

# 양상불 기반의 악취 농도 다지역 통합 예측 모델 개발\*

조성주

충북대학교 빅데이터 협동과정  
(fulans2@naver.com)

최우석

충북대학교 빅데이터 협동과정  
(cdt3017@naver.com)

최상현

충북대학교 경영정보학과/빅데이터 협동과정  
(chois@cbnu.ac.kr)

.....  
전 세계적으로 대기오염 관련 질병 발병률이 상승하고, 2022년 세계보건기구의 보고에 따르면 매년 약 700만 명의 사망자가 발생하고 있다. 또한, 산업 시설 확장과 다양한 배출원 증가, 그리고 악취 물질의 무분별한 방출로 인해 대기오염 문제는 사회적으로 중요성을 띠고 있다. 한국에서도 악취를 독립적인 환경오염으로 정의하며, 지역 주민의 건강에 직접적인 영향을 미치는 문제로 간주하고 있으나 현재까지 악취 관리가 미흡하며 악취 관리 시스템의 개선이 필요하다.

본 연구에서는 악취 관리 시스템 개선을 목표로 충청북도 오창에 설치된 악취 센서에서 수집한 1,010,749개 데이터를 활용하여 양상불 기반의 악취 농도 다지역 통합 예측 모델을 설계하고 분석하였다. 연구 결과, XGBoost 알고리즘을 사용한 모델의 RMSE가 0.0096로 가장 성능이 좋았으며, 단일 지역 모델(0.0146)과 비교하여 평균 오차 크기가 51.9% 낮았다. 이를 통해 서로 다른 지역에서 수집된 악취 농도 데이터를 표준화한 후 다지역 통합 예측 모델을 설계함으로써 데이터의 양을 늘리고 정확도를 높일 수 있으며 또한, 하나의 통합 모델로 다양한 지역에서 예측이 가능함을 확인하였다.

**주제어** : 악취, 트리메틸아민, 통합모델, 머신러닝, 양상불

.....  
논문접수일 : 2023년 9월 6일

논문수정일 : 2023년 9월 19일

게재확정일 : 2023년 9월 21일

원고유형 : Regular Track

교신저자 : 최상현

## 1. 서론

전 세계적으로 대기오염에 의한 질환은 지속적으로 증가하고 있는 가운데(서형준, 이형석, 2019), 세계보건기구는 대기오염에 의한 사망자 수가 연간 약 700만 명에 이른다고 추정하고 있다(WHO, 2022). 또한, 산업시설의 급속한 팽창과 함께 환경사업소, 다양한 축사로 인해 증가하는 대기오염 문제는 사회적으로 중요한 문제가 되고 있으며(Jackson, 2020; Charlotte et al., 2021), 대기오염 중 하나인 악취의 무분별한 유발은 환경소음과 함께 가장

큰 민원의 대상이 되고 있다(Sakawi et al., 2011).

한국의 악취방지법(2005)에서는 악취를 환경정책기본법상 대기오염, 수질오염과 함께 독립된 환경오염으로 정의되며, 두통, 구토 등 불쾌감과 혐오감을 유발하여 신체적, 정신적 스트레스를 증가시키는 등 지역 주민의 건강과 직결되는 문제이다(박진옥 등, 2016). 따라서 주민들의 삶의 질을 개선하기 위해서는 악취로 인한 스트레스를 저감시킬 방안이 필요하다.

해외 연구사례에서 각종 산업시설, 축산업, 폐기물 처리 산업, 담배 연기, 애완동물 배설 등과

\* 본 과제(결과물)는 2023년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 지자체-대학 협력기반 지역혁신사업의 결과입니다.(2021RIS-001)

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2023년 문화체육관광 연구개발사업의 연구결과로 수행되었음 (RS-2023-00227532)

〈표 1〉 한국과 일본의 악취 관리지역 및 민원 현황 비교(2018)

	일본	한국
악취 관리지역 지정 현황	1,283개 지자체(73.8%) * 관리지역을 가지고 있는 지자체 수/ 총 지자체 수	5,116개소(4.65%) * 신고대상 사업장 수/총 사업장 수
관리지역 내 악취 민원 건수	4,870건(39.0%)	1,869건(7.6%)
관리지역 외 악취 민원 건수	6,224건(49.0%)	20,073건(81.0%)

같이 일상 활동에서 발생하는 가스도 생물과 환경에 유해한 영향을 끼칠 수 있음을 시사하였으며(Kiliyankil et al., 2021), 주요 일상 악취 가스에는 트리메틸아민(TMA)이 다수 포함되어 있다는 연구 결과가 나왔다(Wang et al., 2022).

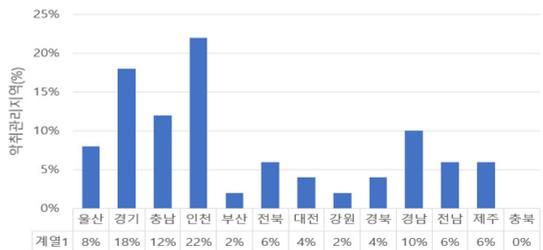
또한, 한국 환경부에서 지정하여 관리하는 주요 악취물질 22종 중 트리메틸아민이 축산(양돈)시설, 돈 및 계분 발효장, 음식물 퇴비화 시설, 위탁폐기물 및 폐수 처리사업장에서 발생하는 악취의 주요 원인 물질로 꼽히고 있다(Phyo et al., 2023).

아민류인 트리메틸아민은 어류가 부패할 때 oxide가 박테리아 또는 효소의 작용으로 분해하면서 발생하며, 0.032ppb 수준으로서 매우 낮은 농도에서 악취를 유발시키는 물질이다(안지원 등, 2007). 트리메틸아민의 냄새 특성은 생선 썩은 냄새와 유사한 비린내 및 불쾌한 냄새가 나며 호흡기, 눈, 피부, 소화기관 등 인체에 많은 유해성을 갖고 있다(Lan et al., 2023). 따라서 악취 물질 중에서도 트리메틸아민의 모니터링이 우선시되어야 한다.

국립환경과학원의 연구 보고서(2018)에 따르면 한국과 미국, 일본은 악취 농도나 강도로 악취의 수준을 평가한다. 구체적으로 한국과 미국은 악취 기준에 대한 만족 여부로 판단하는데 미국은 Odor School에서 현장 조사원을 양성하여 악취 판정사 제도를 운영하는 등 악취 판단 기준에 대해 객관성을 확보하려고 노력하는 반면, 한국은

개인별 전문성에 대해 검증되지 않아 객관성을 확보하기 어려운 실정이다.

〈표 1〉과 같이 국립환경과학원(2018)의 조사에 따르면 일본은 한국과 마찬가지로 인구 밀도가 높고 악취 배출시설과 주거지역이 가까운 거리에 있어 악취와 관련된 민원이 많이 발생하고 있으며(Das et al., 2019), 2012년 이후부터 지자체가 자체적으로 악취규제지역을 지정하여 악취 관리를 실시하고 있다. 반면에 한국은 지자체 계획을 수립하면 환경부 장관 보고 절차를 따르는 복잡한 절차로 인해 일본(73.8%)에 비해 한국(4.7%)의 악취관리 지역이 현저히 부족한 실정이다. 또한, 한국은 악취규제지역 대비 악취규제지역 외에서의 민원 비율이 81%로 일본(49%)에 비해 훨씬 많은 민원이 제기되고 있다.



〈그림 1〉 한국 지역별 악취관리지역 현황

실제로, 환경부(2022) 조사에 따르면 한국의 악취에 대한 민원은 2015년 15,573건에서 2021년

39,397건으로 꾸준히 증가하고 있음에도 불구하고, 한국은 한국환경공단(2023)이 악취 관리 대상으로 지정한 50개 지역으로만 국한되어 관리되고 있다. 즉 한국의 악취관리는 다른 선진 국가에 비해 부족한 실정이다. 특히 농림축산식품부(2021)에 따르면 충청북도는 악취 발생이 제주 지역을 제외하고 전국 10개 지역에서 5위로 높지만, <그림 1>에서 볼 수 있듯이 충청북도에 악취 관리 대상 지역이 하나도 지정되지 않았다.

