Transactions on Energy Systems and Engineering Applications

Article

Relationship between the human development index and the behavior of $PM_{2.5}$ in USA, using multivariate statistics and machine learning

Paul Sanmartin[1], Margarita Castillo[2], Stephanie Galvis[1], Vera Santiago[1], and Alexander Parody[1,*]

[1] Facultad de Ingeniería, Universidad Libre, Barranquilla, Colombia.
[2] Facultad de Ingeniería, Universidad Reformada, Barranquilla, Colombia.
[*] Correspondence: alexandere.parodym@unilibre.edu.co

Received: 21 January 2024; Accepted: 21 October 2024; Published: 04 June 2025

Abstract: : The scientific community has recently shown a rising interest in figuring out how atmospheric pollution affects the behavior of all-encompassing quality-of-life measures. The purpose is to show that bad air quality can have an impact on economic and educational factors in addition to people's health. This research seeks to establish the statistical association between the variables: population, percentage of population at risk of poverty and the Human Development Index (HDI) of the United Nations, with the behavior of the average annual concentration of the pollutant $PM_{2.5}$ in the USA. To achieve this multivariate regression models such as the generalized linear model and the logistic regression model were generated, in addition to generating a Bayesian classifier of neural networks to measure the predictive ability of the variables under study. As a result, the study was able to show that there is a connection between the Human Development Index, population size, and the proportion of the population at risk of poverty, as well as be-tween the average concentration of the pollutant $PM_{2.5}$ and the likelihood that it will exceed the World Health Organization's upper limit. The major finding of the study is that poorer quality of life is related to higher levels of $PM_{2.5}$ pollution concentration in the atmosphere. This is demonstrated by the pollutant's inverse link to both the percentage of the population at risk of poverty and the Human Development Index.

 https://doi.org/10.32397/tesea.vol6.n1.627

1. Introduction

The concern for the environment has been present as a central axis in many research efforts from academia, as well as a central axis in the public policies of many countries around the world, although the efforts should be greater considering the degree of affectation of the planet, which has led to project scenarios that are not very encouraging for the conditions of life on the planet, putting at risk not only the

How to cite this article: Sanmartin, Paul; Castillo, Margarita; Galvis, Stephanie; Santiago, Vera; Parody, Alexander. Relationship between the human development index and the behavior of $PM_{2.5}$ in USA, using multivariate statistics and machine learning. *Transactions on Energy Systems and Engineering Applications*, 6(1): 627, 2024. DOI:10.32397/tesea.vol6.n1.627

fauna and flora of the planet, but also the very existence of humanity itself, as important resources for life, such as fertile land, drinking water and clean air, are affected, It is precisely the study of the latter that is the central purpose of this research, as air quality has gained great relevance in recent years, especially due to the improvement in the capacity to measure air quality in countries that are not part of the major economies of the world, which has allowed us to have relevant.

Over time, countries around the world have sought to improve the performance of their main indicators to show greater growth of their economies. This has presented a growing environmental concern, different researchers have presented works where the existence of an environmental Kuznets curve (CKA) for different countries is evidenced, where the close relationship between economic growth and atmospheric pollution is observed. The United States is not the exception, it is a country with a great economic growth, which today is considered a world power and currently does not have the best air quality [1].

Analysts consider that the economic bonanza that the U.S. has experienced maintains a close relationship with the increase of polluting gases [2]. And that is air quality in the United States is declining dramatically, leaving some 150 million people, nearly half of the country's population, breathing unhealthy and heavily polluted air, according to the "State of the Air" 2020 report [3, 4]. Also, researchers estimate that $PM_{2.5}$ (Fine particles $PM_{2.5}$ refer to inhalable particles with diameters that are generally $_{2.5}$ micrometers or less) is responsible for nearly 48,000 premature deaths in the United States [5].

One factor affecting air quality in the United States is the rise in temperatures and the disruption of short- and long-term climate patterns, as the impact of the climate change being experienced includes an increase in extreme weather events, deterioration of air quality due to heightened ozone formation, and smoke from wildfires [6, 7].

The concentration of $PM_{2.5}$ (particulate matter) in U.S. air is currently 1.9 times higher than the annual value of the WHO (World Health Organization) air quality guidelines [8]. Fine particulate matter, or PM $_{2.5}$, is the smallest and, in turn, one of the most dangerous pollutants. When inhaled, it travels into lung tissue, where it can enter the bloodstream and can contribute to asthma, cardiovascular disease, and other respiratory illnesses [9].

