보다 나은 정부	보 도 자 료	작성과	국가정보자원관리원 빅데이터분석과
행정안전부	2019년 1월 29일(화) 조간 (1. 28. 16:00 이후)부터 보도하여 주시기 바랍니다.	담당자	과 장 하민상 사무관 박인창
		연락처	042-250-5650 042-250-5655

미세먼지, 빅데이터로 예측한다.

- 국가정보자원관리원과 UN글로벌펄스가 협업하여 미세먼지 예측 및 주요 요인 파악 -

□ 세계보건기구(WHO)에서 1급 발암물질로 분류한 미세먼지의 고농도 현상이 잦아지는 등 국민건강과 생명을 직접적으로 위협하는 미세먼지는 이제 온 국민의 관심사이자 국가적 재난의 문제로 대두되고 있다.

□ 행정안전부 국가정보자원관리원(원장 김명희)은 UN 글로벌 펄스(UN Global Pulse)* 자카르타 연구소와 업무협력(MOU)을 체결('18.4.19)하고, **동북아 지역의 미세먼지 예측 및 주요 요인을 데이터에 기반하여 분석**하였다.

* 빅데이터를 이용해 위기 및 재난으로부터 취약계층을 보호하기 위해 마련된 UN사무총장 직속 프로그램으로 현재 뉴욕, 인도네시아 자카르타, 우간다 캄팔라에 Pulse Lab을 운용 중

○ **국내외 요인을 정확히 파악**하기 위하여 **서해안의 인구 밀집지역인 인천지역을 분석대상으로 선정**하였으며,

○ 기존의 수치예측모델과 달리 **머신러닝을 활용하여** ①내일의 미세먼지 예측을 위한 **미세먼지 예측모델을 개발**하고 ②**미세먼지에 영향을 미치는 주요 요인을 파악**한 것이다.

□ 이번 분석에는 '15.1월부터 '18.3월까지의 ①인천 지역 미세먼지·대기오염 데이터(환경부, 28,464건), ②미국항공우주국(NASA)에서 제공하는 동북아 지역의 위성 센서 데이터* 및 ③에어로넷(AERONET)**의 지상 관측 센서 데이터를 활용하였으며,

* NASA Aqua 위성의 MODIS(중간해상도 영상 분광계) 센서 데이터로 미세먼지와 같이 공기 중에 떠 있는 작은 입자인 에어로졸을 관측

** NASA가 운영하는 국제 공동 에어로졸 관측 네트워크로 지상에서 관측

○ **UN 글로벌 펄스 자카르타 연구소**에서는 인도네시아 대기오염 관련 데이터 분석* 경험을 바탕으로 **기술 자문을 제공**하였다.

* Nowcasting Air Quality by Fusing Meteorological Data, Insights from Satellite Imagery and Photos Shared on Social Media Using Deep Learning (2018)

□ 먼저 미세먼지 예보에 최적의 성능을 보인 **그래디언트 부스팅* 기반의 예측 모델을 구현**하였으며,

○ 이를 통해 '18년 1분기를 예측한 결과, **미세먼지(PM_{10}) 84.4%, 초미세먼지($PM_{2.5}$) 77.8%의 정확도**를 보여 <u>**기존 국내 미세먼지 예보에 비해 정확도가 약 15% 높아진 것을 확인**</u>할 수 있었다.

* 약한 예측 모델을 결합하여 예측도를 향상시키는 기계 학습 모델

□ 주요 예측변수로는, **미세먼지의 경우 풍향, 강우량, 서해안 및 중국 산둥성 지역의 에어로졸 농도**로,

초미세먼지의 경우 풍속, 풍향 및 중국 내몽골, 베이징·허베이성 지역의 에어로졸 농도로 나타났다.

○ 상세 분석 결과, **미세먼지가 「나쁨」 일 경우 풍향은 서풍이 불며 산둥성, 산시성, 베이징·허베이성 등의 중국 지역의 에어로졸 농도가 매우 높다**는 것을 확인할 수 있었다.

○ 특히 인천지역 20개 관측소의 미세먼지 예측 연관성을 비교한 결과, **인천 도심 지역이 아닌 백령도 지역의 미세먼지 및 이산화질소(NO_2)가 가장 높은 연관성**을 보였으며, 이는 **국내 요인보다 국외 요인이 상대적으로 높음을 보여주는 결과**이다.

