

Air pollution

대기오염

드랍 더 비트

learn more

D R O P T H E B I T

드랍 더 비트

DROP THE BIT



윤상하



김동규

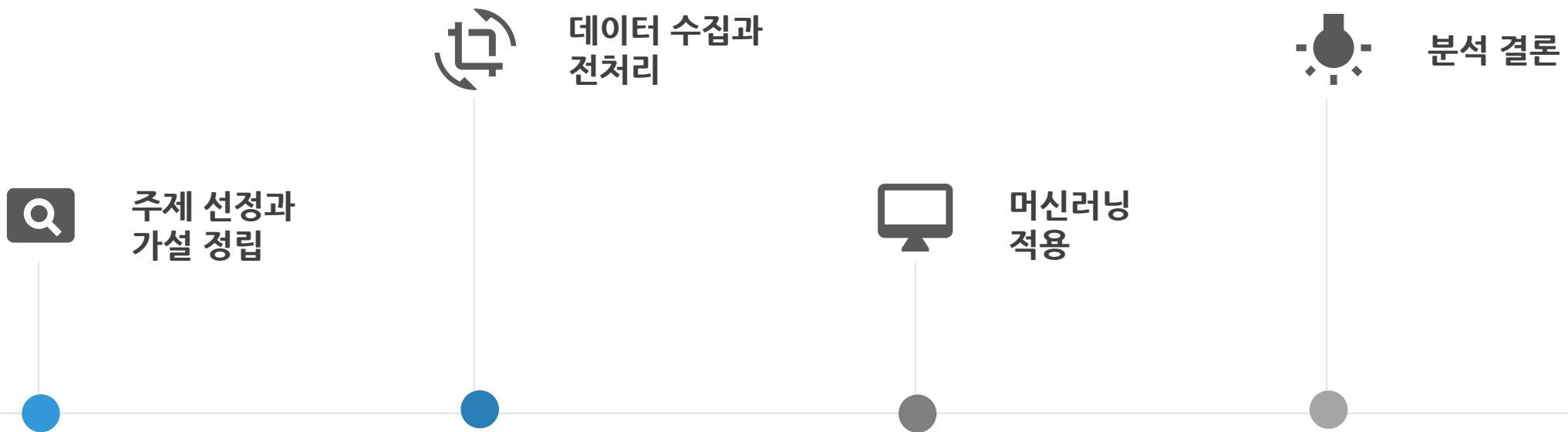


박성훈



김선우

목차



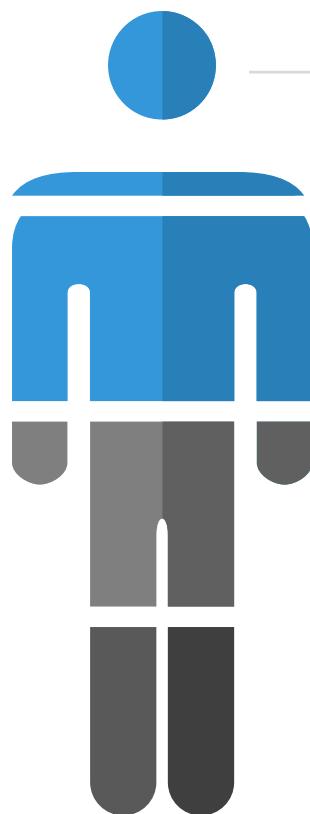
주제 선정



대기 중의 미세먼지, 화학물질, 그리고 다양한 오염 물질들은 우리의 건강과 더불어 지구 전체의 생태계에도 영향을 미치고 있습니다. 대기오염은 우리가 살아가는 환경에서 심각한 문제로 부상하고 있으며, 이는 우리의 삶과 미래에 큰 영향을 미칠 수 있습니다.



주제 선정 : 질병



비염

비루(콧물), 재채기, 가려움증 및 코막힘 중
한 가지 이상의 증상을 동반하는
비점막의 염증성 질환

천식

특정한 유발 원인 물질에 노출되었을 때
기관지의 염증에 의해 기관지가 심하게 좁아져
기침, 천명 (숨쉴 때 쌩쌩거리는 소리), 호흡곤란,
가슴 답답함이 반복적으로 발생하는 질환

