Innovations in Environmental Economics

www.iee.reapress.com

Innov. Environ. Econ. Vol. 1, No. 2 (2025) 96–108.

Paper Type: Original Article

Machine Learning-Driven Predictive Models for Urban Sustainability in the Context of Digital Transformation

Shamina Israr Tithi*

Earth and Environmental Sciences, Brooklyn College, CUNY, USA; Shamina.Tithi@brooklyn.cuny.edu.

Citation:

Received: November 14, 2024 *Revised: January 12, 2025* *Accepted: March 17, 2025*	Tithi, S. I. (2025). Machine learning-driven predictive models for urban sustainability in the context of digital transformation. *Innovations in Environmental Economics, 1*(2), 96-108.

Abstract

Societal considerations make sustainable urban planning in the age of energy and digital revolution paramount. By using state-of-the-art methods for analyzing massive datasets, such as Artificial Intelligence (AI) and Machine Learning (ML), we may get a better grasp of past data and more effectively forecast future occurrences using information gathered from IoT devices. The effects of energy transformation and environmental policy were examined, as were the long-term consequences of specific activities, using a multi-dimensional historical analysis of air pollution that this research used. Predictions of air pollution were also made using ML methods that included geographical considerations. Incorporating data from numerous sites and assessing the effect of neighboring sensors on predictions, this research used many low-cost air sensors categorized as Internet of Things (IoT) devices. Regression models, Deep Neural Networks (DNNs), and ensemble learning were among the ML techniques examined. There was an investigation into the feasibility of using such forecasts in open-source IT mobile systems. The study took place in Kraków, Poland, a city with a long history of air pollution and a UNESCO World Heritage Site. Additionally, Kraków is in charge of creating clean mobility zones and banning the use of solid fuels for heating. According to the study, increasing the city's population has no negative effect on PMx concentrations. The main aspect in bettering air quality, particularly for PMx, is shifting from coal-based to sustainable energy sources. Transportation has a less significant impact. Using linear ML models yields the best results when attempting to forecast infrequent smog episodes. Building a smart city that considers the effects of air pollution on the quality of life may be significantly advanced by acting on the findings of this study.

Keywords: Big data analytics, Sustainable energy shift, Smart urban systems, Machine learning algorithms, Air contaminants, Urban growth.

1 | Introduction

Collaborative efforts across fields such as geography, political science, and environmental engineering are necessary for the establishment of smart cities. A clean environment, high living standards, easy access to education, and efficient policymaking are the four pillars upon which a fulfilled life rests [1–4]. According to studies, people's quality of life is

Corresponding Author: Shamina.Tithi@brooklyn.cuny.edu
https://doi.org/10.48313/iee.v1i2.42

greatly affected by environmental comfort, especially air pollution [5], [6]. It is necessary to have a comprehensive worldwide strategy for urban development in light of the growing urbanisation and environmental problems [7], [8]. Particulate matter analysis is crucial for well-balanced city design, according to Liang et al. [9]. The lack of studies examining the effects of air pollution over the long run in smart city aspirants was highlighted by Jonek-Kowalska [10]. Building more cities might lead to a rise in energy consumption. Air pollution may be worsened depending on certain energy mixtures [11]. In 2021, hard coal accounted for about 22% of Poland's home energy consumption, far more than the 2.5% average throughout the EU [12]. In order to determine how the energy shift will affect air pollution, cutting-edge technology and Internet of Things (IoT) sensors are required.

Air pollution has been a problem in Kraków for a long time. At first, pollution came mostly from the metalworking sector, but now, heating with fossil fuels is a major contributor [13], [14]. The primary pollutant throughout the winter continues to be solid fuels, even if their usage is completely forbidden [15], [16]. Coal accounts for 50% of PM10 in winter and 20% in summer, contributing to its above 40% carbon content. Transportation by automobiles ranks second [17]. The carbon percentage is 30% from natural sources throughout the year [18]. The city is becoming more polluted as outside pollutants make their way in, particularly during the winter months [19]. Part of the reason for this is the city's location, which is influenced by the nearby hills and the Carpathian Mountains to the south [20], [21]. Tourists and locals alike are feeling the effects of the problems plaguing Kraków, a city on the UNESCO World Heritage list. There have been attempts to sustainably revive the tourist business since the COVID-19 epidemic caused a drop in the industry [22]. Kraków is one of the most polluted cities in the world due to its occasional smog occurrences, which may discourage visitors, particularly those worried about contracting the COVID-19 virus [23], [24]. Economic and public health advantages would be added to the city's air quality improvement efforts. By addressing the problems of air pollution, Kraków hopes to become a smart city with contemporary, comfortable living circumstances [25], which will increase its appeal to visitors. Plus, the Mobywatel system [26] was put into place in Poland. It's a digital platform that any person may use, and it provides real-time data on air pollution levels. Thanks to this breakthrough, it may soon be possible to show how zero-emission regulations affect the environment in a concrete way. It has the potential to change people's perspectives and make them think that sustainable city planning is possible. The information gap highlighted by Jonek-Kowalska [10] may be filled with the use of this data.

In this research, we use big data from several sources to identify the most effective methods for forecasting very unusual spikes in pollutant input and to evaluate potential causes impacting air pollution. In this age of energy and digital change, the primary concern is how to steer policymaking towards the most efficient and sustainable use of these technologies. Because of its location in a nation whose energy system is mostly dependent on coal, Kraków provides a valuable example for other places throughout the globe to follow when enacting stringent rules in accordance with EU requirements.

There are two sections to this part, and they both seek to fill a mental void. In order to determine what occurred and its effect on pollution levels during the last decade, we first conducted a historical, descriptive, and diagnostic study of data pertaining to population changes, different kinds of transportation, and heating types. In the context of energy policy, this is of the utmost importance for future informed city planning. From a public health standpoint, the second half is a predictive and prescriptive study, where we examine the possibility of several methodologies to best anticipate smog outbreaks, which are relatively infrequent but substantial. Such instances are anomalies in data analysis, necessitating the use of suitable contemporary approaches due to the frequent ineffectiveness of more conventional procedures in these contexts. Here, we looked at how various neural network designs fared against the autoregressive moving average.