□ 또한 데이터에서 국외 요인을 제거 후 '18년 1분기를 예측한 결과, **「좋음」 등급은 20일에서 30일로 50%나 증가**하는 것으로 나타났다.

□ 향후 국가정보자원관리원은 보다 정확한 예측을 위해 에어로졸 분석 성능이 뛰어난 **국내 정지 위성(천리안 2A·2B) 데이터를 추가로 확보**하고 **다른 분석 모델과의 결합**을 통해 예측 정확도를 높인다는 계획이다.

□ **김명희 행정안전부 국가정보자원관리원장**은 "이번 분석은 국민의 생존권과 직결되는 **미세먼지 문제를 빅데이터로 접근한 아주 의미 있는 사례**"라고 말하며, "**미세먼지 예보에 기계학습 예측모델이 적극적으로 활용되기를 기대**하며 향후에도 재난·안전 등 사회적 가치가 높은 분석과제를 지속적으로 수행하여 **정부정책에 대한 국민의 신뢰**를 얻고 **국민의 삶이 개선되도록 노력**할 계획"이라고 밝혔다.

참고 1	NASA 기상 위성 및 MODIS 센서

□ 미세먼지 분석과 위성 데이터

○ 미세먼지 등 대기 중 오염물질의 발생 및 이동은 특정 지역에 국한되지 않고 넓은 영역을 통해 이동 및 확산되는 특성을 가짐

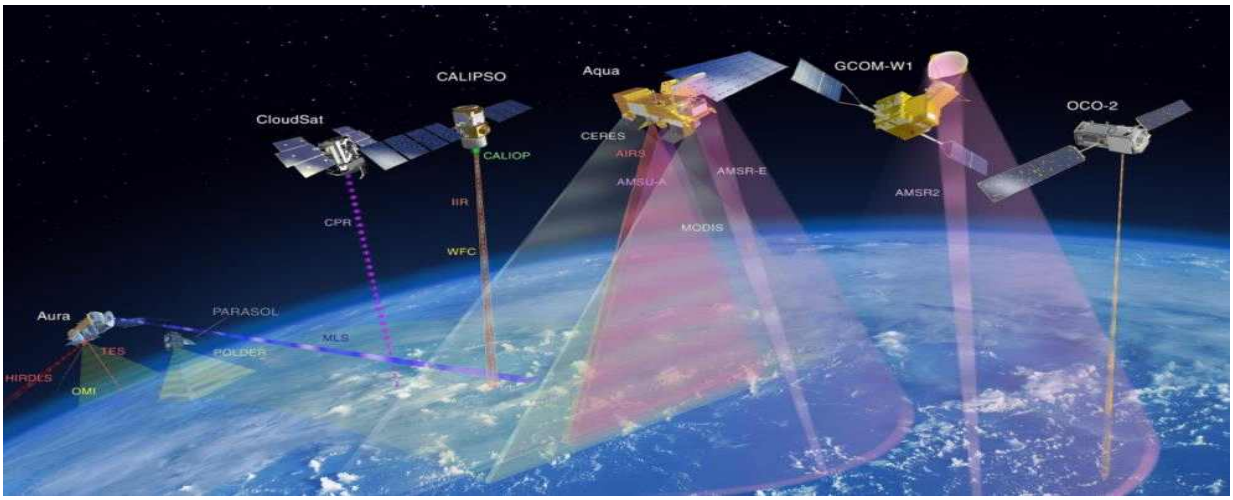
⇒ <u>미세먼지 분석을 위해 위성관측을 통한 광범위한 영역의 데이터 활용 필요</u>

□ NASA 기상 위성

○ 지구 대기 환경 및 기후 연구를 위해 Terra 및 Aqua 위성을 각각 '99년 '02년에 발사하여 현재 운용 중

○ 상기 위성은 특정 시간대에 해당하는 지역을 탐사하도록 설계

- Terra 위성은 10시 30분경 적도 상공을 북→남 방향으로, Aqua 위성의 경우 13시 30분경 남→북 방향 탐사 (탐사범위 : 2,800 km)



□ MODIS란?

