에 해당하는 예측값(val_predictions)을 얻어냄
- 모델 평가를 위해 rmse와 mae를 계산 후 출력
(정규화 된 데이터로 0.2정도의 오차면 예측 성능이 아직 부족하다 판斷)

STEP 4

```
data_matrix = pm_data.to_numpy().T # 데이터를 DMD용으로 전치  
  
# DMD 모델 초기화 및 학습  
svd_rank = 20 # 차원 축소 단계  
dmd = DMD(svd_rank=svd_rank)  
dmd.fit(data_matrix)  
  
# 예측 날짜 설정  
future_steps = 20 # 1일, 2일, 3일 후  
dmd.dmd_time['tend'] = dmd.dmd_time['t0'] + future_steps  
future_data = dmd.reconstructed_data.real[:, -future_steps:]  
  
last_date = pd.to_datetime(pm_data.index[-1])  
predicted_dates = [last_date + pd.Timedelta(days=i + 1) for i in range(future_steps)]
```

- 모델 성능 향상을 위해 svd의 rank를 20으로 축소
- 정확도를 위해 근시일 예측을 하기로 했고 임의로 예측 날짜를 20일로 설정

STEP 5

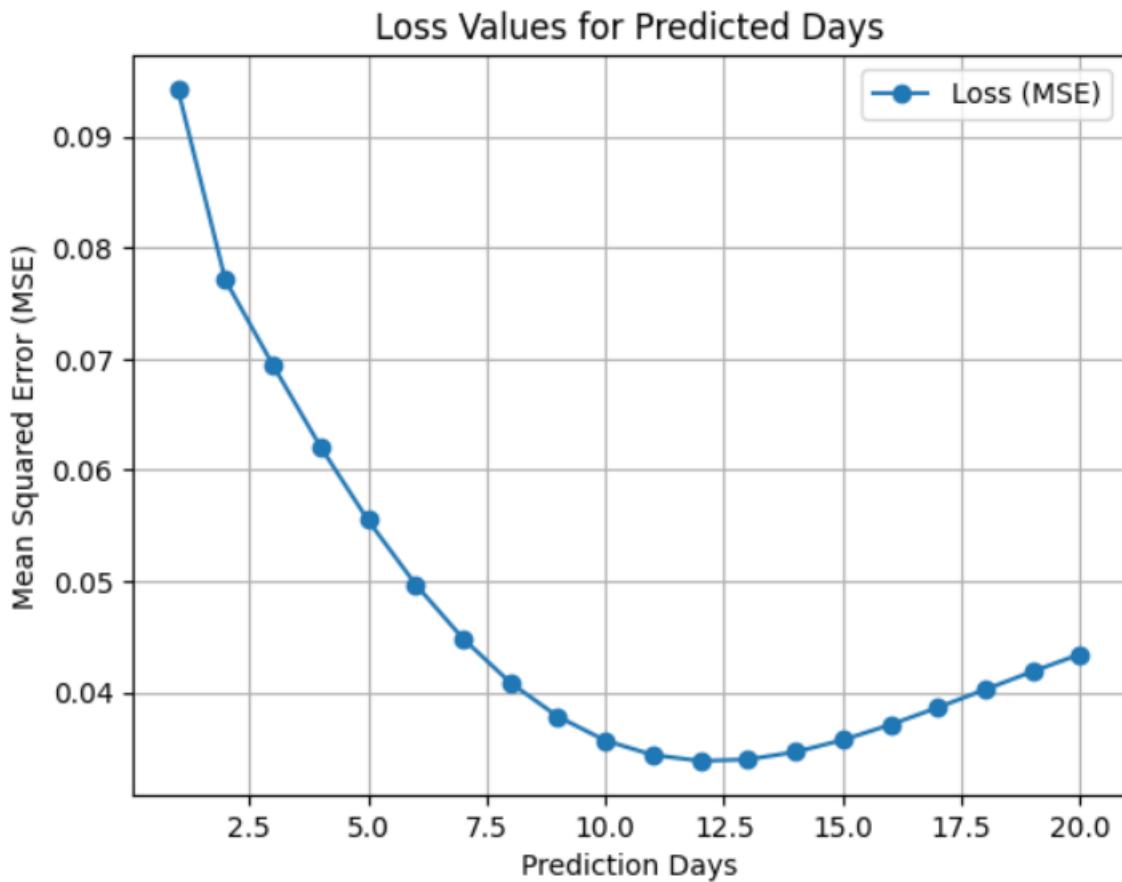
```
# 예측 데이터를 DataFrame으로 변환
predicted_df = pd.DataFrame(future_data.T, columns=pm_data.columns, index=predicted_dates)

# 실제 값: 데이터의 마지막 행
actual_values = pm_data.iloc[-1].values

# 손실 값 계산
loss_values = []
for i in range(future_steps):
    # 예측 값 가져오기
    predicted = future_data[:, i]
    # MSE 계산
    mse = mean_squared_error(actual_values, predicted)
    loss_values.append(mse)
```