A clear example of a country with high economic growth and air quality problems is China. The root of China's environmental problems lies in its spectacular economic growth. For more than 35 years, the country has increased its Gross Domestic Product (GDP) at an annual average of close to 10%, mainly due to the boost of investment, industry, and infrastructure construction, but this has affected its air quality since all these activities require raw materials capable of generating energy, which in turn pollute the air [10–12].

The environmental Kuznets curve (CKA) suggests the existence of a relationship between economic growth and environment. As economies enjoy greater economic growth, environmental degradation occurs. This can be evidenced in a study conducted in Latin America and the Caribbean, where they studied the relationship between economic growth and environmental degradation, it was found that for the period 1970-2008 CO (Carbon Monoxide) and HC (Hydrocarbons) emissions were related to the economic growth of these countries [1].

Even though the United States plays an important role in the economy, it is not common to find studies that show the correlation between economic growth and air pollution. For this reason, it is necessary to conduct studies to determine how different economic indicators are related to air quality. This research seeks to establish, by means of statistical models, the effect of macroeconomic variables on pollutants, $PM_{2.5}$, to provide information of great importance for the design of strategies by local and national environmental authorities, which will allow them to generate contingency plans in the face of possible increases in pollution levels, since the future behavior of macroeconomic variables can be known with great accuracy.

A clear example of a country with high economic growth and air quality problems is China. The root of China's environmental problems lies in its spectacular economic growth. For more than 35 years, the

country has increased its Gross Domestic Product (GDP) at an annual average of close to 10%, mainly due to the boost of investment, industry, and infrastructure construction, but this has affected its air quality since all these activities require raw materials capable of generating energy, which in turn pollute the air [13, 14].

The environmental Kuznets curve (CKA) suggests the existence of a relationship between economic growth and environment. As economies enjoy greater economic growth, environmental degradation occurs. This can be evidenced in a study conducted in Latin America and the Caribbean, where they studied the relationship between economic growth and environmental degradation, it was found that for the period 1970-2008 CO (Carbon Monoxide) and HC (Hydrocarbons) emissions were related to the economic growth of these countries [1].

Even though the United States plays an important role in the economy, it is not common to find studies that show the correlation between economic growth and air pollution. For this reason, it is necessary to conduct studies to determine how different economic indicators are related to air quality. This research seeks to establish, using statistical models, the effect of macroeconomic variables on the pollutant, $PM_{2.5}$, to provide information of great importance for the design of strategies by local and national environmental authorities, which will allow them to generate contingency plans in the face of possible increases in pollution levels, since the future behavior of macroeconomic variables can be known with great accuracy.

2. Methods and Materials

This is a descriptive correlational study that aims to the relationship between macroeconomic variables such as: Population, Percentage of Population at Risk of Poverty and the Human Development Index (HDI), on the behavior of $PM_{2.5}$ in the United States, multivariate statistical models such as the multivariate linear regression model and the logistic regression model were applied, added to a Machine Learning tool such as the Bayesian classifier of neural networks in order to measure the predictive capacity of these variables on the non-compliance of the maximum levels of $PM_{2.5}$.

2.1. Sample

The data were collected between the years 1999 and 2019, excluding the years 2020 and 2021 due to their anomalous nature caused by the COVID-19 pandemic. All variables under study were annual values, both for the independent variables: Human Development Index, population, and percentage of population at risk of poverty, and the dependent variables: average annual concentration of $PM_{2.5}$ and non-compliance with the maximum annual concentration of $PM_{2.5}$ according to the WHO. All data correspond to the most populated states in each of the 9 census zones of the United States (excluding Hawaii and Puerto Rico). These states were: Arizona, California, Florida, Illinois, Iowa, Massachusetts, New York, Tennessee, and Texas, covering states across the country. For the construction of the logistic regression model and the training of the neural network, data from the state of California were used, as it was the only state in the sample that showed non-compliance with the maximum allowed levels of PM2.5 during the years under study.

2.2. Multivariate Models

Two highly distinctive multivariate statistics tools—the multivariate linear regression model and the logistic regression model—were used as part of the research's methodology. When examining the impact of several quantitative variables on a continuous quantitative variable, the multivariate linear regression model is strongly advised [15]. On the other hand, the Logistic Regression Model is helpful for examining how a collection of qualitative and quantitative variables can influence the likelihood that a qualitative variable would be dichotomous [16, 17]. To comprehend the unique impact of each macroeconomic component on the average annual concentration of $PM_{2.5}$, the multivariate linear regression model was utilized. To evaluate the effect of each variable under study on the probability of exceeding the maximum PM2.5 levels, the Logistic Regression Model was used.