There are two sections to this part, and they both seek to fill a mental void. In order to determine what occurred and its effect on pollution levels during the last decade, we first conducted a historical, descriptive, and diagnostic study of data pertaining to population changes, different kinds of transportation, and heating types. In the context of energy policy, this is of the utmost importance for future informed city planning. From a public health standpoint, the second half is a predictive and prescriptive study, where we examine the possibility of several methodologies to best anticipate smog outbreaks, which are relatively infrequent but substantial. Such instances are anomalies in data analysis, necessitating the use of suitable contemporary approaches due to the frequent ineffectiveness of more conventional procedures in these contexts. Here, we looked at how various neural network designs fared against the autoregressive moving average.

This study is unique because it merges two distinct but related fields: Sustainable urban planning and sophisticated Artificial Intelligence (AI)-assisted spatial analysis of massive data related to air pollution, with a focus on energy transition. For two key reasons, we settled on Krakow as our European benchmark city: 1) Krakow is distinct from other cities that have implemented anti-smog laws because it is geographically separated from adjacent places that do not, 2) The city is situated in a nation that has and will continue to have a high priority for moving away from energy production that relies on coal. Problems in properly predicting, communicating, and developing solutions that are good for society must be addressed, and methods must be put in place to evaluate the beneficial impacts of energy transformation on smart city development. The goal of this study is to analyse the best methods for predicting air pollution, which will help with a variety of tasks, including alerting locals, preparing for change, and assessing the effects of that change. Additionally, we looked at past data while keeping in mind a variety of variables that can affect PMx air pollution, including population, vehicle count, and the condition of public transit.

Furthermore, we took into account the results of the initiative to alter the energy composition of residential heating systems in the urban region. In order to build smarter cities that learn from their mistakes and make better ones in the future, this study will use big data and Machine Learning (ML)/AI techniques. In order to do this, we will use ML approaches that are both efficient and dependable to perform statistical analyses on spatio-temporal data. In order to address the question of whether dense spatio-temporal time series can effectively anticipate rare smog episodes, our study will examine the effects of population changes, different forms of transportation, and energy policies on air quality.

2 | Materials and Methods

2.1 | Urban Development and Energy Transition

A quantitative study of the measures implemented by the Kraków City Council and the Chief Inspectorate for Environmental Protection to reduce atmospheric air pollution was carried out using data obtained from Statistics Poland [26], accessed on May 10, 2024. Factors such as the city's population in the relevant years, statistics on private and public vehicle travel, and long-term PM2.5 trends were all included in the research [27], [28]. Public transport metrics used included the number of bus and tram lines and their lengths in kilometres, while data on registered passenger cars was based on the total number of vehicles. We also took into account data on bike traffic, although it was limited to linear infrastructure, i.e., the total length of bike lanes. The research did not include any point infrastructure, such as bike rental stations, since the city does not yet have a bike rental system in place.

A quantitative study of the measures implemented by the Kraków City Council and the Chief Inspectorate for Environmental Protection to reduce atmospheric air pollution was carried out using data obtained from Statistics Poland [26], accessed on May 10, 2024. Factors such as the city's population in the relevant years, statistics on private and public vehicle travel, and long-term PM2.5 trends were all included in the research. Public transport metrics used included the number of bus and tram lines and their lengths in kilometres, while data on registered passenger cars was based on the total number of vehicles [29], [30]. We also took into account data on bike traffic, although it was limited to linear infrastructure, i.e., the total length of bike lanes. The research did not include any point infrastructure, such as bike rental stations, since the city does not yet have a bike rental system in place.

2.2 | Machine-Learning Data Pipeline

The need to build data processing pipelines that are repeatable, maintainable, and modular is growing in today's ML environment. This method facilitates better project management, rapid adaptability to changing needs, and increased project efficiency and scalability [31], [32]. As shown in *Fig. 1*, a complete pipeline was created as part of this project. The four main sections of this pipeline are preprocessing, feature engineering, modelling, and XAI/evaluation.

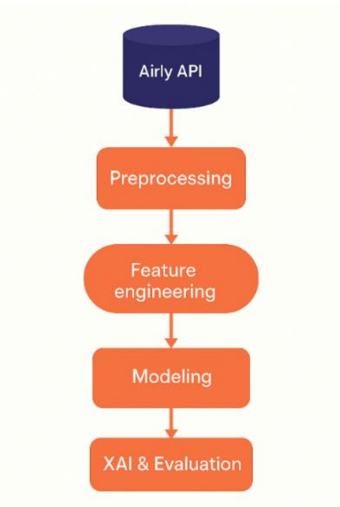

Fig. 1. Overview of the ML data pipeline used in this research.

To make sure the data is resilient against outliers, we used a robust scaler [32] to scale it and interpolate missing values in the preprocessing sub-pipeline. We paid special attention to retaining unusual selections since they are important for our research. In order to capture the intrinsic periodicity in characteristics like wind direction and time of day, cyclic features were constructed during the feature engineering process. Not only that, but Kraków's social and holiday events were included, along with elements like cardinal wind direction and dawn and sunset timings. Additional PM2.5 components, including trend and seasonality, were included based on STL decomposition [27], [33]. To make the prediction model better, we included lag characteristics developed using autocorrelation and Exploratory Data Analysis (EDA). The modelling step included setting up a model factory that could build several models from the Darts library [34] with user-specified settings and parameters, allowing for systematic optimisation and testing. Regression measures were used to evaluate the performance and dependability of the model during backtesting with expanding window optimization [35], [36]. Model residuals and XAI studies, including methods such as Shapley Additive exPlanations (SHAP) [37], were also documented.

2.3 | Machine-Learning Forecasting

Global forecasting models were used in the research. This method allows for the simultaneous building of a single prediction model for several time series that are located in different parts of the world. Its goal is to minimise the noise that each series might bring by capturing the essential patterns controlling the series. This method is stable when extrapolated to other time series, simple to maintain, and computationally efficient. Having a superficial familiarity with the unique traits of each series is a price to pay [38].

