



ISSN: 3022-5388

JKAI website: <https://accesson.kr/jkaia>DOI: <http://dx.doi.org/10.24225/jkaia.2023.1.2.15>

머신러닝을 활용한 내부 발생 요인 기반의 미세먼지 예측에 관한 연구

A Study on Fine Dust Prediction Based on Internal Factors Using Machine Learning

Yong-Joon KIM¹, Min-Soo KANG²

Received: October 28, 2023. Revised: November 16, 2023. Accepted: December 30, 2023.

Abstract

This study aims to enhance the accuracy of fine dust predictions by analyzing various factors within the local environment, in addition to atmospheric conditions. In the atmospheric environment, meteorological and air pollution data were utilized, and additional factors contributing to fine dust generation within the region, such as traffic volume and electricity transaction data, were sequentially incorporated for analysis. XGBoost, Random Forest, and ANN (Artificial Neural Network) were employed for the analysis. As variables were added, all algorithms demonstrated improved performance. Particularly noteworthy was the Artificial Neural Network, which, when using atmospheric conditions as a variable, resulted in an MAE of 6.25. Upon the addition of traffic volume, the MAE decreased to 5.49, and further inclusion of power transaction data led to a notable improvement, resulting in an MAE of 4.61. This research provides valuable insights for proactive measures against air pollution by predicting future fine dust levels.

Keywords : Fine dust, internal factors, prediction, machine learning

Major Classification Code : Artificial Intelligence, etc

1. Introduction

현대 사회에서 화석 연료 사용, 공장 배출가스, 자동차 매연 등으로 인해 대기 오염이 급속하게 증가하고 있다. 이에 따라 국내외에서 대기 오염에 관한 관심이 계속 증가하고 있다. 세계보건기구에 따르면, 전 세계 인구의 92%가 대기 오염으로 인해 영향을 받고 있다고

발표했으며, 해마다 수많은 사람들이 이에 따른 직접적인 피해를 받고 있다고 강조하고 있다(BBC NEWS Korea, 2018). 최근 대기 오염으로 인한 피해가 문제로 대두되면서 대기 오염의 주요 물질 중 미세먼지에 관한 관심이 매우 높아지고 있다. Google Trend로 미세먼지 검색에 따른 관심도 변화를 확인해보면 최근 1년 2022년부터 2023년 기준으로 검색도가 3배 이상 증가된 것을 확인할 수

1 First Author. Undergraduate Student at Eulji University, Medical IT, Republic of Korea, Email: dydwms9118@g.eulji.ac.kr

2 Corresponding Author. Professor at Eulji University, Medical IT, Republic of Korea, Email: mskang@eulji.ac.kr

© Copyright: The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

있으며 주변을 보면 미세먼지 상황을 측정하여 알려주는 미세먼지 알리미 신호등, 미세먼지 알림 어플리케이션 등 사용이 증가되고 있는 것을 알 수 있다. 미세먼지는 세계보건기구에서 1군 발암물질로 지정하였으며, 미세먼지의 농도가 높아짐에 따라 천식이나 기관지염과 같은 호흡기 질환뿐만 아니라 심장박동수 이상 및 심혈관계 질환, 피부질환의 위험요인이 되고 있다. 미세먼지의 악영향은 건강 문제뿐만 아니라 사회 곳곳에 퍼지며, 반도체와 같은 초정밀 고순도를 유지해야 하는 첨단 산업의 공정과 먼지가 존재하지 않는 환경에서 생산해야 하는 주사제의 생산에도 영향을 끼친다. 또한, 미세먼지의 농도가 심각한 날에는 교통수단의 가시거리가 확보되지 않아 지연, 결항, 취소되기도 한다. 미세먼지의 발생은 이와 같이 사회 및 경제적으로 다양한 분야에 부정적인 영향을 미치고 있다(Kim & Moon, 2021). 이러한 미세먼지의 주요 원인으로는 중국을 비롯한 주변 국가들의 영향이 크게 작용하고 있다. 매년 12월부터 봄철인 5월까지, 한국 전역에서는 미세먼지와 같은 인체에 해로운 입자들이 증가하는 현상이 일어난다. 중국을 비롯한 주변 국가들에서 발생한 미세먼지들은 해풍과 계절풍 등을 타고 한국으로 넘어와 실제 한국의 미세먼지 발생률에서 큰 비중을 차지한다. 하지만 국내에서도 미세먼지는 공장, 발전소, 자동차, 선박 등에서 생성되는 대기 오염물질로 인해 발생하고 있다. 특히, 국내에서는 석유, 석탄과 같은 화석 연료가 타면서 발생하는 대기오염물질이 국내 미세먼지의 주요 원인으로 알려져 있다. 이에 국내에서도 미세먼지 오염도를 줄이기 위해 다양한 노력이 필요하다(Park & Shin, 2017). 따라서 외부의 영향에 의해 발생하는 미세먼지뿐만 아니라 내부에서 미세먼지 발생을 유발시키는 요인에도 관심을 가질 필요가 있다.