2.3. Neural Network

To pinpoint the variables that are statistically important regarding $PM_{2.5}$ behavior, multivariate statistical models were used. The factors that were determined to be significant in both models were utilized to train and later validate a crucial piece of machine learning software called a neural network. A Bayesian classifier of neural networks was applied in this instance. When predicting the occurrence of a qualitative variable using quantitative variables, this machine learning technique is used. Based on a region centered on nearby observations of the dependent variable, this classifier calculates a probability density function of the data related to the response variable. The estimated density function, the prior likelihood of falling into any of the categories for the dependent variable, and the cost of forecast error are used as the input parameters for the model's operation [16–18].

2.4. Statistical Methodology

The statistical method created for the research is graphically depicted here.

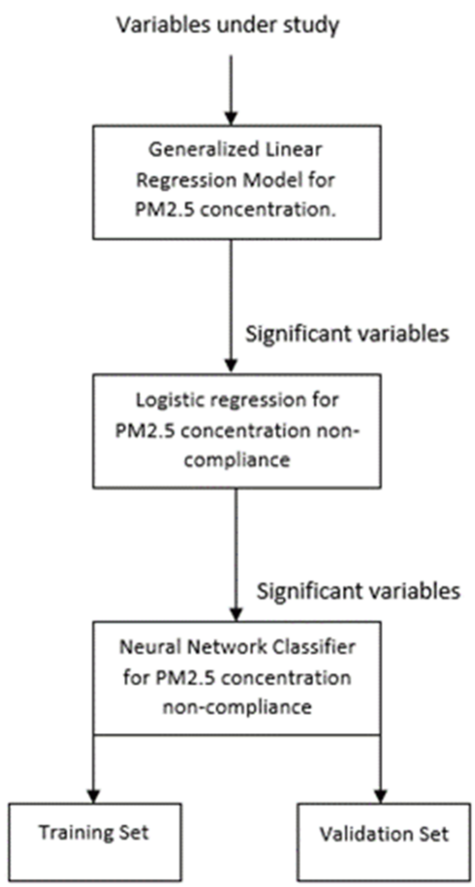

Figure 1. Application of statistical methodology.

Statgraphics XVI software was used for the development of the selected regression models and the training of the Bayesian neural network classifier.

3. Results

3.1. Multivariate Linear Regression Model

As an initial part of the results developed in the research, a multivariate linear regression model was generated to establish whether the variables under study have a statistically significant relationship with the behavior of the average $PM_{2.5}$ concentration.

Table 1. Sum of Square Type III.

Source	Sum of Square	Df	Mean Square	F-Ratio	P-value
HDI	259,615	1	259,615	71.49	0.0000
Population	74,1736	1	74,1736	20.43	0.0000
Poverty Risk Percentage	154,983	1	154,983	42.68	0.0000
Residual	533,813	147	3,63138		
Total (corrected)	889,819	150			

Table 2. 95.0% confidence intervals for coefficient estimates.

Parameter	Estimate	Standard Error	Lower Limit	Upper Limit	V.I.F.
CONSTANT	73,8089	7,11263	59,7526	87,8651	1,0565
HDI	-64,651	7,64621	-79,7617	-49,5403	1,0565
Population	7.23435E-8	1,6007E-8	4.07098E-8	1.03977E-7	1,14091
Poverty Risk Percentage	-40,1506	6,14559	-52,2963	-28,0048	1,17687

From Table 1, which represents the sum of squares type III of the multivariate linear regression model developed, it can be concluded that all the variables under study are related to the average $PM_{2.5}$ concentration, since the p-values obtained are less than 0.05. Furthermore, the adjusted R-squared indicates that the model explains 40% of the variability evidenced in the $PM_{2.5}$ concentrations throughout the years under study.

The mathematical model obtained is as follows:

$$\begin{aligned} \mathrm{PM}_{2.5}\ \text{annual average} = {} & 73,8089 - 64,651 \times \text{HDI} \\ & + 7,23435E - 8 \times \text{Population} \\ & - 40,1506 \times \text{Poverty Risk Percentage} \end{aligned} \tag{1}$$

The generated regression model showed an average forecast error of just -2.3×10^{-7}, following a normal distribution (Shapiro-Wilk p-value of 0.5092). Additionally, there is complete independence among the variables used to build the model, as indicated by the obtained Variance Inflation Factor (V.I.F.) values, which are significantly below the maximum allowed value of 10 (Table 2).

