
딥러닝(TensorFlow)을 활용한 미세먼지 예측

경북대학교 배홍직, 이상목, 이현준

미세먼지 정의



미세먼지는 지름 10μm 이하의 **매우 작은 크기를 가진 입자**를 뜻한다.
그 크기에 따라 10μm 이하는 미세먼지, 2.5μm 이하는 초미세먼지로 나뉜다.

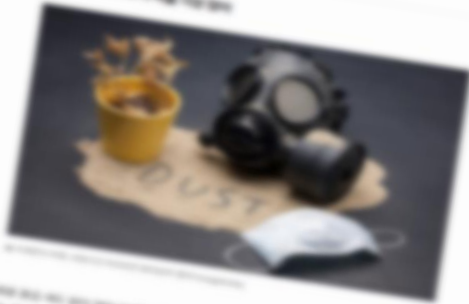
미세먼지 발생원인



중금속과 같은 화학물질 다량 포함

미세먼지 위험성

미세먼지 위험성...경제적 비용은 얼마?



미세먼지를 '1급 발암물질'로 분류

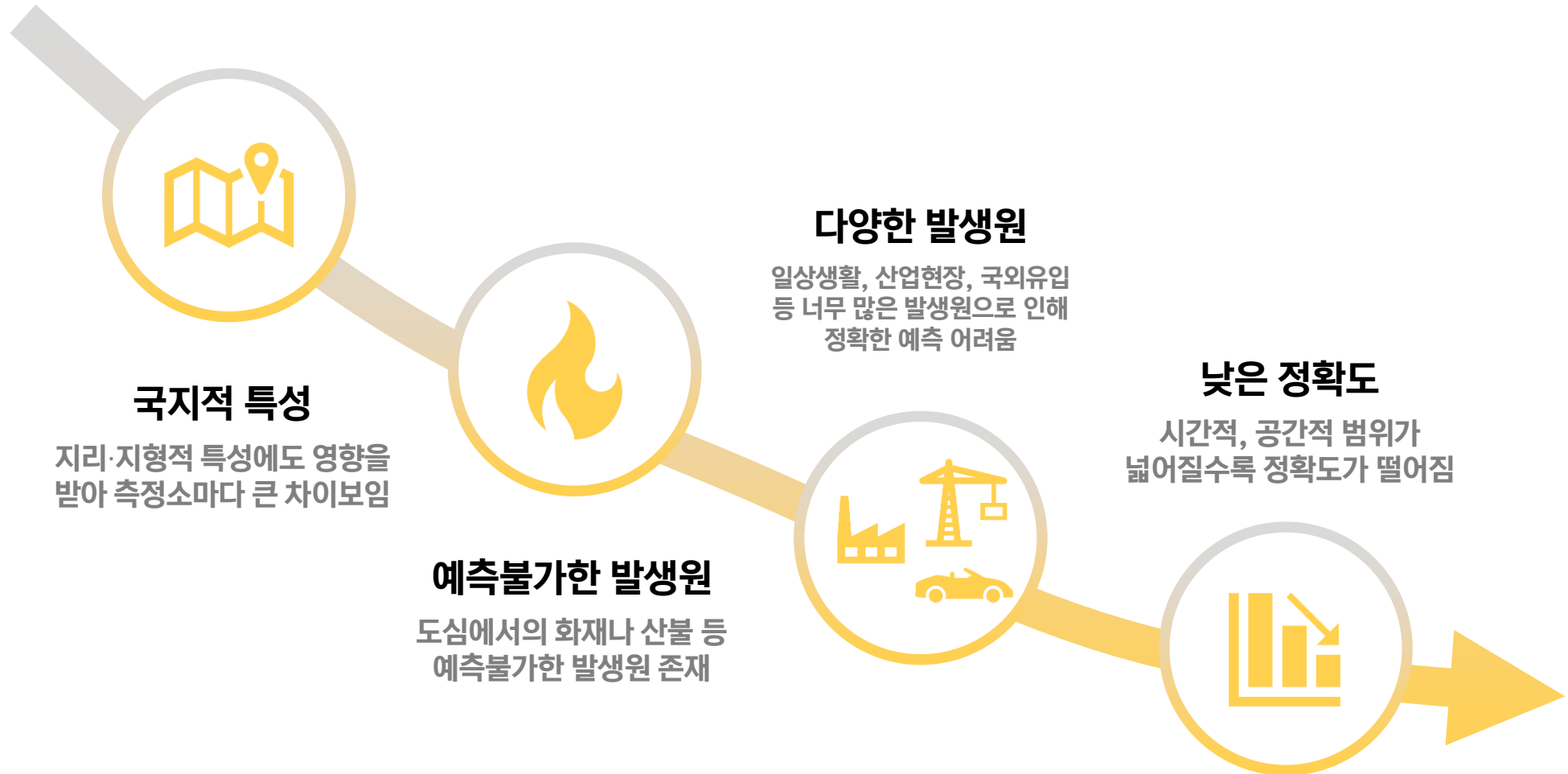
호흡기 질환뿐만 아니라 뇌 질환, 심혈관계 질환, 대사성질환 위험을 높이고 .
루게릭 병까지 심각한 영향을 미치는 것으로 나타났다.

