AIR QUALITY AND URBAN SUSTAINABLE DEVELOPMENT: THE APPLICATION OF MACHINE LEARNING TOOLS

Nidia Isabel Molina-Gómez[1,2], José Luis DíazArévalo[3] , P. Amparo López-Jiménez[2]

[1.]Department of Environmental Engineering, Universidad Santo Tomás, Bogotá, Colombia
[2.]Hydraulic and Environmental Engineering Department, Universitat Politècnica de València, Valencia, Spain
[3.]Department of Civil and Agricultural Engineering, Universidad Nacional de Colombia, Bogotá, Colombia

ABSTRACT

Air quality has an effect on a population's quality of life. As a dimension of sustainable urban development, governments have been concerned about this indicator. This is reflected in the references consulted that have demonstrated progress in forecasting pollution events to issue early warnings using conventional tools which, as a result of the new era of big data, are becoming obsolete. There is a limited number of studies with applications of machine learning tools to characterize and forecast behavior of the environmental, social, and economic dimensions of sustainable development as they pertain to air quality. This article presents an analysis of studies that developed machine learning models to forecast sustainable development and air quality. Additionally, this paper sets out to present research that studied the relationship between air quality and urban sustainable development to identify the reliability and possible applications in different urban contexts of these machine learning tools. To that end, a systematic review was carried out, revealing that machine learning tools have been primarily used for clustering and classifying variables and indicators according to the problem analyzed, while tools such as artificial neural networks and support vector machines are the most widely used to predict different types of events. The non-linear nature and synergy of the dimensions of sustainable development are of great interest for the application of machine learning tools.

Key words: air pollution, sustainability, forecasting, sustainable development goals, and influencing variables.

1 INTRODUCTION

Compatible, socially just, and economically viable ecological development is at the heart of the concept of sustainable development (Mellos 1988). This term has been more widely recognized since the Earth Summit, where it was established that it is essential for development to meet the needs of the present without compromising the ability of future generations to meet their own needs (WCED 1987). Sustainable development is a goal of the global agenda, which nations have pursued since the Millennium Development Goals and currently with the Sustainable Development Goals (SDGs). These set targets that should be integrated into countries' national development plans. In this regard, considering that it is of special concern for nations to achieve the SDGs, there is great interest in forecasting SD behavior. Nations need to make progress in forecasting their territorial behavior and consider indicators' variability in the scope of sustainable development. A future vision of this behavior for the cities and regions that comprise nations is a valuable tool for decision-makers.

Sustainable development (SD henceforth) is supported by three encompassing dimensions; the environmental, social, and economic dimensions. This concept includes environmental health, social equality and economic growth as part of the dimensions' interactions, which on balance, are difficult to measure (Lubell et al. 2009). Any change of any one of those interactions could impact SD, demonstrating that balancing and integrating these dimensions is challenging in the quest for sustainability. In this sense, it is important to mention that greater than 55% of the world population lives in urban zones (United Nations 2019). As a result, air, water and soil quality, species extinctions, worsening health of populations, poverty, among others, all impact SD, specifically urban sustainable development.

Air pollution is one of the effects of great concern at the global level. It is included in the SDGs and may become the main environmental cause of premature mortality (Cruz et al. 2017). A study performed by Krzyzanowski et al. (2014) identified that at least 96% of the population of large cities is exposed to particulate matter smaller than 2.5 micron ($PM_{2.5}$). This study also identified that cities with higher concentrations of particulate matter and the lowest air quality improvement rates over the last decade, tend to be countries with lower levels of economic development. Furthermore, 92% of the world's population lives in places where the air quality exceeds standards established by the World Health Organization (WHO) (WHO 2016). Moreover, the annual global mortality of nearly 3 million people is related to exposure to outdoor air pollution (WHO 2016). Instruments have been developed to monitor and forecast atmospheric pollutants in order to control these correlations. Several studies have applied machine learning (ML) tools to forecast air quality and certain pollutants, primarily particulate matter smaller than 10 and 2.5 micrometers (PM_{10} and $PM_{2.5}$). For example, Antanasijević et al. (2013) and Paas et al.

(2017) applied artificial neural networks (ANN), Karimian et al. (2019) and Zhou et al. (2019) used deep neural networks (DNN), and Oprea et al. (2016) and Wang (2019) utilized decision trees (DT) or their ensembles. Furthermore, support vector machine (SVM) models were developed for the spatio-temporal predictions of $PM_{2.5}$ (Song et al. 2014; de Hoogh et al. 2018).

Machine learning tools are instruments that forecast the behavior of different variables considering large volumes of data and in many cases, substantial quantities of predictors. With several advantages over conventional models, the use of ML has advanced. Different documents on data mining and ML extensively describe the algorithms and their applications (Brink et al. 2016; Lässig et al. 2016). Those studies, which applied conventional models and ML tools, analyzed and forecasted the behavior of pollutants for certain regions and territories. The principal aim has been to provide useful technological tools and results for decision-makers in order to protect populations' health. Nevertheless, these studies have not identified air pollution's impacts on SD. Learning the influence of air pollution on nations' sustainability, with respect to mortality rates and the impacts of air pollution on the world population is of great interest.

Machine learning tools are useful for understanding the impacts of air pollution on the sustainable development of territories. As such, the primary objective of this paper is to analyze the work developed in prior studies, which have used ML tools to forecast air quality and SD. This study aims to identify tools that can closely relate both concepts in order to identify the state of the art ML applications, existing gaps in the field, and determine aspects that can be addressed in the future. This research paper develops its analysis on the following questions: which machine learning tools have been applied to forecast the sustainable development behavior of territories, and which machine learning tools could be applied to identify the influence of air pollution on sustainable development?

This study is of special interest for governments, scientific communities, and societies. It is innovative in that it provides tools for environmental governance. Sustainable development is a challenge for governments as it requires an understanding of each component's behavior, the sustainability dimensions and the territories' evolution in terms of the SDGs. Additionally, this study provides specific information for decision-makers by analyzing the application of machine learning tools where technology and development issues meet.

This paper is structured into four sections; following this introduction, the second section *Materials and Methods*, describes the procedures applied in the systematic revision. The next section introduces the results through a description of the different ML tools employed in various studies to forecast air quality and SD levels. This section contains the results and subsequent discussion of this review, evaluating the tools' functionality based on the requirements for their application in the integration of both concepts. This is followed by a section that analyzes the gaps and proposed future developments. The paper closes with the conclusions section.

2 MATERIALS AND METHODS

This study was developed based on the following four relevant stages: 1) determining the research question and establishing the criteria for exhaustive literature selection; 2) preparing a dataset with the research documents collected in stage 1; 3) researching and evaluating records, and 4) analyzing the selected literature. Figure 1 presents the methodological framework used for this research study.

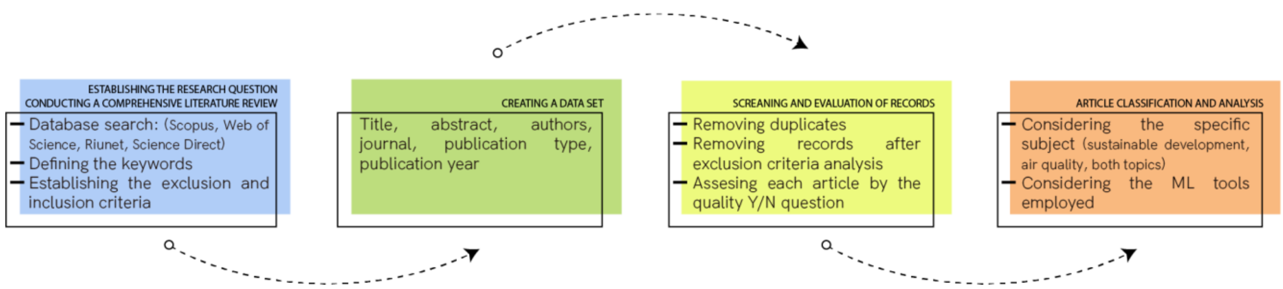

Fig. 1 Methodological framework to understand the relationship between air quality and sustainable development for forecasting.

2.1 ESTABLISHING THE RESEARCH QUESTION AND CONDUCTING A COMPREHENSIVE LITERATURE REVIEW

Sustainable urban development encompasses the integration of three fundamental pillars and their interactions. Identifying the machine learning tools used to forecast sustainable development in territories will enable the identification of a set of tools that could be implemented in other territories with similar characteristics. Furthermore, this identification would be able to guide the combination of ML methods to solve specific problems in each SD dimension, particularly related to air pollution. Therefore, to answer the research question, the criteria for searching and choosing scientific literature have been established as is described in Table 1.