아토피

유아기 혹은 소아기에 시작되는
만성 재발성의 염증성
피부질환으로 소양증(가려움증)과
피부건조증, 특징적인 습진을
동반

Hypothesis

대기오염 데이터 : 활용안



오염도와 질병



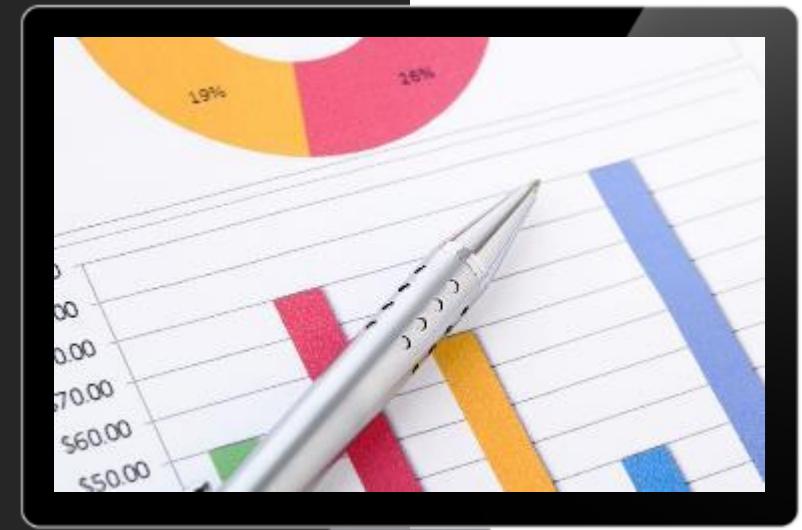
오염물질 간의
상관관계



오염 데이터를
이용한 계절
예측



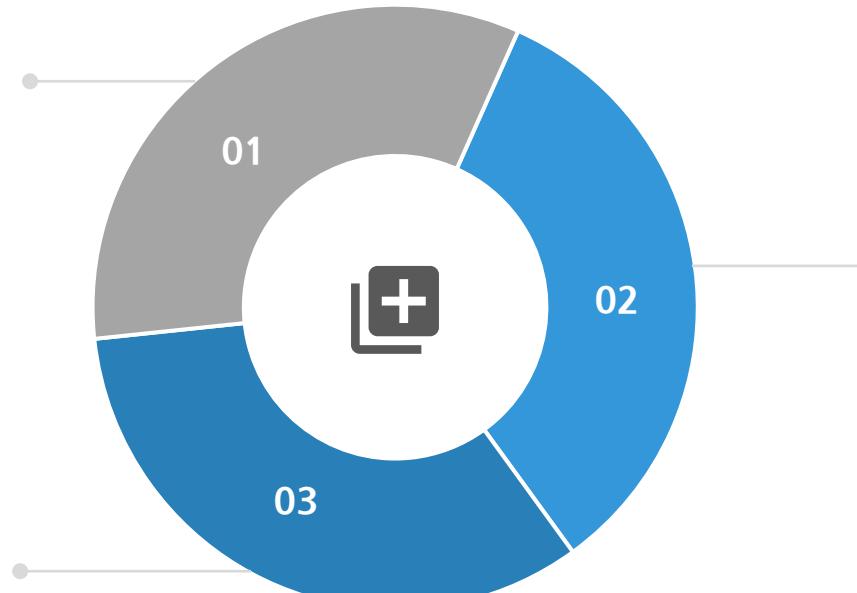
미래의 오염도
예측



수집 데이터

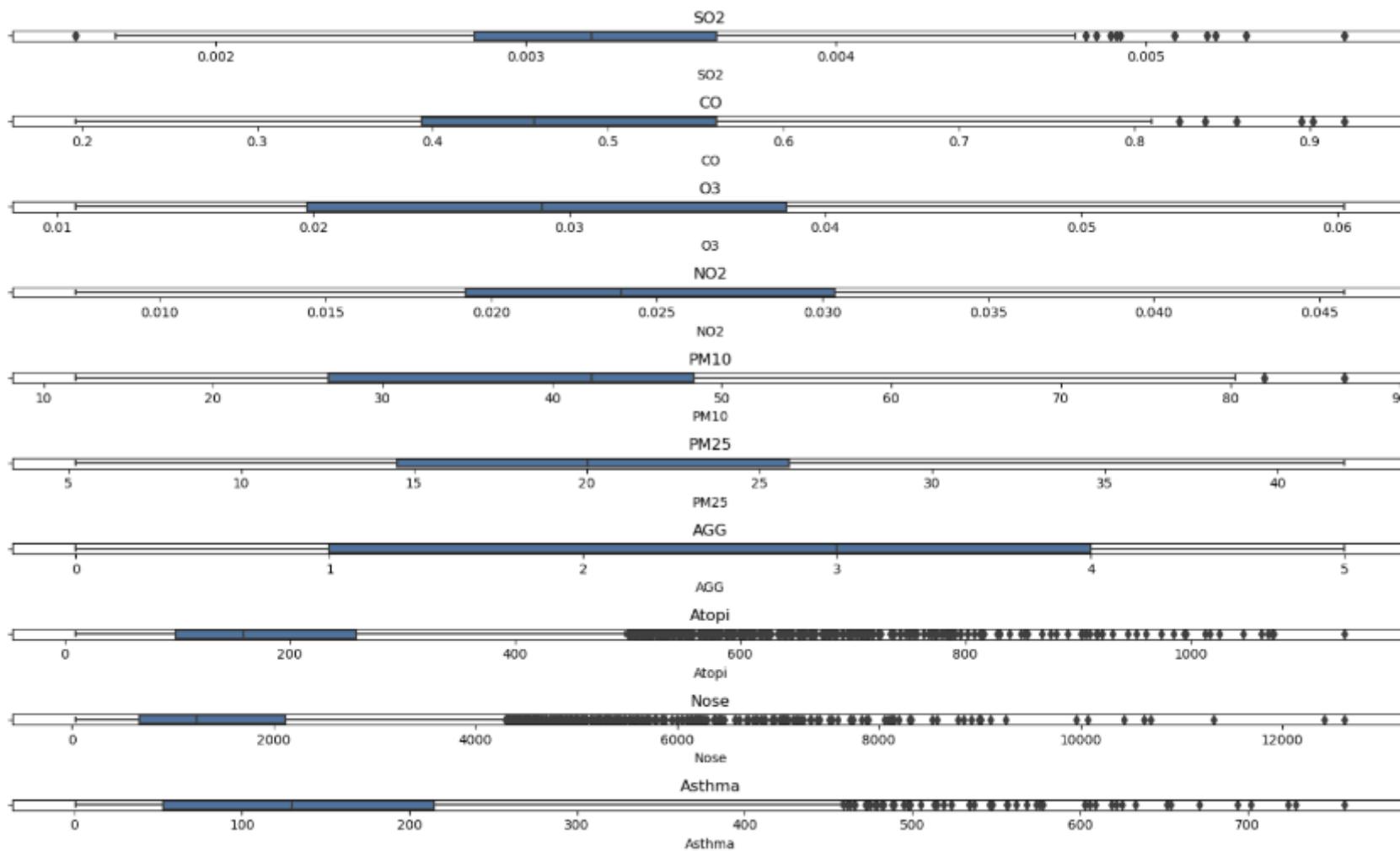
**대기오염도
데이터**
- 에어코리아

날씨 데이터
- 기상청



**질병
데이터**
- 국민건강보험공단
- 건강보험심사평가원

데이터 이상치 확인



| Month | Area | Code | SO2 | CO | O3 | NO2 | PM10 | PM25 | AGG | Atopi | Nose | Asthma |
|---------|--------|----------|----------|----------|----------|----------|----------|----------|-----|-------|------|--------|
| 2020-01 | 서울 강남구 | 111262.0 | 0.003452 | 0.920034 | 0.013274 | 0.038058 | 45.03313 | 28.79224 | 0 | 245 | 848 | 293 |
| 2020-01 | 서울 강남구 | 111262.0 | 0.003452 | 0.920034 | 0.013274 | 0.038058 | 45.03313 | 28.79224 | 1 | 204 | 1422 | 281 |
| 2020-01 | 서울 강남구 | 111262.0 | 0.003452 | 0.920034 | 0.013274 | 0.038058 | 45.03313 | 28.79224 | 2 | 163 | 1463 | 95 |
| 2020-01 | 서울 강남구 | 111262.0 | 0.003452 | 0.920034 | 0.013274 | 0.038058 | 45.03313 | 28.79224 | 3 | 594 | 9008 | 519 |
| 2020-01 | 서울 강남구 | 111262.0 | 0.003452 | 0.920034 | 0.013274 | 0.038058 | 45.03313 | 28.79224 | 4 | 284 | 6435 | 428 |