"초미세먼지가 코로나 감염 높인다"

미세먼지가 뼈건강에도 영향...골다공증 위험성 첫
확인

"실내서도 극성인 발암물질"...미세먼지

미세먼지 예측 한계



미세먼지 예측 설정 범위 및 방법



경북대학교

경북대학교 근처 신암동
대기측정소 데이터 활용

공간적 범위



한 시간 후

현재 미세먼지 농도를
기반으로 한 시간 후 농도 예측

시간적 범위



농도 예측

파이썬 텐서플로우를
활용한 DNN 모델 사용

예측 방법

수집 데이터



참고문헌: 대구지역 PM10과 PM2.5 농도 특성 및 기여인자 분석 연구

수집 데이터 출처



기상청

Korea Meteorological
Administration

대구 대기질 자료(2003~2017)

CO SO₂ NO₂ O₃ PM₁₀ PM_{2.5}

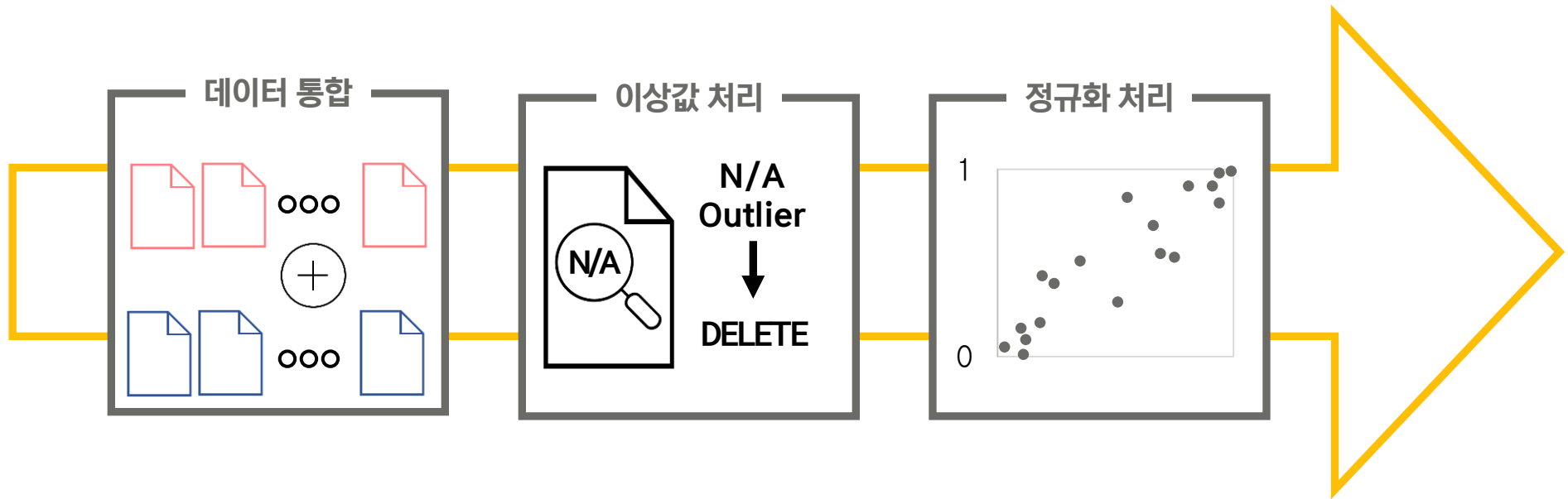
매 시간 데이터 수집

대구 기상 자료(2003~2017)