Table 1. Primary aspects to identify scientific literature in the framework of this study

Database	Keywords and equation used in the search	Inclusion and exclusion criteria
- Scopus - Web of Science (WoS) - Riunet	((("machine learning" OR "data mining" OR "pattern recognition") AND ("air quality" OR "air pollution" OR "atmospheric pollution" OR "atmosphere" OR "atmosp*") AND ("sustainable development" OR "sustainability" OR "sustainable" OR "sustain*")))	**Date:** Last 20 years (2000-2019) **Publication type:** Articles, reviews, doctoral theses, conference papers. **Language:** English **Peer reviewed publications:** Articles, reviews, doctoral

Database	Keywords and equation used in the search	Inclusion and exclusion criteria
Science Direct	((("machine learning" OR "data mining") AND ("air quality" OR "air pollution" OR "atmospheric pollution" OR atmosp) AND ("sustainable development" OR sustain)))	theses, conference papers published in a peer reviewed journals or conference proceedings. **Publications' objective:** The primary goal of the publication is forecasting sustainable development OR air quality OR atmospheric pollution with machine learning tools. Only studies that strictly entailed the use of machine learning tools in both sustainable development and air quality were included.

The databases identified in Table 1 were the primary sources for data collection. Additionally, the search engine of the Universitat Politècnica de València was used based on the criteria outlined in Table 1.

2.2 CREATING A DATA SET FOR THE RECORDED INFORMATION

A template was designed for the inclusion of each of the records gathered by the searches. In addition to specific information identifying each record, the information base included the identification of the machine learning tools used in the studies, their use in each case, the tools utilized in processes to train and validate the data for forecasting, and each study's analysis objective: sustainable development, air quality according to pollutants or sustainable development dimensions, in the event that the study focused on specific dimensions.

2.3 SCREENING AND EVALUATION OF RECORDS

Once the data set was created based on the criteria established in Table 1, the identified records were reviewed. The information was verified by reading abstracts of each study. Duplicate records in the databases were eliminated, as well as records that did not meet the quality criteria established for inclusion in the data set to be analyzed.

2.4 ARTICLE CLASSIFICATION AND ANALYSIS

The set of records was then analyzed by reading each document and identifying essential information to answer the following research questions: which machine learning tools have been applied to forecast the sustainable development behavior of territories, and which machine learning tools could be applied to identify air pollution's impact on sustainable development? The classification and analysis of the scientific literature were guided by the subjects covered in the documents, the machine learning tools applied, how they were used, the evaluation metrics employed, and their performance.

3 RESULTS AND DISCUSSION

Following the methodological framework described above, a great number of scientific works were found, mainly in the fields of forecasting air pollutants, air quality indexes to issue early warnings, ensembles of tools to identify future spatio-temporal behavior of pollutants, as well as the identification of factors that induce the presence of atmospheric pollutants at a specific level. Although scientific production has increased with the integration of sustainable development goals, research on the application of machine learning tools in this field is limited. The following are the results found in each case.

3.1 APPLICATION OF MACHINE LEARNING TOOLS IN THE FRAMEWORK OF SUSTAINABLE DEVELOPMENT

Different studies have evaluated the state of progress of the environmental, social, and economic dimensions of SD and their respective indicators. Machine learning tools have analyzed the behavioral patterns of conjugate variables, as well as the lack of a self-learning capacity, the treatment of non-linear relationships, and possible bias to estimate the weight of each dimensions of sustainable development (Zhang and Huan 2006; Zhang et al. 2009).

Machine learning encompasses tools and techniques to identify patterns within data or process time series in some cases. ML is an integral piece of predictive modeling; whose principal aim is developing reliable models to forecast variable behavior or that of specific problems. A model’s reliability and accuracy are determined by specific metrics whose application is closely tied the model's purpose. The root mean square error (RMSE), mean square error (MSE), mean absolute error (MAE), maximum mean absolute error (MMAE), mean percentage error (MPE), mean absolute percentage error (MAPE), mean bias error (MBE), accuracy (Acc), relative error ratio (RER), correlation coefficient (R), and index of agreement (IA) are among the metrics used to evaluate the performance of ML models. ML has been successfully applied to specific subjects regarding the dimensions of SD. Certain studies have examined the dimensions and specific indicators to a greater degree (Li et al. 2006; Gounaridis et al. 2018), while others employ a more holistic vision (Zhang and Huan 2006; Zhang et al. 2009). Studies that assessed environmental, social and economic policies are included, and their results provide an understanding of the variables to be reinforced to achieve certain SDGs and increase eco-efficiency and socio-efficiency levels in the territory analyzed. The application of ML tools depends on the objectives established and information availability.

3.1.1 Information required for the studies

The application of ML tools both in the scope of the indicator and in a broader vision of SD, requires information from macroeconomic indicators, such as gross domestic product (GDP), level of financial development, education, rate of employment generated by industry, and environmental quality indexes (Li et al. 2006; Zhang et al. 2009; Pérez-Ortíz et al. 2014). This information is centered on indexes, indicators, and variables according to the dimension analyzed. To illustrate, for the environmental dimension; the concentration of atmospheric pollutants, water consumption and volume of waste-water per unit of GDP, the biochemical oxygen demand, noise level, the rate of treated industrial water, solid waste use, and reforested areas are taken into consideration (Li et al. 2006; Zhang and Huan 2006). With respect to the social dimension; the student rate, population density, unemployment rate, distance to hospitals and transportation stations are used (Zhang and Huan 2006; Gounaridis et al. 2018). To analyze the economic dimension; the GDP, the ratio of industry to GDP, foreign investment, energy consumption per unit of GDP, among others, are taken into account (Zhang and Huan 2006; Zhang et al. 2009; Wang and Xiao 2017). However, this data is not available in every territory with the necessary quality and timeliness, and it is generally recorded at the national level. Most studies in this field analyze the behavior of cities and provinces in China, others performed behavioral analysis of different European Union nations, primarily applying SVM, with ANNs and DTs employed to a lesser degree.

SD assessments are limited due to the scarcity and quality of data (Toumi et al. 2017), as it tends to present atypical values, additional noise, missing data and even errors (Gibert et al. 2016). Countries in Latin America and the Caribbean have deficiencies in the standardized information management and collection. Some cities in the region are characterized by high population densities, disjointed urban development, and interesting geographic conditions for the study of air quality and sustainable urban development. Forecasting the behavior in a territory as it pertains to SD in these types of urban areas calls for the application of tools to understand their growth and development, as well as to complete and standardize the information to be included in a ML model.

The scarcity of information on the characteristics required to calculate and forecast sustainability progress, demonstrates the need to use tools to assign data to the evaluated territories, as presented in the study developed by Phillis et al. (2017). Other studies have proposed the spatial identification of different variables (Holloway and Mengersen 2018), but they are conditioned upon the information visible in satellite images or georeferenced data. These studies provide useful tools to solve the problem of data quality and territorial scope. Nevertheless, it is important to solve the problem of data frequency generation to feed and validate machine learning models. In this sense, intervention from, and synergy with, the government sector plays an important role.

Data capturing, recording and quality assurance is another challenge that impacts SD as it limits the effectiveness and accuracy of ML application results. Territories have centralized their information through reports or international management platforms. There are different types of software and programs that provide support for informed decision-making (Gibert et al. 2012; Kadiyala and Kumar 2017a), but a minimum quality standard must be ensured for the information used.

3.1.2 Machine learning models

A selected group of machine learning tools is described below. This group of tools focuses on classification and regression problems in the framework of sustainability analysis and air quality forecasting. Through an analysis of different studies developed in the last 20 years related to forecasting sustainable development, it was found that SVMs, ANNs, Random Forest (RF), and DNNs are the primary machine learning tools used to forecast sustainable development or their dimensions in different territories.

- Support Vector Machine (SVM)

SVMs are a group of supervised learning algorithms related to classification and regression problems (Sierra 2006). The classification categorizes new objects in two or more separate groups based on their properties and a set of observations (de Hoogh et al. 2018). SVM draws a vector to separate classes; the greater their separation, the greater the recognition of different groups. For problems regarding more than two dimensions, SMV finds a hyperplane that maximizes the separation margin between classes. The classification could be supported by Kernel functions. The regression, referred to as support vector regression (SVR), enables it to be applied to forecast continuous variables; as such, the data needs to be trained (Suárez et al. 2011). Both classification and regression require data for validation and testing purposes.