○ 미국 NASA의 지구감시계획(EOS : Earth Observing System)에 의해 Aqua(해양)와 Tera(지형·대기) 위성에 탑재된 관측 센서

○ MODIS* 데이터는 저작권이 없어 다양한 글로벌 연구에 활용되며 육상과 해양의 표면 온도, 해류의 흐름, 대기 관측 등의 데이터로 구성

* MODerate resolution Imaging Spectroradiometer : 중간해상도 영상 분광계

참고 2 NASA 위성 데이터

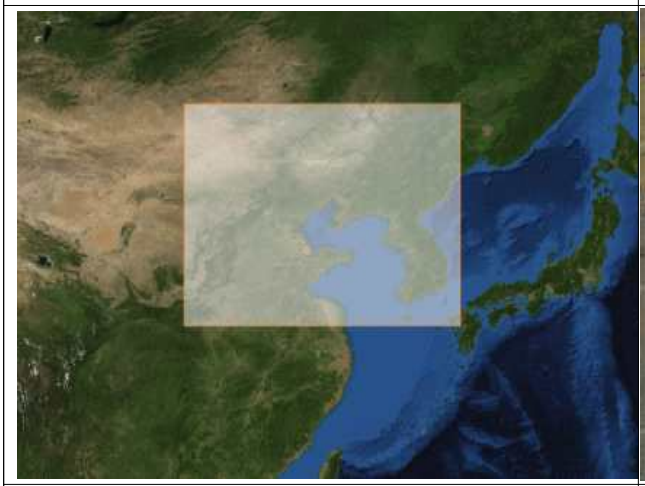
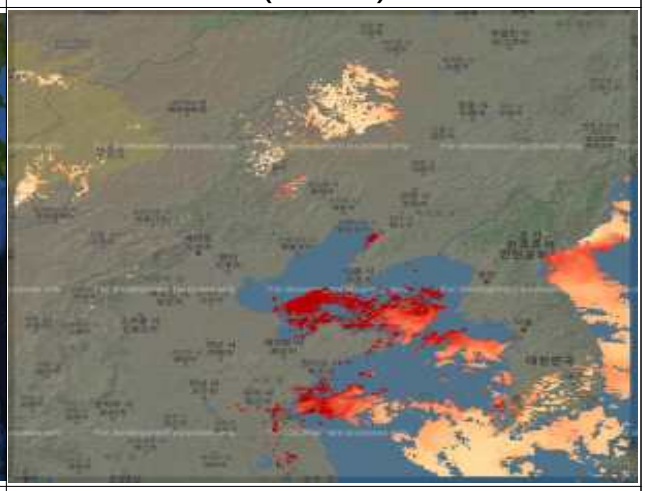
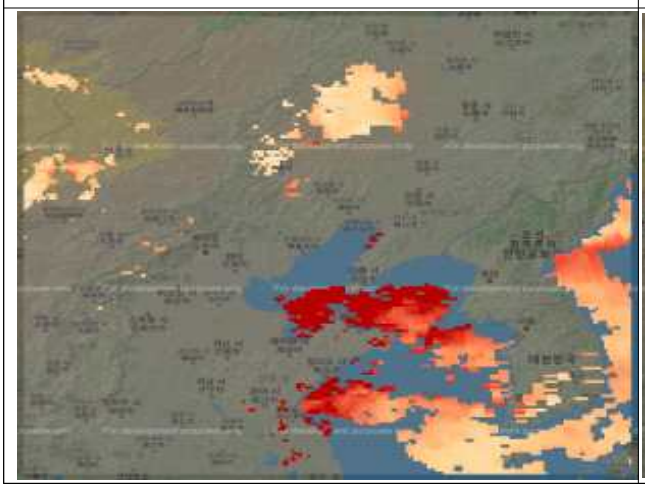
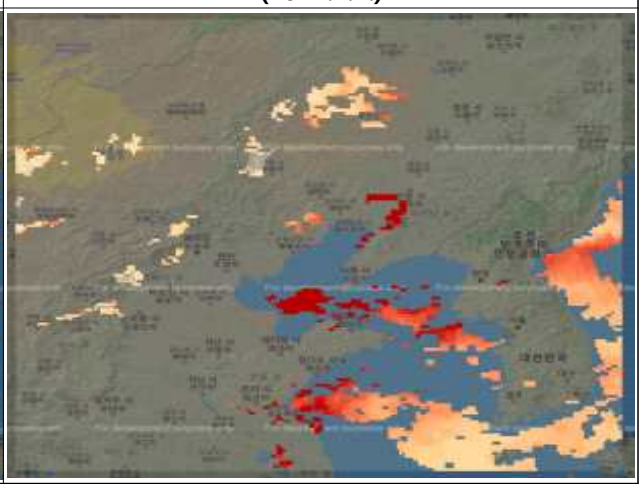
□ 위성 관측 데이터