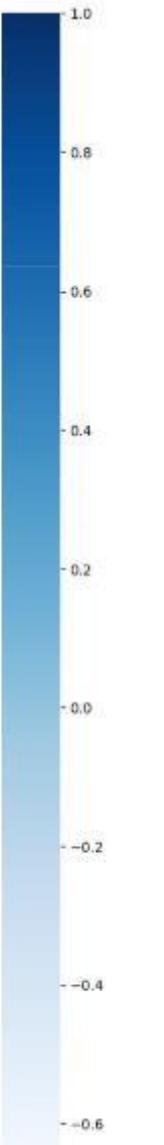
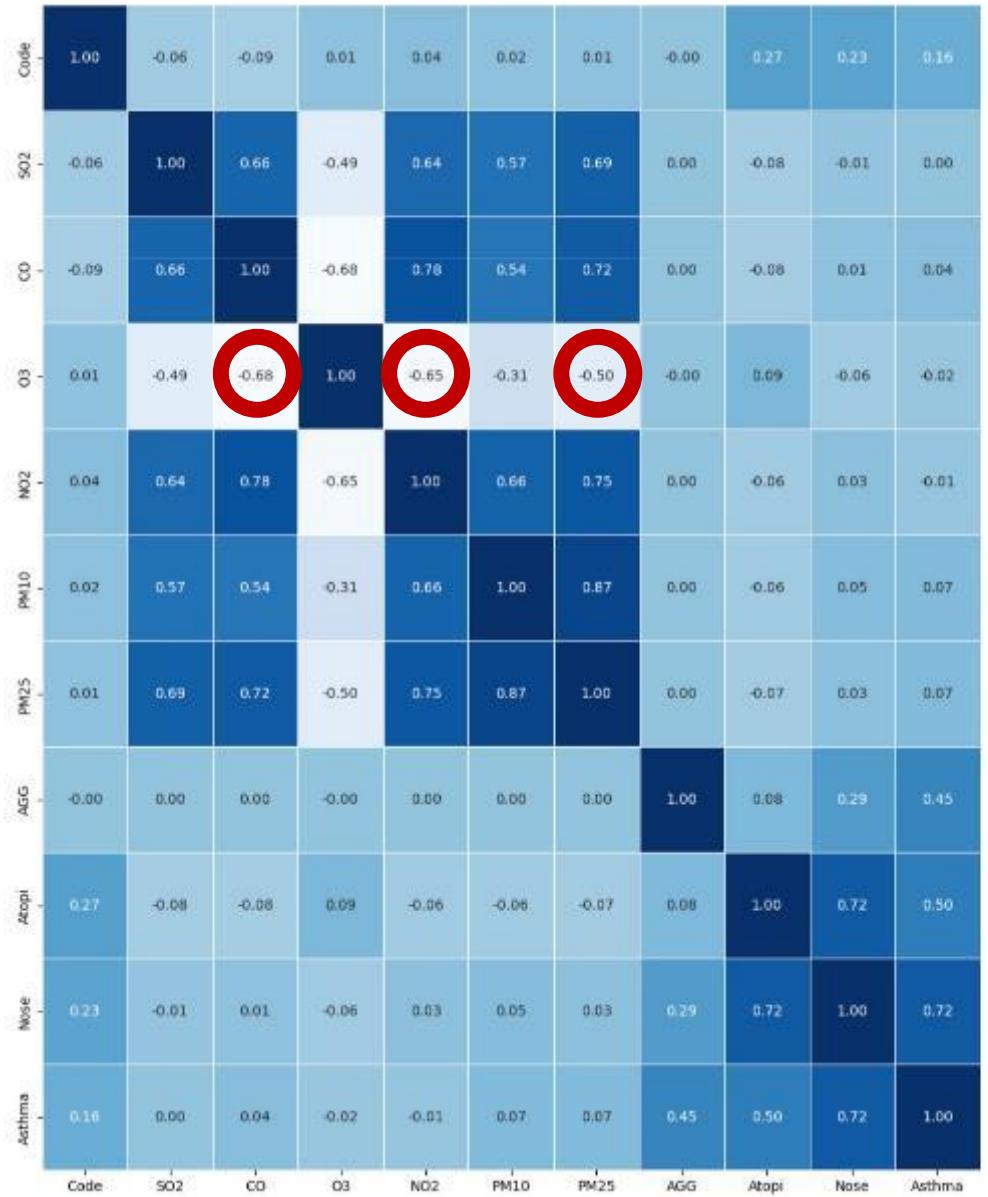
데이터 전처리