SVMs have advantages in treating small samples and non-linear data patterns (Wang and Xiao 2017; de Hoogh et al. 2018). They are used to predict nations' progress levels in terms of the environmental, social, economic, and institutional dimensions of SD, and to forecast eco-efficiency, including the analysis of spatial data behavior and mixed frequency data modeling (Li et al. 2006; Pérez-Ortíz et al. 2014; Wang and Xiao 2017). Table 2 summarizes the principal studies developed in the framework of SD forecasting by SVM methods. It is important to note that the studies compared different ML tools to find a reliable model for their defined objective.

Table 2. Applications of SVM in the field of sustainable development

Specific tools and applications in the machine learning model	Objective	Information regarding the training process	Model accuracy metrics	Study
Cobweb algorithm for hierarchical clustering to identify regions with similar indicator behavior in the framework of social, economic, environmental, and institutional dimensions, followed by an ordinal classifier: SVM or logistic regression (LR) SVM and LR classifiers and their reformulation to ordinal regression, were compared with their ensemble versions. The Gaussian function was applied as a kernel function	Classification and ranking countries according to its SD level.	Two methods developed for the training process: 1. A dataset with 75% of the patterns for training and 25% for testing, applied 30 times. 2. Three different data partitions were used with different years for training and testing.	MMAE= 0.28 Acc= 92.5 for the SVM trainable ensemble with ordinal coding	(Pérez-Ortíz et al. 2014)
Principal component analysis (PCA): for the dimension to reduce of environmental indicators, which was followed by SVR or ANN.	Forecast model for eco-efficiency up to 1 year in advance.	A training dataset with 13 environmental indicators measured over 13 years and a	RER= 1) 0.48%; 2) 2.9%	(Li et al. 2006)

Specific tools and applications in the machine learning model	Objective	Information regarding the training process	Model accuracy metrics	Study
1. SVM-SVR, the Gaussian radial basis function was applied as a kernel function and, a sequential minimal optimization algorithm for data training. 2. ANN - radial basis neural network function (•)		validation dataset with 13 indicators measured for 1 year		
Support vector spatial dynamic and mixed data sampling (SVSD-MIDAS) to consider spatial correlation, sampling data frequency, and the non-linear relationship between eco-efficiency and the factors. Additionally, the radial basis function kernels were adopted	Prediction of the regional eco-efficiency with spatial mixed-frequency panel data	A training and validation dataset with information from 1998 to 2013	Average MPE = <1% in different spatial settings Average MSE = <1% in different spatial settings	(Wang and Xiao 2017)

The dots (•) in Table 2 identify the machine learning tool(s) that are not support vector machines. This type of ML was compared with the SVM employed in the study of interest.

- Artificial Neural Networks (ANN)

ANNs are non-linear operators formed by a set of neurons connected to each other and to an external environment through weight-determined connections (Sierra 2006). They generalize non-linear relationship patterns between input and output variables with noise information (Kadiyala and Kumar 2017b), and are used for industrial, energy, agriculture, environmental, transportation, economic and water conservation purposes (Chen et al. 2018). ANNs are composed of the following three layers: 1) an input layer that receives the predictor variables; 2) one or more hidden layers, which usually share the same information, and are interconnected in different ways; and 3) an output layer; for regression, the output layer may be a single layer while in a classification case the output will consist of a node of possible output classes (Holloway and Mengersen 2018). It is important to mention that each node or neuron in the hidden layers represents an activation function that acts on a weighted input of the previous layers' outputs (Holloway and Mengersen 2018). Neural networks can be classified based on their number of layers (monolayer or multilayer networks) and according to the direction that the information flows (recurrent networks, feedforward networks). The most widely known ANN is the multilayer perceptron (MLP). The back-propagation (BP) algorithm performs better the ANN given its capacity to identify the correct weight of the nodes in the ANN.

In SD modeling, artificial neural networks simulate non-linear relationships among indexes and prevent bias found in traditional weight design methods (Zhang et al. 2009). They have been used in evaluating SD by applying the BP algorithm to establish the degree of relationship between indexes (Zhang and Huan 2006; Zhang et al. 2009), and to determine the behavior of the environmental dimension and activity levels in an emissions inventory through the use of sustainability indicators (Antanasijević et al. 2013). Table 3 outlines the aforementioned studies and indicates the reliability of the machine learning structures employed.

Deep neural networks (DNNs) are ANN, which cover complex architectures and have more than one layer of hidden units, which improves the computer's learning capacity. They are used in the energy field through multivariate analyses of environmental, social and economic factors with big data (Ifaei et al. 2017). They are also employed in the field of greenhouse gas mitigation, as a basis for developing SD plans (Ifaei et al. 2017; Madu et al. 2017), and as a key tool for air quality forecasts that considers particulate matter (Karimian et al. 2019; Zhou et al. 2019).

A little number of publications that employed ANN to predict sustainable development, or a combination of its dimensions, was found in the studies analyzed. Table 3 summarizes those studies and the application of ANN over the last 20 years, indicating their use and the resulting metrics used to evaluate the performance of the developed ML models.

Table 3. Applications of ANN in the field of sustainable development

Specific tools and applications in the machine learning model	**Objective**	**Information regarding the training process**	**Model accuracy metrics**	**Study**
BP neural network: definition of the degree of relationship among indexes in the same layer.	Evaluation of sustainable development levels for society, economy, zoology, resource, science, and technological factors	10 years of statistical data used in the evaluation (1993-2003)		(Zhang and Huan 2006)
BP neural network to estimate the index weight of sustainable development and forecasting SD factors.	Forecasting sustainable development factors	A training dataset with 33 factors from 8 years, and another with 1 year of data for testing.	Average error 0.037	(Zhang et al. 2009)
Six annual hourly consumption variables predicted by recurrent and deep neural networks as machine learning tools. The prediction was part of the information of the Techno-Econo-Socio-Environmental Multivariate Analysis (TESEMA) model which additionally encompassed: 1. PCA: Identifying linear modeling between variables. 2. k-Nearest Neighbors (k-NN): Clustering zones according to modeling results. 3. Multivariate data analysis (TESEMA): Reducing data variability. 4. Partial least squares (PLS): investigate any correlation between major variables of TESEMA and population density	DNN: prediction of annual hourly electrical demand load using the maximum domestic power consumption at peak intervals, industrial power consumption, stored power at power plants, and total energy trade. TESEMA model: Analyzing the correlation between variables and population density in the study zone.	Two-thirds of available data as a training set and the rest as the testing set.	RSME = 73.15% for DNN	(Ifaei et al. 2017)

- Decision Trees

Decision trees (DT) are an inductive inference method of machine learning that develops classification rules or regression tasks. The former consists of a supervised learning method that uses class-labeled training examples; the leaves are the labeled classes and the branches are the features that conduct the classification tasks. This kind of tree creates classification rules to be applied for new observations to be classified. Decision tree regression enables the forecasting of continuous variables based on input predictors. For example, decision trees can to be applied for pattern recognition with regard to specific characteristics of an area, which was the case in a study developed by Zeng et al. (2019), or to forecast pollution levels according to the behavior of pollution sources (Zalakeviciute et al. 2020). M5P, C4.5 and random forests are decision tree algorithms employed in studies reviewed. The C4.5 algorithm can be used for classification tasks and builds decision trees using the information gain concept (Tzima 2011), while the M5P algorithm combines trees with linear regression models at the leaves (Shaban 2016). The C5.0 algorithm considers the information entropy to select the best classification method. To improve the robustness of the model, C5.0 introduces boosting technology that consists of repeating sampling simulation of existing weighted samples (Zeng et al. 2019).

Random forests (RF) correspond to a collection of DT that are applied to classification and regression problems (Brink et al. 2016). They make statistical predictions by averaging a set of non-correlated regression or classification trees, and are capable of computing non-linear relationships and interaction effects (Zhan et al. 2018). For example, they can be used to create potential land use transition maps by applying the RF classification algorithm to satellite images, combining dynamic, biophysical, socioeconomic and legislative factors (Gounaridis et al. 2018). Table 4 outlines the aforementioned studies and the application of DT over the last 20 years.