2.3.1 | Models

Twelve different models were first tested in this investigation; five were ultimately chosen for parameter adjustment according to their performance. A total of 455,520 measurements were used to train the models, including data gathered from 52 sensors throughout the world. From 00:00:00 on January 1, 2022, until 00:00:00 on January 1, 2023, each sensor logged 8760 readings. Linear models like ridge regression, which add L2 norm loss functions to traditional linear regression, are part of the first category [32]. We looked at both standard linear models and more advanced ones tailored to time series, including DLinear and NLinear. In order to make a final forecast, the DLinear model breaks down the input data into its trend and seasonal components. It then processes each component using single-layer linear transformations. The model works especially well with data that is heavy on trends. The NLinear model increases data adaptability and overall model performance by reintegrating the final value of the sequence after transformation, which improves adaptation to data changes [39]. This is done before processing the data via a linear layer.

Linear models like ridge regression, which add L2 norm loss functions to traditional linear regression, are part of the first category [32]. We looked at both standard linear models and more advanced ones tailored to time series, including DLinear and NLinear. In order to make a final forecast, the DLinear model breaks down the input data into its trend and seasonal components. It then processes each component using single-layer linear transformations. The model works especially well with data that is heavy on trends. The NLinear model increases data adaptability and overall model performance by reintegrating the final value of the sequence after transformation, which improves adaptation to data changes [25], [40], [41]. This is done before processing the data via a linear layer. A variety of sophisticated deep learning models were also included in the assessment. These included transformers like the Temporal Fusion Transformer (TFT) and variations of Recurrent Neural Networks (RNNs), including GRU, LSTM, and NBEATS. With fewer trainable parameters, the GRU—a simplified version of the LSTM—can save training time but may suffer in complicated situations due to its reliance on update and reset gates to control the flow of information. To accurately represent temporal dependencies, the TFT makes use of the transformer architecture, which incorporates a probabilistic component that is particularly helpful when predicting time series data with inherent uncertainties and an attention mechanism that is good at combining both global trends and local fluctuations [42], [43]. The transformer architecture has been shown to outperform traditional RNNs in certain time series forecasting tasks.

2.3.2 | Evaluation

Backtesting with expanding window optimisation was used to assess the models, as seen in *Fig. 2*. The quantity of data will grow over time since training is done on a dataset that contains past information. When fresh data is included in the model, backtesting enables a more accurate evaluation of its efficiency [44], [45].

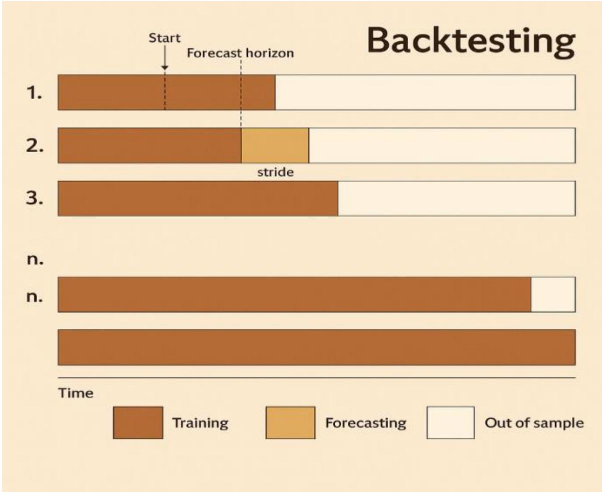

Fig. 2. Fundamental understanding of backtesting with expanding window optimization.

With the use of measures like Mean Absolute Error (MAE), Root Mean Square Error (RMSE), R^2, and Mean Absolute Percentage Error (MAPE), the predicted data during backtesting was confirmed. By providing a simple way to grasp the inaccuracy in units of the predicted variable, the MAE makes it easier to comprehend the accuracy of predictions. RMSE is useful for identifying problems in prediction models when outliers are substantial, since it is more sensitive to outliers than MAE. To further our knowledge, we may use the Multi-Analytical Relative Risk Error (MARRE) to evaluate how well a model is doing in comparison to a base or benchmark model. Another metric that looks at how well the model fits the data and how accurate its predictions are in comparison to the real values is the MAPE [32], [34].

$$\text{MAE}\ (y, y^{\sim}) = \frac{1}{N}\sum_{i=1}^{N}(y_i - y_i^{\sim}). \tag{1}$$

$$\text{RMSE}\ (y, y^{\sim}) = \sqrt{\frac{\sum_{i=0}^{N}(y_i - y_i^{\sim})^2}{N}}. \tag{2}$$

$$R^2\ (y, y^{\sim}) = 1 - \frac{\sum_{i=1}^{N}(y_i - y_i^{\sim})^2}{\sum_{i=1}^{N}(y_i - \overline{y_i})^2}. \tag{3}$$

$$\text{MAPE}\ (y, y^{\sim}) = \frac{100\%}{N}\sum_{i=0}^{N-1}\frac{y_i - y_i^{\sim}}{y_i}. \tag{4}$$

$$\text{MARRE}\ (y, y^{\sim}) = 100.\frac{1}{N}\sum\frac{(y_i - y^{\sim})}{\max(y_i) - \max\,(y_i^{\sim})}, \tag{5}$$

where y is the mean of the dependent variables, max (y_i) is the highest value in the series, min (y_i) is the lowest value in the series, y_i is the actual value of the i-th observation, y_i is the predicted value of the i-th observation, and N is the number of observations.

3 | Results

3.1 | Urban Development

Along with a demographic chart of the city, *Fig. 3* shows the evolution of PM2.5 concentration in Kraków over a decade. There has been a noticeable uptick in the population of Kraków. There were two distinct periods of relatively slow population growth over the time period under consideration (2010–2014 and 2015–2019, respectively). In contrast, the PM2.5 trend line shows a downward trajectory over the last few years until seeing a reversal in 2015 and 2016. The years 2013, 2014, 2017, and 2019 had the greatest drops in PM2.5 concentration.