데이터 제거 및 수정

월별 취합

질병 데이터와 취합





데이터 히트맵



오존 / 일산화탄소



오존 / 초미세먼지



오존 / 이산화질소

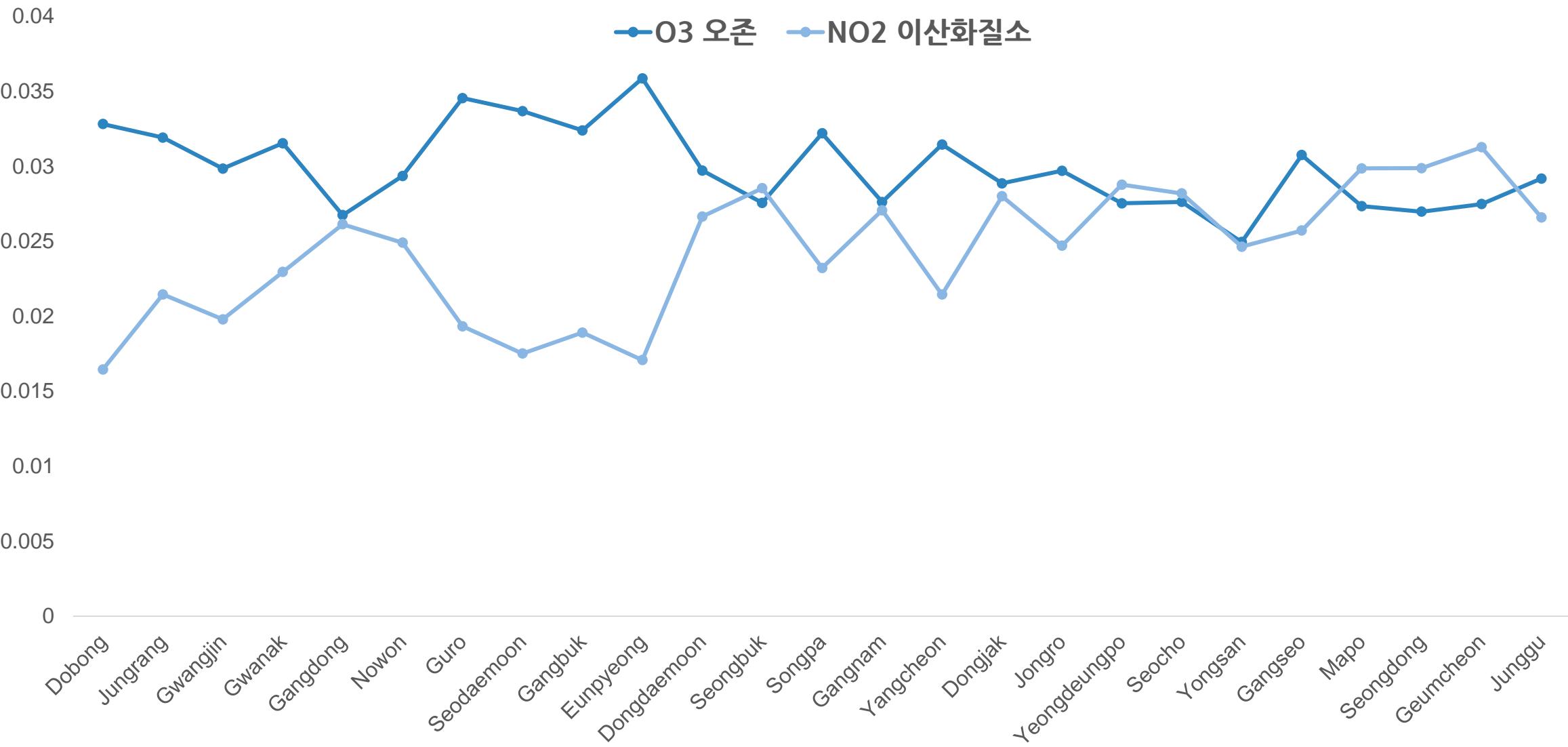
오존과 일산화탄소



오존과 초미세먼지



오존과 이산화질소



| Area | O3 오존 | RANK | NO2 이산화질소 | RANK |
|--------------|-------------|------|-------------|------|
| Geumcheon | 0.027480101 | 21 | 0.0312703 | 1 |
| Seongdong | 0.026973322 | 23 | 0.029874502 | 2 |
| Mapo | 0.027342212 | 22 | 0.02986125 | 3 |
| Yeongdeungpo | 0.027529382 | 20 | 0.028776024 | 4 |
| Seongbuk | 0.027557646 | 19 | 0.028535018 | 5 |
| Seocho | 0.027626949 | 17 | 0.028191428 | 6 |
| Dongjak | 0.028858333 | 16 | 0.027996329 | 7 |
| Gangnam | 0.027618503 | 18 | 0.027068459 | 8 |
| Dongdaemoon | 0.029715358 | 12 | 0.026649019 | 9 |
| Junggu | 0.029180898 | 15 | 0.026589707 | 10 |
| Gangdong | 0.026743324 | 24 | 0.026134439 | 11 |
| Gangseo | 0.030745323 | 10 | 0.025711173 | 12 |
| Nowon | 0.029351067 | 14 | 0.024905178 | 13 |
| Jongro | 0.029698397 | 13 | 0.024698874 | 14 |
| Yongsan | 0.024949601 | 25 | 0.024639257 | 15 |
| Songpa | 0.032195804 | 6 | 0.023216831 | 16 |
| Gwanak | 0.031534024 | 8 | 0.022949144 | 17 |
| Jungrang | 0.031915785 | 7 | 0.021452174 | 18 |
| Yangcheon | 0.031451515 | 9 | 0.021447684 | 19 |
| Gwangjin | 0.029845453 | 11 | 0.019791074 | 20 |
| Guro | 0.03454227 | 2 | 0.019328547 | 21 |
| Gangbuk | 0.03239077 | 5 | 0.01890932 | 22 |
| Seodaemoon | 0.033678615 | 3 | 0.017514847 | 23 |
| Eunpyeong | 0.035860619 | 1 | 0.01708297 | 24 |
| Dobong | 0.032816583 | 4 | 0.01645232 | 25 |

오존과 이산화질소



대체적으로 이산화질소의 농도가 높은 곳은 오존의 농도가 낮고 이산화질소의 농도가 낮은 지역은 오존의 농도가 높다



이산화질소는 대기 중의 휘발성 유기화합물(VOCs)과 반응하여 오존을 생성하는 전구물질 역할을 한다.