Table 4. Applications of decision trees in the field of sustainable development

Specific tools and applications in the machine learning model	Objective	Information regarding the training process	Model accuracy metrics	Study
Two decision tree models that applied the C5.0 growth and pruning algorithm for the 1) development stage target and 2) region target.	Identifying patterns and characteristics of sustainability of coal-mining cities considering data from 55 prefectures and 34 indicators	The development stage of the 55 coal-mining cities was the classification attribute	Acc= 92.73% for the stage model, Acc=94.55% for region model.	(Zeng et al. 2019)

Specific tools and applications in the machine learning model	Objective	Information regarding the training process	Model accuracy metrics	Study
1. Remote sensors and Landsat images: determining changes in urban dynamics. 2. Random forests: 3 predictor variables randomly sampled at each decision tree split and 500 trees to be built 3.Cellular automata: Forecasting land use changes through 2045	Development of potential transition surfaces by combining 20 dynamic, biophysical, socio-economic and legislative factors.	A set of randomly distributed points was plotted against the Landsat images and high-resolution images available via Google Earth.	Acc= 90.9-93.1% with respect to the resulting yearly map compared against reference samples	(Gounaridis et al. 2018)

Combining ML with other tools, such as those presented in Tables 2 – 4, has improved pattern recognition efficiency and data clustering to more accurately forecast the behavior of an analyzed data set.

3.1.3 Important variables that influence sustainable development

The relationship between the atmospheric component and SD has been demonstrated through the development of sustainable energy plans based on dimensional impact analyses and energy consumption (Ifaei et al. 2017), in addition to forecasting air quality by using certain sustainability indicators (Antanasijević et al. 2013). Studies predict and/or analyze atmospheric pollutant behavior based on information registered by monitoring stations and other sources, while advancing the integration of geospatial information, which includes the relationship between geospatial data and mixed frequency data analysis through SVM (Wang and Xiao 2017). The same occurs with the application of the BP algorithm in ANNs, establishing the degree of relationship of indexes in the same layer (Zhang et al. 2009). The results of the relationship between different variables and information characteristics are emphasized in order to understand future land use behavior in a territory (Gounaridis et al. 2018).

The scope of most studies is both from a holistic vision of SD and the specific analysis of a variable or indicator in the framework of SD. There is a lack of studies that identify the most influencing variables or indicators of SD progress. Urban territories are influenced by different variables and identifying the progress of each and their synergetic behavior in the scope of SD is necessary for decision-makers. Identifying variables' level of importance should consider the data that feeds the models. In this vein, tools like analytic hierarchy process and clustering could guide the knowledge of influencing variables in sustainability categories. Hybrids or ensembles of tools offer an improvement of the models' exactitude or reduced computational costs, as demonstrated in the studies developed by Peng et al. (2017) and Zhan et al. (2018). Therefore, the ensembles, which include hierarchical and clustering tools, could support the identification of important variables regarding the sustainable development behavior of territories, as well as its forecast.

Even though there are a limited number of studies associated with the forecast of SD and its influential variables, research for air quality forecasting through ML tools provides guidelines for causality analyses of urban sustainable development progress. Tools like proportion-based causality tests and sequential forward feature selection applied by Wang (2019) facilitate the evaluation and forecasting of urban sustainable development.

3.2 APPLICATION FOR AIR QUALITY

ANNs, SVMs and DTs have been utilized to forecast air quality and their influencing variables. Studies primarily focus on Asia and Europe. Countries such as India, China and Saudi Arabia, which according to the WHO have higher registered levels of pollution, stand out due to their application of ML tools to forecast future pollution events.

ANNs have been employed for air quality and atmospheric emission applications, with MLP and BP predominantly used as training algorithm. To identify the best forecasting tool, ANNs have been compared with conventional statistical models, DTs, and SVMs. Multilayer perceptron neural networks have been extensively analyzed and have been combined with the Manhattan propagation algorithm (Souza et al. 2015), with self-organizing map (SOM) and hierarchical clustering (Tamas et al. 2016), linear regression statistical techniques, or in comparison with an extreme learning machine (ELM) (Peng et al. 2017). Hybrid models provide synergetic results, in particular to establish early warnings, and have improved behavior in forecasting specific atmospheric pollutants. The application of ELM has surpassed the limitations of MLP with respect to the non-linearity of data and operational costs of traditional ANN methods (Peng et al. 2017).

To identify pollution risk, clustering algorithms such as the K-means algorithm have been used by applying the AQ algorithm to defined groups (Cervone et al. 2008). Furthermore, MLP was applied to forecast PM_{10} and $PM_{2.5}$, while the K-mean and PCA algorithms facilitated the identification of input variables for the model (Franceschi et al. 2018).

For the ozone (O_3) and PM_{10} forecasts, lazy learning networks (LL) perform better than pruned neural networks (PNN) and feed-forward neural networks (FFNN), while PNNs effectively detect the increase of alarm and attention thresholds for the pollutants analyzed (Corani 2005). To forecast the concentration of nanoparticles, FFNNs are combined with BP, in which the inclusion of all types of variables enabled the ANN to precisely map the non-linear relationship between the measured and expected nanoparticles (Al–Dabbous et al. 2017). However, to predict $PM_{2.5}$, it was concluded that increasing the number of data points does not necessarily result in better estimates, as it is more a correlation between the main factor and those related to it (Ni et al. 2017).

Additionally, the effectiveness of different ML algorithms was compared, as well as integration with the Gaussian dispersion model for an emission source, establishing excellent behavior in forecasting the dispersion of atmospheric pollutants, which is an applicable method for predicting and identifying source parameters (Ma and Zhang 2016). Table 5 lists the results of the studies analyzed and how ANNs have been applied and combined with algorithms, see the artificial neural network model in Table 5 for model training and selecting input variables, while also presenting that a single study compared different tools.

Table 5. Application of ANNs to air quality and atmospheric pollutants

Study	Artificial neural network model	Objective	Information regarding the training process	Model accuracy metrics
(Souza et al. 2015)	MLP configured with hyperbolic tangent activation function + Manhattan propagation algorithm	Forecast daily concentrations of PM_{10}.	80% of data used for training (daily samples collected from 0.7/2009-06/2013) and 20% for testing the data set (daily samples collected from 07.2013-06.2014).	Improvements of MSE compare with individual MPL and ensembles = 8.85% (4 neurons in the hidden layer).
(Tamas et al. 2016)	1. MLP + Levenberg-Marquardt training algorithm + hybridized with hierarchical clustering. 2. MLP hybridized with SOM and k-mean clustering.	Hourly concentrations of O_3, NO_2 and PM_{10}, 24 hours in advance.	Clustering methods were used to subdivide the data set, with MLP trained on each subset: 60% for training data (3 years of data), 20% for validation (1 year of data), and 20% for testing (1 year of data).	RMSE =18.65 (O_3); 12.1 (NO_2); 7.4 (PM_{10}) μgm^3 MAE =14.69 (O_3); 8.58 (NO_2); 5.77 (PM_{10}) μgm^3 IA = 0.87 (O_3); 0.8 (NO_2); 0.74 (PM_{10})
(Peng et al. 2016)	ELM based on MLP+ hill-climbing algorithm to determine the optimal number of hidden nodes. ELM conducted the training task, followed by an online sequential ELM (OSELM) which updated the model. The input data of the model consisted of meteorological variables, O_3, $PM_{2.5}$, NO_2, and physical variables. An online sequential multiple linear regression and the MLP configured with hyperbolic tangent activation function were compared together with the OSELM.	Hourly concentrations of O_3, NO_2 and $PM_{2.5}$, 48 hours in advance.	2 years of information for training and validation (2009/07-2011/07), three for testing (2011/08-2014/07) as well as model updating.	The models were ranked using the forecast scores averaged over all forecast lead times, for each pollutant. For MAE and correlation scores the OSELM outperformed MLP and MLR for O_3, NO_2 and $PM_{2.5}$
(Franceschi et al. 2018)	A combination of PCA to identify the most influencing predictors, K-means for data grouping, and MLP neural network + BP as the training algorithm.	Hourly and daily prediction of PM_{10} and $PM_{2.5}$.	Ratios for training and validating the data set: 80% for training and 20% for validation.	RSME= 15.62 (PM_{10}); 5.79 ($PM_{2.5}$) μg/m^3 MAE=,13.39 (PM_{10}); 4.72 ($PM_{2.5)}$ μg/m^3