기온 풍속 습도 강수량

매 시간 데이터 수집

데이터 전처리



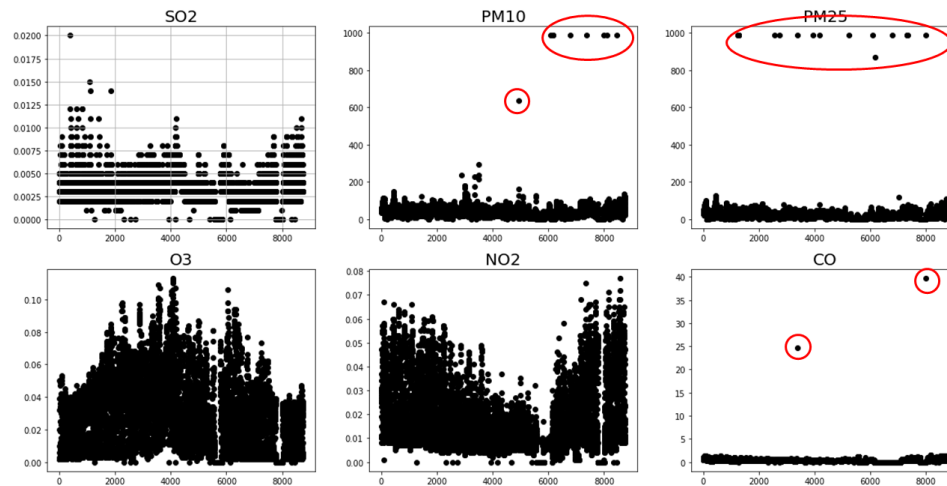
이상값 처리

측정소	년	월	일	시	SO2	PM10	PM25	O3	NO2	CO
신암동	2017	11	26	20	0	0	0	0	0	0
신암동	2017	11	26	21	0	0	0	0	0	0
신암동	2017	11	26	22	0	0	0	0	0	0
신암동	2017	11	26	23	0	0	0	0	0	0
신암동	2017	11	26	24	0	0	0	0	0	0
신암동	2017	11	27	1	0	0	0	0	0	0
신암동	2017	11	27	2	0	0	0	0	0	0
신암동	2017	11	27	3	0	0	0	0	0	0
신암동	2017	11	27	4	0	0	0	0	0	0
신암동	2017	11	27	5	0	0	0	0	0	0
신암동	2017	11	27	6	0	0	0	0	0	0
신암동	2017	11	27	7	0	0	0	0	0	0
신암동	2017	11	27	8	0	0	0	0	0	0
신암동	2017	11	27	9	0	0	0	0	0	0
신암동	2017	11	27	10	0	0	0	0	0	0
신암동	2017	11	27	11	0	0	0	0	0	0

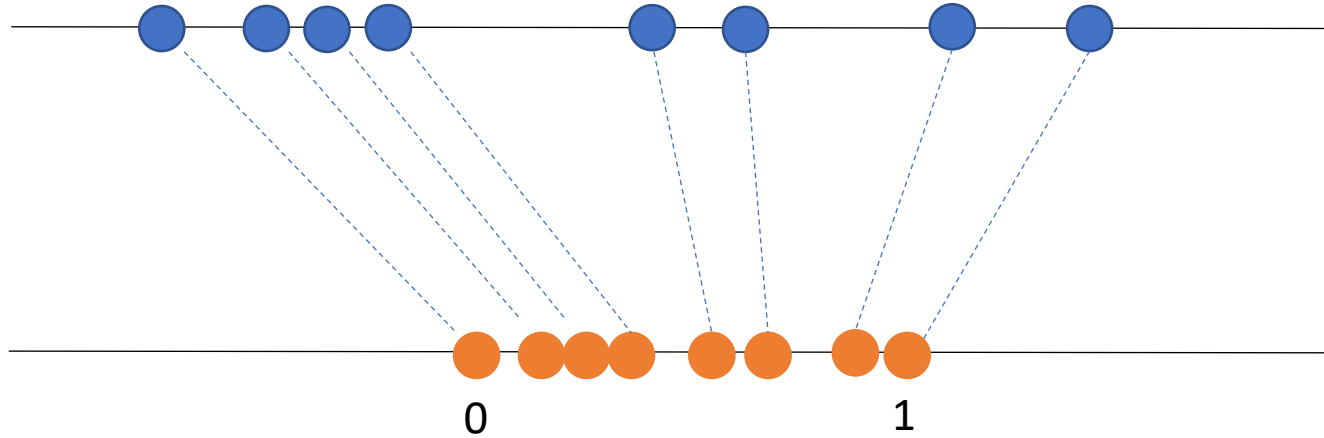
기기
미작동

기기
오작동

총 데이터 8,759 中
이상값 1,449개
DELETE



정규화 처리



$$\tilde{x} = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Min-Max Normalization

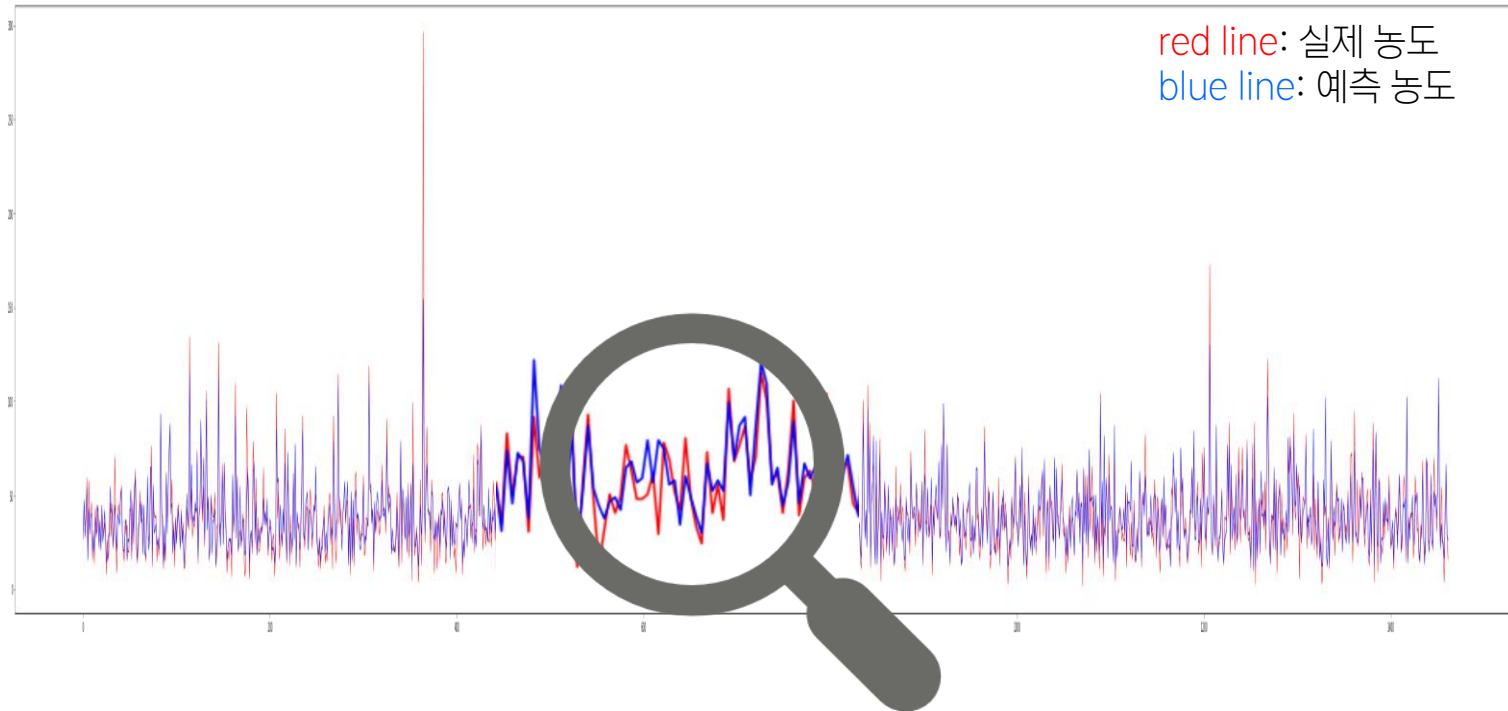
서로 다른 변수의 값 범위를 일정한 수준으로 맞추기 위해
최소값을 0, 최대값을 1로 맞추고 범위를 정규화 시키는 방법