머신러닝 적용

learn more

머신러닝 모델



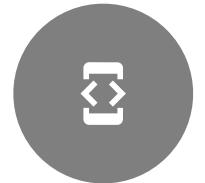
미세먼지 기반 초미세먼지 예측



사용모델

라쏘

Lasso



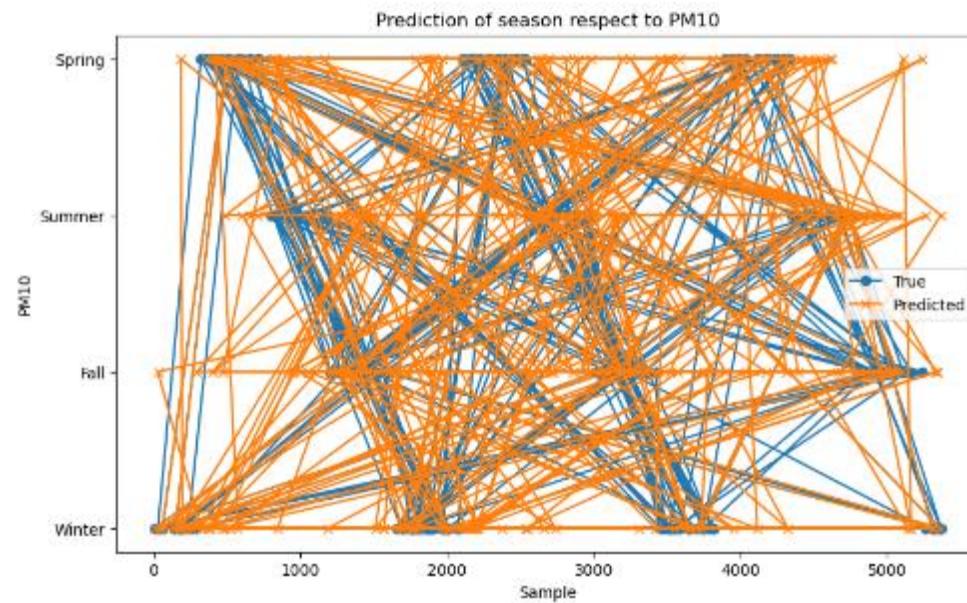
Train_score

0.0



Test_score

-0.0005890658879792452



미세먼지 기반 초미세먼지 예측



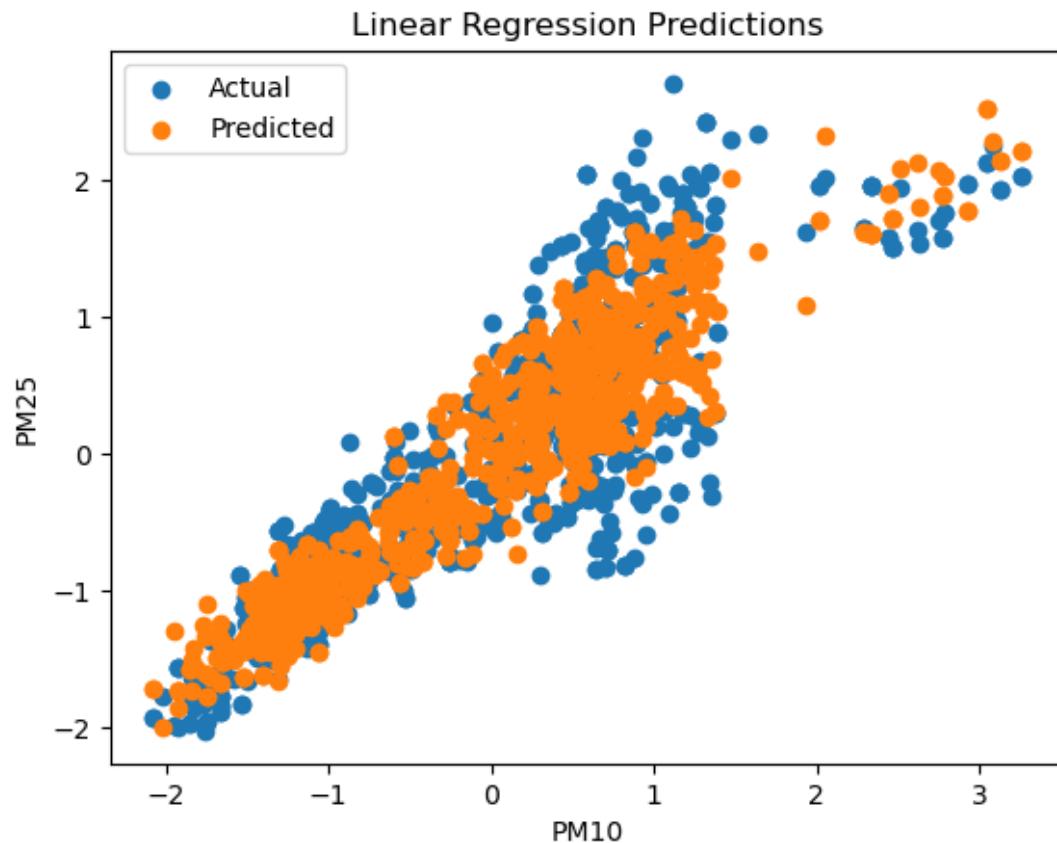
사용모델
선형회귀
LinearRegression



Train_score
0.8558119129636244



Test_score
0.8538708714645833



미세먼지 기반 초미세먼지 예측



사용모델

릿지

Ridge



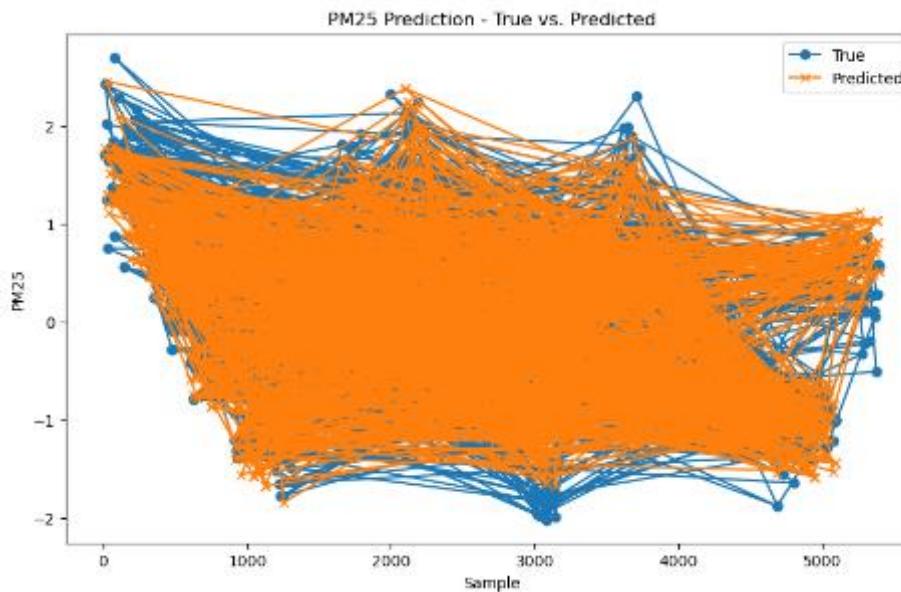
Train_score

0.8704021840152497



Test_score

0.8685239720392697

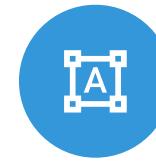
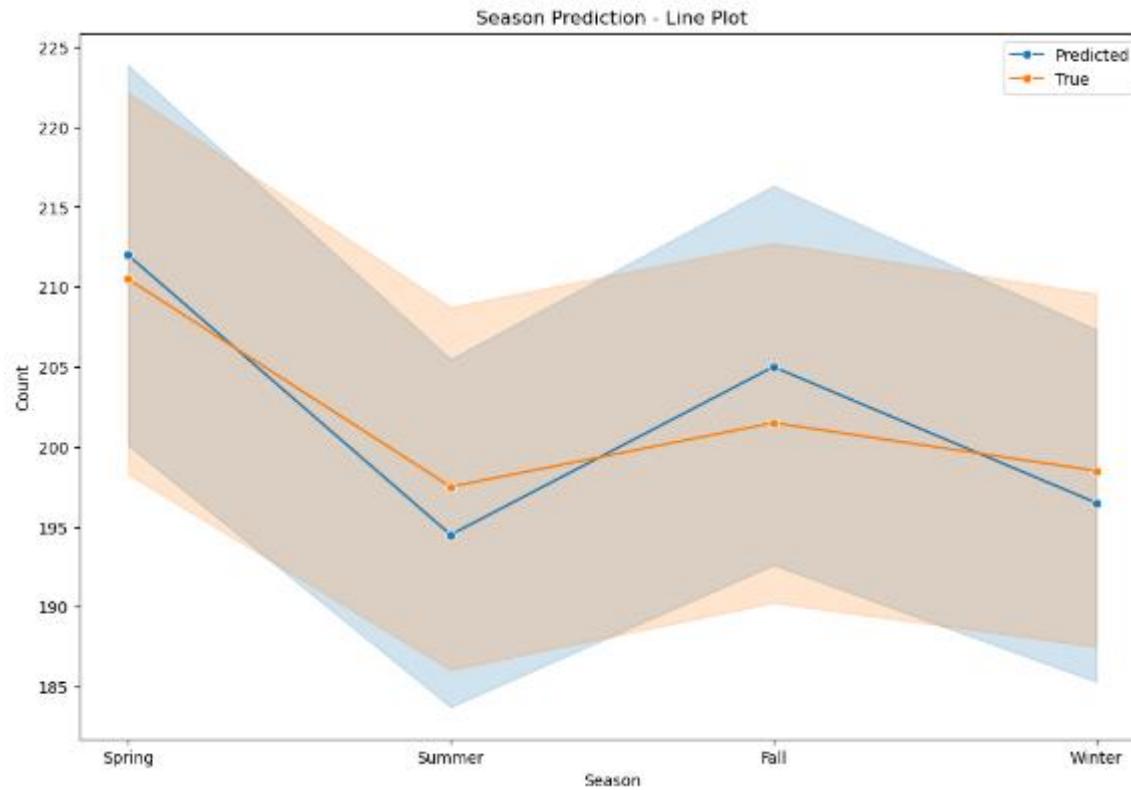


| Month | Area | SO2 | CO | O3 | NO2 | PM10 | PM25 | AGG | Atopi | Nose | Asthma | Season |
|---------|---------|----------|----------|----------|----------|----------|----------|-----|-------|------|--------|--------|
| 2020-01 | Gangnam | 0.003452 | 0.920034 | 0.013274 | 0.038058 | 45.03313 | 28.79224 | 0 | 245 | 848 | 293 | Winter |
| 2020-01 | Gangnam | 0.003452 | 0.920034 | 0.013274 | 0.038058 | 45.03313 | 28.79224 | 1 | 204 | 1422 | 281 | Winter |
| 2020-01 | Gangnam | 0.003452 | 0.920034 | 0.013274 | 0.038058 | 45.03313 | 28.79224 | 2 | 163 | 1463 | 95 | Winter |
| 2020-01 | Gangnam | 0.003452 | 0.920034 | 0.013274 | 0.038058 | 45.03313 | 28.79224 | 3 | 594 | 9008 | 519 | Winter |