Study	Artificial neural network	Objective	Information regarding	Model accuracy metrics
(Paas et al. 2017)	MLP configured with a hyperbolic tangent transfer function + BP training algorithm.	Prediction of ($PM_{0.25\text{-}10}$) mass concentrations and particle number concentrations.	Data split through stratified random sampling with self-organizing map. 70% of the subset was used for training, 20% for validation, and 10% for testing.	RMSE= 7.78 µg/m^3
(Ni et al. 2017)	Back propagation neural network.	A correlation analysis of $PM_{2.5}$, meteorological data, pollutant concentration data and social media.	Ratios for modeling and testing the data set: 70% for training and 30% for testing.	RMSE=24.06 µg/m^3
(Chen et al. 2018)	Measurement of partial mutual information to select significant variables + ensemble ANN-based output estimation + KNN regression output estimation error.	Forecasting an air quality index one day in advance considering the following predictors: PM_{10}, $PM_{2.5}$, and SO_2	2 years of information for training, one for validation.	The model cannot simulate well for large AQI values, but has good precision for medium and small AQI values
(Antanasijević et al. 2013)	General regression neural network + genetic algorithm for training.	Forecasting PM_{10} up to two years in the future with the following predictors: GDP, gross inland energy consumption, wood incineration factor, motorization rate, paper production, and processing of certain minerals.	5 years of data from 26 countries for training and validation, 2 years for testing.	MAE = 10%
(Corani 2005)	1. FFNN configured with a hyperbolic tangent transfer function and the Levenberg-Marquardt training algorithm. 2. LL 3. PNN	Prediction of O_3 and PM_{10} at 9:00 am for the current day and detection of exceedance.	Cross-validation approach	True/predicted correlation=0.85 (O_3); 0.9 (PM_{10}) Success index=0.6 (O_3); 0.75 (PM_{10}) MAE=15.87 (O_3) and 8.25 for LL which outperformed the FFNN and PNN in the prediction.
(Al–Dabbous et al. 2017)	FFNN configured with a hyperbolic sigmoid transfer function + BP training algorithm based on Levenberg Marquardt optimization.	Prediction of nanoparticles regarding different input variables.	Ratios for modeling and testing the data set: 80% for training and 20% for testing.	R^2 value = 0.79 IA = 0.94
(Ma and Zhang 2016)	1. RBF + Gaussian dispersion model	Emission source parameters identification and pollutant dispersion forecasting.	Ratios for modeling and testing the data set: 47% for training and 53% for testing.	MSE=482.78 R=0.89
	2. BP + Levenberg-Marquardt training algorithm + Gaussian			MSE=371.82 R=0.91
	3. SVR+ Gaussian dispersion model			MSE=300.37 R=0.92
(Zhou et al. 2019)	1. Shallow multi-output long short-term neural network memory (SM-LSTM)	Regional multi-step-ahead air quality ($PM_{2.5}$, PM_{10}, NO_x) forecast horizon: t+1 up to t+4	Mini-batch gradient decent algorithm, dropout neuron algorithm, and L2 regularization algorithm for the training process.	MSE= 0.87
	2. Deep multi-output LSTM (DM-LSTM) neural network			MSE= 0.72
(Karimian et al. 2019)	1. Long short-term neural network (LSTM)	Prediction of $PM_{2.5}$ up to 48 hours.	Ratios for training and validating the data set: 60% for training, 20% for validation, and 20% for testing.	RMSE=8.91µg/m^3 MAE= 6.21 µg/m^3 R^2= 0.8
	2. DNN: Deep feed forward neural network (DFNN)			RMSE=19.45 µg/m^3 MAE= 14.52 µg/m^3 R^2= 0.49

Study	Artificial neural network	Objective	Information regarding	Model accuracy metrics
	3. Multiple additive regression trees (•)			RMSE=12.83μg/m^3 MAE= 8.99 μg/m^3 R^2= 0.56

The dots (•) in Table 5 identify the machine learning tool(s) that are not artificial neural networks. This type of ML was compared with the ANN employed in the study of interest.

MLP was compared with SVR, in which the generalization capacity acquired in a relatively small amount of learning data and a large number of entry nodes demonstrated better behavior for SVR (Suárez et al. 2011; García et al. 2013; Shaban et al. 2016). This capacity was also demonstrated in forecasting air quality indexes (Liu et al. 2017; Weizhen et al. 2014) (see Table 6). Furthermore, SVR was utilized for spatio-temporal modeling of $PM_{2.5}$, taking into account missing information (Song et al. 2014), and to forecast atmospheric pollutants by developing an online method (Wang et al. 2008). SVR and the Gaussian function kernel have been used to capture the non-linearity and interaction between predictors (Wang et al. 2008; Weizhen et al. 2014). In the prediction of hourly and daily levels of carbon monoxide (CO), the hybrid model of SMV and partial least squares (PLS), which was used to reduce input data, performed better with less modeling time and better performance metrics than the just SVM (Yeganeh et al. 2012).

Table 6. Application of SVMs to forecast air quality and atmospheric pollutants

Study	SVM model	Objective	Information regarding the training process	Model accuracy metrics
(García et al. 2013; Suárez et al. 2011)	1. MLP neural network configured with a sigmoid activation function (•) 2. SVR + sequential minimal optimization algorithm, and kernel function variants	Obtain a relationship between concentrations of CO, SO_2, NO, NO_2, PM_{10}, O_3	10-fold cross-validation	SVM outperformed MLP with correlation coefficients that ranged from 0.62 to 0.9 for the pollutants.
(Liu et al. 2017)	SVR	AQI forecasting up to 24 in advance, with information ($PM_{2.5}$, PM_{10}, SO_2, CO, NO_2 and O_3 levels, AQI, temperature, weather, wind force and direction) on cities with similar urban air pollution.	4 folds, each with 25% of the data. 3 folds were selected for training, the reaming fold was used for model testing.	RMSE = 6.54 MAPE = 0.0534
(Weizhen et al. 2014)	SVR + successive over relaxation algorithm + Gaussian kernel function	Prediction of PM_{10} and $PM_{2.5}$ for the following 24h and hourly forecasting based on daily average aerosol optical depths and meteorological parameters.	k-fold cross validation. 83% of the data used as a training dataset, the remaining as a testing dataset.	R^2=0.87 Average error=12.66 ug/m^3
(Song et al. 2014)	Spatial data aided incremental SVR	Daily average spatio-temporal prediction of $PM_{2.5}$ based on hourly PM_{10} levels.	The split for training and testing the ML model encompasses data from August 2006 to 2009 for training and data from 2011 to 2012 as a testing dataset.	RMSE= 1.0775 MAE = 0.81 MBE = 0.18 IA= (0.51-0.68)

Study	SVM model	Objective	Information regarding	Model accuracy metrics
(Wang et al. 2008)	SVR model + Gaussian kernel function	Prediction of respirable particulate matter, NO_X and SO_2 concentrations, up to 24 hours and 1 week in advance by using data found online.	59% for training and 41% for testing the dataset.	RMSE=25.89ug/m^3 MAE=19.29 ug/m^3
(Yeganeh et al. 2012)	Partial least square (PLS) for data selection and to reduce the amount of input data + SVR + radial-basis function as a kernel function.	Hourly and daily CO concentration forecast by using data on PM_{10}, total hydrocarbons (THC), NOx, CH_4, SO_2, O_3 and meteorological parameters.	75% for training and 25% for testing the model.	RMSE = 0.711 R^2 =0.654

The dots (•) in Table 6 identify the machine learning tool(s) that are not support vector machine. This type of ML was compared with the SVM employed in the study of interest.

On the other hand, the M5P decision tree algorithm outperformed both SVM and ANNs, given the efficiency of its tree structure and generalization capacity (Shaban et al. 2016) (see Table 7). The M5P algorithm performed well in forecasting $PM_{10,}$ when heuristic rules were applied (Oprea et al. 2016). A comparable situation was found when applying the RF algorithm, which performed better than the SVM algorithms (Pandey et al. 2013). RFs outperform other classifiers because of their capacity to assimilate forecasts from a variety of simple tree classifiers based on more predictive variables, resulting in low levels of bias and variance (Pandey et al. 2013). By applying different predictors, it was found that different variables have different types of impacts on the levels of ultrafine particulate matter in the atmosphere. The same case was presented in a comparison of five types of algorithms: LR, MLP, DT (C4.5 algorithm), SVM, and a variant of "ZCS-DM" decision trees, which belong to a class of ML tools, termed learning classifier systems, which use conditional rules. SVM and ZCS-DM performed well, the latter algorithm identified extreme cases and forecasted pollution episodes (Tzima et al. 2011).