본 연구는 내부에서 미세먼지를 발생시키는 데 영향을 미치는 다양한 변수들을 분석하고 효과적인 예측을 위한 최적의 알고리즘을 제안하는 것을 목적으로 한다. 이러한 과정을 진행하기 위해서는 미세먼지 예측 모델의 성능 비교가 왜 중요한지에 대한 설명이 필요하다. 미세먼지 농도를 예측하는 것은 여러 가지 측면에서 중요한데 첫째, 미세먼지 농도 예측은 건강을 보호하는 데 중요한 역할을 한다. 공기 중의 고농도 미세먼지는 앞서 말했듯이 호흡기 문제를 포함해 심혈관 질환, 알레르기 반응 등 다양한 건강 문제와 연관이 있다. 정확한 미세먼지 예측은 개인들에게 미세먼지 농도에 대한 경고를 제공하고 외부활동을 자제하거나 호흡 보호장비를 권장하는 등의

예방조치를 취할 수 있도록 돕는다. 두 번째, 미세먼지 예측 모델은 미세먼지 농도가 위험 수준에 도달할 때 공공 건강 경보를 발령하는 데 도움을 준다. 이는 대중이 건강을 위협받는 상황을 미리 예측할 수 있고 해당 지역에서의 활동을 제한하거나 예방조치를 취할 수 있도록 안내할 수 있다. 세 번째, 정책 제정에도 영향을 미친다. 황사, 연무 등의 대기 상태를 더 정확하게 예측할 수 있는 모델이 개발된다면, 정확도가 1% 향상될 때마다 한계지불의사액(Marginal Willingness to Pay)이 48원씩 증가한다(Kim et al., 2020). 즉 개선된 미세먼지 예측 모델을 적용한다면, 보다 경제적이고 효과적인 미세먼지 관리 정책을 수립할 수 있을 것으로 기대된다. 또한, 정확한 데이터와 모델을 토대로 환경 정책 결정자들은 미세먼지 관리와 대기질 향상을 위한 효과적인 조치를 수립할 수 있다. 네 번째, 미세먼지 예측 정보는 시민들이 환경 보호에 참여하는 데 도움을 준다. 시민들은 개인적으로 미세먼지 농도에 대한 정보를 받아들이고 활동을 조정하여 건강을 보호하며, 환경 보호에 기여할 수 있다. 이와 같은 목적성을 가지고 본 연구에서는 기계학습을 통한 회귀 분석을 활용해 데이터를 분석하고, 미세먼지 농도를 예측할 수 있는 방법을 찾고자 한다. 회귀 분석은 다른 독립변수들에 의해 하나의 종속변수가 어떻게 설명되는지 알아보기 위한 통계적인 기법으로 변수 간의 관계를 파악하고 이를 기반으로 예측 모델을 구축할 수 있어 예측을 수행하는 데 유용하다(D. F. Specht, 1991). 또한, 기계학습을 활용하면 일반적인 선형 예측 모델에서 필요한 선형 가정의 제약에서 벗어날 수 있다. 이는 모델이 데이터의 비선형 관계를 더 효과적으로 학습할 수 있음을 의미하며 기계학습은 예측력이 높은 모델을 개발할 수 있어서 데이터로부터 미묘한 패턴이나 복잡한 상호작용을 감지하는 데 강점을 가진다(Lee & Jeon, 2018).