충청북도 청주시에 위치한 오창지역은 충북 제조업의 중심지인 오창과학산업단지가 조성되어 공장이 많이 지어져 있으며, 이에 따라 인근지역의 악취 민원이 매우 많은 편이다(장동열, 2019). 따라서 본 연구에서는 우선적으로 충청북도의 오창읍에서 발생하는 트리메틸아민 물질을 대상으로 하였으며, 도심에서 악취를 유발하는 물질은 대기 중에 부유하기 때문에 온도 및 습도 등과 같은 기상환경과 밀접한 관계를 지니고 있다. 이를 반영하기 위해, 기상 자료 개방 포털에서 Open API로 개방한 방재 기상 관측자료(AWS)를 활용하였으며, 또한, 시간의 흐름에 따라 악취 농도가 변하는 특성을 고려하여 시간속성을 반영한 변수를 생성함으로써, 머신러닝 기반의 양상불 모델을 활용한 악취 농도 다지역 통합 예측 모델을 개발하고자 한다.

본 연구자는 특정 지역을 대상으로 한 악취 농도 예측 모델이 아닌, 다른 지역에서도 활용할 수 있는 악취 농도 다지역 통합 예측 모델을 개발함으로써, 여러 지역에도 적용할 수 있는 효율적인 악취의 체계적인 관리 방법을 제안하고자 한다. 뿐만 아니라 한국 악취 관리 규제 대상 외 지역인 충청북도 오창읍의 악취 농도를 예측하여, 시민들에게 사전 관련 정보 제공이 가능할 것으로 보이며, 이에 따라 시민들의 민원감소 및 삶의 질을 향상시키고자 한다.

## 2. 선행 연구

### 2.1 악취 오염원 분석 연구

악취를 포함한 미세먼지 등 대기오염이 심각해지면서 국내·외 전문가 및 정부 기관에서는 많은 관심과 다양한 연구를 진행하고 있으며(유지현, 2020; Liu, 2021), 이전부터 악취 배출시설에서의 악취 발생 원인 물질에 대한 연구사례가 나오고 있다.

도우곤(2020)은 생곡산업단지 인근의 주요 악취 유발의 원인이 되는 악취물질을 측정하여 발생 현황을 파악한 후, 주요 배출원별 악취 발생 특성 분석 및 주요 원인 물질들의 악취 기여율을 도출하여 그에 맞는 물질 별 저감 대책 방안을 제안하였다. 구체적으로 22개의 지정된 냄새 화합물은 3개의 주요 배출원, 4개의 공장 지역 경계 지점 및 4개의 영향을 받은 지역의 출구에서 측정하였으며, 그 결과 19개의 성분을 검출하였다. 공업 지역에 위치한 경계지역에서는 프로피온산과 n-부틸산이 배출기준을 초과하였으며 악취 활성 값을 이용한 악취 기여 분석 결과, 휘발성 지방산은 공업 지역의 경계 지점 2곳과 주요 북쪽 배출원 1곳에서 기여도가 가장 높았고, 알데하이드류는 음식물류 폐기물처리 시설 등과 같이 다른 사업장에서 기여도가 상대적으로 높았다.

Shan et al. (2019)는 퇴비화 조건 및 공정의 최적화를 위해, 하수 슬러지 퇴비화 공정에서 발생하는 암모니아와 휘발성 황화합물(VSCs)의 농도를 측정하여 시간에 따른 냄새 기여도와 악취 생성의 원인을 분석하였다. 연구 결과, 암모니아가 모든 작업장에서 지배적인 주요 물질이었으며 상대적으로 높은 농도의 VSCs가 슬러지 적재장과 퇴비화 작업장에서 검출되었다. 악취 오염은

퇴비화 과정의 첫 15일 동안 주로 발생했으며, 암모니아의 냄새 기여도는 시간이 지남에 따라 증가했고 VSC는 처음 5일 동안 크게 증가했다. 암모니아의 생산은 온도, 질소 함량 및 상태 변화와 관련이 있지만 VSC의 반응 및 변형으로 인해 생산이 복잡하다는 것을 알 수 있었다.

위와 같은 연구는 지역별, 사업장별 악취 원인 물질을 찾아내거나 악취물질의 농도를 측정함으로써 그에 따른 물질 별 저감 대책을 마련하는 등 주로 오염원을 분석하였다. 하지만, 악취는 배출 원과의 거리, 기상환경, 지형과 같은 주변 환경 요인으로 인해 민감도나 농도가 달라진다(Conti et al., 2020). 따라서 악취 농도를 고려하기 위한 통계적인 접근이 필요하며(Invernizzi et al., 2021), 이에 따라 대기오염물질의 확산을 표현하는 가장 기본적인 방법은 가우시안 모델로, 해당 모델은 특정 배출원으로부터 일정 거리까지에 있는 오염 물질의 농도를 추정하는 데 사용하는 수학적 모델이다(석희정 등, 2020).

Invernizzi et al. (2021)는 악취 농도 평균, 분산 및 농도의 강도 측면을 계산하기 위해서 악취 농도 변동의 가변성을 고려한 가우시안 플룸 모델을 사용하였으며, 악취 평균 농도 값의 72% 정도가 해당 요인에 속한 반면에 표준편차나 농도에 대한 결과에 대해서는 데이터를 과대평가하는 것을 확인하였다.

하지만, 가우시안 모델은 매우 정교하고 정상 상태의 기상 조건이 제공되어야 하며 시시각각 변하는 배출량도 고려해야 하고, 기상 조건이 비 정상 상태를 고려할 수 있지만, 과거 상태에 대한 데이터를 고려하기 어려우며, 3차원 공간 특성을 반영하기 힘들기 때문에 이러한 모델의 접근법에는 한계점이 존재한다(김강한 등, 2020; Conti et al., 2020). 따라서 본 연구에서는 통계 기반의

접근법의 한계를 보완하기 위하여 빅데이터 기술인 머신러닝 기법으로 악취 수준을 예측하고자 한다.

## 2.2 빅데이터 기반의 악취 수준 예측 연구

최근 빅데이터 수집이 용이해지면서 빅데이터를 활용하여 과거 속성 데이터를 반영할 수 있으며, 악취의 시계열적인 속성과 기상 데이터를 반영한 악취 농도 예측 모델을 개발하고 악취 농도 저감을 위한 효율적인 관리를 위해 관련한 머신러닝과 딥러닝 기법의 연구가 고안되고 있다.

황두환(2021)은 머신러닝의 여러 가지 회귀 기법들을 활용하여 돈사 시설 인근의 황화수소의 농도 데이터를 바탕으로 기상 상황을 고려하여 악취 농도 예측 연구를 진행함으로써, 부대 내 근무자 및 거주자에게 사전에 정보를 제공하기 위하여 가장 높은 정확도와 활용성을 보이는 모델을 선정하였다. 그 결과, K-Nearest Neighbor Regression(KNR) 알고리즘이 가장 적합한 모델이었으며 정확도는 95.9%를 보였다.

Jiang et al. (2023)은 생활 폐기물에서의 악취 가스를 처리하기 위해 생활폐기물의 4가지 요인(무게, 습윤 성분, 온도, 발효 시간)을 고려하여 악취 가스 예측지표로 채택하였고, 512가지의 악취 가스의 농도 값을 이용한 예측 연구를 진행하였다. 연구 결과 Random Forest 모델이 89%로 가장 정확한 악취 가스 발생량을 예측하였다.

딥러닝을 활용한 연구에서는 Xu et al. (2022)는 기상 조건과 폐기물 특성을 사용하여 ANN 모델을 구축하여 매립지 작업표면의 악취 배출률을 예측하기 위한 연구를 진행하였으며, 연구 결과 폐기물의 특성은 배출률에 민감한 매개변수이며 기상학적 매개변수는 불확실성에 상당한 요인을 가지는 것을 확인하였다.