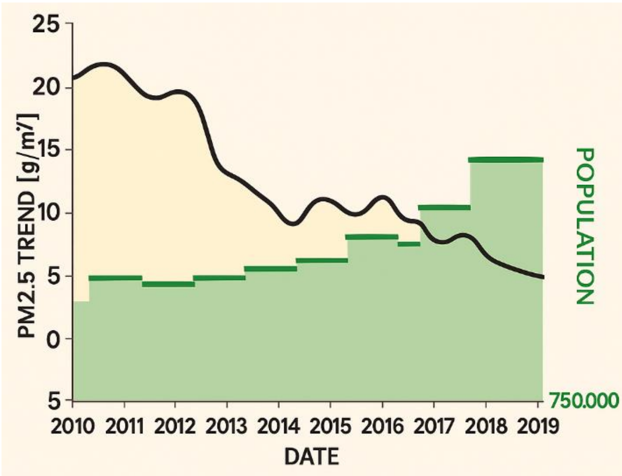

Fig. 3. The population of Kraków (green) and the evolution of PM2.5 levels (black/yellow) from 2010 to 2019.

Fig. 4 shows the evolution of city infrastructure from 2010 to 2019. The number of passenger vehicles on the road has steadily increased over the last several years. Notably, in 2019, there were more than 650,000 vehicles registered in Kraków alone. Of them, somewhat more than 6,000 were electric cars, with about 3,000 being hybrids [42]. The city's network of bike lanes has grown substantially, with a total of more than 50 km added in the last decade. Aside from a steep drop in 2012, the number of miles of tram lines is fairly constant throughout the studied period [46]. There is a discernible downward trend in the overall length of bus lines (also given in km) up to 2014, after which there is a definite upward trend in the years that follow.

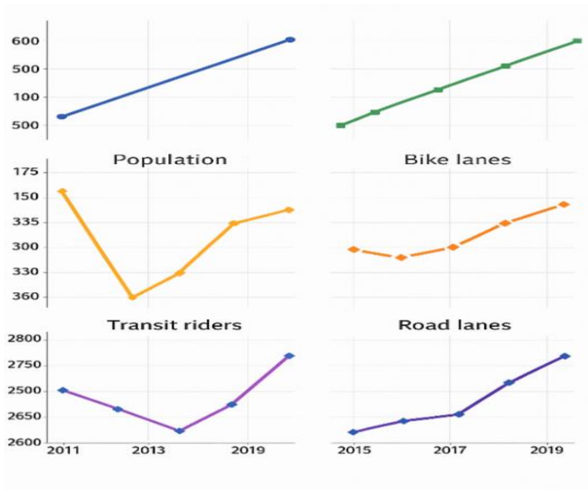

Fig. 4. Trends in Krakow's urban infrastructure from 2010 to 2019, illustrating annual changes in automobile registrations (blue), bike lanes (green), tram lines (orange), and bus lanes (purple).

3.2 | Energy Transition

Nearly 24,000 coal-fired combustion chambers, furnaces, and boilers were recorded in Kraków in 2015 [37]. Both the ones used for water heating and the ones connected to building insulation have been retired; the quantity of these furnaces is shown in *Fig. 5*. The number of renewable energy installations is also shown on the graph. From 2014 to 2016, the number of furnaces and boilers decommissioned was modest, but in 2017 and later years, it increased significantly. There is no discernible rising trend in the number of installations powered by renewable energy sources during the studied period; rather, their number stays modest and largely constant year-to-year, with the exception of 2018, when it almost reached 500.

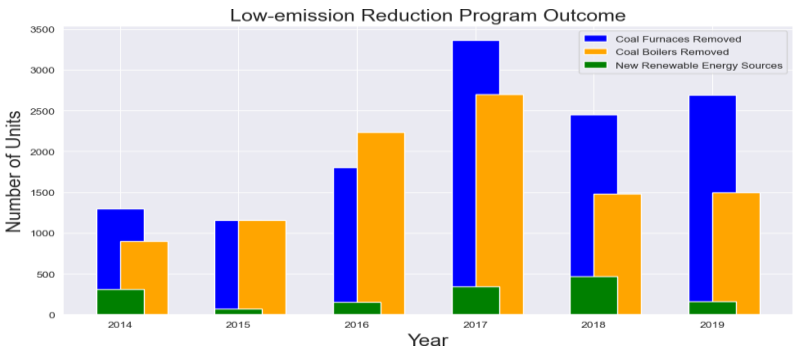

Fig. 5. Kraków's initiative to reduce emissions (PONE) from 2014 to 2019. Green represents the amount of new renewable energy sources, whereas blue represents the number of coal boilers that have been dismantled

3.3 | Machine-Learning Forecasts

Fig. 2 shows that the backtesting method with expanding window optimisation was used to train and assess all ML models. Eighty percent of the available data, or the date October 20, 2022, at 01:00, was used as the beginning point for the backtest. With a continuous 24-hour prediction horizon, the horizon was methodically moved by one time step. When looking at *Table 1* findings for several models, it's clear that linear models like DLinear, NLinear, and Ridge have the most promise for PM2.5 forecasting. Overfitting was avoided during training by keeping an eye on the loss function, in this case, MSE Loss. The DLinear model finished training after five epochs (*Fig. 6*).

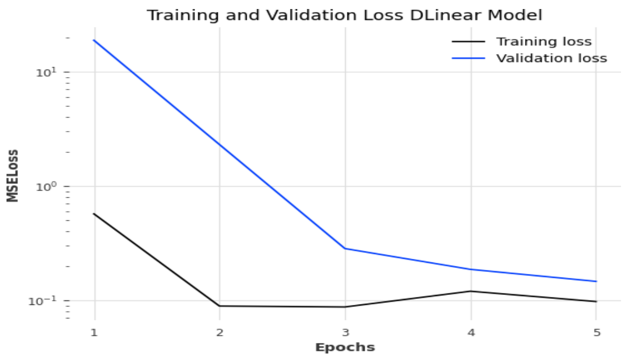

Fig. 6. Loss function DLinear.