따라서 본 연구에서는 미세먼지에 영향을 미칠 수 있는 변수 중 내부에서 미세먼지를 발생시키는 변수들을 분석하여 그 변수들의 미세먼지 예측에 대한 영향력을 파악하고 알아보고자 했다. 또한, 다양한 기계학습 기법을 활용해 미세먼지 농도의 예측을 진행하고 기계학습 예측 모델 간의 예측 성능 비교를 통해 미세먼지 농도를 예측하는 데 가장 적합한 기계학습 모델을 파악했다.

2. Related Research

2.1. Machine Learning

기계학습은 컴퓨터 시스템이 데이터로부터 학습하여 경험을 쌓고, 새로운 데이터에 대해 예측이나 의사 결정을 수행할 수 있는 능력을 갖추는 분야이다. 기계학습은 인공지능의 한 분야로, 데이터와 통계적 모델링을 기반으로 패턴을 학습하여 문제를 해결하고 의사 결정을 수행하는 데 사용된다. 기계학습의 주요 단계는 학습에 필요한 데이터를 수집하고 준비, 데이터를 정제하고, 특성을 추출하거나 변환하여 모델 학습에 적합한 형태로 가공하는 데이터 전처리, 해결하고자 하는 문제의 특성에 맞는 적절한 모델을 선택, 선택한 모델을 훈련 데이터로 학습, 훈련된 모델을 평가 데이터에 적용하여 모델의 성능을 평가, 훈련된 모델을 새로운 데이터에 적용하여 예측 순으로 이루어진다. 이와 같은 단계 중 모델을 훈련 데이터로 학습시키는 단계에서 레이블이 있는 데이터는 지도학습을 적용하고 레이블이 없는 데이터는 비지도 학습을 적용한다. 지도학습이란 기계학습의 주요 패러다임 중 하나로, 훈련 데이터에 레이블이 포함되어 있어 모델이 입력과 출력 간의 매핑을 학습하는 방식이다. 즉, 주어진 입력과 그에 대응하는 출력 간의 관계를 학습하여 새로운 입력에 대한 출력을 예측하는 것이 목적이다. 지도학습은 분류와 회귀로 나뉘며, 분류란 출력이 범주형 변수인 경우로, 주어진 입력을 미리 정의된 클래스 또는 범주 중 하나로 할당하는 문제를 의미하고, 회귀란 출력이 연속형 변수인 경우로, 주어진 입력에 대해 실수 값을 예측하는 문제를 의미한다(Lee, 2022). 본 연구에서는 지도학습 기법 중 XGBoost(Extreme Gradient Boosting), 랜덤 포레스트(Random Forest), 인공신경망(Artificial Neural Network; ANN) 기법을 활용해 미세먼지 농도를 예측하고자 한다.

2.2. Factors Affecting PM Levels

미세먼지의 원인을 찾기 위한 다양한 연구가 진행되고 있다. 선행연구에 따르면 대기과 기후에 관련된 요인이 미세먼지의 주요한 원인으로 제시되고 있다. 2001년부터 2016년까지 15년간의 한반도 지역의 대기 환경자료와 기상자료를 분석하여 미세먼지와 대기 환경자료, 기상자료의 상관성을 검증하였으며 이를 통해 한반도의 미세먼지는 풍속과 풍향, 온도의 변화에 따라 영향을 받고 있음을 확인했다(Han et al., 2017). 한 연구에서는 강원도 지역의 배출량, 영향권 분석, 미세먼지 성분, 배출원 추정,