하이퍼 파라미터 설정

- Train, Validation, Test 셋은 6:2:2의 비율로 나누었다.
- Learning rate, Epochs 등과 같은 하이퍼 파라미터들은 팀에서 임의의 값을 설정한 뒤, 결과치가 가장 좋은 하이퍼 파라미터로 결정하였다.

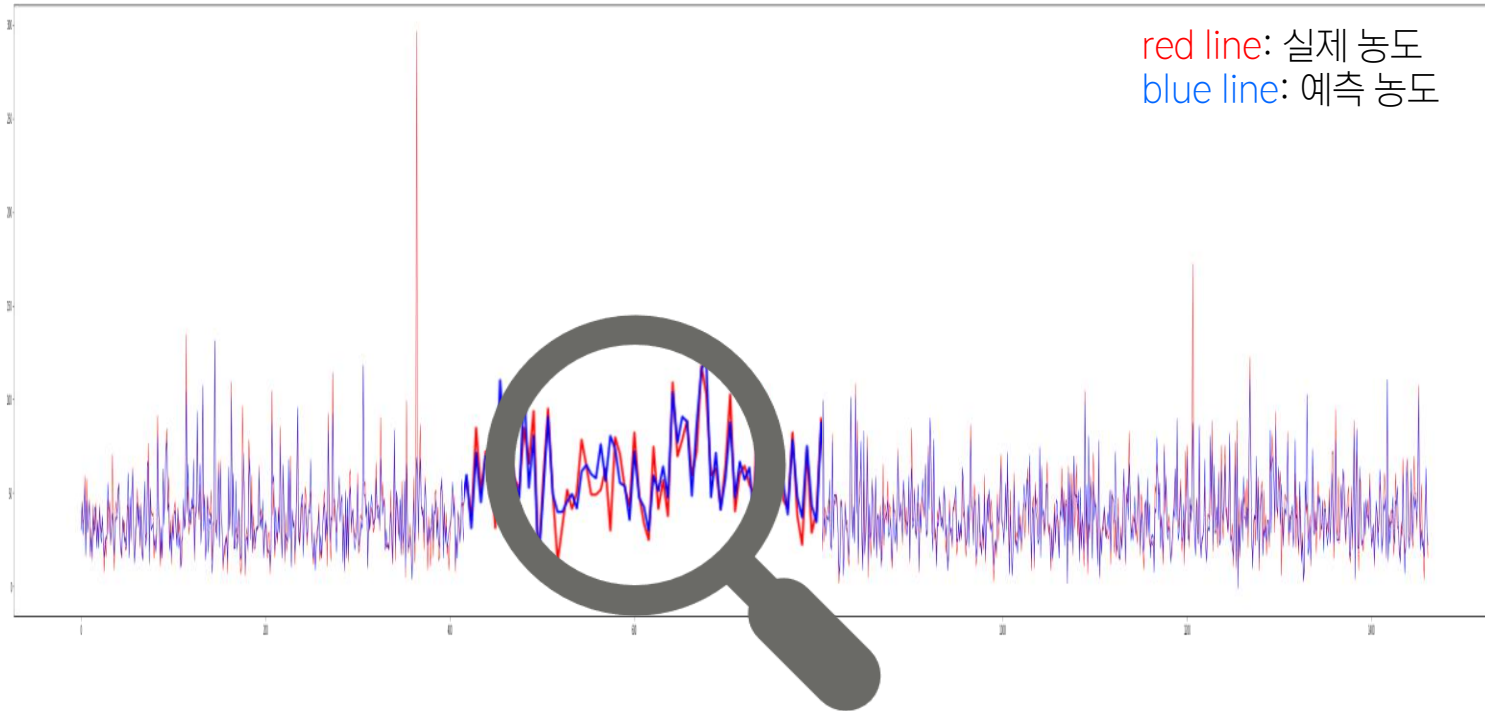
구분	값
Training Data	60%
Validation Data	20%
Testing Data	20%
Optimizer	Adam
Activation	relu
Learning rate	0.001
Epochs	100
Loss function	mse
Callback method	Early Stopping(patience=10)