They have a high coefficient of determination (R^2) and low levels of average errors (MAE and RMSE, for example). The Ridge Regression model (R^2=0.956) and the DLinear model (R^2=0.947) are particularly noteworthy. With a time frame of 168 samples (7 days), the deep learning and linear models that were adjusted for time series produced the best results.

Table 1. Model performance table.

Model	MAE	RMSE	R2	MAPE	MARRE
Ridge	2.666	3.867	0.956	15.621	1.859
ARIMA	6.688	9.282	0.755	36.417	4.622
XGBoost	4.104	6.763	0.879	19.234	2.747
CatBoost	3.839	6.236	0.897	18.463	2.569
LGBM	4.863	7.445	0.85	27.175	3.34
GRU	5.17	7.79	0.831	25.855	3.582
LTSM	5.258	7.704	0.83	27.206	3.682
NBEATS	12	17.915	0.079	76.314	8.547
TCN	13.276	19.651	−0.108	68.585	9.448
TFT	3.915	5.971	0.9	17.675	2.71
NLinear	3.356	4.695	0.932	20.706	2.418
DLinear	2.947	3.888	0.947	20.354	2.21

In the case of fast PM2.5 concentration peaks, nonlinear models, including tree-based, gradient boosting, and deep learning models, perform poorly. A comparison of the DLinear and XGBoost models is seen in *Fig. 7*. These abrupt changes are best handled by linear models. The bigger the peak, however, the more inaccurate the models were.

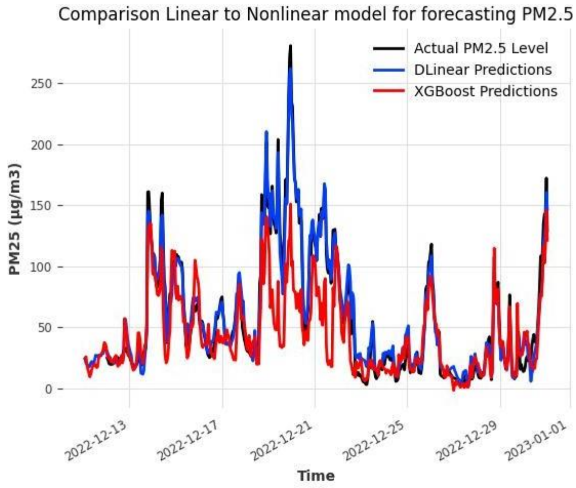

Fig. 7. Evaluation of XGBoost and DLinear for PM2.5 forecasting.

4 | Discussion

This indicates that the examined metropolitan area's air pollution levels are unaffected by the city's fast population expansion, especially after 2014. It is possible to build highly populated metropolitan areas while ensuring healthy, high-quality air, according to this encouraging finding. Seeing individuals moving to Kraków right now makes it all the more crucial. The first anti-smog legislation in the nation was adopted by the provincial parliament of Malopolska in 2013, which may have a direct correlation to the considerable decrease in pollution in Krakow throughout 2013 and 2014 [44]. One of the purposes of this law was to specify which fuels might be used inside the Kraków Municipal District. Another initiative that started at the same time was the PONE project, which provides subsidies to help pay for the replacement of heating systems that use coal. The 2012 launch of the Kraków Smog Alarm was a direct outcome of the many social demonstrations and grassroots initiatives that preceded these legislative amendments [43], [47], [48].

Regardless of the rate of population increase, a correlation between household heating fuel type and air pollution is obvious. However, a multi-dimensional analysis of the city's mobility system in relation to trends in population growth and lower PM2.5 levels reveals fascinating insights. The number of registered cars, the majority of which have combustion engines, has a positive correlation and is clearly a cause and effect. Nevertheless, there has been a downward trend in PM2.5 pollution. The quantity of public transportation routes is directly proportional to the growth in individual transportation. People may be less inclined to use public transit, particularly buses, if the number of people driving alone continues to rise at such a high rate. The idea of a smart city calls for a methodical expansion of the already steady network of tram lines. Because bicycling not only offers extra health advantages from cardiovascular activity but also actually reduces emissions from propulsion, the growth of bike networks has been a net gain. The negative link between the number of passenger automobiles and the quantity of PM2.5 may, however, be readily misunderstood. The maximum number of automobiles in a city is not always equal to the number of cars registered there. With more and more public transport options, people may only need to utilise their automobiles for occasional outings rather than their daily commutes. Second, research using isotopic [49] and geostatistical [50] methods demonstrated that heating with coal was the most critical component.

Transport clearly has little effect on the overall trend of PM2.5 in Kraków, according to these numbers. Various research, including big data analysis, has proven that the rise in PM2.5 concentrations is substantially impacted by yearly seasons and is driven by the combustion of solid fuels, particularly coal, for heating houses and water. It should be noted that these studies only look at airborne particulate matter that is suspended; they don't take into account other volatile chemicals that are harmful to health, the majority of which may be produced by vehicles. Above, we see that transportation has a much lesser effect on PMx concentrations than home heating. Consequently, it is reasonable to use data analysis and ML methods to forecast and optimally administer the city throughout the fall, winter, and spring seasons in order to prevent adverse public health occurrences. In addition to addressing concerns about expanding access to public transit, which, as previously said, does not significantly impact this kind of pollution, these measures should also take into account the need to restrict pollution from nearby cities, which is a major contributor [19], [43].

For the most part, when dealing with fast concentration peaks, linear models like DLinear, NLinear, and Ridge Regression do a better job of forecasting PM2.5 pollution levels. These models are more robust against extreme values because of the L2 regularisation method, which limits the magnitude of the coefficients. This is particularly important in winter, when pollution data might show abrupt shifts owing to unexpected input of pollutants. Data like this may be challenging for nonlinear models like tree-boosted, Deep Neural Networks (DNNs), and gradient-boosted models because they tend to overfit certain characteristics in the training data, such as unusual pollution patterns. Predicting future changes in PM2.5 concentrations, particularly under dynamically changing external circumstances, may be challenging due to the complex patterns that might be modelled as a consequence of overfitting. Therefore, environmental planning and industrial laws may benefit from linear models' increased stability and predictability when analysing time series data related to PM2.5 pollution.