윤주상(2022)은 제조 산업 현장에서의 사물인터넷 단말기로부터 실시간 센서 데이터를 수신하고 이 데이터를 모델에 적용하여 화재 예측 및 악취 감지, 예측에 실시간으로 사용할 수 있는 딥러닝 기반의 LSTM 알고리즘 모델을 개발하였다. 암모니아와 휘발성 유기화합물과 습도 데이터를 입력하여 모델의 평균 예측 정확도는 90%의 성능을 보였지만, 시계열 데이터의 한계로 예측 시점이 멀어질수록 정확도가 떨어지는 것을 확인하였다.

다만, 악취의 확산 범위 및 지속시간 등은 공간적 특성에 따라 많은 영향을 받기 때문에, 정확한 악취 농도 수준 진단이 어렵다는 한계점이 있다(김병욱 등, 2020). 이러한 한계점을 해결하기 위해 시간과 공간을 결합한 시공간 분석을 통해 대기질의 농도 예측의 정확도를 높이려는 연구가 있다.

이와 관련된 연구로 이종영(2018)은 악취와 미세먼지 농도를 예측하기 위해서 시간 예측 변수에는 딥러닝 기반의 RNN-LSTM 기법의 앙상블 모델을 사용하였고, 공간 예측 변수로는 지리 가중 회귀 모델을 기반으로 하였다. 두 가지 예측 모델의 결과를 시간 단위로 데이터를 생성하여 다시 LSTM 모델의 입력 값으로 최종 대기질의 예측치를 산출하였다. 또한, 측정소가 없는 지역은 물론 미래 악취 및 미세먼지를 예측하였다.

하지만 위와 같은 빅데이터 기반의 악취 수준 예측 모델은 특정 지역에 국한된 예측 모델이기 때문에 학습한 모델을 타 지역에 적용시키는 것이 어렵다. 악취 관리는 여러 지역을 포괄적으로 관리해야 하기 때문에 여러 지역에 적용할 수 있는 통합모델을 개발하는 것이 중요하다. 따라서 본 연구에서는 데이터 변환 기법 중 Data Scaling 기술을 적용하여 지역 간 규모를 일치시켜 여러 지역에 적용할 수 있는 앙상블 기반의 악취 농도 다지역 통합 예측 모델을 설계 및 제안하고자 한다.

### 2.3 앙상블(Ensemble)

앙상블 기법은 하나의 네트워크가 아닌 여러 개의 네트워크에서 도출된 결과들을 결합하여 예측 성능을 향상시키는 방법이다. 즉, 여러 학습이 적용된 서로 다른 예측모형을 생성 후, 예측 모형의 분석 결과를 종합하여 최종 예측 결과를 도출해 내는 방법을 말한다. 앙상블 기법은 다양한 종류의 모델을 병합함으로써 단일 편향에 영향을 받지 않아 과적합의 문제를 해결할 수 있다. 무엇보다도 단일모형보다 여러 개의 모형을 구축하는데 효율적이며, 단일모형에서 놓칠 수 있는 작은 패턴 반영이 가능한 장점이 있다(이경수 등, 2022). 일반적으로 앙상블 학습은 약한 학습(Weak Learner)을 활용하여 강건한 학습(Strong Learner)을 만드는 것으로 단일모형을 이용하기보다 다양한 모형을 조합하는 것이 더 좋은 결과를 도출할 수 있다는 것을 가정한다(이경수 등, 2022). 분석 과정에 따라 의사결정나무, 배깅(Bagging), 랜덤 포레스트(Random Forest), 부스팅(Boosting) 기법으로 구분할 수 있다. 첫째, 배깅(Bagging) 방법은 Breiman(1996)에 의해 개발된 분류 앙상블 방법이다. 배깅은 전체 데이터에서 일부 데이터를 복원 추출한 후 부스트랩(Bootstrap)을 통해 N개의 부분 데이터 집합(Subset data)을 만들어 내고, 이러한 부분 데이터 집합에 모형을 N 번 적합하여 예측값을 각각 계산한다. 마지막으로 이러한 예측 평균값을 최종 예측치를 정하게 된다. 둘째, 부스팅(Boosting)은 이전 모형 학습이 다음 모형 학습으로 이루어지면서 순차적으로 학습한다. 이 같은 방식으로 학습데이터의 샘플 가중치를 조정해 학습을 진행하는 것이 특징이라 할 수 있다. 즉, 최초의 모형을 구성한 후, 종속변수가 아닌 잔차를 업데이트하는 방식으로 모형을 수정하게

된다(Zhou & Zhou, 2021). 이러한 앙상블 모델을 활용함으로써, 여러 지역에서 수집된 데이터를 모두 융합하여 모델의 예측 능력을 향상시킬 수 있을 것으로 보이며, 개별 모델의 알고리즘의 한계인 특정 데이터 패턴에 민감하게 반응하여 예측 불안정성을 초래하는 약점을 보완할 수 있다. 따라서 앙상블 모델을 활용하여 여러 모델의 평균 또는 다수결을 통해 안정적인 예측 정확도를 확보하고, 예측의 정확도를 높임으로써 신뢰할 수 있는 결과를 제공할 수 있을 것으로 보인다.

### 2.3.1 랜덤포레스트(Random Forest)

랜덤포레스트는 복수의 결정 트리들을 학습하는 방법으로 무작위 노드 최적화와 배깅을 결합한 방법인 분류 회귀 트리(CART)를 이용하여 상관관계가 없는 트리들로 포레스트를 구성한다. 여기서, CART는 예측인자 또는 설명변수의 비선형성과 상호작용을 활용하여 반응 변수에 대한 영향을 측정하는 기법으로, 설명변수에 대해 중요도 기준에 따라 줄기(branch)를 만들며, 최종 노드에서 반응변수에 관한 판단을 내린다(엄하늘 등, 2020).

랜덤포레스트의 작동 방식은 다음과 같다. 먼저, 훈련 데이터로부터  $N$  크기를 갖는 부트스트랩 샘플  $Z$ 를 복원 추출한다. 그리고 이를 이용하여 나무  $T_b$ 를 성장시키는데, 독립변수 총  $p$ 개에서 랜덤으로  $m$ 개의 변수를 지정하고 그중 분할 개선도(goodness of split)가 가장 큰 변수와 그 분기점을 찾아 이진 분할(binary split)하는 두 하위 노드를 형성한다. 각 노드마다 이러한 분할 과정을 거치다가 단말 노드의 크기가  $n_{\min}$ 에 도달하면 나무  $T_b$ 의 성장을 종료한다. 그리고 위의 과정을  $b=1$ 부터  $B$ 까지 반복하여 나무 앙상블이 형성

되면 새로운 데이터 값  $x$ 에 대한 예측 값은 식(1)과 같이 앙상블을 이루고 있는 모든 나무의 평균값을 취해 도출된다(Rigatti, 2017).

$$\widehat{f}_{rf}^B(X) = \frac{1}{B} \sum_{b=1}^B T_b(X) \quad (1)$$

### 2.3.2 XGBoost

Extreme Gradient Boosting(XGBoost)는 여러 개의 의사결정트리로 구성된 부스팅 앙상블 기반의 모델 중 하나로, 의사 결정 규칙을 트리 구조로 표현하여 분류와 예측을 수행할 수 있다(김영훈 등, 2020). 부스팅 기법은 입력 데이터( $x$ )와 Target 데이터( $y$ )가 주어질 때, 식(2)와 같이  $f_1(X)$  모델 학습 도중의 에러( $Error_1$ )가 발생하면 기중치를 부여한 후 식(3)과 같이  $f_2(X)$  모델로 학습에서 발생한 에러( $Error_2$ )를 합쳐 나머지 학습을 진행하고 이를 반복하여 여러 개의 모델( $f_1(X), f_1(X), f_1(X), \dots, f_T(X)$ )을 사용해 목적 함수의 Loss를 줄여나가는 기법이다. 최종적으로 식(4)처럼 여러 개의 Weak Learner들로 구성된 모델( $f_1(X), f_1(X), f_1(X), \dots, f_T(X)$ )들로 하나씩 학습을 진행하는 구조를 갖는다(Chen et al., 2015).

$$y = f_1(X) + Error_1 \quad (2)$$

$$Error_2 = Error_1 - f_2(X) \quad (3)$$

$$y = f_1(X) + f_2(X) + \dots + f_T(X) + Error_T \quad (4)$$

### 2.3.3 lightGBM

LightGBM(Light Gradient Boosting Machine)은 알고리즘 구조가 깊은 XGBoost의 학습 시간이 오래 걸린다는 한계점을 보완하기 위해 Microsoft에서 고안한 GBDT(Gradient Boosted Decision Trees)

기반 모델이다(Ke et al., 2017). GBDT는 다른 부스팅 알고리즘과 마찬가지로 회귀 및 분류에 사용되며 약한 예측 모델의 양상불 형태로 예측 모델을 제공한다.