- 예측 데이터를 저장
- 실제 값과 예측 값의 mse 계산 후 loss 값 분석

STEP 6

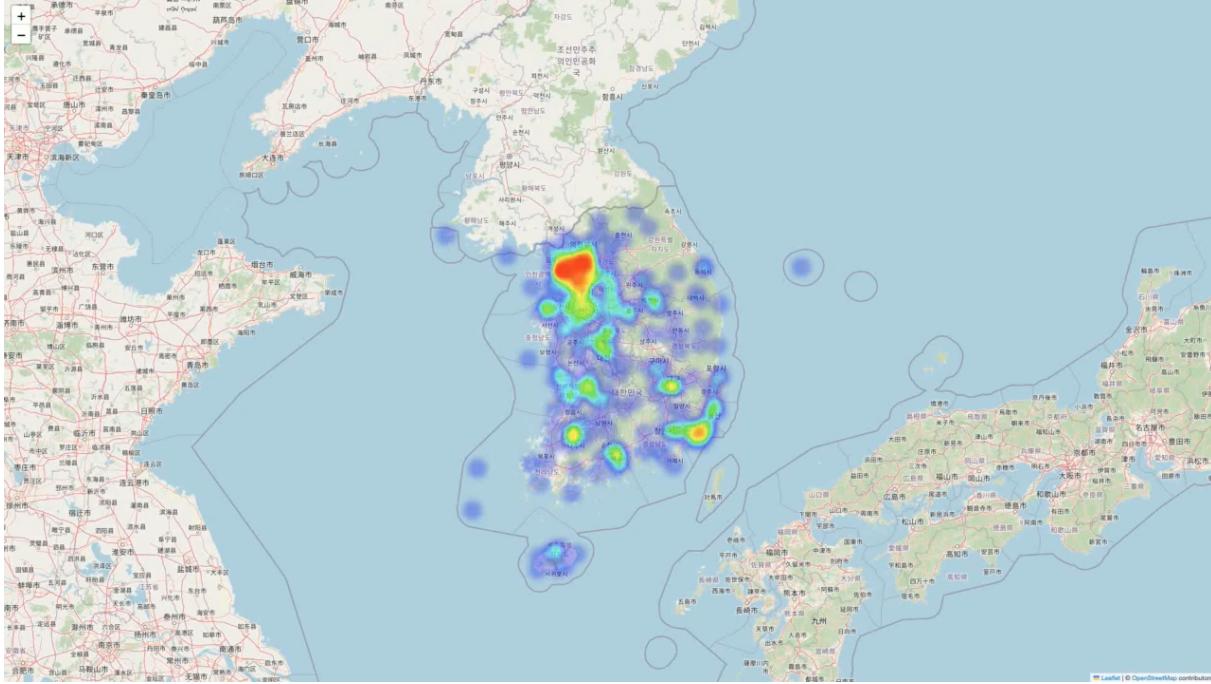


- 오차 값이 0.1 미만으로 모델 성능을 향상시키는데 성공