To determine the influence of the variables under study on the annual average behavior of $PM_{2.5}$, graphs of the fitted model were generated to visually illustrate the pollutant's behavior in the environment with respect to the values of statistically significant independent variables:

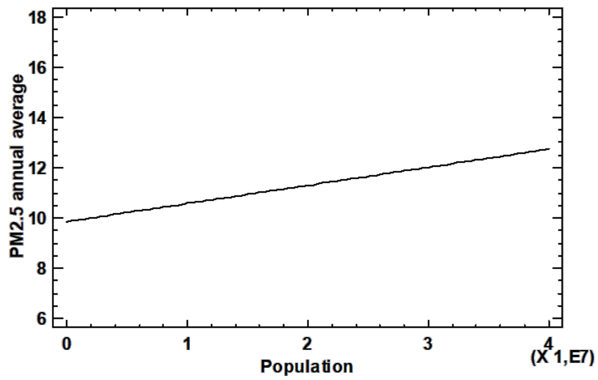

Figure 2. $PM_{2.5}$ annual average vs Population.

Figure 2 confirms a similar trend in PM2.5 growth with the increasing population of the cities under study, mirroring the pattern observed in the population density graph.

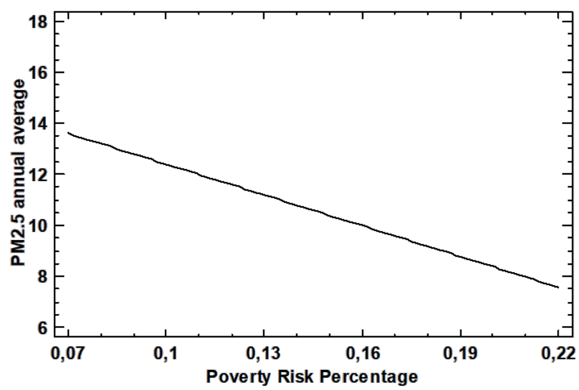

Figure 3. $PM_{2.5}$ annual average vs Poverty Risk Percentage.

In Figure 3, an inversely proportional relationship is noticeable between the annual average concentration of $PM_{2.5}$ and the percentage of the population at risk of poverty. A higher percentage of the population below the poverty threshold corresponds to lower $PM_{2.5}$ pollution, likely due to reduced industrial and commercial activity in areas with higher poverty rates, and possibly a decrease in vehicular traffic, leading to lower pollution.

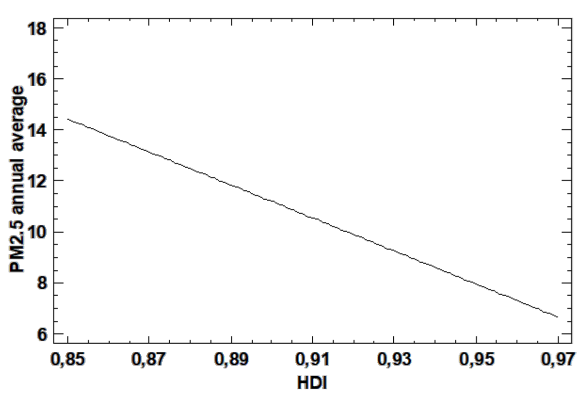

Figure 4. $PM_{2.5}$ annual average vs HDI.

Finally, Figure 4 illustrates the relationship between the Human Development Index (HDI) and the annual average concentration of $PM_{2.5}$. Like Figure 3, an inversely proportional relationship is evident, indicating that a higher quality of life for the population is associated with lower pollution levels. This result supports the idea that striving for a higher overall quality of life for the population (health, education, and economy) is crucial for reducing environmental pollution.

3.2. Logistic Regression Model

To determine which values of the variables under study most increase the probability of non-compliance with the maximum $PM_{2.5}$ concentration, a logistic regression model was applied in which the dependent variable was the non-compliance or non-compliance with the maximum annual permitted levels of $PM_{2.5}$ logistic regression model was applied where the dependent variable was the non-compliance or non-compliance with the maximum annual permitted levels.

Table 3. Likelihood Ratio Tests.

Factor	Chi-Square	Df	P-value
Poverty Risk Percentage	22,7748	1	0,000
Population	8,50934	1	0,003
HDI	31,2489	1	0,000

Table 3 shows the p-value of the likelihood ratio tests for each variable under study, the model presents a percentage of explanation of the dependent variable of 21%. To determine the cut-off values that significantly increase the probability of non-compliance with $PM_{2.5}$ values in each of the variables that were statistically significant, graphs of the adjusted logistic model were generated.