LightGBM은 주어진 데이터가  $X=(x_i, y_i)_{i=1}^n$  이라고 할 때, 식(8)과 같이 특정 손실 함수  $L(y, f(x))$  의 기댓값을 최소화하는 특정 함수  $f(x)$  에 대한 근사치  $\hat{f}(x)$  를 찾는 것을 목표로 한다. 그리고 식(9)처럼 다수의  $T$  회귀 트리  $\sum_{t=1}^T f_t(X)$  를 통합하여 최종 모형에 근사치를 부여한다. 회귀 트리는  $w_{q(x)}, q \in 1, 2, \dots, J$  로 표현할 수 있다. 여기에서  $J$  는 잎의 수를 나타내고 트리의 결정 규칙에 대한  $qstands$  는 잎 노드의 표본 가중치를 나타내는 벡터이다. 마지막으로, LightGBM은 식(10)과 같이  $t$  단계를 거치며 학습이 이루어진다(Ke et al., 2017).

$$\hat{f} = \arg \min E_{y, X} L(y, f(x)) \quad (5)$$

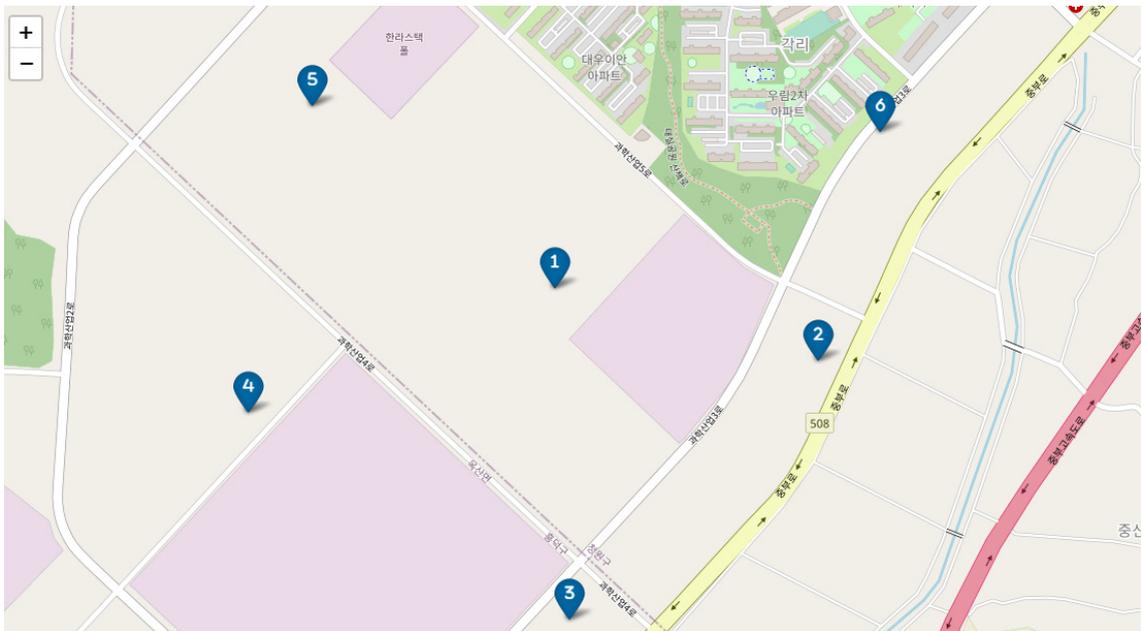
$$f_T(X) = \sum_{t=1}^T f_t(X) \quad (6)$$

$$\gamma_t = \sum_{i=1}^n (y_i, F_{t-1}(x_i) + F_t(x_i)) \quad (7)$$

### 3. 연구 방법

#### 3.1 데이터 수집

본 연구에서는 양상불 기반의 악취 농도 다지역 통합 예측 모델을 개발하기 위하여 충청북도 오창에 설치된 6개의 악취 센서에서 수집되는 데이터를 확보하였다. 각 센서의 위치는 <그림 2>와 같다. 구체적으로 해당 센서는 트리메탈아민(TMA), 황화



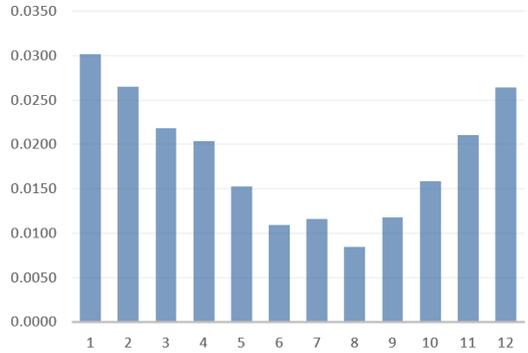
<그림 2> 6개 분석 대상 센서 위치 시각화(충청북도 오창읍)

수소(H<sub>2</sub>S), 암모니아(NH<sub>3</sub>), 에탄올(ETHANOL), 메탄에티올(CH<sub>3</sub>SH), 이산화질소(NO<sub>2</sub>), 이산화황(SO<sub>2</sub>), 일산화탄소(CO), 휘발성유기화합물(VOCs) 등 9종의 악취물질에 대한 측정값을 10분 단위로 저장한다. 분석에 사용한 데이터는 2019년 1월 1일부터 2023년 4월 12일까지 약 4년 4개월간 수집된 1,010,749개(지역별 168,458개, 6개 지역) 데이터를 대상으로 하였다.

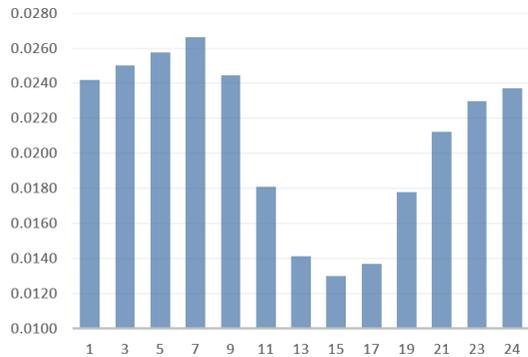
도심에서 악취를 유발하는 물질은 대기 중에 부유하기 때문에 시간에 따라 서서히 증감하는 특성이 있다. 또한, 온도와도 밀접한 관계가 있는데 일반적으로 악취물질은 26~30℃에서 강한 영향을 미치며, 온도가 낮아질수록 악취의 세기가 감소하는 경향이 있다. 그리고 습도의 영향도 커서 60~80℃의 상대습도에서 악취에 더 민감하게 반응한다. 따라서, 본 연구에서는 악취 수준 예측을 위해 시간 데이터와 기상 데이터를 수집 및 생성하였다.

기상 데이터는 한국 기상청에서 운영하는 기상자료 개방 포털에서 Open API로 개방한 방재 기상관측(AWS) 자료를 활용하였다. 제공하는 데이터는 기온, 풍속, 풍향, 강수량, 습도, 이슬점 온도, 운량, 시정거리 등 기상 정보를 포함한다. 또한 악취 데이터와 결합하기 위해 1분 단위로 수집된 기상 데이터를 10분 단위로 변환하여 1,010,749개 데이터로 맞추었다.

그 다음, 악취의 특성을 분석하여 시계열 정보와 상관성을 확인하였다. 악취물질은 상대적으로 온도가 낮고 습도가 높을수록 농도가 진한 특성을 띠는 것을 확인하였다. 온도와 습도는 계절과 시간대 와도 밀접한 연관성이 있기 때문에 계절에 따른 평균 악취와 시간에 따른 평균 악취를 분석하였다.