5 | Conclusions

Important insights have been uncovered by our study of the effects of many variables, including population expansion, on PMx concentrations. There was a 3% rise in population in the region we looked at within the last decade. Even though PMx concentrations fell by 40% due to this expansion, our statistics show that it had no adverse effect on air quality. In spite of the increasing urban density since 2014, the region seems to have effectively controlled air quality. The shift away from coal-based energy sources to renewable energy or natural gas has been the most important component in improving air quality, especially with regard to PMx, according to our research. To everyone's surprise, transit is not the main driver of changes in PMx pollution levels in the city, despite early speculation. Our research also shows that there seems to be a two-speed energy transition happening for PMx mitigation: One speed is the change in heating energy sources, which is a major contributor, and the other speed is the change in transportation. In addition, smart city crisis response planning may greatly benefit from big data and automated prediction systems. When comparing linear and nonlinear models for PM2.5 level prediction, the former often performs better. Linear models include DLinear, NLinear, and Ridge Regression, whereas nonlinear models include TCN and traditional ARIMA. When there are sudden shifts in the data, this performance really stands out. They are less vulnerable to outliers in pollution data because of their stability and resistance to overfitting, which is caused by techniques like L2 regularisation, which limit the magnitude of the coefficients.

Author Contribution

The author was solely responsible for the conception and design of the study, development of the methodology, implementation of the computational framework, validation of the results, sensitivity analyses, and preparation of the manuscript.

Funding

This work was conducted without any financial support from funding agencies in the public, commercial, or non-profit sectors.

Data Availability

All data generated or analyzed during this study are included in this published article.

Conflicts of Interest

The author declares that there are no conflicts of interest relevant to the content of this article.

References

[1] Zhang, L., Su, D., Guo, W., & Li, S. (2022). Empirical study on urban sustainable development model based on identification of advantages and disadvantages. *Frontiers in sustainable cities, 4*, 894658. https://doi.org/10.3389/frsc.2022.894658

[2] Raihan, A., Ridwan, M., Salehi, M., & Zimon, G. (2025). Influences of financial development and energy price on renewable energy: An Italian case. *Energy engineering*, *122*(2), 1–22. https://doi.org/10.32604/ee.2025.059016

[3] Raihan, A., Voumik, L. C., Ridwan, M., Akter, S., Ridzuan, A. R., Wahjoedi, … & Ismail, N. A. (2024). Indonesia's path to sustainability: Exploring the intersections of ecological footprint, technology, global trade, financial development and renewable energy. In *Alareeni, B. & Elgedawy, I. (Eds.), Opportunities and risks in AI for business development: Volume 1* (pp. 1–13). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-65203-5_1

[4] Kurniawati, T., Rahmizal, M., Ridwan, M., Aspy, N. N., Mahjabin, T., Eleais, M., & Ridzuan, A. R. (2025). Reassessing the load capacity curve hypothesis in ASEAN-5: Exploring energy intensity, trade, and financial inclusion with advanced econometric techniques. *International journal of energy economics and policy*, *15*(2), 195–208. https://www.econjournals.com/index.php/ijeep/article/download/17328/8596/43212

[5] Wesz, J. G. B., Miron, L. I. G., Delsante, I., & Tzortzopoulos, P. (2023). Urban quality of life: A systematic literature review. *Urban science*, *7*(2), 56. https://doi.org/10.3390/urbansci7020056

[6] Urbee, A. J., Hasan, M. A., Ridwan, M., & Dewan, M. F. (2025). Adaptation and resilience in the face of climate-induced migration: Exploring coping strategies in the urban economy of barishal metropolitan city. *Environment, innovation and management*, *1*, 2550005. https://doi.org/10.1142/S306090112550005X

[7] Ridwan, M., Tahsin, M. S., Al-Absy, M. S. M., Eleais, M., Ridzuan, A. R., & Mukthar, K. P. J. (2025). Economic Alchemy: Unraveling the nexus between trade openness, inflation, exchange rates, and economic growth in bangladesh. *International journal of economics and financial issues*, *15*(3), 244. https://doi.org/10.32479/ijefi.17779

[8] Raihan, A., Ridwan, M., Sarker, T., Atasoy, F. G., Zimon, G., Bari, A. B. M. M., … & Mohajan, B. (2025). The influence of different environmental factors toward Vietnam's net-zero emissions goal. *Innovation and green development*, *4*(3), 100229. https://doi.org/10.1016/j.igd.2025.100229

[9] Liang, L., & Gong, P. (2020). Urban and air pollution: A multi-city study of long-term effects of urban landscape patterns on air quality trends. *Scientific reports*, *10*(1), 18618. https://doi.org/10.1038/s41598-020-74524-9

[10] Jonek-Kowalska, I. (2023). Assessing the effectiveness of air quality improvements in polish cities aspiring to be sustainably smart. *Smart cities*, *6*(1), 510–530. https://doi.org/10.3390/smartcities6010024

[11] Raihan, A., Hasnat, M. Al, Rahman, S. M., Ridwan, M., Rahman, M. M., Islam, M. T., … & Bari, A. B. M. M. (2025). Recent advancements in alternative energies, technological innovations, and optimization strategies for seaport decarbonization. *Innovation and green development*, *4*(3), 100252. https://doi.org/10.1016/j.igd.2025.100252

[12] Raihan, A., Zimon, G., Ridwan, M., Rahman, M. M., & Salehi, M. (2025). Role of mineral resource rents, renewable energy, and energy efficiency toward carbon neutrality in China. *Energy nexus*, *17*, 100394. https://doi.org/10.1016/j.nexus.2025.100394

[13] Gharbi, I., Rahman, M. H., Muryani, M., Esquivias, M. A., & Ridwan, M. (2025). Exploring the influence of financial development, renewable energy, and tourism on environmental sustainability in Tunisia. *Discover sustainability*, *6*(1), 127. https://doi.org/10.1007/s43621-025-00896-5

[14] Raihan, A., Ridwan, M., Zimon, G., Rahman, J., Tanchangya, T., Bari, A. B. M. M., … & Akter, R. (2025). Dynamic effects of foreign direct investment, globalization, economic growth, and energy consumption on carbon emissions in Mexico: An ARDL approach. *Innovation and green development*, *4*(2), 100207. https://doi.org/10.1016/j.igd.2025.100207

[15] Ridwan, M., Akther, A., Al Absy, M. S. M., Tahsin, M. S., Bin Ridzuan, A. R., Yagis, O., & Mukhta, K. P. (2024). The role of tourism, technological innovation, and globalization in driving energy demand in major tourist regions. *International journal of energy economics and policy*, *14*(6), 675–689. 10.32479/ijeep.17344.