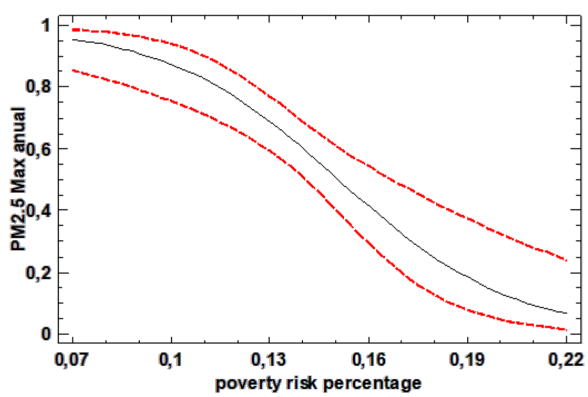

Figure 5. Adjusted logistic model for percentage at risk of poverty.

From Figure 5 it can be seen that when the percentage of the population at risk of poverty is below 13% (0.13 in proportion) the probability of not complying with the maximum levels of $PM_{2.5}$ decreases visibly, this fact allows establishing that there is an inverse relationship between human development and the behavior of the concentration of the pollutant $PM_{2.5}$ in the city of Los Angeles, a fact that will allow generating interesting conclusions associated with the effect of increasing the quality of life of the people in a population and its interaction with air quality.

From Figure 6 the probability of not complying with the maximum $PM_{2.5}$ levels increases almost linearly as the population increases.

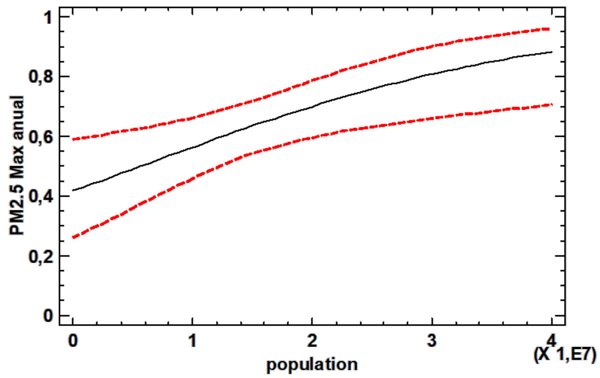

Figure 6. Logistic model fitted for population size.

Figure 7 shows that when HDI values are above 0.89, the probability of not complying with the maximum levels of $PM_{2.5}$ decreases quite markedly. This allows establishing that the efforts to improve the quality of life in a comprehensive manner in the inhabitants have a direct or indirect impact on the behavior of the concentration of the pollutant $PM_{2.5}$ in the environment, this fact is interesting because in the calculation of the Human Development Index of the United Nations the Gross Domestic Product participates and this is formed by the amount of products and services that are generated in a region, therefore this relationship shows that there can be an increase in human development without this meaning that the amount of particulate matter in the environment increases, which undoubtedly becomes an interesting bet in search of social development in a sustainable manner.

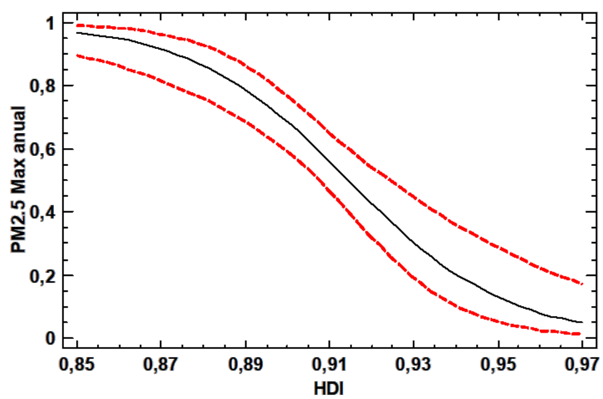

Figure 7. Adjusted logistic model for Human Development Index.

As a final part of the analysis, a Bayesian classifier of neural networks was applied to measure the predictive capacity that the variables under study would have when forecasting a non-compliance with the maximum annual permitted levels of $PM_{2.5}$ in the USA.

3.3. Neural Network

Finally, the results of the application of a Bayesian classifier of neural networks are presented. This tool will complement the conclusions reached through the multivariate linear regression model and the logistic regression model, where these models allowed not only to identify a statistically significant relationship between the variables analyzed and the concentration of $PM_{2.5}$, but through this artificial intelligence tool it is possible to establish the prognostic capacity of these variables in order to make future projections of non-compliance with the maximum permitted levels of $PM_{2.5}$ concentrations, but through this artificial intelligence tool it is possible to establish the prognostic capacity of these variables to make future projections of non-compliance with the maximum permitted levels of PM2.5 based on the behavior of the economic variables analyzed.

Table 4. Training Set.

$PM_{2.5}$ Max Annual	Members	% Correct Classification
No	49	88
Yes	73	90
Total	122	89

Table 5. Validation Set.