〈그림 3〉 계절에 따른 평균 TMA 농도



〈그림 4〉 시간에 따른 평균 TMA 농도

실제로 <그림 3>을 보면 기온이 낮은 12월, 1월, 2월이 기온이 높은 6월, 7월, 8월에 비해 평균 TMA 농도가 약 두 배 이상 높은 것을 확인할 수 있다. 또한, <그림 4>를 보면 낮 시간대(11시~19시)에 비해 밤 시간대(21시~03시)와 새벽 시간대(05시~09시)에 평균 TMA 농도가 약 두 배 이상 높은 것을 확인할 수 있다. 따라서 시간 데이터는 시간 정보와 악취 데이터 간의 상관분석 결과를 반영하여 6개 구간으로 구분하는 Hour 변수(02시~06시, 06시~10시, 10시~14시, 14시~18시, 18시~22시)와 3개 구간으로 구분하는 Month 변수(겨울 : 11월~

2월, 여름 : 6월-9월, 간절기 : 3월-5월, 10월), 주말 여부 변수(N:토, 일, Y:월, 화, 수, 목, 금) 등 3개 파생 변수를 생성하였다. 특히, 본 연구에서는 Hour 변수, Month 변수를 만들 때, 통상적인 구분 방식을 사용하지 않고 악취의 특성을 반영할 수 있도록 데이터 기반으로 적절하게 범주화하였다.

<표 2>는 수집 및 가공한 데이터셋에 대한 정리 표이다. 구체적으로 6개 지역별 10분 단위로 수집한 악취 센서 데이터 10종, 기상 관측 데이터 5종, 시간 데이터 3종으로 총 18종 변수로 구성하였다.

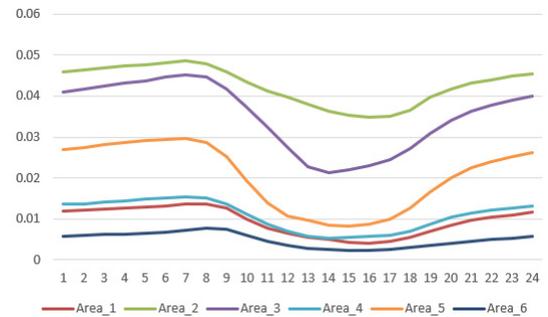
<표 2> 분석 데이터셋 변수

구분	변수명	비고
악취 센서 데이터	OU	복합악취 농도
	H2S	황화수소 농도
	NH3	암모니아 농도
	VOCS	휘발성유기화합물 농도
	ETHANOL	에탄올 농도
	CH3SH	메탄에티올 농도
	CO	일산화탄소
	SO2	이산화황
	NO2	이산화질소
	기상 관측 데이터	Temp
Precipitation		강수량
Wind_dir		풍향
Wind_speed		풍속
Humid		습도
시간 데이터	Month_C	범주화한 계절
	Hour_C	범주화한 시간
	week_C	주말 여부
목표 변수(Y)	TMA	트리메틸아민 농도

양상불 기반의 악취 농도 다지역 통합 예측 모델의 목표 변수(Y)는 TMA로 지정하였으며, 예측하려는 시점에서 10분 단위로 총 6칸(60분)을 예측하기 위해 10분 후, 20분 후, ..., 60분 후까지의 트리메틸아민(TMA) 농도로 설정하였다.

### 3.2 다지역 통합 예측 모델 개발을 위한 데이터 표준화

악취의 농도는 센서가 설치된 지역의 공간적 요인에 따라 차이가 발생할 수 있다. 예를 들어 주변에 녹지가 조성된 장소에서는 상대적으로 평균 악취 수준이 낮은 반면에, 환경사업소 또는 공장, 고속도로 근접 지역은 평균 악취 수준이 매우 높다. 그렇기에 일반적으로 특정 지역을 대상으로 악취 농도 예측 모델을 개발한다면 다른 지역에서 활용하기에는 적절하지 않다.



<그림 5> 지역별 시간대에 따른 평균 TMA 농도

분석에 활용하는 6개 지역에서도 공간적 요인에 차이가 있으며 <그림 5>와 같이 이에 따른 평균 TMA 농도의 차이가 존재한다. 구체적으로 <그림 2>에서 3번 지역은 환경사업소가 인접해 있어 다른 지역보다 평균 악취 수준이 심하며 2번

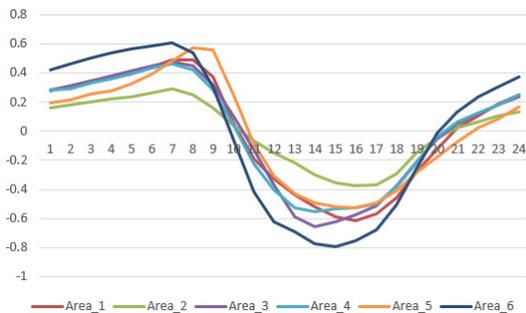
지역의 경우 차도 옆에 위치하여 평균 악취 수준이 심한 편이다. 반면 6번 지역의 경우 근처에 공원 등 녹지가 조성되어 있어 악취 수준이 상대적으로 낮다. 따라서 일반적인 방법으로 모델을 설계한다면 악취 농도 예측 모델의 정확도는 떨어질 수밖에 없으며, 지역별로 독립적인 악취 농도 예측 모델을 설계하여야 한다.

본 연구에서는 지역마다의 악취 데이터의 범위를 일치시키기 위해 표준화 알고리즘(Standard Scaler)을 적용하였다. 모든 지역의 악취 농도는 범위의 차이는 있으나 시간대별로 변화하는 악취 농도의 변화율은 일정하기 때문에 표준화 작업을 진행함으로써 악취 농도의 차이를 제거할 수 있다.

구체적으로 악취 농도를 일치시키기 위해 사용한 표준화 알고리즘은 식(8)과 같다.

$$x_{scaled} = \frac{x - \bar{x}}{\sigma} \quad (8)$$

표준화 알고리즘은 실제 값을 평균값으로 빼 다음 표준편차로 나눠서 평균이 0, 표준편차가 1이 되도록 데이터의 범위를 가공한다. 표준화 알고리즘을 적용함으로써 모든 지역의 평균과 표준편차를 일치시킴으로써 데이터의 범위를 균일화할 수 있다.



<그림 6> 표준화한 지역별 시간대에 따른 평균 TMA 농도

실제로 표준화한 지역별 시간대에 따른 평균 TMA의 농도는 <그림 6>과 같다. <그림 6>을 보면 <그림 5>와 달리 지역과 상관없이 시간대에 따라 비슷한 수치를 나타내는 것을 확인할 수 있으며 표준화한 데이터는 하나의 모델로 학습하기에 용이하다는 것을 알 수 있다. 따라서 본 연구에서는 데이터를 표준화함으로써 다지역 통합 예측 모델을 개발할 수 있으며, 지역별 단일 모델을 만드는 것(1개 지역 데이터 수 : 168,458개)에 비해 학습에 사용하는 데이터(6개 지역 데이터 수 : 1,010,749개)를 6배 이상 늘릴 수 있기 때문에 모델의 성능도 높일 수 있다.

## 4. 연구 결과

### 4.1 앙상블 모델 학습

본 연구에서는 10분 단위로 한 시간 후까지 총 6개 미래 시점의 악취 농도를 예측하는 모델을 설계하였다. 우선, 수집한 악취 데이터 1,010,749개를 7:2:1 비율로 학습데이터(709,540개), 검증데이터(200,132개), 시험데이터(101,077개)로 분리하였다. 다만, 본 연구에 활용하는 데이터셋은 6개 지역을 포괄하기 때문에 지역별로 데이터를 나누고 취합하였다.

앙상블 기반의 악취 농도 다지역 통합 모델을 학습하기 위해 랜덤포레스트, XGBoost, lightGBM 3종의 알고리즘을 활용하였으며 <표 3>과같이 모델별 하이퍼 파라미터를 설정하였다.