- Loss 값이 최소가 되는 예측 날짜가 12일이므로 2024-02-12까지 예측

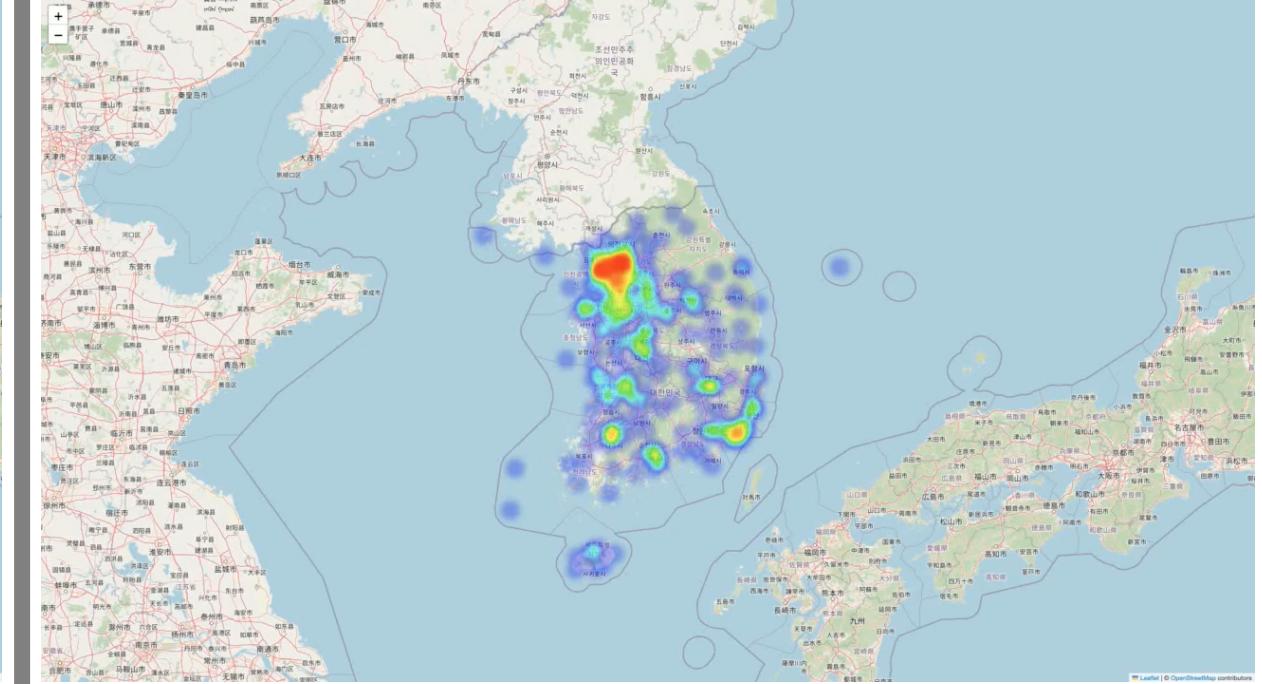
Part 4
시각화

예측한 데이터 시각화 후 비교

실제 미세먼지 데이터(pm10)