| 2020-01 | Gangnam | 0.003452 | 0.920034 | 0.013274 | 0.038058 | 45.03313 | 28.79224 | 4 | 284 | 6435 | 428 | Winter |

계절 칼럼 추가



대기 오염도를 이용한 계절예측



로지스틱

Train Score : 0.7575436739015352
Test Score : 0.7358024691358025



경사하강법

Train Score : 0.6296982530439386
Test Score : 0.6154320987654321



결정트리

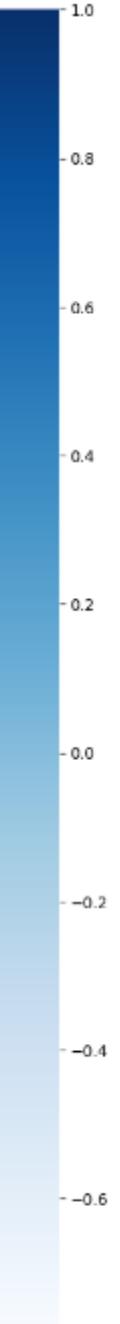
Train Score : 0.9915299100052938
Test Score : 0.9864197530864197 1

| Month | Area | SO2 | CO | O3 | NO2 | PM10 | PM25 | AGG | Atopi | Nose | Asthma | Season | AVG (°C) | MAX (°C) | MIN (°C) |
|---------|---------|---------|---------|---------|---------|---------|---------|-----|-------|------|--------|--------|----------|----------|----------|
| 2020-01 | Gangnam | 0.00345 | 0.92003 | 0.01327 | 0.03805 | 45.0331 | 28.7922 | 0 | 245 | 848 | 293 | Winter | 1.6 | 11.9 | -6.5 |
| 2020-01 | Gangnam | 0.00345 | 0.92003 | 0.01327 | 0.03805 | 45.0331 | 28.7922 | 1 | 204 | 1422 | 281 | Winter | 1.6 | 11.9 | -6.5 |
| 2020-01 | Gangnam | 0.00345 | 0.92003 | 0.01327 | 0.03805 | 45.0331 | 28.7922 | 2 | 163 | 1463 | 95 | Winter | 1.6 | 11.9 | -6.5 |
| 2020-01 | Gangnam | 0.00345 | 0.92003 | 0.01327 | 0.03805 | 45.0331 | 28.7922 | 3 | 594 | 9008 | 519 | Winter | 1.6 | 11.9 | -6.5 |
| 2020-01 | Gangnam | 0.00345 | 0.92003 | 0.01327 | 0.03805 | 45.0331 | 28.7922 | 4 | 284 | 6435 | 428 | Winter | 1.6 | 11.9 | -6.5 |