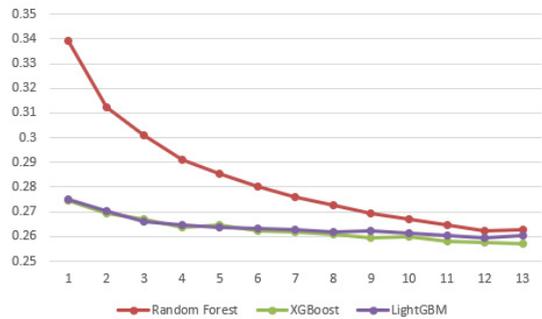
<표 3> 알고리즘 별 사용한 하이퍼 파라미터

Algorithm	Hyper parameter
Random Forest	<ul style="list-style-type: none"> <li>• n_estimators = 100</li> <li>• min_samples_leaf: 32</li> <li>• min_samples_split: 32</li> </ul>
XGboost	<ul style="list-style-type: none"> <li>• eta = 0.2</li> <li>• n_estimators = 100</li> <li>• num_leaves = 31</li> <li>• subsample = 0.5</li> </ul>
LightGBM	<ul style="list-style-type: none"> <li>• n_estimators = 100</li> <li>• num_leaves = 31</li> <li>• subsample = 0.5</li> <li>• boost_from_average = True</li> </ul>

또한, 양상블 모델의 하이퍼 파라미터 중 트리의 최대 깊이를 설정하는 Max Depth는 3에서 15까지 1씩 증가시키면서 모델의 성능을 확인하고 가장 우수한 Max Depth를 선정하였다.

<표 4> Max Depth별 RMSE 표

max depth	Random Forest	XGB	LGBM
5	0.0907	0.0712	0.0707
6	0.0849	0.0696	0.07
7	0.0814	0.0701	0.0695
8	0.0786	0.0687	0.0692
9	0.0761	0.0686	0.0691
10	0.0743	0.068	0.0686
11	0.0726	0.0674	0.0689
12	0.0713	0.0675	0.0683
13	0.0701	0.0667	0.0677
14	0.0689	0.0663	0.0673
15	0.0691	0.0662	0.0679



<그림 7> Max Depth별 RMSE 변화 그래프

<표 4>와 <그림 7>을 보면 XGBoost가 Max Depth=15일 때, RMSE 값이 0.0662로 가장 낮았으며 lightGBM이 Max Depth=14일 때 0.0673, Random Forest가 Max Depth=14일 때 0.0689로 좋았다. 또한, 세 개의 모델 모두 Max Depth가 15보다 커지면 과적합이 발생하였다. 따라서 알고리즘 별 Max Depth Size는 각각 14, 15, 14로 설정하였다.

#### 4.2 다지역 통합 예측 모델 성능 평가

다음으로, 본 연구에서 제안하는 표준화 알고리즘을 적용한 약취 농도 다지역 통합 모델의 성능을 확인하기 위해 단일 지역 모델과의 성능을 비교하였다. 단일 지역 모델은 <그림 2>에서 가운데에 있는 3번 지역으로 선정하였으며, 3번 지역 데이터셋으로 학습한 후 3번 지역을 포함한 6개 지역의 약취 농도를 예측하도록 모델을 설계하였다. 구체적으로 통합 모델과 단일 모델 모두 미리 준비한 학습 데이터와 검증 데이터를 활용하여 모델을 학습하였으며, 성능 비교는 학습에 쓰이지 않은 101,077개의 시험 데이터를 이용하였다. 단일 지역 모델 학습에 사용한 알고리즘은 다지역 통합 모델에 사용한 것과 동일하게 Random Forest,

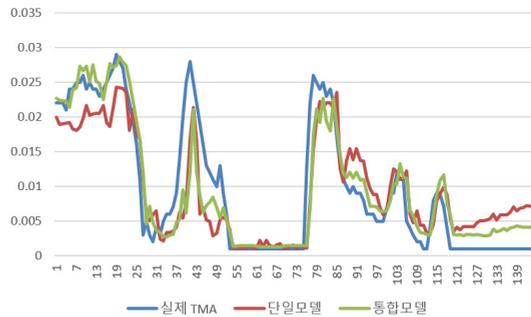
XGBoost, lightGBM 3가지 모델을 활용하였다. 또한, 하이퍼 파라미터 역시 위에서 선정한 값과 같은 값으로 설정하였다. 모델을 비교하기 위한 성능지표로는 RMSE를 사용하였으며 RMSE는 0에 가까울수록 오차가 작다는 것을 의미한다.

〈표 5〉 단일 지역 모델과 통합모델 성능 비교

구분	사용 알고리즘	RMSE
단일지역 모델	Random Forest	0.01689
	XGBoost	0.01459
	lightGBM	0.01563
다지역 통합 모델	Random Forest	0.01490
	XGBoost	0.00960
	lightGBM	0.00973

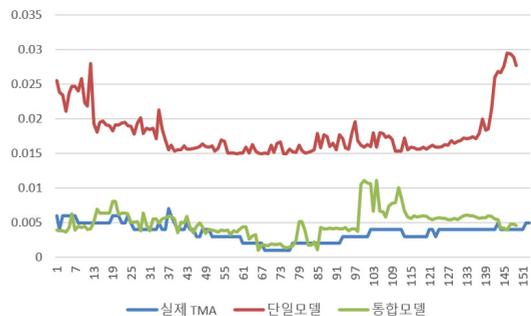
〈표 5〉는 10분 단위로 한 시간 후까지 악취 농도를 예측하는 단일 지역 모델과 다지역 통합 모델의 성능 결과표이다. 단일 지역 모델과 다지역 통합 모델 모두 XGBoost의 RMSE 값이 0.01459, 0.00960으로 가장 우수하였다. 두 모델의 RMSE를 비교한 결과 단일 지역 모델(0.01459)보다 다지역 통합 모델(0.0096)이 0.00499 낮았으며, 상대적으로 RMSE 값이 51.9% 낮은 것을 확인할 수 있었다.

그리고 자세하게 비교하기 위하여 실제 악취 농도와 모델별 예측한 악취 농도를 비교하였다. 실제 데이터 비교는 단일 지역 학습에 사용한 6번 지역과 학습에 사용하지 않은 5번 지역을 대상으로 비교하였으며 이를 통해 단일 지역 모델이 다른 지역을 예측한 결과와 통합 모델이 예측한 결과를 비교함으로써 본 연구에서 제안하는 모델의 성능을 검증하고자 하였다.



〈그림 8〉 3번 지역 실제값과 예측값 비교

〈그림 8〉은 3번 지역의 2023년 4월 11일, 실제 측정된 악취 농도와 두 모델의 악취 농도 예측값을 시각화한 그래프이다. 그래프를 확인해보면 통합모델의 예측 결과가 단일 지역 모델보다 실제 TMA 농도와 근접한 것을 볼 수 있다. 즉, 해당 지역의 데이터로 학습하는 것보다 표준화 알고리즘을 이용하여 다른 지역의 데이터까지 학습데이터에 포함하여 데이터의 양을 늘리는 것이 정확도가 더 높아진다는 것을 의미한다.



〈그림 9〉 6번 지역 실제값과 예측값 비교

〈그림 9〉는 단일 지역 모델의 학습에 사용하지 않은 6번 지역의 2023년 4월 11일, 실제 측정된 악취 농도와 두 모델의 악취 농도 예측값을

시각화한 그래프이다. 해당 그래프를 확인해 본 결과, 단일 지역 모델에서는 전체적으로 실제 TMA보다 매우 높은 값으로 예측하였다. 이는 실제로 인접한 지역임에도 불구하고 5번 지역의 공간적 요인으로 6번 지역에 비해 평균 악취 수준이 높아 모델이 제대로 작동하지 않았다고 볼 수 있다. 즉, 동일한 기상 환경 조건을 갖더라도 악취 농도의 범위가 달라 같은 모델로 예측하는 것이 불가능하다는 것을 의미한다.

반면, 통합 모델의 경우, 모델 학습 이전에 종속 변수 TMA 농도의 범위를 평균이 0, 표준편차가 1이 되게끔 표준화하여 6번 지역에서도 모델의 성능이 우수하였다.

결론적으로 서로 다른 지역에서 수집한 악취 농도 데이터를 표준화한 후 다지역 통합 예측 모델을 설계함으로써 모델 학습에 사용할 수 있는 데이터의 양이 늘어남으로써 정확도를 향상시키고 또한, 여러 지역을 하나의 통합 모델로 예측이 가능하다는 것을 확인하였다.