[16] Ridwan, M., Akther, A., Tamim, M. A., Ridzuan, A. R., Esquivias, M. A., & Wibowo, W. (2024). Environmental health in BIMSTEC: The roles of forestry, urbanization, and financial access using LCC theory, DKSE, and quantile regression. *Discover sustainability*, *5*(1), 429. https://doi.org/10.1007/s43621-024-00679-4

[17] Ridwan, M., Aspy, N. N., Bala, S., Hossain, M. E., Akther, A., Eleais, M., & Esquivias, M. A. (2024). Determinants of environmental sustainability in the United States: Analyzing the role of financial development and stock market capitalization using LCC framework. *Discover sustainability*, *5*(1), 319. https://doi.org/10.1007/s43621-024-00539-1

[18] Onwe, J. C., Ridzuan, A. R., Uche, E., Ray, S., Ridwan, M., & Razi, U. (2024). Greening Japan: Harnessing energy efficiency and waste reduction for environmental progress. *Sustainable futures*, *8*, 100302. https://doi.org/10.1016/j.sftr.2024.100302

[19] Danek, T., Weglinska, E., & Zareba, M. (2022). The influence of meteorological factors and terrain on air pollution concentration and migration: a geostatistical case study from Krakow, Poland. *Scientific reports*, *12*(1), 11050. https://doi.org/10.1038/s41598-022-15160-3

[20] Zaręba, M., Danek, T., & Zając, J. (2020). On including near-surface zone anisotropy for static corrections computation—Polish Carpathians 3D seismic processing case study. *Geosciences*, *10*(2), 66. https://doi.org/10.3390/geosciences10020066

[21] Raihan, A., Atasoy, F. G., Coskun, M. B., Tanchangya, T., Rahman, J., Ridwan, M., … & Yer, H. (2024). Fintech adoption and sustainable deployment of natural resources: Evidence from mineral management in Brazil. *Resources policy*, *99*, 105411. https://doi.org/10.1016/j.resourpol.2024.105411

[22] Raihan, A., Ibrahim, S., Ridwan, M., Rahman, M. S., Bari, A. B. M. M., & Guneysu Atasoy, F. (2025). Role of renewable energy and foreign direct investment toward economic growth in Egypt. *Innovation and green development*, *4*(1), 100185. https://doi.org/10.1016/j.igd.2024.100185

[23] Rahman, M., Al Amin, M., Hasan, R., Hossain, S. M. T., Rahman, M. H., & Rashed, R. A. M. (2025). A Predictive AI framework for cardiovascular disease screening in the us: Integrating EHR data with machine and deep learning models. *British journal of nursing studies*, *5*(2), 40–48. https://doi.org/10.32996/bjns.2025.5.2.5

[24] Rahman, M. D. H., Rahaman, M., Arafat, Y., Rahat, S. K. R. U. I., Hasan, R., Rimon, S. M. T. H., & Dipa, S. A. (2025). Artificial intelligence for chronic kidney disease risk stratification in the USA: Ensemble vs. deep learning methods. *British journal of nursing studies*, *5*(2), 20–32. https://doi.org/10.32996/bjns.2025.5.2.3

[25] Breque, M., De Nul, L., & Petridis, A. (2021). European commission, directorate-general for research and innovation. *Industry 5.0: Towards a sustainable, human-centric and resilient european industry*. https://apre.it/wp-content/uploads/2025/06/Shaping-the-future.pdf

[26] Włodyka, E. M. (2024). Implementation of e-government and artificial intelligence in polish public administration. *TalTech journal of european studies*, *14*(2), 118–136. https://doi.org/10.2478/bjes-2024-0019

[27] Poland, S. (2024). *Accessed on May 10 2024*. https://stat.gov.pl/en/

[28] Rahat, S. K. R. U. I., RAHMAN, M. D. H., Arafat, Y., Rahaman, M., Hasan, M. M., & Amin, M. Al. (2025). Advancing diabetic retinopathy detection with ai and deep learning: Opportunities, limitations, and clinical barriers. *British journal of nursing studies*, *5*(2), 01–13. https://doi.org/10.32996/bjns.2025.5.2.1

[29] Khan, M. R. I., Tahsin, M. S., Al Mamun, A., Al-Samad, K., Ahmed, H., Al Wahid, S. K. A., … & Khan, M. R. I. (2025). Multimodal machine learning for proactive supply chain risk management: An AI-driven framework integrating sensor, operational, and external intelligence. *Journal of computer science and technology studies*, *7*(9), 657–673. https://doi.org/10.32996/jcsts.2025.4.1.77

[30] Rahman, M. H., Ansar, M. T. Bin, Hossain, S., Saha, U. S., Imam, H., & Ahsan, I. T. (2025). AI-powered QA in healthcare software: Leveraging predictive analytics and digital twins for safe, cost-effective, and Agile medical systems. *Journal of computer science and technology studies*, *7*(9), 619–628. https://doi.org/10.32996/jcsts.2025.4.1.70

[31] Rahman, M. H., Anwar, M. M., & Hossain, F. (2025). AI-driven big data and business analytics: Advancing healthcare, precision medicine, supply chain resilience, energy innovation and economic competitiveness. *Journal of medical and health studies*, *6*(3), 205–215. https://doi.org/10.32996/jmhs.2025.6.3.30