구분		Dataset	제공 해상도
NASA위성 Aqua, Terra	MODIS AOD	Image Optical Depth Land And Ocean	3KM
		Dark Target Deep Blue Combined (분석활용)	10KM
		Image Optical Depth Land And Ocean	

○ 수집 기간 / 범위 : 2015~2018년 3월 / 몽골, 중국, 한국 지역

※ 수집 좌표 : N 46.7, S 32.9, W 110.4, E 130.4

○ 데이터 형태 / 단위 : 이미지 파일(TIFF) 또는 HDF 파일 / 일 단위

수집 데이터 관측 범위	MODIS 3km Image Land & Ocean data (2017.1.1.)
	
MODIS 10km Image Land & Ocean data (2017.1.1.)	**MODIS 10km Dark Target & Deep Blue Data (2017.1.1.)**
	

참고 3 미세먼지 예보

□ 미세먼지 예보

○ 환경부에서 하루 4회 (오전 5, 11시, 오후 5, 11시) 미세먼지 예보

- 예보 등급은 좋음, 보통, 나쁨, 매우나쁨의 4단계로 구분

미세먼지 예보등급 및 예보내용

예보 내용		등급(㎍/㎥)			
		좋음	보통	나쁨	매우나쁨
예보 물질	미세먼지 (PM_{10})	0~30	31~80	81~150	151 이상
	미세먼지 ($PM_{2.5}$)	0~15	16~50	51~100	101 이상
행동 요령	민감군	-	실외활동시 특별히 행동에 제약을 받을 필요는 없지만 몸 상태에 따라 유의하여 활동	장시간 또는 무리한 실외활동 제한, 특히 천식을 앓고 있는 사람이 실외에 있는 경우 흡입기를 더 자주 사용할 필요가 있음	가급적 실내활동, 실외활동시 의사와 상의
	일반인	-	-	장시간 또는 무리한 실외활동 제한, 특히 눈이 아픈 증상이 있거나 기침이나 목의 통증으로 불편한 사람은 실외활동을 피해야 함	장시간 또는 무리한 실외활동제한, 목의 통증과 기침 등의 증상이 있는 사람은 실외활동을 피해야 함

※ 미세먼지 예보등급은 PM_{10}과 $PM_{2.5}$ 중 높은 등급을 기준으로 발표

※ 민감군 : 어린이, 노인, 천식같은 폐질환과 심장질환을 앓고 있는 어른

- 광역시도 기준으로 PM10과 PM2.5 예보

내일 [01월22일]

예보등급

△[미세먼지] 경기남부·중남·호남권은 '나쁨', 그 밖의 권역은 '보통'으로 예상됨. 다만, 서울·경기북부·중북은 오전과 밤에 '나쁨' 수준의 농도가 나타날 수 있음.