$PM_{2.5}$ Max Annual	Members	% Correct Classification
No	12	75
Yes	19	95
Total	31	87

Prior probabilities: proportional to occurrence in training set
Error costs: equal for all classes
Number of cases in training set: 122
Number of cases in validation set: 31
Spacing parameter used: 0,0492188 (optimized by jackknifing during training)

It is important to note that the data used for the construction of the network and its validation were randomly selected, thereby avoiding the impact of temporal dependencies on the performance of the neural network validation.

Table 4, which corresponds to the results of the training set of the neural network developed, and Table 5, which represents the results of the validation set of the model's forecasting capacity, show the excellent performance of the variables under study when forecasting the occurrence of non-compliance with the maximum permitted levels of $PM_{2.5}$, where in the training set a 89% correct classification of the cases of both compliance and non-compliance with the maximum permitted levels was achieved; it should be noted that 90% of the occurrences of non-compliance were correctly classified. The validation set showed that the neural network correctly classified 87% of the cases and that 95% of the occurrences of non-compliance with the maximum $PM_{2.5}$ levels were correctly identified. These results allow us to identify that a correct projection of the variables: poverty risk, population, and the human development index (HDI) could be sufficient to predict the future occurrence of emergencies due to exceeding the maximum permitted concentration levels of $PM_{2.5}$ particulate matter.

Next, Figure 8 shows the network graph generated by the Bayesian classifier of neural networks, being this graph a way to visualize the number of nodes that are generated in the neural network, in which a constant activity of minimization of the result in the cost function of the forecast error is developed, such optimization allows generating coefficients for the base mathematical model that decrease the expected value of error for the generated model.

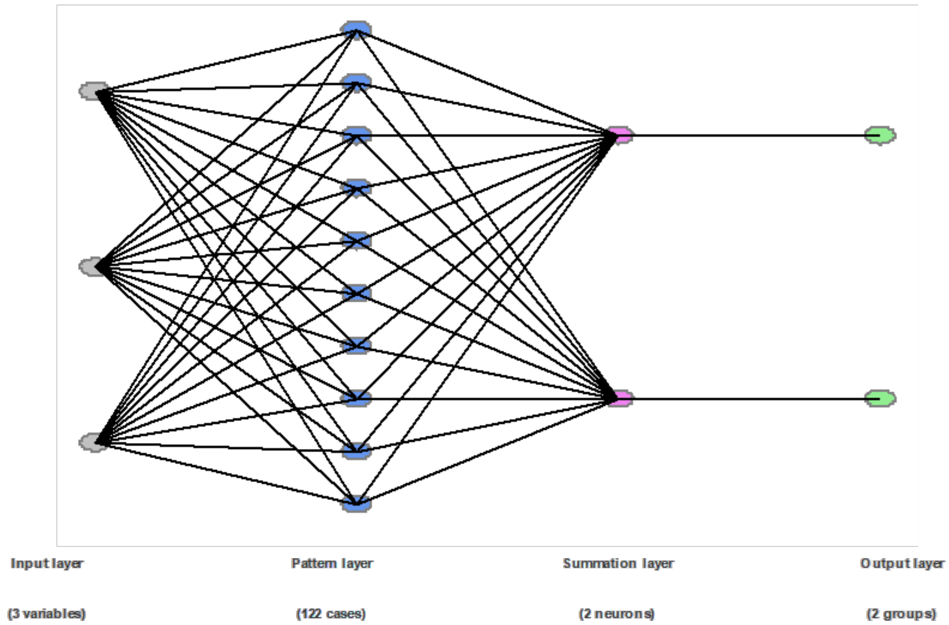

Figure 8. Network graph for Bayesian neural network classifier.

4. Conclusions

With the development of this research, a statistically significant relationship was established between economic variables such as population, percentage of population at risk of poverty, and the Human Development Index (HDI) with the behavior of the average concentration of the $PM_{2.5}$ pollutant, through mathematical models such as multivariate linear regression. Additionally, through a logistic regression model, it was possible to establish how each of the significant variables influenced the probability of exceeding the maximum levels set by the World Health Organization (WHO) in the concentration of the evaluated pollutant. It was evident that the variables analyzed in this study could be sufficient to predict the behavior and non-compliance with the maximum levels of $PM_{2.5}$ pollutant concentration, as these economic indices are susceptible to projection. Therefore, it could not only predict the short-term economic behavior of a country but also anticipate increases in $PM_{2.5}$ concentration. This information could assist environmental authorities, municipalities, and risk management offices in decision-making. Furthermore, this type of information could be linked to contingency plans for atmospheric pollution events and establish measures to contribute to the reduction of this pollutant, thereby reducing its impact on the health of the population.

