

기상 상태에 따른 대기환경지수(AQI) 변화와 중요 요인에 대한 연구

김대양(*), 태정수(**), 아타케 마사아키(***), 이영한(****), 정영애(*****)

(*) 선문대학교 AI소프트웨어학과, daeyangkim124@gmail.com

(**) 선문대학교 AI소프트웨어학과, xowjdt@gmail.com

(***) 선문대학교 AI소프트웨어학과, awuy316@gmail.com

(****) 선문대학교 AI소프트웨어학과, hans0209@sunmoon.ac.kr

(*****) 선문대학교 AI소프트웨어학과, dr.youngae.jung@gmail.com

1. 연구 배경

과거부터 현재까지 산업화와 도시화가 진행되면서 미세먼지의 위험은 계속해서 증가하고 있다. 미세먼지를 유발하는 요인은 다양하지만, 인위적인 요인으로는 공장과 같은 작업장의 영향이 무시될 수 없다[1].

지난 10년간 국내 공장 증가율은 2002년부터 2012년까지 연평균 5.3%에서 2012년부터 2021년까지 3.1%로 둔화되었으며, 현재는 2%대에 머물러 있다[2]. 그러나 가동 중인 공장은 여전히 우리의 삶에 영향을 미친다.

본 연구에서는 우리의 삶에 영향을 미치는 미세먼지의 주요 요인과 그 효과를 분석하기 위해 영국, 미국, 중국의 한 지점을 기준으로 기상 데이터를 수집 및 분석하였다. 이 분석을 통해 얻은 시사점을 바탕으로 산업단지 조성을 고려하는 기업들이 참고할 수 있는 요인을 도출하는 것을 목표로 한다.

2. 연구 내용

2.1 분석 배경

본 연구를 진행하기에 앞서, 분석 대상 지역을 비교하는 척도가 필요하다. 본 연구에서는 AQI(Air Quality Index; 대기질 지수)를 사용한다[3].

$$AQI = \left(\frac{I_{high} - I_{low}}{C_{high} - C_{low}} \right) (C - C_{low}) + I_{low} \quad (\text{식 1})$$

I는 AQI의 값에 따른 범주의 상한과 하한을 나타내며 C는 오염물질의 농도이며 각각 상한과 하한을 뜻한다.

AQI는 나라에 따라 기준이 다르기 때문에 분석대상이 되는 나라에 맞춰서 계산하였다.

2.2 데이터 수집

위도 경도를 통해 특정 지역의 오염물질, 기상데이터를 수집하는데 특화된 사이트 Open-Meteo에서 API를 통해 수집을 진행하였다[4].

2.3 기술 분석

수집한 데이터셋은 오염물질관련 특성 7개, 기상관련 23개, 기타 11개, 종속변수(AQI) 1개로 총 42개의 변수로 구성하였다. 상관분석을 진행한 결과, 같은 범주에 속하는 데이터 중에서 측정된 기준이 다른 경우(측정 고도, 단위) 상관관계에 차이가 존재하였다. 이에 서로 상관이 보이는 값의 평균을 계산하거나, 분석 목적에 맞지 않는 변수를 제거하는 방식으로 변수의 수를 축소하였다.

분석 중 정상 데이터와 이상치에 해당하는 데이터 간 분포에 차이가 있는 변수를 토대로 분석을 진행하였고, 이상치는 Tukey Fences 방법을 사용하여 판별하였다.

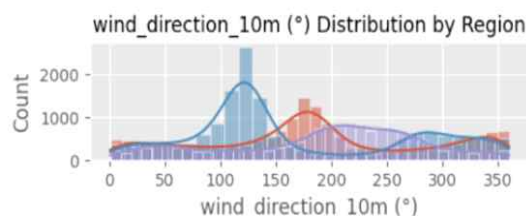


그림 1. Wind Direction 변수의 분포(정상)

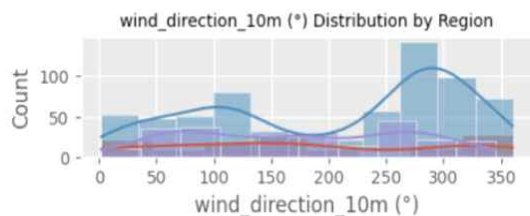


그림 2. Wind Direction 변수의 분포(이상치)

위와 같이 분포의 차이를 보이는 특성을 선별하여 변수 구성을 진행하였다.