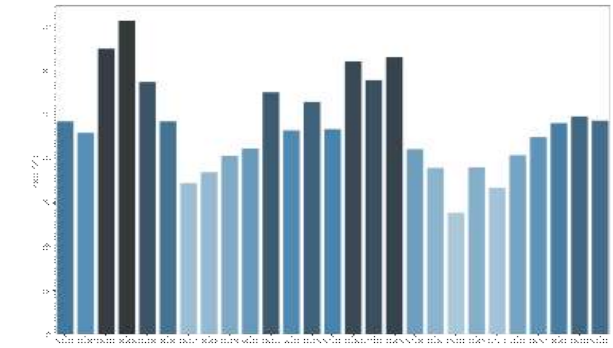
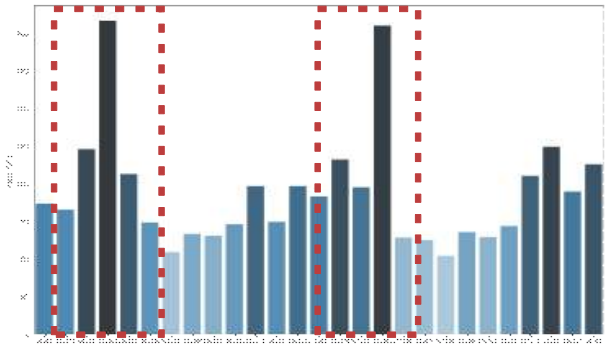
구분	서울	인천	경기 북부	경기 남부	강원 영서	강원 영동	대전	세종	충북	충남	광주	전북	전남	부산	대구	울산	경북	경남	제주
미세먼지	보통	보통	보통	나쁨	보통	보통	보통	보통	보통	나쁨	나쁨	나쁨	나쁨	보통	보통	보통	보통	보통	보통
PM_{10}	보통	보통	보통	보통	보통	보통	보통	보통	보통	보통	보통	보통	보통	보통	보통	보통	보통	보통	보통
$PM_{2.5}$	보통	보통	보통	나쁨	보통	보통	보통	보통	보통	나쁨	나쁨	나쁨	나쁨	보통	보통	보통	보통	보통	보통

※미세먼지예보 등급은 PM_{10}과 $PM_{2.5}$ 중 높은 등급을 기준으로 발표됨.

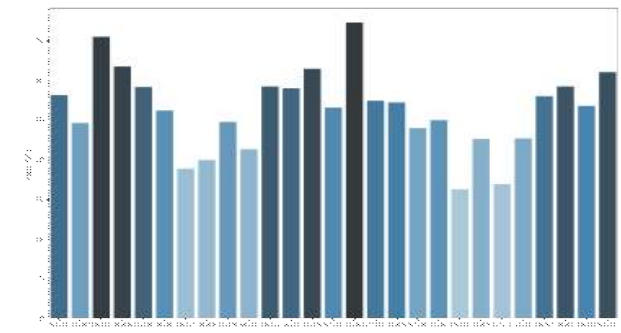
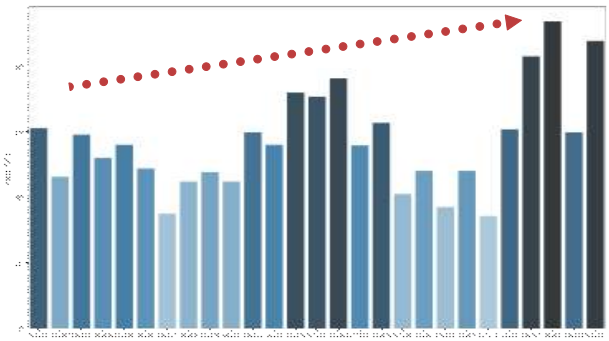
참고 4 인천지역 미세먼지

□ 미세먼지 농도

○ (PM10) 3~5월의 미세먼지 평균 농도가 높고, 특히 해당기간의 월별 최대 미세먼지 농도가 크게 높아지는 특징을 보임

PM10 월별 평균	PM10 월별 최대
	

○ (PM2.5) PM10과 유사하게 3~5월의 미세먼지 평균 농도가 높고, 월별 최대 미세먼지 농도가 점차 높아지는 추세임

PM25 월별 평균	PM25 월별 최대
	

□ 미세먼지 예보 등급

○ 데이터 수집 기간('15~'18. 3) 중 미세먼지 등급 비율은 「보통」이 PM10 75.7%, PM2.5 56.0%로 가장 많은 편임

구분		예보 등급 (일평균, ㎍/㎥)			
		좋음	보통	나쁨	매우나쁨
PM 2.5	기준	0~15	16~35	36~75	76~
	건수	238 (20.1%)	**663 (56.0%)**	277 (23.4%)	5 (0.4%)
PM 10	기준	0~30	31~80	81~150	151~
	건수	196 (16.6%)	**895 (75.7%)**	86 (7.3%)	6 (0.5%)