지형, 기상을 분석하여 낮은 풍속과 태백산맥의 지형 효과에 따른 대기 정체 등의 이유로 원주시의 미세먼지 고농도 현상을 설명하였다(Park et al., 2020). 또 다른 선행연구에서는 미세먼지의 원인 중 하나로 차량에서 배출되는 오염물질에 대해 연구하였다. 브라질의 상파울루에서 PROCONVE (Programa de controle de emissões veiculares)라는 배출량 제어 정책을 시행하여 차량에서 배출되는 오염물질을 규제하였다. 상파울루는 PROCONVE 정책을 통해 대기 중에 일산화탄소와 질소산화물이 감소하는 결과를 보였으며, 미세먼지의 정책 감축 효과는 대기 중 미세먼지의 33%인 것으로 평가되었다(Jang et al., 2018). 또한, 우리나라의 연료 소비량에 따라 미세먼지 수치에 간접적으로 영향을 끼칠 수 있는 것으로 확인됐다(Kim, 2018). 선행연구를 통해 미세먼지에 대한 원인으로 풍속, 대기 오염물질 등 대기 환경 요인과 자동차 배기가스, 매연 등 화석 연료로 인한 요인으로 제시되었다. 본 연구는 두 가지 요인을 모두 고려하여 미세먼지 농도 예측을 진행한다.

2.3. Predicting PM Concentrations

미세먼지 농도를 예측하는 것은 미세먼지에 대한 피해를 예방하는 데 중요한 역할을 하기 때문에 다양한 연구가 진행되고 있다. 대부분의 선행연구에서는 기온, 강수량, 풍속 등의 기상 데이터와 NO₂, SO₂, CO, O₃, PM₁₀ 등 대기 오염 데이터를 활용한 예측을 하는 경우가 많았다. 한 연구에서는 기상청에서 제공하는 기상 데이터와 에어코리아에서 제공하는 대기오염물질 데이터를 이용하여 미세먼지 농도 예측을 진행하여 다양한 머신러닝 예측 성능 비교를 하고자 하였다. 황사가 주로 유입되는 경로인 산둥반도와 가장 인접한 인천광역시의 데이터를 추출하였고, 인천시의 다양한 기상 요인 및 대기오염물질들의 상관관계 확인 후, 모형을 구축하였다. 모형으로는 MLP, RNN, LSTM, GRU 그리고 CNN을 사용하였고, 기본적인 하이퍼파라미터와 단일층으로 구성하여 예측 성능을 비교하여 GRU모형의 성능이 가장 우수하다는 것을 확인했다(Kim, 2023). 또 다른 연구에서는 기상 데이터와 미세먼지 관련 데이터를 활용하여 다중회귀분석을 통해 미세먼지 발생에 영향을 미치는 변수를 추출하였다. 이를 토대로 머신러닝 회귀학습기 모형으로 학습하여 예측력이 높은 모형을 추출하였고, 검증용 데이터를 이용하여 예측 모형의 성능을 검증하였다.

그 결과, 예측력이 높은 모형은 선형회귀모형, 가우스 과정 회귀모형, 서포트 벡터 머신으로 나타났으며, 훈련용 데이터의 비율과 예측력은 비례하지 않은 것으로 나타났다. 또한, 예측치와 실측치 차이의 평균치는 크지 않지만, 미세먼지 실측치가 높을 때, 예측력이 다소 떨어지는 것으로 나타났다(Kim & Moon, 2021). 반면 대기 환경 데이터뿐만 아니라 다른 데이터를 활용한 선행연구도 있었다. 2010년 1월부터 2019년 11월까지 전국의 대기 환경 데이터, 항만의 화물 입항 현황 데이터 그리고 발전량 데이터를 활용하여 미세먼지 농도 예측을 진행하였으며 대기 환경데이터에 다른 변수 데이터를 추가함에 따라 유의미한 예측 성능 향상을 보였다(Sung et al., 2020).