SVM has been used to compare ML algorithm performance. SVM was compared with single decision trees (SDT), decision tree forests (DTF), and decision tree boosting (DTB) for forecasting air quality indexes (Singh et al. 2013). It was concluded that as the result of incorporating aggregation and optimization algorithms, DTF and DTB outperformed SVM both in classification and regression (Singh et al. 2013). RFs outperform chemical transport models, which require input variable information from emission inventories that may not be accurate (Zhan et al. 2018).

DTB was used to forecast PM_{10} concentrations, revealing that when comparing the multiple linear regression model (MLRM), the quantile regression model (QRM), and the generalized additive model (GAM), the capacity of the QRM to capture contributions from covariates in different quantiles produces better forecasting, when compared to procedures in which a single central trend is considered for a set of independent variables (Sayegh et al. 2014).

To improve the exactitude of the models and reduce computational costs in data treatment, analysis and forecasting, hybrids or ensembles of tools have been made (Peng et al. 2017; Zhan et al. 2018). The choice was made to combine DTs in their different degrees of complexity, which have been compared to the performance of conventional statistical tools, SVMs and ANNs. Furthermore, DTs have been combined with linear regression tools to improve their accuracy. Table 7 presents the studies analyzed, including the decision tree model developed in each study, the objective of the ML model, information regarding the training and testing datasets, and each model's performance. Some studies evaluated the performance of the ML model in comparison with other learnings architectures.

Table 7 Application of decision trees to air quality and/or atmospheric pollutants

Study	Decision tree model	Objective	Information regarding the training process	Model accuracy metrics
(Zhan et al. 2018)	RF	Prediction of the spatio-temporal distribution of daily 8h maximums of O_3 concentrations.	10-fold cross-validation (9 groups for training and 1 for testing).	RMSE=26 μg/m^3 R^2 = 0.69
(Sayegh et al. 2014)	Boosted regression trees	Prediction of hourly PM_{10} concentration lev-els.	10-fold cross validation; 11 months of data used as a training data set; 1 month as a testing dataset.	MBE=-41.1 μg/m^3 RMSE=125.6 MAE= 80.4 R=0.54 IA=0.66
(Singh et al. 2013)	1. Decision tree forest	• Seasonal air quality discrimination • Air quality index prediction	Kennard-Stone approach for a uniform scatter of training data around the training do-main (70% for training and 30% for testing), k-folds-cross validation.	Acc=96.68% RMSE=6.58 MAE=5.24 R=0.90
	2. Decision tree boost-ing			Acc=96.45% RMSE=6.59 MAE=5.26 R=0.90
(Tzima et al. 2011)	3. Tree induction algorithm C4.5	Air quality episode fore-casting	10-fold cross validation	Overall average rank regarding the kappa coefficient for pollu-tants=4.25 (C4.5); 2.5 (ZCS-DM); 1.75 (SVM); 3.4(MLP)
	4. Rule induction ZCS-DM algorithm			
	5. SVM (•)			
	6. MLP (•)			
(Shaban et al. 2016)	1. M5P decision tree	Prediction of NO_2, SO_2, and O_3 concentrations up to 24 hours in advance.	Time windowing (forecasting horizon) = 2 windows are used for training and testing.	RMSE=5.8
	2. SVM (•)			RMSE=6.4
	3. ANN (•)			RMSE=16.4
(Oprea et al. 2016)	1. M5P decision tree	Prediction of PM_{10} con-centrations levels up to 3 days in advance.		RMSE=8.38; MAE=6.52; R=0.87
	2. Reduced error prun-ing tree algorithm			RMSE=12.74; MAE=9.34; R=0.65
(Wang 2019)	1. DT	Air quality classification	k-cross validation	Acc= 75.1%
	2. Ensemble (boosted and bagged trees)			Acc= 90.2% and 89.1%
	3. k-NN			Acc= 80.2%
	4. SVM (linear and gaussian) (•)			Acc= 82.3% and 85.6%

The dots (•) in Table 7 identify the machine learning tool(s) that are not decision trees. This type of ML was compared with the DT employed in the study of interest.

The machine learning tools presented above enabled the forecasting of air quality or specific indicators (see Tables 5 – 7). Their performance and accuracy primarily depend on the data set and feature selection. Additionally, the machine learning tools listed in Tables 2 – 4 showed the first advances related to forecasting sustainable development or its dimensions. Nevertheless, hybrid or ensembles like those used by Karimian et al. (2019) and Wang (2019) improve specific ML tools, and could be applied in forecasting sustainable development, possibly even to identify influencing variables.

However, a specific analysis of the influence of air quality or atmospheric pollution on sustainable development was not found in any of the studies analyzed; nor a study that identifies the degree of importance of variables on sustainable development to achieve the targets defined in the SDGs. Specific aspects and results of the reviewed studies provided an understanding that there are two fundamental factors that need to be addressed in order to advance sustainable development forecasting; the information required for training and model validation, and the identification of important variables that influence sustainable development.

3.3 GAPS AND FUTURE DEVELOPMENTS

The application of ML must consider the characteristics of the data and its behavior. The atmosphere and reactions within it are non-linear complex processes, whose analysis is necessary to forecast SD and understand its impact. As such, SVMs and DTs are tools that can support this analysis.

The application of DNNs could be explored in conjunction with tools that have overcome data management difficulties. The majority of studies reviewed perform an analysis of redundancies to include variables with the most forecasting potential, avoiding the use of variables that introduce errors to the modeling results, which include over fitting.

One of the objectives in an air quality assessment is determining the environmental quality and reflecting human demands on the same (Chen et al. 2018). Some areas have geographic and atmospheric conditions that are not conducive to adequate air-circulation, coinciding with populations located in zones that are characterized by cities that are not optimally designed (Saeed et al. 2017). Given their effect on a population's health, these factors must be analyzed as they are part of the dynamic of SD.

As ultrafine particles are atmospheric pollutants that are not widely monitored, it is possible to learn their impact on the health of a population and on SD by applying ML and correlating them to other atmospheric pollutants.

Machine learning tools, their ensembles, and using techniques such as analyzing satellite imaging, offer decision-makers the possibility of having a better understanding of their territories. In the framework of sustainable development and its targets, decision-makers everywhere have the responsibility of identifying the most effective actions to achieve the SDGs. The economic, environmental, and social dimensions are not static and require continuous support to maintain or improve their behavior. For example, any change to the environment and health dimensions of a population, both in an urban territory or country, could influence every sustainability dimension. Furthermore, the use of ensembles of machine learning tools that include hierarchical and clustering tools applied by Franceschi et al. (2018) and Tamas et al. (2016) with MLP neural networks or by Oprea et al. (2016) and Singh et al. (2013) who used different kinds of decision trees, or by Peng et al. (2017) who applied SVR to identify subsets of the most relevant attributes for the prediccion taks, may facilitate early decisions from leaders.

Several studies have successfully applied machine learning tools to forecast air quality, with accuracies ranging from 70-95%. In this sense, the methodology and ML tools described by the studies mentioned above are a reference for the analysis of sustainability and its forecasting. Furthermore, research studies developed by Wang (2019) outline the relevant methods and tools replicable to evaluate the influence of air quality on urban sustainable development.

4 CONCLUSIONS

Forecasting the level of SD and its variation in function of certain input parameters, enables decision-makers to establish the emphasis variable for increasing the sustainability of an analyzed territory. Through the analysis of the references consulted, the use of ML in defining sustainability indexes was evident, supporting the relevance of its application, as biases are prevented in determining the weights of parameters and indicators.

Regarding the research questions of this study, it is important to note that SVMs, ANNs, RFs, and DNNs were the machine learning tools used either to forecast sustainable development or their dimensions in different territories. The most widely used tools in the field of SD are SVMs and to a lesser degree, ANNs and DTs. However, it is important to note that no study was found which identified the influence of air quality on sustainable development. Nevertheless, different approaches and distinct machine learning tools have been applied, which enable further research to make progress in determining the influence of air quality on sustainable urban development, as well as in its forecasting. In this vein, combining ML with other tools, such as hybrid systems or ensembles that include hierarchical and clustering tools, is useful in identifying the influence of different variables on urban sustainable development.