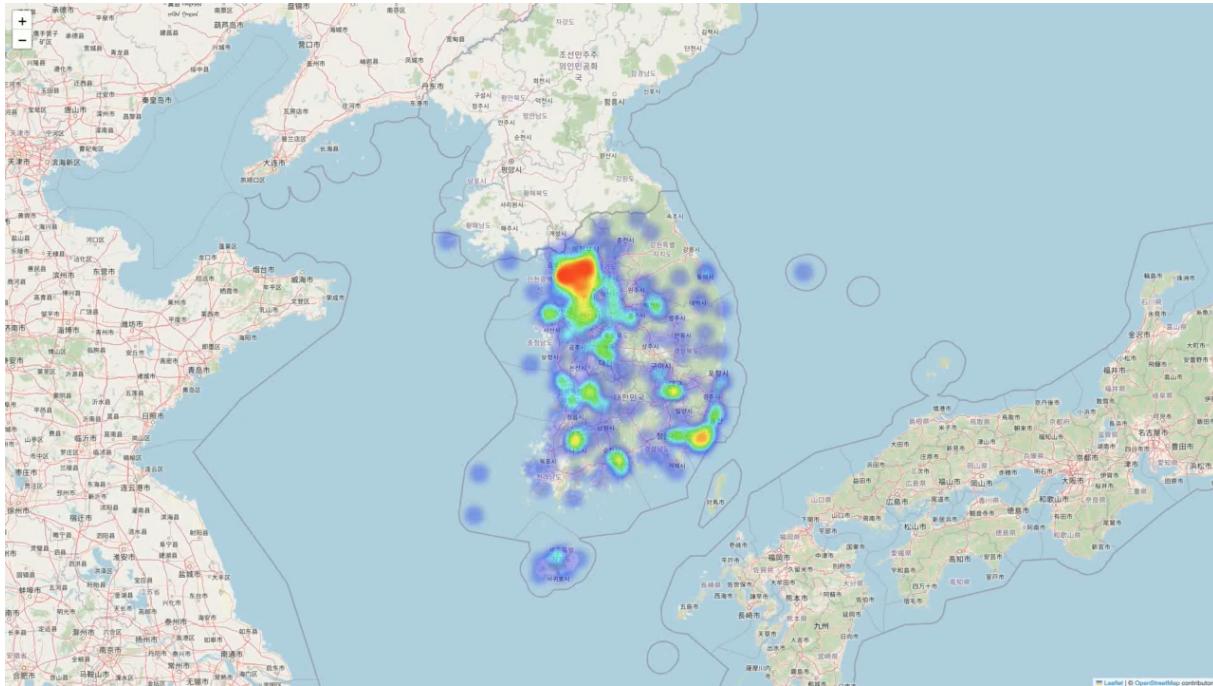


예측한 미세먼지 데이터(pm10)

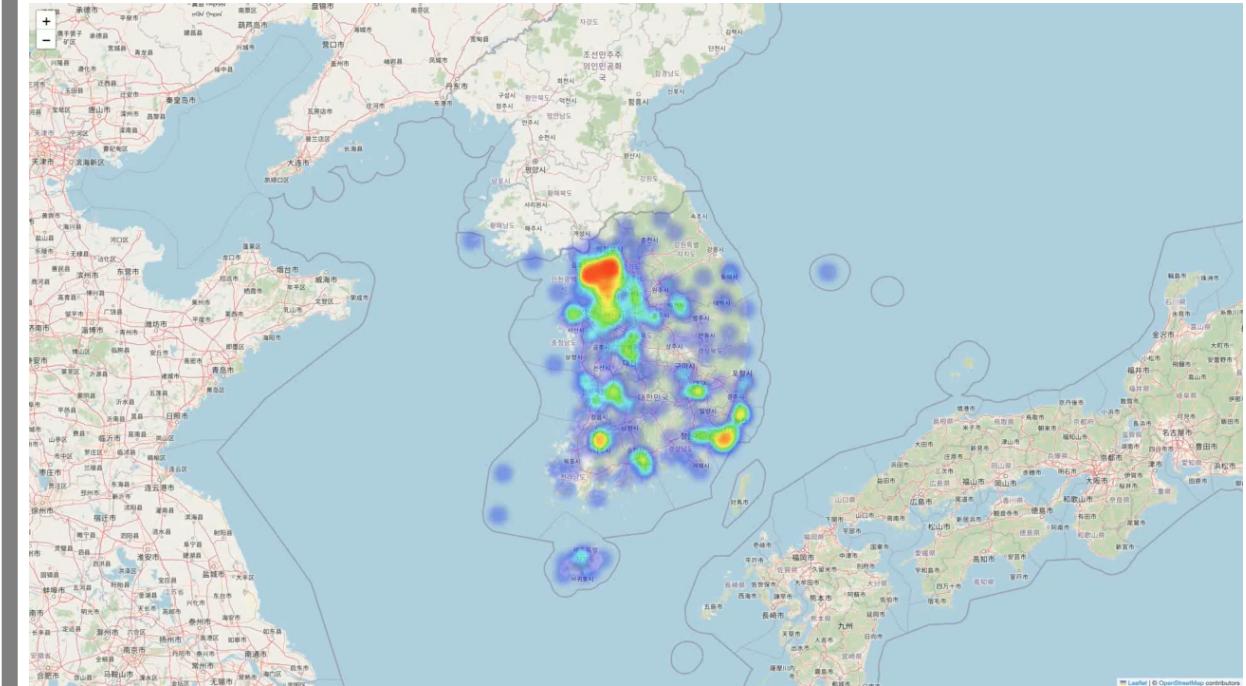


예측한 데이터 시각화 후 비교

실제 미세먼지 데이터($\text{pm}2.5$)



예측한 미세먼지 데이터($\text{pm}2.5$)





THANK YOU