본 연구는 선행연구에서 활용된 변수를 참고하여 적절한 변수를 선택하여 미세먼지의 농도를 예측하고자 한다. 미세먼지 수치에 대한 예측 정확성을 높이기 위해 다양한 영향을 미치는 변수들을 추가로 고려했다.

3. Data and Analysis Methodology

본 연구에서는 기상 데이터, 대기 오염 데이터, 교통 데이터, 전력거래량 데이터를 활용하여 미세먼지 수치를 예측했다. 서울시는 교통 및 산업 활동이 집중되어 있고 인구 밀도가 높아 다양한 영향을 받는 도시이기 때문에 교통량과 전력거래량이 대기 환경에 미치는 영향을 더 정확히 이해할 수 있다. 또한, 서울시 데이터로 먼저 모델을 구축하고 이를 기반으로 예측한 결과를 다른 지역의 데이터와 비교하거나 전국적인 데이터와 함께 활용하여 모델을 보완하는 실험을 할 수 있으며 이를 통해 모델의 정확성을 개선할 수 있다. 예측을 위한 미세먼지 데이터는 서울시의 미세먼지 농도 월별 평균 측정치며 대기오염물질 데이터는 아황산가스, 오존, 이산화질소, 일산화탄소에 대한 월별 평균 측정치로 구성되어 있다. 서울 열린데이터 광장에서 제공하는 서울시 월별 대기 오염도 평균 데이터를 활용하였으며 2016년 1월부터 2020년 12월까지의 데이터를 수집했다. 기상 데이터는 기상청에서 제공하는 기온, 강수량, 풍속에 대한 월별 평균 측정치로 구성되어 있다. 2016년 1월부터 2020년 12월까지의 데이터를 수집했으며, 서울특별시 월별 평균 데이터를 활용했다. 교통량 데이터는 서울시 교통량에 대한 데이터를 활용했다. 서울시설공단에서 제공하고 있으며, 조사 지점에 따른 서울시 유입, 유출 차량

데이터를 나타내고 있다. 월별 서울시 유입, 유출 차량 총계 데이터를 활용했으며 연도별 조사 지점 수가 달라, 예측에 사용된 데이터는 지점 수에 따라 평균으로 전처리 작업을 실시한 후 사용하였다. 수집된 데이터는 2016년 1월부터 2020년 12월까지의 데이터로 구성되어 있다. 전력거래량 데이터는 KPX 전력 거래소에서 제공한 데이터이며 서울시 일별 전력거래량에 대한 데이터이다. 수집 데이터는 2016년 1월부터 2020년 12월까지로 구성되어 있으며 LNG, 폐기물 등의 대기 오염을 발생시키 미세먼지에 영향을 주는 원료를 사용해 발전한 전력거래량을 월별 평균으로 전처리 작업을 실시한 후 사용하였다.

Microsoft사의 Azure를 활용하여 XGBoost, 랜덤 포레스트, 인공신경망 3가지 기계학습 기법을 통해 미세먼지 농도 예측을 실시하였다. 가장 먼저 기상 데이터와 대기 오염 데이터로 변수를 지정하고 교통량 데이터, 전력거래량 데이터 순으로 변수를 추가하여 미세먼지 농도 예측을 진행하였다. 기계학습 모델의 예측 성능은 MAE와 RMSE 기준으로 판단하였다.

4. Results

Table 1: The accuracy of the PM prediction model

		XGBoost	RF	ANN
Air Quality	MAE	6.04	6.22	6.25
	RMSE	7.63	7.86	7.83
Add Traffic Volume	MAE	5.49	6.03	5.49
	RMSE	6.92	8.04	6.91
Add Power Volume	MAE	5.47	5.50	4.61
	RMSE	6.90	7.12	5.93

미세먼지 예측 모델 비교는 기상 데이터와 대기 오염에 대한 데이터를 시작으로, 기존 데이터에 교통 데이터를 추가한 데이터, 전력거래량 데이터를 추가한 데이터

순서대로 데이터를 변경하며 미세먼지 예측 모델 비교가 진행되었다. 측정 값은 소수점 네 번째 자리로 반올림을 실시하였다.