This project allowed establishing the nature of the relationship between variables as important as the Human Development Index (HDI) and the behavior of the $PM_{2.5}$ pollutant, allowing the identification of an inverse relationship between these variables. This, in turn, affirms that the direction of public policies aimed at improving people's living conditions, expressed in terms of achieving higher levels of education, increasing life expectancy, and obtaining increasingly higher results in the country's gross domestic product, will be associated with lower levels of $PM_{2.5}$ concentration. This establishes that efforts to improve the quality of life of people also include the pursuit of actions leading to the reduction of atmospheric pollution through air quality improvement strategies.

The results obtained in this research align with the conclusions of the systematic review conducted by Jackson in 2020, where, after analyzing 178 studies associated with the analysis of psychological, economic, and social effects of air pollution, it was established that it affects labor productivity and market behavior, significantly impacting the quality of life of citizens [19]. These findings also align with the results of a study conducted in India in 2019, supported by the Bill & Melinda Gates Foundation, which highlighted how atmospheric pollution negatively affected India's gross domestic product, impacting the quality of life of the population [20]. These results also align with studies on quality of life and human development, where the increase in air pollution hinders proper human development in populations, especially $PM_{2.5}$ [20–22]. Even in developing countries, the increase in pollutant gas emissions is associated with an increase in poverty [23], This is consistent with the findings of the presented article, where it was shown that a higher population at risk of poverty correlates with higher levels of particulate matter in the environment. In general terms, the relationship between economic poverty and environmental impact is established [24–26], where greater economic inequality leads to more population exposure to adverse environmental conditions. Therefore, the pursuit of sustainable economic alternatives can become a driver of human development, especially in developing nations.

This research was limited to a single country, in this case, the United States, analyzing nine states: Arizona, California, Florida, Illinois, Iowa, Massachusetts, New York, Tennessee, and Texas, with the latter being selected for the application of statistical models, as it exhibited an increase above WHO-accepted levels for $PM_{2.5}$ during the analyzed time period, allowing for comparison with macroeconomic indices. For this reason, it is recommended for future research to include other countries with different levels of development to determine if the same interactions occur between $PM_{2.5}$ particles and macroeconomic variables associated with the quality of life of the population.

Finally, the article was able to demonstrate that classical multivariate statistical models, such as multivariate linear regression and logistic regression models, can be used simultaneously with Machine Learning tools, such as the Bayesian classifier of neural networks, to explain the relationship between air quality and macroeconomic indicators. With classical models, it is possible to understand the nature of the relationship between the variables under study, and through Machine Learning, generate prediction models with high reliability, in addition to establishing the predictive capacity of the statistically significant variables.

Funding: This research received no external funding.

Author contributions: The authors of the article had an active participation in its construction, with their contributions distributed as follows: Conceptualization, Alexander Parody and Margarita Castillo.; Methodology, Alexander Parody.; Software, Paul Sanmartin; Formal Analysis, Alexander Parody and Stephanie Galvis; Investigation, Stephanie Galvis; Resources, Margarita Castillo; Data Curation, Alexander Parody; Writing – Original Draft Preparation, Alexander Parody and Stephanie Galvis; Writing – Review & Editing, Vera Santiago and Paul Sanmartin; Supervision, Alexander Parody.; Project Administration, Alexander Parody; Funding Acquisition, Alexander Parody.

Disclosure statement: The authors declare no conflict of interest.

References

[1] H. Briceño Scremin and G. Díaz Clavo. Relationship between the increase in Carbon Dioxide (CO2) emissions and the growth of GDP per capita in the US and Brazil (1960-2014) based on the Environmental Kuznets Curve. *Universidad Católica Andrés Bello*, 2019.

[2] D. Alandete. Pollution in the united states reaches its maximum in eight years. *ABC*, 2019. [last accessed: 10 diciembre 2021].

[3] U.S. Environmental Protection Agency. Integrated Science Assessment (ISA) for Particulate Matter (Final Report). Technical report, U.S. EPA, Washington, DC, 2019.

[4] S. LaMotte. Air quality in the US is dramatically worse than in previous years, says new "State of the Air" report. *CNN Español*, 2020. [last accessed: 10 diciembre 2021].

[5] Health Effects Institute. State of Global Air 2020: A Special Report On Global Exposure To Air Pollution And Its Health Impacts. Technical report, Health Effects Institute, Boston, 2020.

[6] American Lung Association. Health impact of air pollution. *American Lung Association*, 2021. [last accessed: 03 octubre 2021].