## 5. 결론

전 세계적으로 대기오염 문제가 지속적으로 증가하고 있으며, 도심에서 악취의 무분별한 유입은 환경소음과 함께 가장 큰 민원의 대상이 되고 있다. 악취에 대한 문제는 두통, 구토 등 불쾌감과 혐오감을 유발하여 지역 주민의 건강과 직결되는 문제이기 때문에 악취를 모니터링하고 관리하는 방안이 필요하다. 따라서 본 연구에서는 하나의 모델로 여러 지역의 악취 농도를 예측할 수 있는 양상불 기반의 악취 농도 다지역 통합 예측 모델을 설계하고 방안을 제안하였다.

악취 농도 다지역 통합 예측 모델을 개발하기

위하여 2019년 1월 1일부터 2023년 4월 12일까지 충청북도 오창에 설치된 6개의 악취 센서에서 수집된 1,010,749개 데이터를 활용하였으며, 기상청 API를 통해 악취에 영향을 미치는 기상 데이터도 확보하였다.

Random Forest, XGBoost, lightGBM 등 3개의 양상불 알고리즘을 활용하여 10분 단위로 1시간 미래를 예측하는 모델을 설계한 결과, XGBoost의 Max Depth가 15일 때 RMSE가 0.2572로 가장 우수하였다. 또한, 본 연구에서 제안하는 악취 농도 다지역 통합 예측 모델의 성능을 확인하기 위해 단일 지역 모델과 비교한 결과, 단일 지역 모델(0.01459)보다 다지역 통합 모델(0.0096)의 RMSE가 0.00499 낮았으며, 상대적으로 RMSE 값이 51.9% 낮은 것을 확인할 수 있었다. 그뿐만 아니라 학습에 사용한 지역 외에는 제대로 된 예측을 할 수 없는 단일 지역 모델과 달리 다지역 통합 모델의 경우 6개 지역 모두 안정적인 예측 성능을 보인 것을 확인하였다.

본 연구에서는 지역마다 악취의 평균 수준이 다르더라도 표준화 기법을 통해 악취 수준의 범위를 평균 0, 표준편차가 1인 분포로 적합 시킴으로써 서로 다른 지역의 데이터를 하나의 데이터로 통합하고 통합 모델을 설계하는 방안을 제안하였다. 제안한 방법은 학습에 사용할 수 있는 데이터의 양적 크기를 키워 학습한 모델의 성능을 높일 수 있었으며, 지역마다 독립적인 학습 모델을 만드는 기존 방법과 달리 단 하나의 모델로 여러 지역을 예측하는 것이 가능하다는 데에 의의가 있다. 본 연구 모델을 활용한다면 도심의 악취를 모니터링하고 관리하는 데에 효과적인 성능을 보일 수 있을 것이라 기대한다.

다만, 연구에 활용한 지역은 센서 간 최대 직선거리가 약 1.5km로 인접한 위치에 있어 지역

간 악취 농도 사이에 상관성이 있었으며, 다지역 통합 모델 성능의 신뢰성을 확보하기에 한계가 있었다. 따라서, 향후 연구에서는 충청북도 오창 뿐만 아니라 더 많은 지역을 대상으로 연구를 진행하고 모델의 성능을 확인하고자 한다.

## 참고문헌(References)

### [국내 문헌]

- 김병욱, 현근우, 배선학, 홍영균, 이영섭, 이건호, 최승봉. (2020). 악취의 시공간적 발생 특성 및 분포도 분석-강원지역을 대상으로, *한국환경보건학회지*, 46(4), 376-387. DOI : 10.5668/JEHS.2020.46.4.376
- 김영훈, 최홍식, 김선웅.(2020). XGBoost를 활용한 리스크패리티 자산배분 모형에 관한 연구. *지능정보연구*, 26(1), 135-149.
- 도우곤, 정우식. (2020). 도시공단지역 복합 악취 배출원에서 발생하는 악취물질별 기여율 분석, *한국대기환경학회지*, 36(3), 360-374. DOI : 10.5572/KOSAE.2020.36.3.360
- 안지원, 오경용, 박신영, 이기환, 김용휘, 김기현. (2007). 음식물의 부패정도에 따른 악취물질의 발생특성 연구: 트리메틸아민, 암모니아를 중심으로, *한국냄새환경학회지*, 6(3), 172-182.
- 박진욱, 최일수, 나명환. (2016). 공간 자료를 이용한 대기오염이 순환기계 건강에 미치는 영향 분석. *품질경영학회지*, 44(3), 677-688. DOI : 10.7469/JKSQM.2016.44.3.677
- 서형준, 이형석. (2019). 대기오염물질이 환경성 질화에 미치는 영향: 수도권 지역의 차치구 및 시·군을 중심으로. *서울도시연구*, 20(3), 39-59. DOI : 10.23129/seouls.20.3.201909.39

엄하늘, 김재성, 최상욱. (2020). 머신러닝 기반 기업부도위험 예측모델 검증 및 정책적 제언: 스테킹 앙상블 모델을 통한 개선을 중심으로. *지능정보연구*, 26(2), 105-129.

유지현. (2020). 미세먼지 확산 모델링을 이용한 대기질 예측 시스템에 대한 연구, *전기전자학회논문지*, 24(4), 216-220. DOI : 10.7471/ikeee.2020.24.4.1136

윤주상, 이태진. (2022). 옛지 시스템을 위한 LSTM 기반 화재 및 악취 예측 모델, *정보처리학회 논문지*, 컴퓨터 및 통신시스템, 11(2), 67-72.

이경수, 박예린, 신윤중, 손권상, 권오병. (2022). 효율적 수입식품 검사를 위한 머신러닝 기반 부적합 건강기능식품 탐지 방법. *지능정보연구*, 28(3), 139-159.

이종영, 최명진, 주영인, 양재경. (2019). 미세먼지, 악취 농도 예측을 위한 앙상블 방법, *산업경영시스템학회지*, 42(4), 203-210. DOI : 10.11627/jkise.2019.42.4.203

황두환, 조경근. (2021). 머신러닝 기반의 돈사 악취농도 예측(황화수소 농도를 중심으로), *실내환경 및 냄새학회지*, 20(4), 311-321.

### [국외 문헌]

- Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., & Zhou, T. (2015). Xgboost: extreme gradient boosting. *R package version 0.4-2*, 1(4), 1-4. <https://doi.org/10.1145/2939672.2939785>
- Conti, C., Guarino, M., & Bacenetti, J. (2020). Measurements techniques and models to assess odor annoyance: A review. *Environment international*, 134, 105261. <https://doi.org/10.1016/j.envint.2019.105261>
- Das, S., Lee, S. H., Kumar, P., Kim, K. H., Lee, S. S., & Bhattacharya, S. S. (2019). Solid waste