기상 데이터와 대기 오염에 대한 데이터를 활용한 미세먼지 예측 모델 분석 결과, 평균 절대 오차인 MAE(Mean Absolute Error)를 기준으로 XGBoost의 MAE는 6.04로 가장 높은 예측 정확도를 보였으며, 평균 제곱근 오차인 RMSE(Root Mean Squared Error)를 기준으로도 XGBoost의 RMSE가 7.63로 가장 높은 예측 정확도를 보인 것을 확인할 수 있다. 예측 정확도는 MAE 기준 XGBoost, 랜덤 포레스트, 인공신경망 순으로 우수하게 나타났고, RMSE 기준으로는 XGBoost, 인공신경망, 랜덤 포레스트 순으로 우수하게 나타났다. 기상 데이터와 대기 오염에 대한 데이터에서 교통 데이터 추가해서 미세먼지를 예측한 결과, MAE를 기준으로 모든 모델의 성능이 향상되었으며 특히, XGBoost와 인공신경망의 경우 성능 향상 폭이 랜덤 포레스트와 비교하여 큰 것을 확인할 수 있다. 이번에도 MAE 기준으로 XGBoost의 MAE가 5.49로 가장 성능이 좋은 수치를 보였으나 인공신경망의 MAE도 5.49으로 매우 근소한 차이를 보였다. RMSE 기준으로는 XGBoost와 인공신경망의 경우 성능이 향상되었지만 랜덤 포레스트의 경우 성능이 근소하게 하락하였다. 이번 경우에는 인공신경망의 RMSE가 6.91로, XGBoost의 RMSE인 6.9187보다 매우 근소하게 좋은 성능의 수치를 보였다. 예측 정확도는 MAE 기준 XGBoost, 인공신경망, 랜덤 포레스트 순으로 우수하게 나타났고, RMSE 기준으로는 인공신경망, XGBoost, 랜덤 포레스트 순으로 우수하게 나타났다. 기상 데이터와 대기 오염에 대한 데이터와 교통 데이터에 전력거래량 데이터를 추가해서 미세먼지를 예측한 결과, MAE와 RMSE 기준에서 모든 모델의 성능이 향상되었다. 특히, 인공신경망의 경우 다른 두 모델과 비교하여 높은 예측 정확도를 보여주고 있다. 예측 정확도는 MAE와 RMSE 기준으로 인공신경망, XGBoost, 랜덤 포레스트 순으로 우수하게 나타났다.

기상 데이터와 대기 오염에 대한 데이터를 활용한 모형의 경우 XGBoost가 가장 좋은 결과를 보였다. MAE 기준 6.04, RMSE 기준 7.63의 오차를 보여주고 있다. 랜덤 포레스트와 인공신경망은 MAE 기준 6.22, 6.25의 오차를 보였으며, RMSE 기준 7.86, 7.83의 오차를 보였다. MAE 기준 인공신경망의 정확도 성능이 가장 낮은 것을 확인할 수 있었고 RMSE 기준 랜덤 포레스트의 정확도 성능이 가장 낮은 것을 확인할 수 있었다. 교통 데이터가 추가된