PM₁₀ 정보 활용 DNN 모델



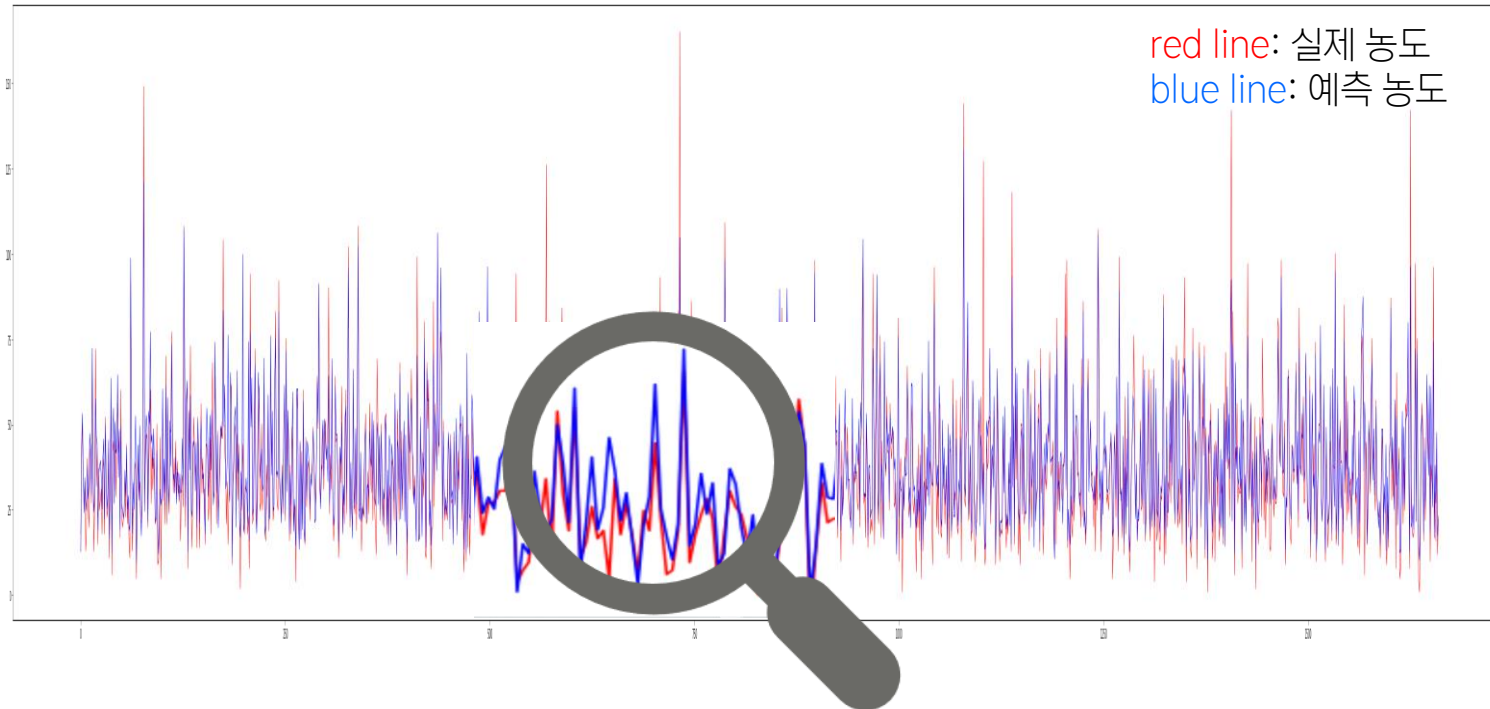
DNN 모델을 테스트 셋에 적용시킨 결과
정확도 약 76%

PM₁₀ + 대기 정보 활용 DNN 모델



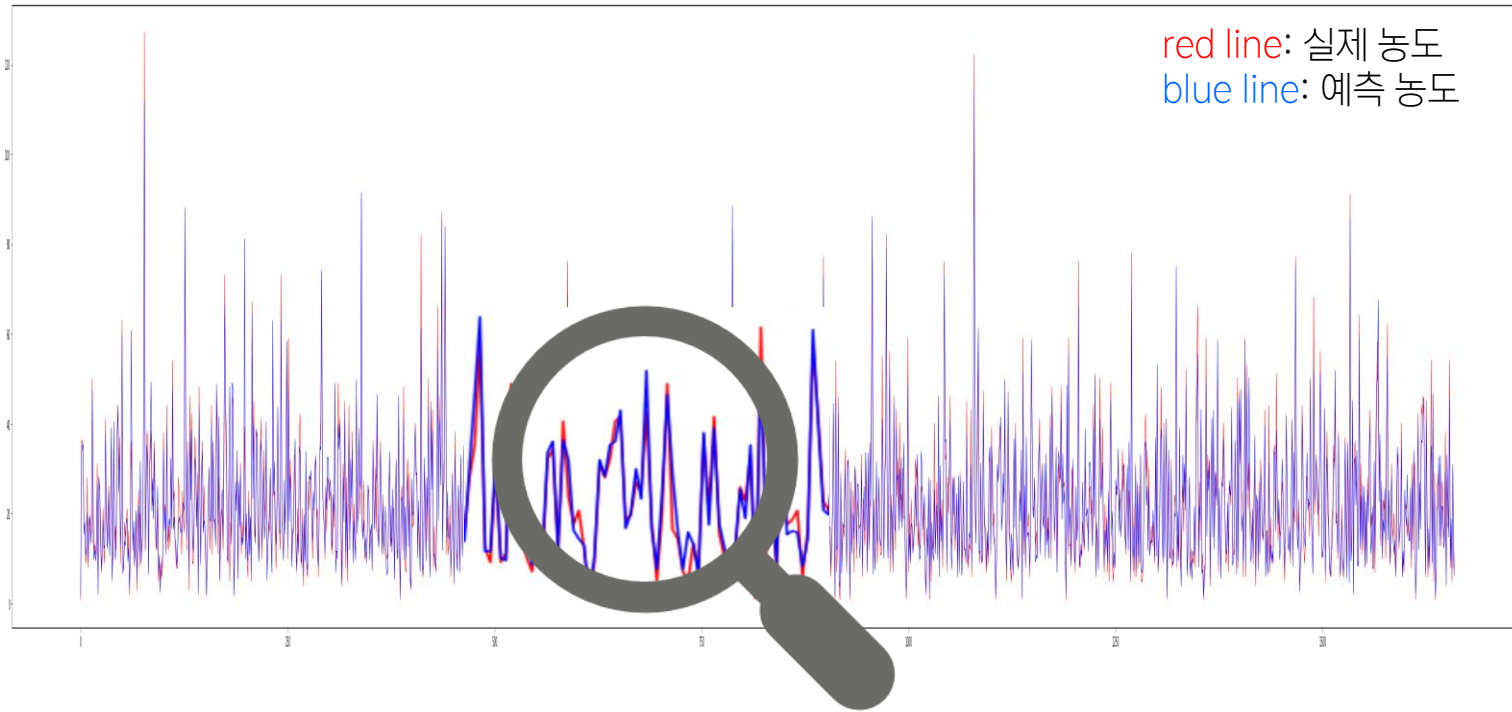
DNN 모델을 테스트 셋에 적용시킨 결과
정확도 약 74%

PM₁₀ + 대기 + 기상 정보 활용 DNN 모델



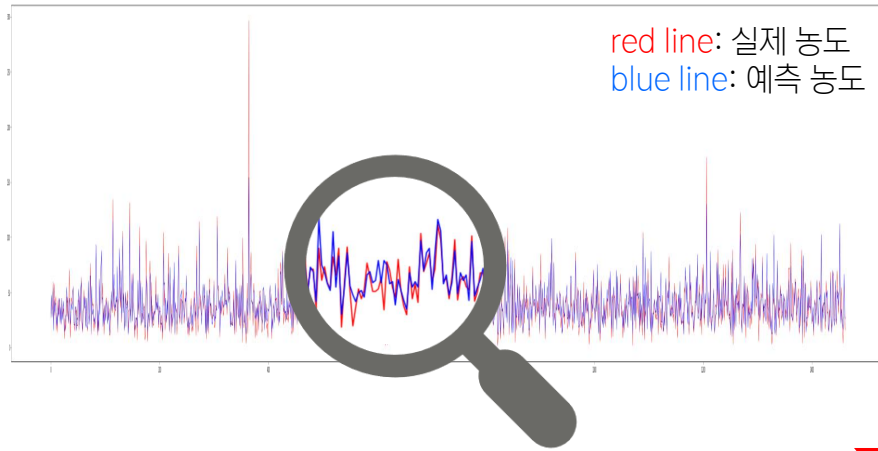
DNN 모델을 테스트 셋에 적용시킨 결과
정확도 약 78%

PM_{2.5} + 대기 + 기상 정보 활용 DNN 모델

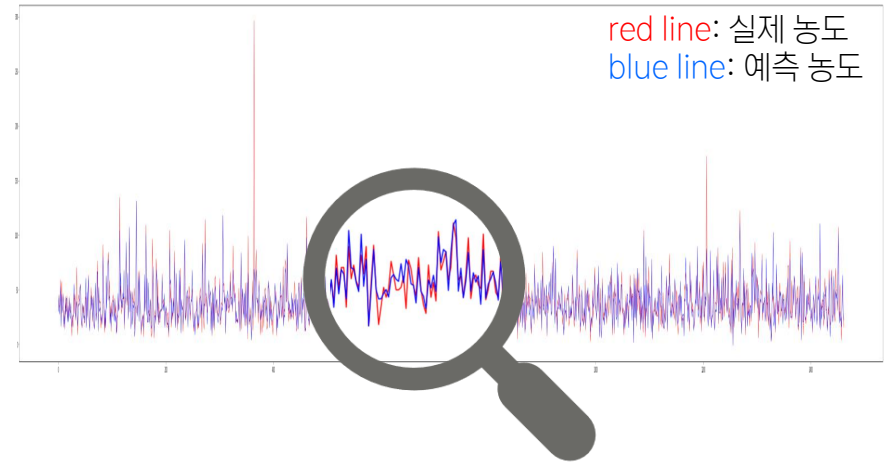


DNN 모델을 테스트 셋에 적용시킨 결과
정확도 약 89%

모델 간 성능 비교