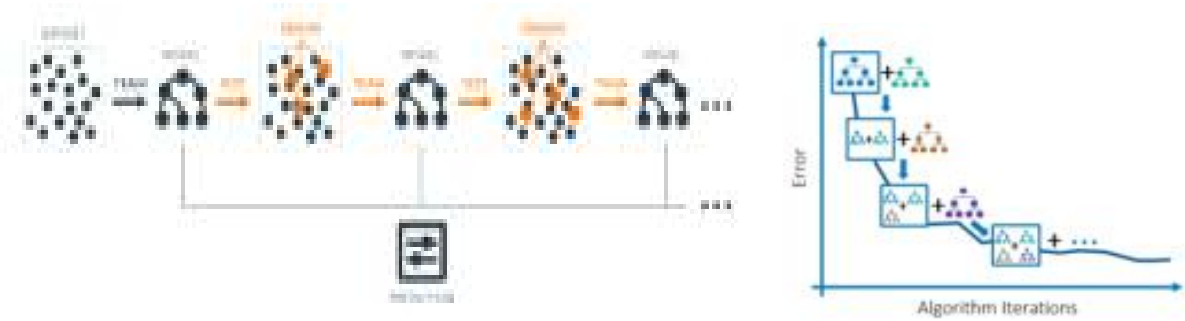
※ 인천지역 23개 측정망 중 도시 대기 15개소의 PM10 및 PM25 수치를 평균하여 대기 질을 4단계로 예보. 미세먼지 농도(일평균, ㎍/㎥) 예보 기준

참고 5 GBM 기반 미세먼지 예측 모델

□ GBM (Gradient Boosting Machine)

○ 여러 개의 의사 결정 모델을 연결하여 강력한 모델을 만드는 부스팅 방식의 앙상블 기법

- 이전에 만들어진 의사결정 모델의 오차를 보완하는 방식으로 순차적으로 모델이 연결되어 뒤로 갈수록 오차가 작아짐



□ GBM 기반 미세먼지 예측 모델

○ NASA Aqua 위성 MODIS 센서 데이터 및 인천지역 관측 데이터를 전처리하여 GBM 모델을 학습하여 최적의 모델 구성

참고 6 PM10 예측 모델 결과

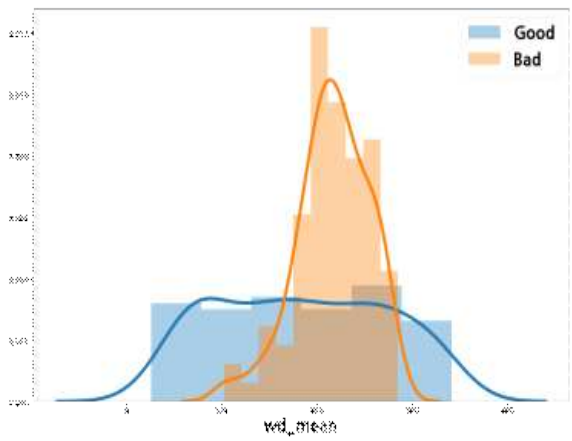
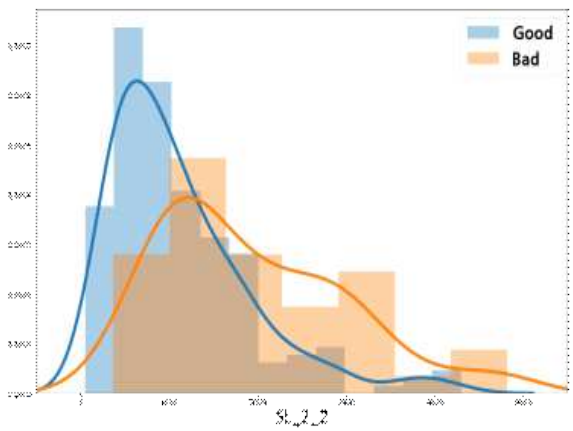
□ 변수 중요도