모형의 경우, MAE 기준 XGBoost가 5.49의 오차로 전과 같이 가장 좋은 결과를 보였으나 RMSE 기준으로는 인공신경망이 6.91로 가장 좋은 결과를 보였다. 그러나 두 기준 모두에서 매우 근소한 차이를 보여주고 있기 때문에 오차의 순위보다는 두 모델 모두 오차의 크기가 크게 감소한 것에 더 큰 의미가 있다고 보여진다. 그에 반해 랜덤 포레스트의 경우 MAE 기준으로는 6.03로 성능이 개선되었으나, RMSE 기준에서는 8.04로 성능이 떨어졌다. 전력거래량까지 추가된 모형의 경우 인공신경망이 가장 좋은 결과를 보였다. MAE 기준 4.61, RMSE 기준 5.93로 두 기준 모두에서 오차의 크기가 크게 감소하였으며, 다른 두 모델과 비교하여 가장 낮은 오차 결과를 보여주었다. XGBoost와 랜덤 포레스트는 MAE 기준 5.47, 5.50의 오차를 보였으며, RMSE 기준 6.90, 7.12의 오차를 보였다. XGBoost와 랜덤 포레스트 모두 이전보다 성능이 개선되었지만 XGBoost의 경우 그 효과가 매우 미비하였다.

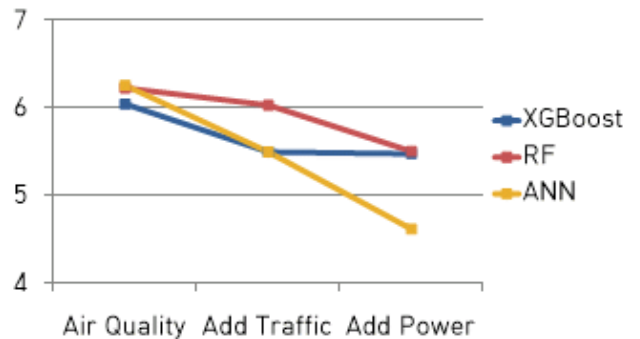


Figure 1: The model accuracy based on MAE

교통 데이터를 추가한 경우에서 RMSE 기준 랜덤 포레스트의 성능이 감소한 것을 제외하면 변수가 추가됨에 따라 모든 모델의 성능이 개선된 것을 확인할 수 있다. 특히 인공신경망의 경우, 교통 데이터와 전력거래량 데이터를 각각 추가한 결과 모두 모델의 성능이 상당히 향상된 것을 확인할 수 있었다. 인공신경망 모델에서 나타난 이러한 추세를 고려할 때, 모델의 개선 여지가 상당히 크다고 할 수 있다. 다시 말해, 미세먼지에 영향을 미칠 수 있는 유의미한 변수가 모델에 추가될 경우, 미세먼지 예측의 정확도를 더 향상시킬 수 있을 것으로 판단된다.

5. Conclusion

미세먼지로 인해 발생하는 문제점을 파악하고 피해가 발생하지 않도록, 또는 피해가 줄어들 수 있도록 예방하기 위해 미세먼지 예측은 매우 중요하다고 할 수 있다. 본 연구는 미세먼지 농도 예측을 위해 다양한 미세먼지 관련 변수를 활용하여 연구를 진행하였다. 연구결과 미세먼지 수치에 영향을 주는 변수들이 추가될수록 예측 모형의 예측 성능이 향상되는 것을 확인할 수 있었다. 본 연구의 결과를 통해서 미세먼지 농도 예측을 실시할 경우 내부에서 자체적으로 미세먼지를 발생시키는 요인을 고려하는 것이 예측 성능을 개선하는 데 큰 도움이 될 수 있다는 점을 확인했고, 미세먼지를 줄이기 위해서는 지역 내부에서 미세먼지를 발생시키는 원인을 찾아 그에 대한 대책을 마련해야 한다는 점을 알 수 있었다. 알고리즘 모델 간의 성능을 비교한 결과는 인공지능경망이 본 연구에서 가장 적합한 것으로 파악되었다.