표 1. 변수 구성표

순	변수	변수 설명
1	Carbon	탄소($\mu\text{g}/\text{m}^3$)
2	Nitrogen	질소($\mu\text{g}/\text{m}^3$)
3	Sulphur	황($\mu\text{g}/\text{m}^3$)
4	Ozone	오존($\mu\text{g}/\text{m}^3$)
5	Dust	먼지($\mu\text{g}/\text{m}^3$)
6	UV Index	자외선 지수
7	Temperature	기온($^{\circ}\text{C}$)
8	Wind Data	바람 관련 데이터
9	Soil Data	토양 관련 데이터

구성한 데이터를 토대로 추가적인 탐색적 데이터 분석을 진행한 결과, 오염물질 뿐 아니라 해당 지역의 바람의 세기, 기온, 습도 등이 상당한 영향을 주는 것으로 분석되었다. 그 외에도 계절의 변화에 따라서 부는 바람의 방향에 따라서도 AQI에 대한 변화를 확인할 수 있었다.

표 2. 풍향에 따른 AQI 평균값

순	풍향	AQI 평균값
1	NE	34.232485
2	NW	28.592089
3	SE	35.158607
4	SW	31.397365

2.4 예측분석

예측분석을 위해 사용한 모델은 Extra Tree 회귀모델을 사용하였다. fold를 10으로 설정하여 교차검증을 진행하면서 총 9개의 회귀모델에서 각

교차검증 스코어를 파악하여 모델을 선정하였다.

파라미터는 max_depth에서 7, max_feature에서 10, min_sample_split에서 3, n_estimators에서 300으로 설정하여 학습을 진행하였다.

직접적인 영향을 미치는 오염물질을 제외한 변수 중요도는 토양의 습도나 온도이며 고도 100~255cm의 측정값이 가장 중요한 것으로 보인다.

풍향의 경우 풍향 자체로 영향을 미치기보다 불어오는 방향에 무엇이 존재하는지에 따라서 측정지역의 AQI가 변하는 것으로 미루어 지리적 위치에 대한 고려도 필요하다고 판단된다.

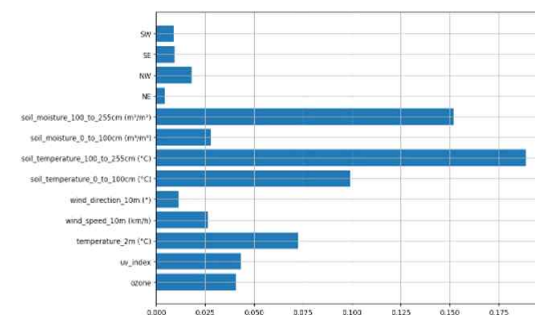


그림 3. 변수 중요도

4. 결론 및 향후 연구 계획

본 연구에서는 오염물질의 존재뿐만 아니라 공기 또는 토양 습도, 풍속이나 정체, 풍향과 같은 기상 및 계절적 요인도 고려해야 한다는 점을 확인하였다. 미국에서는 분석 지점의 남쪽에 위치한 오클라호마시티에서 남풍이 AQI에 영향을 미치는 것으로 나타났다.

산업계가 이러한 요인을 고려할 경우, 산업단지 건설에 있어 더 나은 의사 결정을 내릴 수 있을 것이다.

향후 계획으로는 의미 있는 지리적 요구를 반영한 데이터 수집을 통해 데이터 품질을 개선하고, 지리적 및 인구 통계적 요소를 포함하여 연구의 한계를 극복하는 방향으로 요인을 확장하는 것을 목표로 하고 있다.

감사의 글

본 연구는 2024년 과학기술정보통신부 및 정보통신
기획평가원의 SW중심대학사업 지원을 받아 수행되었
음(2024-0-00023)

참고 문헌

- [1] 심창섭 . 미세먼지의 주요원인과 대응 전략 . 에너지포
커스 2019 봄호(제16권 제1호 통권71호)
- [2] 편집국, “국내 제조업 성장 둔화 심각...글로벌 경쟁력
저하 우려”, 2023.3.12., [https://biz.newdaily.co.kr/
site/data/html/2023/03/12/2023031200038.html](https://biz.newdaily.co.kr/site/data/html/2023/03/12/2023031200038.html)
- [3] 황광일 외 1인, “실시간 통합 실내 공기질 알람지수 결
정 방법 및 장치”, South Korea. Patent
KR101771053B1, Aug. 25, 2017.
- [4] 데이터 셋, <https://open-meteo.com/>