- management: Scope and the challenge of sustainability. *Journal of cleaner production*, 228, 658-678. <https://doi.org/10.1016/j.jclepro.2019.04.323>
- Invernizzi, M., Capra, F., Sozzi, R., Capelli, L., & Sironi, S. (2021). Development and Evaluation of a Fluctuating Plume Model for Odor Impact Assessment. *Applied Sciences*, 11(8), 3310. <https://doi.org/10.3390/app11083310>
- Jiang, Y., Huang, J., Luo, W., Chen, K., Yu, W., Zhang, W., & Huang, Y. (2023). Prediction for odor gas generation from domestic waste based on machine learning. *Waste Management*, 156, 264-271. <https://doi.org/10.1016/j.wasman.2022.12.006>
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30.
- Kiliyankil, V. A., Fugetsu, B., Sakata, I., Wang, Z., & Endo, M. (2021). Aerogels from copper (II)-cellulose nanofibers and carbon nanotubes as absorbents for the elimination of toxic gases from air. *Journal of Colloid and Interface Science*, 582, 950-960. <https://doi.org/10.1016/j.jcis.2020.08.100>
- Kim, G., Lee, M. I., Lee, S., Choi, S. D., Kim, S. J., & Song, C. K. (2020). Numerical modeling for the accidental dispersion of hazardous air pollutants in the urban metropolitan area. *Atmosphere*, 11(5), 477. <https://doi.org/10.3390/atmos11050477>
- Lan, K., Liu, S., Wang, Z., Wei, J., & Qin, G. (2023). High-performance olfactory receptor-derived peptide sensor for trimethylamine detection on the pyramid substrate structure. *Sensors and Actuators A: Physical*, 114452. <https://doi.org/10.1016/j.sna.2023.114452>
- Liu, Z. (2021). Urban sewage treatment odor gas release characteristics and regional differences. *Environmental Technology & Innovation*, 21, 101190. <https://doi.org/10.1016/j.eti.2020.101190>
- Lu, J. G. (2020). Air pollution: A systematic review of its psychological, economic, and social effects. *Current opinion in psychology*, 32, 52-65. DOI: 10.1016/j.copsyc.2019.06.024
- Noel, C., Vanroelen, C., & Gadeyne, S. (2021). Qualitative research about public health risk perceptions on ambient air pollution. A review study. *SSM-Population Health*, 15, 100879. <https://doi.org/10.1016/j.ssmph.2021.100879>
- Phyo, S., Kim, S., Lee, H., Han, S. S., Oh, Y., & Lee, J. (2023). Adsorption enhancement of hazardous odor gas using controlled thermal oxidation of activated carbon. *Journal of Cleaner Production*, 393, 136261. <https://doi.org/10.1016/j.jclepro.2023.136261>
- Rigatti, S. J. (2017). Random forest. *Journal of Insurance Medicine*, 47(1), 31-39. DOI: 10.17849/inism-47-01-31-39.1
- Sakawi, Z., Sharifah, S. A., Jaafar, O., & Mahmud, M. (2011). Community perception of odor pollution from the landfill. *Research. Journal of Environmental and Earth Sciences*, 3(2), 142-145.
- Seok, H. J., Kahn, J. B., Kim, Y. L., Seo, J. W., Hong, S. Y., & Kim, H. J. (2020). Discharge impact analysis of air pollutants for integrated environmental management. *Journal of Environmental Analysis, Health and Toxicology*, 23(4), 240-247. DOI: <https://doi.org/10.36278/jeaht.23.4.240>
- Wang, Y., Lü, F., Kang, X., Xu, X., Chen, W., Chai, H., & He, P. (2022). Odor characteristics and

- health risks during food waste bioconversion by housefly (*Musca domestica* L.) larvae. *Journal of Cleaner Production*, 376, 134343. <https://doi.org/10.1016/j.jclepro.2022.134343>
- Xu, A., Li, R., Chang, H., Xu, Y., Li, X., Lin, G., & Zhao, Y. (2022). Artificial neural network (ANN) modeling for the prediction of odor emission rates from landfill working surface. *Waste Management*, 138, 158-171. <https://doi.org/10.1016/j.wasman.2021.11.045>
- Zhao, S., Yang, X., Zhang, W., Chang, J., & Wang, D. (2019). Volatile sulfide compounds (VSCs) and ammonia emission characteristics and odor contribution in the process of municipal sludge composting. *Journal of the Air & Waste Management Association*, 69(11), 1368-1376. <https://doi.org/10.1080/10962247.2019.1629356>
- Zhou, Z. H., & Zhou, Z. H. (2021). Ensemble learning . *Springer Singapore*. 181-210.
- 장동열. (2019, March 15). 청주 오창 ‘14년 악취 미스터리’ 이번엔 풀릴까. 뉴스1. <https://www.news1.kr/articles/?3579333>
- 대기관리과. (2023, 2, 27). 악취관리지역 (2023년 01월 기준). 환경부. Retrieved March 8, 2023 from [http://me.go.kr/home/web/public\\_info/read.do;jsessionid=4gh7NFmkrxf09CK9XzBHlwEb.mehome1?pagerOffset=20&maxPageItems=10&maxIndexPages=10&searchKey=&searchValue=&menuId=10357&orgCd=&condition.publicInfoMasterId=2&condition.deleteYn=N&publicInfoId=24&menuId=10357](http://me.go.kr/home/web/public_info/read.do;jsessionid=4gh7NFmkrxf09CK9XzBHlwEb.mehome1?pagerOffset=20&maxPageItems=10&maxIndexPages=10&searchKey=&searchValue=&menuId=10357&orgCd=&condition.publicInfoMasterId=2&condition.deleteYn=N&publicInfoId=24&menuId=10357)
- 미래정책연구실. (2021. 5, 17). 주간 농업농촌 식품 동향\_Vol.19\_2021.5.17. 한국농촌경제연구원. Retrieved March 10, 2023 from <https://www.krei.re.kr/krei/selectBbsNttView.do?bbsNo=76&key=271&nttNo=140969>

## [URL]

- 공부주, 강경희, 조강남, 김정훈, 박정민, 강대일, 김나래, 김희진, 유정웅, 김대곤. (2018). 현장 후각 측정법을 이용한 악취측정법 개선(II). 국립환경과학원. Retrieved January 10, 2023, from <https://scienceon.kisti.re.kr/commons/util/originalView.do?dbt=TRKO&cn=TRKO201900002639>
- 국립환경과학원. (2018. 12). 제 2차 악취방지 종합시책 수립을 위한 연구. 국립환경과학원. Retrieved January 10, 2023 from <https://scienceon.kisti.re.kr/commons/util/originalView.do?dbt=TRKO&cn=TRKO201900002638>
- 대기관리과. (2021, 1. 5.). 악취방지법. 환경부. Retrieved March 8, 2023 from <https://www.law.go.kr/LSW/lsInfoP.do?efYd=20210105&lsiSeq=228123#0000>

Abstract

## Development of an Ensemble-Based Multi-Region Integrated Odor Concentration Prediction Model

Seong-Ju Cho\* · Woo-seok Choi\* · Sang-hyun Choi\*\*

Air pollution-related diseases are escalating worldwide, with the World Health Organization (WHO) estimating approximately 7 million annual deaths in 2022. The rapid expansion of industrial facilities, increased emissions from various sources, and uncontrolled release of odorous substances have brought air pollution to the forefront of societal concerns. In South Korea, odor is categorized as an independent environmental pollutant, alongside air and water pollution, directly impacting the health of local residents by causing discomfort and aversion. However, the current odor management system in Korea remains inadequate, necessitating improvements.

This study aims to enhance the odor management system by analyzing 1,010,749 data points collected from odor sensors located in Osong, Chungcheongbuk-do, using an Ensemble-Based Multi-Region Integrated Odor Concentration Prediction Model. The research results demonstrate that the model based on the XGBoost algorithm exhibited superior performance, with an RMSE of 0.0096, significantly outperforming the single-region model (0.0146) with a 51.9% reduction in mean error size. This underscores the potential for increasing data volume, improving accuracy, and enabling odor prediction in diverse regions using a unified model through the standardization of odor concentration data collected from various regions.

**Key Words** : Odor, Trimethylamine, Integrated Model, Machine Learning, Ensemble

Received : September 6, 2023 Revised : September 19, 2023 Accepted : September 21, 2023

Corresponding Author : Sang-Hyun Choi

---

\* Dept. of Bigdata, Chungbuk National University  
\*\* Corresponding Author: Sang-Hyun Choi  
Dept. of MIS/Bigdata, Chungbuk National University  
1, Chungdae-ro, Seowon-gu, Cheongju-si, Chungcheongbuk-do, Korea  
Tel: \*\*\*-\*\*\*\*-\*\*\*\* E-mail: chois@cbnu.ac.kr

## 저 자 소개



**조성주**

현재 충북대학교 빅데이터 협동과정 석사과정에 재학 중이다. 주요 연구 분야는 Data Science, Machine Learning, Deep Learning 등이다.



**최우석**

현재 충북대학교 빅데이터 협동과정 박사과정에 재학 중이며, 주요 연구 분야는 Time Series Analysis, Text Analysis, Image Analysis, Deep Learning 등이다.



**최상현**

현재 충북대학교 경영정보학과와 빅데이터 협동과정 교수로 재직 중이다. KAIST에서 경영정보공학을 전공으로 박사학위를 취득하였으며, 주요 연구 분야는 Business Process, Data Analysis, Machine Learning, Deep Learning 등이다.