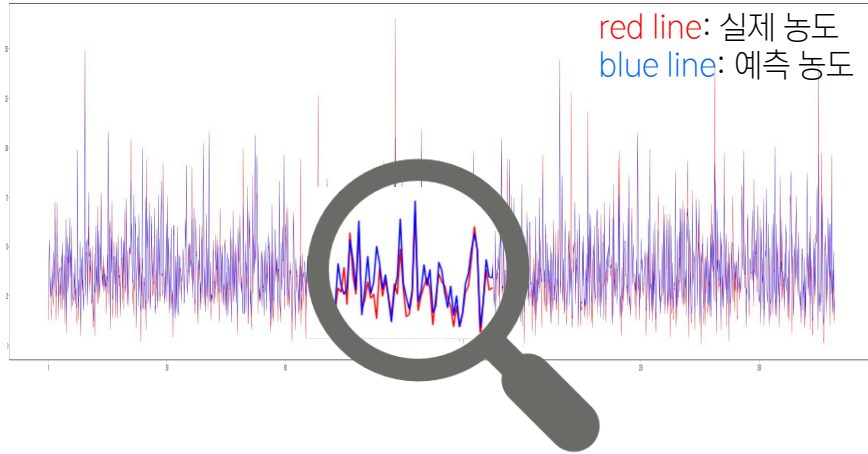
PM₁₀ 만 Feature로 사용한 모델
정확도 약 76%



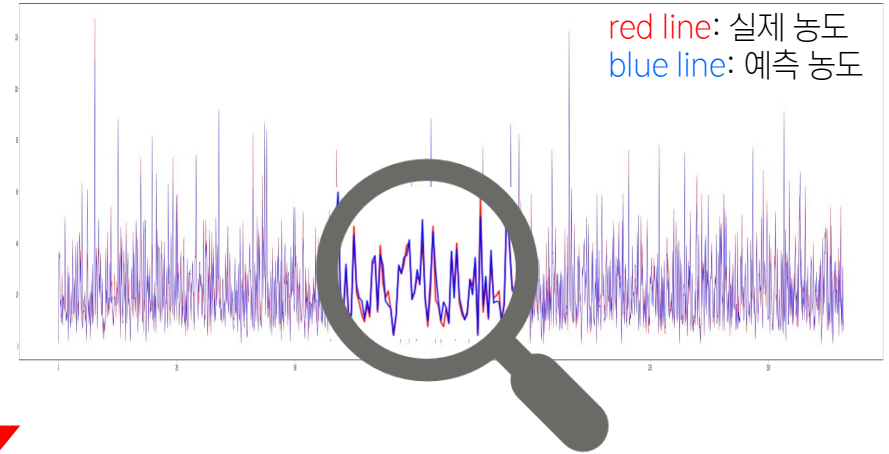
PM₁₀, 대기정보를 Feature로 사용한 모델
정확도 약 74%

- 대기정보만 feature로 추가해줄 경우, PM₁₀만을 feature로 주었을 때보다 정확도가 떨어지는 결과가 나타났음
- 대기인자들의 미세먼지 2차 발생에 필요한 화학반응이 온도, 습도와 같은 기상인자에 영향을 받는데 이가 누락됨에 따라 오히려 정확도가 떨어진 것으로 사료됨

모델 간 성능 비교



PM₁₀, 대기, 기상정보를 Feature로 사용한 모델
정확도 약 78%



PM_{2.5}, 대기, 기상정보를 Feature로 사용한 모델
정확도 약 89%

- 대기, 기상정보를 함께 추가해줄 경우, 앞선 모델들에 비해 정확도가 조금이나마 증가하는 모습을 관찰할 수 있었음
- PM_{2.5}의 경우, PM₁₀보다 더 높은 정확도를 보여주었음
- 이는 PM_{2.5} 농도 변화의 폭이 PM₁₀에 비해 작고 빈도도 적어 나타난 결과로 사료됨

결론

- 미세먼지 농도 예측 모델은 **미세먼지 정보와 대기, 기상 정보**를 함께 features로 주었을 경우, **가장 정확도가 높게** 나타났음.
- 미세먼지 농도가 저농도에서 고농도로, 고농도에서 저농도로 **농도가 급변하는 경우 오차 값이 크게** 나타났음. 이는 앞서 미세먼지 예측 한계점에서 말했던 발생원들에 의한 농도 변화로 수집 데이터 부족으로 추측됨.
- **PM_{2.5}가 PM₁₀보다 더 높은 정확도**를 보여주었음. 이는 위에서 말한 농도의 급변화 정도가 적기 때문에 나타난 결과로 추측됨.