[32] Ridzuan, A. R., Rahman, N. H. A., Singh, K. S. J., Borhan, H., Ridwan, M., Voumik, L. C., & Ali, M. (2024). Assessing the impact of technology advancement and foreign direct investment on energy utilization in malaysia: An empirical exploration with boundary estimation. *Technology and business model innovation:*

challenges and opportunities (pp. 1–12). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-55911-2_1

[33] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O. (2011). Scikit-learn: Machine learning in Python. *The journal of machine learning research*, *12*, 2825–2830. https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf?source=post_page

[34] Voumik, L. C., Akter, S., Ridwan, M., Pujiati, A., Handayani, B. D., Keshminder, J. S. (2023). Exploring the factors behind renewable energy consumption in Indonesia: Analyzing the impact of corruption and innovation using ARDL model. *International journal of energy economics and policy*, *13*(5), 115–125. https://doi.org/10.32479/ijeep.14530.

[35] Herzen, J., Lässig, F., Piazzetta, S. G., Neuer, T., Tafti, L., Raille, G. (2022). Darts: User-friendly modern machine learning for time series. *Journal of machine learning research*, *23*(124), 1–6. https://www.jmlr.org/papers/volume23/21-1177/21-1177.pdf

[36] Rahman, M. H., Hossin, M. E., Hossain, M. J., Uddin, S. M. M., Faruk, M. I., Anwar, M. M., & Hossain, F. (2024). Harnessing big data and predictive analytics for early detection and cost optimization in cancer care. *Journal of computer science and technology studies*, *6*(5), 278–293. https://doi.org/10.32996/

[37] Hossain, T. (2022). A comparative analysis of adversarial capabilities, attacks, and defenses across the machine learning pipeline in white-box and black-box settings. *Appl res artif intell cloud comput*, *5*(1), 195–212. https://core.ac.uk/download/pdf/613211540.pdf

[38] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *31st conference on neural information processing systems (NIPS 2017), long beach, Ca, USA.* (Vol. 30, pp. 4768–4777). NIPS. https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

[39] Zareba, M., Cogiel, S., Danek, T., & Weglinska, E. (2024). Machine learning techniques for spatio-temporal air pollution prediction to drive sustainable urban development in the era of energy and data transformation. *Energies*, *17*(11), 2738. https://doi.org/1073/17/11/2738

[40] Urbee, A. J., Ridwan, M., & Raihan, A. (2024). Exploring educational attainment among individuals with physical disabilities: A case study in Bangladesh. *Journal of integrated social sciences and humanities*, 1–16. https://doi.org/10.62836/jissh.v1i1.181

[41] Mozumder, M. A. S., Nguyen, T. N., Devi, S., Arif, M., Ahmed, M. P., Ahmed, E., … & Uddin, A. (2024). Enhancing customer satisfaction analysis using advanced machine learning techniques in fintech industry. *Journal of computer science and technology studies*, *6*(3), 35–41. 10.32996/jcsts.2024.6.3.4

[42] Hossain, M. N., Anjum, N., Alam, M., Rahman, M. H., Taluckder, M. S., Al Bony, M. N. V., … & Jui, A. H. (2024). Performance of machine learning algorithms for lung cancer prediction: a comparative study. *International journal of medical science and public health research*, *5*(11), 41–55. https://doi.org/10.37547/ijmsphr/Volume05Issue11-05

[43] Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. *International journal of forecasting*, *37*(4), 1748–1764. https://doi.org/10.1016/j.ijforecast.2021.03.012

[44] Arif, M., Hasan, M., Al Shiam, S. A., Ahmed, M. P., Tusher, M. I., Hossan, M. Z. (2024). Predicting customer sentiment in social media interactions: Analyzing Amazon help Twitter conversations using machine learning. *International journal of advanced science computing and engineering*, *6*(2), 52–56. https://doi.org/10.62527/ijasce.6.2.211

[45] Yang, R. (2018). Omphalos, uber's parallel and language-extensible time series backtesting tool. *Acceso el*, *19*. https://www.uber.com/blog/omphalos/

[46] Tasneem, H., Chowdhury, A. A. A., Sayem, S. S. U., & Emon, A. K. (2025). Exploring the impact of FinTech and IT management on sustainability: Evidence from the United States using ARDL analysis. *Environment, innovation and management*, *1*, 1–22. https://dx.doi.org/10.1142/S3060901125500140

[47] Chowdhury, A. A. A., Sayem, S. S. U., Hossain, T., Emon, A. K., Rafi, A. H., & Alam, S. (2025). Unlocking load capacity potential: The synergy of financial technology and ICT in the United States. *Information sciences and technological innovations*, *2*(1), 57–71. https://doi.org/10.48314/isti.v2i1.35

[48] Holnicki, P., Kałuszko, A., & Nahorski, Z. (2022). Scenario analysis of air quality improvement in Warsaw, Poland, by the end of the current decade. *Atmosphere*, *13*(10), 1613. https://doi.org/10.3390/atmos13101613

[49] Rahman, M. H., Das, A. C., Shak, M. S., Uddin, M. K., Alam, M. I., Anjum, N., … & Alam, M. (2024). Transforming customer retention in fintech industry through predictive analytics and machine learning. *The american journal of engineering and technology*, *6*(10), 150–163. https://doi.org/10.37547/tajet/Volume06Issue10-17

[50] Chojnacka, K. J. (2024). The climate-neutral economy and the air pollution in Poland after 2000: A descriptive and model approach based on the energy market and the industry sector. *Myśl ekonomiczna i polityczna*, *81*(2), 35–55. 10.26399/meip.2(81).2024.08/k.j.chojnacka

[51] Kurasz, A., Lip, G. Y. H., Święczkowski Michałand Tomaszuk-Kazberuk, A., Dobrzycki, S., & Kuźma, Ł. (2025). Air quality and the risk of acute atrial fibrillation (EP-PARTICLES study): A nationwide study in Poland. *European journal of preventive cardiology*, 1–22. https://doi.org/10.1093/eurjpc/zwaf016