기온 칼럼 추가



| | SO2 | CO | OS | NO2 | PM10 | PM25 | AGG | Atopi | Nose | Asthma | Avg (°C) | Max (°C) | Min (°C) |
|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------|----------|----------|----------|
| SO2 | 1.00 | 0.66 | -0.49 | 0.64 | 0.57 | 0.69 | 0.00 | -0.08 | -0.01 | 0.00 | -0.67 | -0.67 | -0.65 |
| CO | 0.66 | 1.00 | -0.68 | 0.78 | 0.54 | 0.72 | 0.00 | -0.08 | 0.01 | 0.04 | -0.73 | -0.76 | -0.72 |
| OS | -0.49 | -0.68 | 1.00 | -0.65 | -0.31 | -0.50 | -0.00 | 0.09 | -0.06 | -0.02 | 0.69 | 0.74 | 0.68 |
| NO2 | 0.64 | 0.78 | -0.65 | 1.00 | 0.66 | 0.75 | 0.00 | -0.06 | 0.03 | -0.01 | -0.69 | -0.68 | -0.68 |
| PM10 | 0.57 | 0.54 | -0.31 | 0.66 | 1.00 | 0.87 | 0.00 | -0.06 | 0.05 | 0.07 | -0.69 | -0.61 | -0.69 |
| PM25 | 0.69 | 0.72 | -0.50 | 0.75 | 0.87 | 1.00 | 0.00 | -0.07 | 0.03 | 0.07 | -0.77 | -0.75 | -0.75 |
| AGG | 0.00 | 0.00 | -0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.08 | 0.29 | 0.45 | -0.00 | -0.00 | -0.00 |
| Atopi | -0.08 | -0.08 | 0.09 | -0.06 | -0.06 | -0.07 | 0.08 | 1.00 | 0.72 | 0.50 | 0.08 | 0.08 | 0.08 |
| Nose | -0.01 | 0.01 | -0.06 | 0.03 | 0.05 | 0.03 | 0.29 | 0.72 | 1.00 | 0.72 | -0.09 | -0.10 | -0.09 |
| Asthma | 0.00 | 0.04 | -0.02 | -0.01 | 0.07 | 0.07 | 0.45 | 0.50 | 0.72 | 1.00 | -0.10 | -0.10 | -0.09 |
| Avg (°C) | -0.67 | -0.73 | 0.69 | -0.69 | -0.69 | -0.77 | -0.00 | 0.08 | -0.09 | -0.10 | 1.00 | 0.98 | 0.98 |
| Max (°C) | -0.67 | -0.76 | 0.74 | -0.68 | -0.61 | -0.75 | -0.00 | 0.08 | -0.10 | -0.10 | 0.98 | 1.00 | 0.94 |
| Min (°C) | -0.65 | -0.72 | 0.68 | -0.68 | -0.69 | -0.75 | -0.00 | 0.08 | -0.09 | -0.09 | 0.98 | 0.94 | 1.00 |



기온과 미세먼지로 초미세먼지 예측

Graphic Design



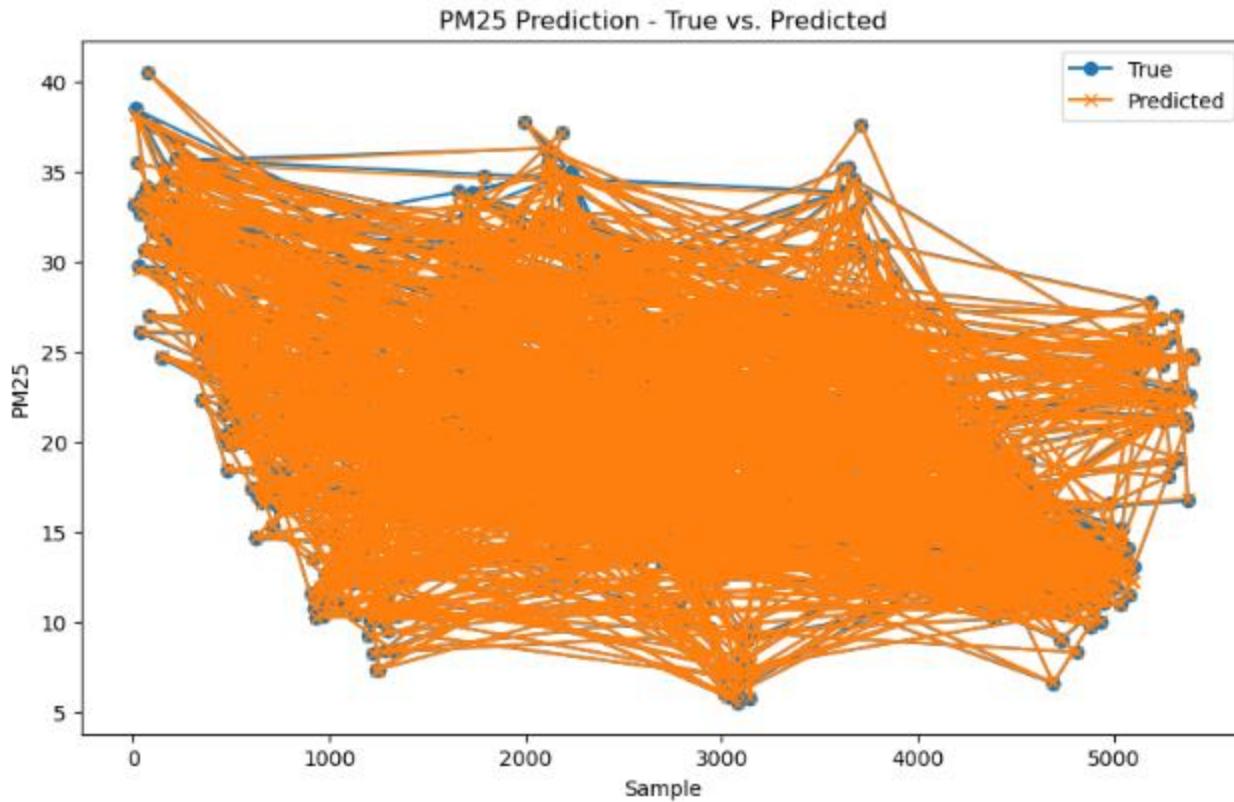
Make a type specimen book unknown printer took a galley of type and scrambled it to make a type specimen book.