본 연구를 통해 미세먼지 농도 예측 모델의 성능을 개선시킬 수 있는 방법론 및 미세먼지로 인한 피해를 줄이고 예방할 수 있는 대책을 마련할 수 있을 것으로 기대된다. 또한, 인공지능경망 기법을 통해 앞으로의 미세먼지 수치도 예측하여 미세먼지를 예방하고 적절한 대처를 할 수 있을 것으로 보이며 미세먼지 농도 예측에 유의미한 변수를 추가로 발견하여 다양한 변수를 함께 고려한다면 예측 모델은 보다 심층적이고 포괄적인 정보를 반영하게 되어, 예측의 정확도를 크게 향상시킬 수 있을 것으로 기대된다. 이를 통해 서울시뿐만 아니라 다른 지역의 데이터와 비교하거나 전국적인 데이터와 함께 활용하여 다양한 환경 정책과 규제지원으로 도시 및 지역 단위에서 미세먼지 농도를 효과적으로 관리 또는 대응이 가능할 것으로 보인다. 더 나아가 에너지 소비를 최적화하고, 환경에 미치는 영향을 최소화하여 에너지 효율을 향상시키고 비용을 절감할 수 있는 기회를 제공할 것으로 기대된다.

References

BBC NEWS Korea. (2018, March 26). Retrieved December 10, 2023, from <https://www.bbc.com/korean/news-43524873>

Han, J. M. & Jo, G. H. & Kim, J. G. (2017), Verify a Causal Relationship between Fine Dust and Air Condition-Weather Data in Selected Area by Contamination Factors, *Korean Big Data Society*, 2(1), pp. 17-26.

Jang, Y. W. & Lee, K. W. & LEE, T. H. & GHIM, Y. S. & Kim, J. H. & Kim, B. G. (2018), The Study of Air Quality in Sao Paulo, Brazil, *Hankuk University of Foreign Studies Center for Latin American Studies*, 37(2), pp. 137-158.

Kim, H. G. (2018), Estimating Changes of Causative Factors' Influences: Focusing on Diesel, *Korean Society of Data Science*, 20(2), pp. 747-757.

Kim, H. L. & Moon, T. H. (2021), Machine learning-based Fine Dust Prediction Model using Meteorological data and Fine Dust data, *Journal of Korean Geographic Information Society*, 24(1), pp. 99-111.

Kim, H. M. & Lee, S. W. & Kim, I. G. & Lee, D. G. & Yoo, S. H. (2020), Measuring the Economic Value of Improving Observation Systems Use and Prediction Model Accuracy_Conjoint Analysis, *Korean Journal of Innovation*, 15(1), pp. 301-320.

Kim, S. Y. (2023), A study on PM10 forecasting model using machine learning, *Korea Data Information Science Society*, 34(5), pp. 763-773.

Lee, S. B. (2022), An Introduction to Machine Learning Focusing on Predictive Models Using Supervised Learning, *Ewha Womans University Institute of Education Science*, 53(3) pp. 1-43.

Lee, T. H. & Jeon, M. J. (2018), Prediction of Seoul House Price Index Using Deep Learning Algorithms with Multivariate Time Series Data, *Journal of Housing and Urban Research*, 8(2), pp. 33-56.

Park, S. A. & Shin, H. J. (2017), Analysis of the Factors Influencing PM2.5 in Korea: Focusing on Seasonal Factors, *Korea Institute of Environmental Policy and Evaluation*, 25(1), pp. 227-248.

Park, Y. S. & Park, J. H. & Kang, S. Y. & Lee, S. H. & Son, J. S. & Yoo, C. & Lee, S. B. & Kim, J. S. (2020), Analysis of High PM₁₀, PM_{2.5} Concentration in Kangwon, *Proceedings of the Korean Society for Atmospheric Environment Annual Conference*, pp. 146-146.

Specht, D. F. (1991), A General Regression Neural Network, *IEEE Trans. on Neural Networks*, 2(6), pp. 568-576.

Sung, S. H. & Kim, S. J. & Ryu, M. H. (2020), A Comparative Study on the Performance of Machine Learning Models for the Prediction of Fine Dust: Focusing on Domestic and Overseas Factors, *Korean Society for Innovation*, 15(4), pp. 339-357.