[7] American Lung Association. State of the Air 2021. Technical report, American Lung Association, 2021.

[8] IQAir. Air quality index (AQI) and PM2.5 air pollution in the United States. Technical report, IQAir, Goldach, 2021.

[9] R. Ramirez. Millions of deaths could be prevented by meeting new air quality guidelines, says WHO, 2021. [last accessed: 10 diciembre 2021].

[10] D. Méndez. 10 reasons why China is so polluted, 2014. [last accessed: 11 diciembre 2021].

[11] Xiaolan Li, Yanjun Ma, Yangfeng Wang, Ningwei Liu, and Ye Hong. Temporal and spatial analyses of particulate matter (PM 10 and PM 2.5) and its relationship with meteorological parameters over an urban city in northeast China. *Atmospheric Research*, 198:185–193, December 2017.

[12] Xiaoyu Qi, Gang Mei, Salvatore Cuomo, Chun Liu, and Nengxiong Xu. Data analysis and mining of the correlations between meteorological conditions and air quality: A case study in Beijing. *Internet of Things*, 14:100127, June 2021.

[13] Alexander von Lühmann, Antonio Ortega-Martinez, David A. Boas, and Meryem Ayşe Yücel. Using the General Linear Model to Improve Performance in fNIRS Single Trial Analysis and Classification: A Perspective. *Frontiers in Human Neuroscience*, 14, February 2020.

[14] Ping Wang, Junji Cao, Xuexi Tie, Gehui Wang, Guohui Li, Tafeng Hu, Yaoting Wu, Yunsheng Xu, Gongdi Xu, Youzhi Zhao, Wenci Ding, Huikun Liu, Rujin Huang, and Changlin Zhan. Impact of Meteorological Parameters and Gaseous Pollutants on PM2.5 and PM10 Mass Concentrations during 2010 in Xi'an, China. *Aerosol and Air Quality Research*, 15(5):1844–1854, 2015.

[15] Abhik Das. *Logistic Regression*, page 1–2. Springer International Publishing, 2021.

[16] E. J. D. L. H. Granadillo, J. M. Gómez, and L. L. Polo. Evaluación de la competitividad comercial del sector industrial colombiano mediante el coeficiente de apertura exportadora. *Aglala*, 10(1):180–195, 2019.

[17] P. Chiarvetto, L. Lourdes, F. Saravia, and N. Brignole. Aplicación de redes neuronales artificiales para la predicción de calidad de aire. 2008.

[18] Jessie María Ramírez Suárez. Modelo de predicción de la calidad del aire a partir de datos meteorológicos e información del tráfico automovilístico, 2018.

[19] Jackson G Lu. Air pollution: A systematic review of its psychological, economic, and social effects. *Current Opinion in Psychology*, 32:52–65, April 2020.

[20] A. Pandey, M. Brauer, M. L. Cropper, K. Balakrishnan, P. Mathur, S. Dey, and L. Dandona. Health and economic impact of air pollution in the states of India: the Global Burden of Disease Study 2019. *The Lancet Planetary Health*, 5(1):e25–e38, 2021.

[21] Andrew Adewale Alola, Edmund Ntom Udemba, Chikaodinaka Iwuagwu, and Ibrahim Abdallah. Assessing the human development aspects of CO, PM2.5, PM10, NOX, and SO2 in the United States. *Heliyon*, 9(7):e18072, July 2023.

[22] Halil Altıntaş and Yacouba Kassouri. Is the environmental Kuznets Curve in Europe related to the per-capita ecological footprint or CO2 emissions? *Ecological Indicators*, 113:106187, June 2020.

[23] Muhammad Awais Baloch, Danish, Salah Ud-Din Khan, Zübeyde Şentürk Ulucak, and Ashfaq Ahmad. Analyzing the relationship between poverty, income inequality, and CO2 emission in Sub-Saharan African countries. *Science of The Total Environment*, 740:139867, October 2020.

[24] Kizito Uyi Ehigiamusoe, Muhammad Tariq Majeed, and Eyup Dogan. The nexus between poverty, inequality and environmental pollution: Evidence across different income groups of countries. *Journal of Cleaner Production*, 341:130863, March 2022.

[25] Salim Khan and Wang Yahong. Income inequality, ecological footprint, and carbon dioxide emissions in Asian developing economies: what effects what and how? *Environmental Science and Pollution Research*, 29(17):24660–24671, November 2021.

[26] Atif Awad and M. Hersi Warsame. The poverty-environment nexus in developing countries: Evidence from heterogeneous panel causality methods, robust to cross-sectional dependence. *Journal of Cleaner Production*, 331:129839, January 2022.