Web Design



Make a type specimen book unknown printer took a galley of type and scrambled it to make a type specimen book.

기온과 미세먼지로 초미세먼지 예측



랜덤 포레스트
Random Forest



Train Score
0.9997415733612642



Test Score
0.998506326067875



기온과 미세먼지로 초미세먼지 예측



미세먼지 수치
RM10 : 25



평균기온
AVG (°C) : 4



예측 초미세먼지
PM25 : 23.33751772

대기오염도별 유병률 확인



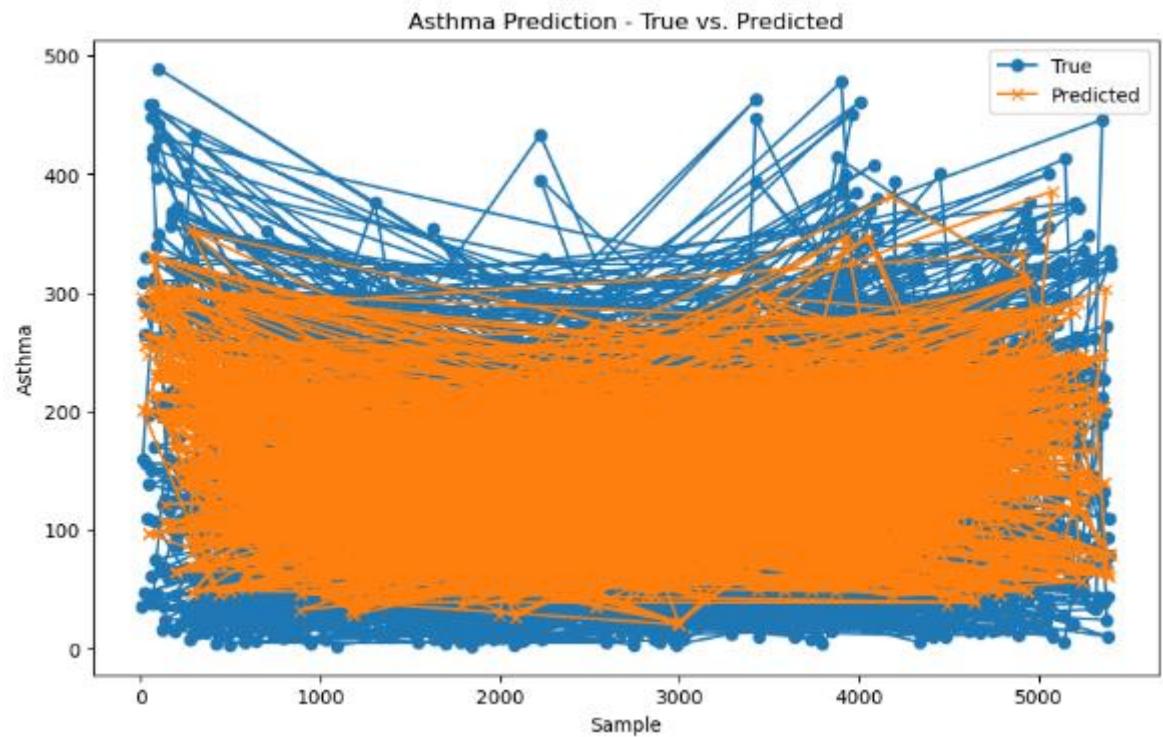
사용모델
랜덤 포레스트
Random Forest



Train Score
0.449271677327567

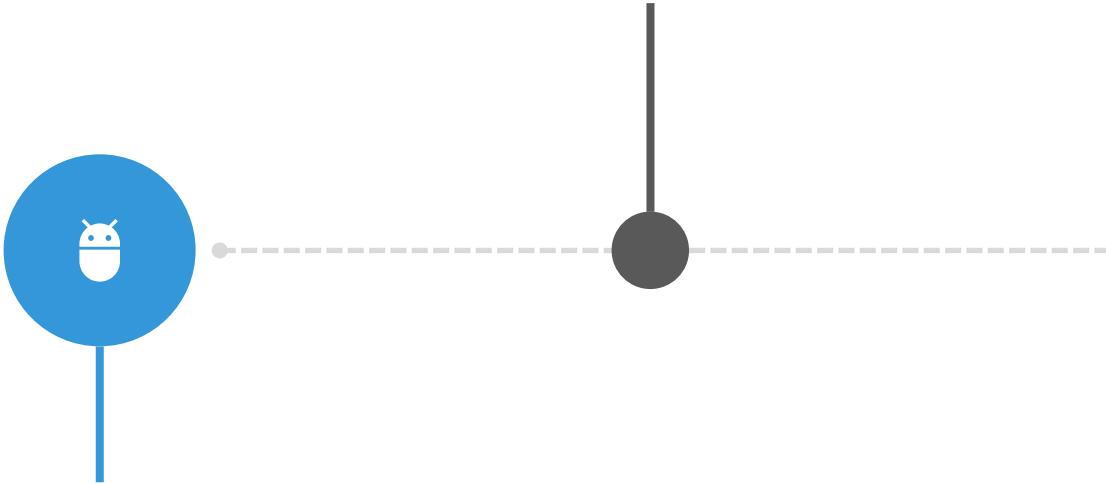


Test Score
-0.017135254258737



프로젝트 결론

서울내의 구별 평균적인
대기오염도는 지역의 영향보다는
계절의 영향이 큰것을 확인 할 수 있다.



기존 가설 중 하나였던 대기오염
물질간의 상관관계는 오염데이터로만
분석하기는 어렵다

기온에 따라서 미세먼지의 예측 스코어의 차이가 판이하므로 계절마다 미세먼지 변동은 확실하게 확인할 수 이었다.



대기오염 데이터로만은
아토피 / 천식 / 비염
유병율과 연관짓기는 어렵다

느낀점

A black and white photograph of a man with glasses and a beard, wearing a light blue shirt, sitting at a wooden desk. He is looking at a computer screen showing a bar chart. A laptop on the desk also displays a chart. The background features a brick wall and a window with a lamp.

감사합니다 : >