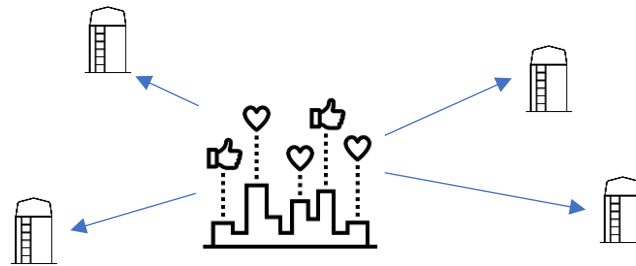
추가 개발계획

○ 시간적 범위 확장



기존의 모델에 교통량, 발전소 주변대기정보 및 풍향 등의 features를 추가로 추가하고 1시간보다 넓은 범위의 시간에서 높은 정확도를 가진 미세먼지 예측 모델을 개발

○ 공간적 범위 확장

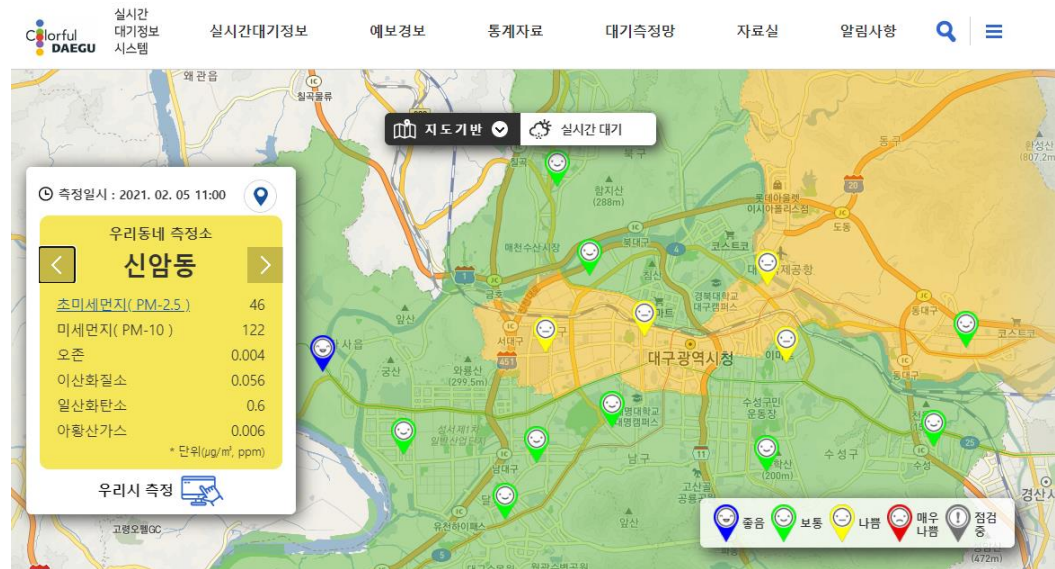


기존의 신암동 측정소뿐만 아니라 대구에 위치하는 각각의 측정소별로 모델을 개발하여 경북대학교에서 대구 전체로 공간적 범위를 확장

추가 개발계획

○ 웹·앱 어플리케이션 개발

정확도가 높은 미세먼지 예측 모델을 개발, OpenAPI 활용하여 실시간 대기정보와 함께 실시간 예측정보를 함께 제공해주는 웹·앱 어플리케이션 개발 제공



실시간 대기정보 시스템(출처: 에어코리아 대구, <http://air.daegu.go.kr>)

참고문헌

[학위 논문]

- 안지민. "대구지역 PM10과 PM2.5 농도 특성 및 기여인자 분석 연구." 국내석사학위논문 계명대학교 대학원, 2018. 대구
- 이기혁. "복합 신경망 구조를 이용한 미세먼지 위험 단계 예측 모델 설계 및 분석." 국내석사학위논문 한양대학교 대학원, 2020. 서울
- 조경우. "미세먼지 예측을 위한 딥러닝 기반 농도별 분리 예측 모델." 국내박사학위논문 한국기술교육대학교 일반대학원, 2020. 충청남도
- 채상미. "학교 미세먼지 대응 위한 RNN-LSTM 순환 신경망 미세먼지 예측." 국내석사학위논문 아주대학교 대학원, 2020. 경기도

[학술지]

- 성상하,김상진,and 류민호. "미세먼지 예측을 위한 기계학습 모델 간 성능 비교 연구:국내 발생 데이터를 중심으로." 한국혁신학회지 15.4 (2020): 339-357.
- 전영태,유숙현,and 권희용. "Outlier 데이터 제거를 통한 미세먼지 예보성능의 향상." 멀티미디어학회논문지 23.6 (2020): 747-755.
- 조경우,정용진,강철규,오창헌,Cho Kyoung-woo,Jung Yong-jin,Kang Chul-gyu,and Oh Chang-heon. "미세먼지 예측을 위한 기계 학습 알고리즘의 적합성 평가." 한국정보통신학회논문지 23.1 (2019): 20-26.

Q & A

THANK YOU