With respect to ML tools and the non-linearity of information that characterizes the dimensions of sustainable development, SVMs and DTs have performed the best in computing this type of data. However, despite the existence of ML approaches with certain characteristics that makes them suitable for forecasting, boundary conditions establish an opportunity for their use. It is necessary to consider the tools' favorable aspects, applying those that adjust to specific conditions, including the availability of information with the required continuity and quality that enables the proper reflection of behavioral patterns of variables in a given territory, which is the key to forecasting.

Only the studies that applied ML in the field of air quality established the required time for forecasting or data treatment as an analytical value. This is a possible parameter of interest given the importance of making effective use of time in the development of these types of studies.

This study is useful in the field of environmental, social and economic dimension analysis, as it provides a map of different ML tools and the authors who have used them in specific case studies, either to address a specific challenge or in integrating the dimensions.

There is a need for studies that identify the behavior of territories concerning indicator variation and sustainable development variables. The environmental, social, and economic dimensions are interconnected, in which each all have an effect on the others. Furthermore, to understand and identify the most influencing actions on health, the environment, and economic effects, it is necessary to forecast behaviors and identify influencing variables, to develop scenarios and make the best decisions possible. The environmental dimension affects the social dimension in the framework of the population's health.

This study is unique and innovative, it recognizes the importance of air quality in sustainable development and seeks to identify the behavior of the ML tools applied in different studies to forecast both concepts. It also identifies the tools used to understand the influence of variables on sustainable development or the dimensions of sustainability, in addition to those used to establish the association of predictors. Even when documentation was found regarding machine learning tools, it is necessary to collect the experiences developed for decision making. Few studies have taken the initiative of forecasting sustainable development and analyze its most-influencing possible variables. This work identifies the degree of progress in this sense, as it supports decision-making that could be undertaken by governments as part of their goals and objectives and those tied to the 2030 Agenda for Sustainable Development.

5 REFERENCES

Al–Dabbous A, Kumar P, Khan A (2017) Prediction of airborne nanoparticles at roadside location using a feed–forward artificial neural network. Atmos Pollut Res 8:446–454. https://doi.org/10.1016/j.apr.2016.11.004

Antanasijević D, Pocajt V, Povrenović D, Ristić M, Perić-Grujić A (2013) PM_{10} emission forecasting using artificial neural networks and genetic algorithm input variable optimization. Sci Total Environ 443: 511–519. https://doi.org/10.1016/j.scitotenv.2012.10.110

Brink K, Richards J, Fetherolf M (2016) Real-world machine learning. Richards & Fetherolf, Eds. Manning Publications.

Cervone G, Franzese P, Ezber Y, Boybeyi Z (2008) Risk assessment of atmospheric emissions using machine learning. Nat Hazard Earth Sys 8: 991–1000. https://doi.org/https://doi.org/10.5194/nhess-8-991-2008

Chen S, Kan G, Li J, Liang K, Hong Y (2018) Investigating China’s Urban Air Quality Using Big Data, Information Theory, and Machine Learning. Pol J Environ Stud 27:565–578. https://doi.org/10.15244/pjoes/75159

Corani (2005) Air quality prediction in Milan: Feed-forward neural networks, pruned neural networks and lazy learning. Ecol Model 185:513–529. https://doi.org/10.1016/j.ecolmodel.2005.01.008

Cruz C, Gómez A, Ramírez L, Villalva A, Monge O, Varela J, Quiroz J, Duarte H (2017) Calidad del aire respecto de metales (Pb, Cd, Ni, Cu, Cr) y relación con salud respiratoria: Caso Sonora, México. Rev Int Contam Ambie 33:23–34. https://doi.org/10.20937/RICA.2017.33.esp02.02

de Hoogh K, Héritier H, Stafoggia M, Künzli N, Kloog I (2018) Modelling daily $PM_{2.5}$ concentrations at high spatio-temporal resolution across Switzerland. Environ Pollut 233:1147–1154. https://doi.org/https://doi.org/10.1016/j.envpol.2017.10.025

Franceschi F, Cobo M, Figueredo M (2018) Discovering relationships and forecasting PM_{10} and $PM_{2.5}$ concentrations in Bogotá, Colombia, using Artificial Neural Networks, Principal Component Analysis, and k-means clustering. Atmos Pollut Res 9:912–922. https://doi.org/10.1016/j.apr.2018.02.006

García N, Combarro E, del Coz J, Montañes E (2013) A SVM-based regression model to study the air quality at local scale in Oviedo urban area (Northern Spain): A case study. Appl Math Comput 219:8923–8937. https://doi.org/10.1016/j.amc.2013.03.018

Gibert K, Sànchez-Màrre M, Sevilla B (2012) Tools for Environmental Data Mining and Intelligent Decision Support. In *iEMSs*. Leipzig, Germany. http://www.iemss.org/society/index.php/iemss-2012-proceedings. Accessed 26 November 2018

Gibert K, Sànchez-Marrè M, Izquierdo J (2016) A survey on pre-processing techniques: Relevant issues in the context of environmental data mining. Ai Commun 29:627–663. https://doi.org/10.3233/AIC-160710

Gounaridis D, Chorianopoulos I, Koukoulas S (2018) Exploring prospective urban growth trends under different economic outlooks and land-use planning scenarios: The case of Athens. Appl Geogr 90:134–144. https://doi.org/10.1016/j.apgeog.2017.12.001

Holloway J, Mengersen K (2018) Statistical machine learning methods and remote sensing for sustainable development goals: A review. Remote Sens 10:1-21. https://doi.org/10.3390/rs10091365

Ifaei P, Karbassi A, Lee S, Yoo Ch (2017) A renewable energies-assisted sustainable development plan for Iran using techno-econo-socio-environmental multivariate analysis and big data. Energ Convers Manage 153:257–277. https://doi.org/10.1016/j.enconman.2017.10.014

Kadiyala A, Kumar A (2017a) Applications of R to Evaluate Environmental Data Science Problems. Environ Prog Sustain 36:1358–1364. https://doi.org/10.1002/ep.12676

Kadiyala A, Kumar A (2017b) Vector Time Series-Based Radial Basis Function Neural Network Modeling of Air Quality

Inside a Public Transportation Bus Using Available Software. Environ Prog Sustain 36:4–10. https://doi.org/10.1002/ep.12523

Karimian H, Li Q, Wu Ch, Qi Y, Mo Y, Chen G, Zhang X, Sachdeva S (2019) Evaluation of different machine learning approaches to forecasting $PM_{2.5}$ mass concentrations. Aerosol Air Qual Res 19:1400–1410. https://doi.org/10.4209/aaqr.2018.12.0450

Krzyzanowski M, Apte J, Bonjour S, Brauer M, Cohen A, Prüss-Ustun A (2014) Air Pollution in the Mega-cities. Current Environmental Health Reports 1:185–191. https://doi.org/10.1007/s40572-014-0019-7

Lässig Kersting and Morik (2016). *Computat Sustainability*. Springer. https://doi.org/10.1007/978-3-319-31858-5

Li Y, Wu Y.-X, Zeng Z.-X, Guo L (2006) Research on forecast model for sustainable development of Economy-Environment system based on PCA and SVM. In *Proceedings of the 2006 International Conference on Machine Learning and Cybernetics* (Vol. 2006, pp. 3590–3593). Dalian, China: IEEE. https://doi.org/10.1109/ICMLC.2006.258576

Liu B.-Ch, Binaykia A, Chang P.-Ch, Tiwari M, Tsao Ch.-Ch (2017) Urban air quality forecasting based on multi-dimensional collaborative Support Vector Regression (SVR): A case study of Beijing- Tianjin-Shijiazhuang. *Plos One* 12:1–17. https://doi.org/10.1371/journal.pone.0179763

Lubell M, Feiock R, Handy S (2009) City adoption of environmentally sustainable policies in California’s Central Valley. J Am Plann Assoc 75:293–308. https://doi.org/10.1080/01944360902952295

Ma D, Zhang Z (2016) Contaminant dispersion prediction and source estimation with integrated Gaussian-machine learning network model for point source emission in atmosphere. J Hazard Mater 311:237–245. https://doi.org/10.1016/j.jhazmat.2016.03.022

Madu C, Kuei N, Lee P (2017) Urban Sustainability Management: A Deep Learning Perspective. Sustain Cities Soc 30:1–17. https://doi.org/10.1016/j.scs.2016.12.012