○ GBM 모델에서 미세먼지 4단계의 정확한 예측을 위해 필요한 주요 변수의 중요도 제공

변수명	변수설명
wd_mean	풍향 평균
rain_f	비 예보
SL_2_4	서해안 및 인천
SL_2_2	중국 산둥성 위성 센서데이터
wd_std	풍향 표준편차
SL_0_0	중국 내몽골 자치구 위성 센서데이터
ws	풍속
SL_2_0	중국 산시성 위성 센서 데이터
meanPM10_hourkurtosis	PM10 일 척도
SL_1_3	중국 랴오닝성 위성 센서 데이터

□ 주요변수 특성

○ 예보 등급이「나쁨」일 경우 전날 풍향은 서풍이며, 산둥성 지역의 위성 센서 데이터의 분포도 높은 편임

예보 등급별 풍향 분포	예보 등급별 산둥성 위성 센서 데이터 분포
	

참고 7 PM2.5 예측 모델 결과

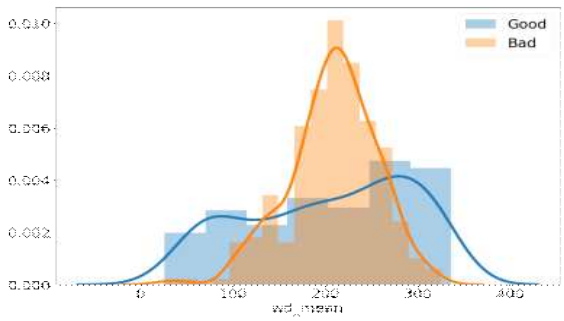
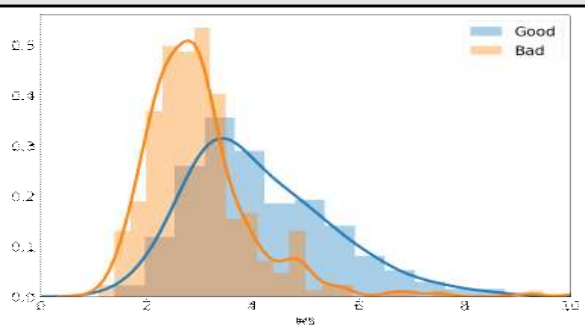
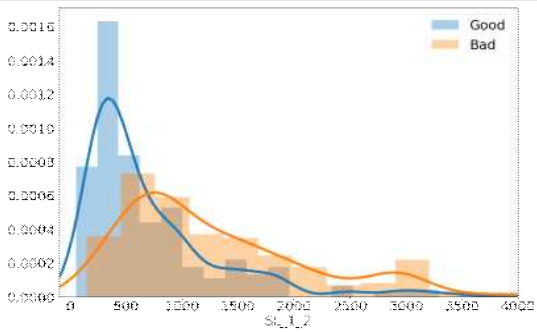
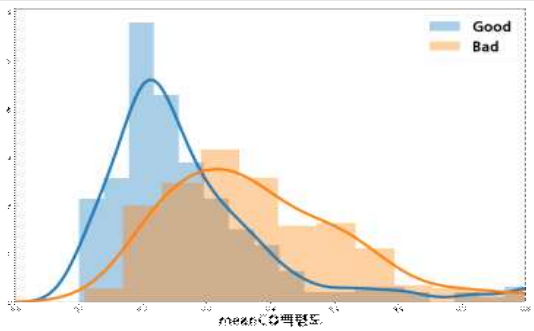
□ 변수 중요도

○ PM2.5 미세먼지 4단계의 정확한 예측을 위해 필요한 주요 변수 중요도

변수명	변수설명
ws	풍속 평균
ws_f	풍속 예보
wd_mean	풍향 평균
meanCO백령도	백령도 CO
rain	강수량
SL_0_1	중국 내몽골 자치구 위성 센서 데이터
wd_f	풍향 예보
SL_1_2	중국 베이징 및 허베이성 위성 센서 데이터
meanPM25_hourskew	PM10 일 왜도
wd_std	풍향 편차

□ 주요변수 특성

○ 예보 등급이「나쁨」일 경우 전날 풍향은 서풍이며, 풍속도 상대적으로 느림. 또한 베이징·허베이성 센서 데이터와 백령도 지역의 일산화탄소 데이터 값이 높음

예보 등급별 풍향 분포	예보 등급별 풍속 분포
	

예보 등급별 베이징·허베이성 센서 데이터 분포	예보 등급별 백령도 일산화탄소 데이터 분포