Mellos K (1988) Theory of Eco-development. In *Perspectives on Ecology* (pp. 59–74). London: Palgrave Macmillan UK. https://doi.org/10.1007/978-1-349-19598-5_4

Ni X.Y, Huang H, Du W.P (2017) Relevance analysis and short-term prediction of $PM_{2.5}$ concentrations in Beijing based on multi-source data. Atmos Environ 150:146–161. https://doi.org/10.1016/j.atmosenv.2016.11.054

Oprea M, Dragomir E, Popescu M, Mihalache S (2016) Particulate Matter Air Pollutants Forecasting using Inductive Learning Approach. Rev Chim 67:2075–2081. Retrieved from http://www.revistadechimie.ro

Paas B, Stienen J, Vorländer and Schneider Ch (2017) Modelling of Urban Near-Road Atmospheric PM Concentrations Using an Artificial Neural Network Approach with Acoustic Data Input. *Environments* 4:1-25. https://doi.org/10.3390/environments4020026

Pandey G, Zhang B, Jian L (2013) Predicting submicron air pollution indicators: a machine learning approach. Environ Sci-Proc Imp 15:996–1005. https://doi.org/10.1039/c3em30890a

Peng H, Lima A, Teakles A, Jin J, Cannon A, Hsieh W (2017) Evaluating hourly air quality forecasting in Canada with nonlinear updatable machine learning methods. Air Qual Atmos Health 10:195–211. https://doi.org/10.1007/s11869-016-

0414-3

Pérez-Ortíz M, de La Paz-Marín M, Gutiérrez P.A, Hervás-Martínez C (2014) Classification of EU countries' progress towards sustainable development based on ordinal regression techniques. Knowl-Based Syst 66:178–189. https://doi.org/10.1016/j.knosys.2014.04.041

Phillis Y, Kouikoglou V, Verdugo C (2017) Urban sustainability assessment and ranking of cities. Comput Environ Urban 64:254–265. https://doi.org/10.1016/j.compenvurbsys.2017.03.002

Saeed S, Hussain L, Awan I, Idris A (2017) Comparative Analysis of different Statistical Methods for Prediction of $PM_{2.5}$ and PM_{10} Concentrations in Advance for Several Hours. Int J Comput Sci Netw Secur 17:45–52.

Sayegh A, Munir S, Habeebullah T (2014) Comparing the Performance of Statistical Models for Predicting PM_{10} Concentrations. Aerosol Air Qual Res 14:653–665. https://doi.org/10.4209/aaqr.2013.07.0259

Shaban K, Kadri A, Rezk E (2016) Urban Air Pollution Monitoring System With Forecasting Models. *IEEE SENSORS JOURNAL*, 16:2598–2606. https://doi.org/10.1109/JSEN.2016.2514378

Sierra B (2006) *Aprendizaje automático conceptos básicos y avanzados Aspectos prácticos utilizando el software Weka*. Madrid, España: Madrid Pearson Prentice Hall.

Singh K, Gupta S, Rai P (2013) Identifying pollution sources and predicting urban air quality using ensemble learning methods. Atmos Environ 80:426–437. https://doi.org/10.1016/j.atmosenv.2013.08.023

Song L, Pang S, Longley I, Olivares G, Sarrafzadeh A (2014) Spatio-temporal $PM_{2.5}$ Prediction by Spatial Data Aided Incremental Support Vector Regression. In *International Joint Conference on Neural Networks* (pp. 623–630). Beijing: IEEE. https://doi.org/10.1109/IJCNN.2014.6889521

Souza R, Coelho G, da Silva A, Pozza S (2015) Using Ensembles of Artificial Neural Networks to Improve PM_{10} Forecasts. Chem Eng Trans 43:2161–2166. https://doi.org/10.3303/CET1543361

Suárez A, García P.J, Riesgo P, del Coz J.J, Iglesias-Rodríguez F.J (2011) Application of an SVM-based regression model to the air quality study at local scale in the Avilés urban area (Spain). Math Comput Modell 54:453–1466. https://doi.org/10.1016/j.mcm.2011.04.017

Tamas W, Notton G, Paoli C, Nivet M, Voyant C (2016) Hybridization of air quality forecasting models using machine learning and clustering: An original approach to detect pollutant peaks. Aerosol Air Qual Res 16:405–416. https://doi.org/10.4209/aaqr.2015.03.0193

Toumi O, Le Gallo J, Ben Rejeb J (2017) Assessment of Latin American sustainability. Renew Sust Energ Rev 78:878–885. https://doi.org/10.1016/j.rser.2017.05.013

Tzima F, Mitkas P, Voukantsis D, Karatzas K (2011) Sparse episode identification in environmental datasets: The case of air quality assessment. Expert Syst Appl 38:5019–5027. https://doi.org/10.1016/j.eswa.2010.09.148

United Nations, Department of Economic and Social Affairs (2019). *World urbanization prospects The 2018 Revision*. New York. https://doi.org/10.18356/b9e995fe-en

Wang B (2019). Applying machine-learning methods based on causality analysis to determine air quality in China. Pol J

Environ Stud 28:3877–3885. https://doi.org/10.15244/pjoes/99639

Wang W, Men C, Lu W (2008) Online prediction model based on support vector machine. Neurocomputing 71:550–558. https://doi.org/10.1016/j.neucom.2007.07.020

Wang X, Xiao Z (2017) Regional eco-efficiency prediction with Support Vector Spatial Dynamic MIDAS. J Clean Prod 161:165–177. https://doi.org/10.1016/j.jclepro.2017.05.077

WCED (1987). *Report of the World Commission on Environment and Development: Our Common Future: Report of the World Commission on Environment and Development*. Oslo. https://doi.org/10.1080/07488008808408783

Weizhen H, Zhengqiang L, Yuhuan Z, Hua X, Ying Z, Kaitao L, Donghui L, Peng W, Yan M (2014) Using support vector regression to predict PM_{10} and $PM_{2.5}$. In *IOP Conference Series: Earth and Environmental Science* (Vol. 17). IOP. https://doi.org/10.1088/1755-1315/17/1/012268

WHO (2016). OMS | La OMS publica estimaciones nacionales sobre la exposición a la contaminación del aire y sus repercusiones para la salud. *WHO*. http://www.who.int/mediacentre/news/releases/2016/air-pollution-estimates/es/Accesed Accessed 26 November 2018

Yeganeh N, Shafie M. P., Rashidi Y, Kamalan H (2012) Prediction of CO concentrations based on a hybrid Partial Least Square and Support Vector Machine model. Atmos Environ 55: 357–365. https://doi.org/10.1016/j.atmosenv.2012.02.092

Zalakeviciute R, Bastidas M, Buenaño A, Rybarczyk Y (2020) A Traffic-based method to predict and map urban air quality. Appl. Sci. 10. https://doi.org/10.3390/app10062035

Zeng L, Guo J, Wang B, Lv J, Wang Q (2019) Analyzing sustainability of Chinese coal cities using a decision tree modeling approach. Resour. Policy 64, 101501. https://doi.org/10.1016/j.resourpol.2019.101501

Zhan Y, Luo Y, Deng X, Grieneisen M, Zhang M, Di B (2018) Spatiotemporal prediction of daily ambient ozone levels across China using random forest for human exposure assessment. Environ Pollut 233:464–473. https://doi.org/10.1016/j.envpol.2017.10.029

Zhang Y, Huan Q (2006) Research on the evaluation of sustainable development in Cangzhou city based on neural-network-AHP. In *Proceedings of the Fifth International Conference on Machine Learning and Cybernetics* (Vol. 2006, pp. 3144–3147). https://doi.org/10.1109/ICMLC.2006.258407

Zhang Y, Shang W, Wu Y (2009) Research on sustainable development based on neural network. In *2009 Chinese Control and Decision Conference* (pp. 3273–3276). IEEE. https://doi.org/10.1109/CCDC.2009.5192476

Zhou L, Zhou C, Yang F, Che L, Wang Bo, Sun D (2019) Spatio-temporal evolution and the influencing factors of $PM_{2.5}$ in China between 2000 and 2015. J Geogr Sci 29:253–270. https://doi.org/10.1007/s11442-019-1595-0

Zhou Y, Chang F-J, Chang L-Ch, Kao I-F, Wang Y.S (2019) Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts. J Clean Prod 209:134–145. https://doi.org/10.1016/j.jclepro.2018.10.243