

Research Article

Analytical, Dynamic, and Interactive Platform for Generation and Managing Predictive Models Focused on Energy Sector

Inés Romero ¹ and Alberto Ochoa-Zezzati ²

¹Programa de Desarrollo Sostenible y Energías Renovables, Universidad Internacional Iberoamericana, Campeche 24560, Mexico

²Departamento/Facultad: Inteligencia Artificial Aplicada, Universidad Autónoma de Ciudad Juárez, Chihuahua 32310, Mexico

Correspondence should be addressed to Inés Romero; ines.romero@doctorado.unini.edu.mx

Received 7 February 2022; Revised 7 May 2022; Accepted 11 May 2022; Published 6 June 2022

Academic Editor: Antonio Bracale

Copyright © 2022 Inés Romero and Alberto Ochoa-Zezzati. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Ensuring the supply of electricity in a reliable and safe way is not an easy task, especially when considering renewable and clean energy generated with wind turbines given the intermittency or variability of the wind; also considering different time horizons increases complexity. Mexico has great potential for wind energy in the Eastern region and, to meet this challenge, a platform capable of generating forecast models automatically through mathematical techniques and artificial intelligence and managing them is proposed aimed at providing support based on knowledge and presenting the information graphically through a flexible dashboard, which is customizable and accessible when and where required. In this investigation, components related to the generation of electrical energy in this area are identified and a centralized system is proposed, with information segmentation, management of 3 user profiles, 6 KPIs, 5 configurable parameters, 7 different forecast models using statistical techniques, support vector machines, and automatic and deep learning, with 2 ways of visualization, to carry out analyses at 3 different time horizons. It is built in a modular way with free and open-source software. The results in the energy sector show that it allows focusing on priority activities avoiding rework, ensures reliability and completeness, is scalable, avoids duplication, allows resources to be shared, responds quickly to hypotheses, and has a global and summarized view of relevant data according to the interested party for different periods of time in an agile way, reducing times and offering support to the decision maker.

1. Introduction

The variability of renewable energies influences the operation of the electrical system in terms of system reliability and the quality of the energy delivered to users. The operation of the electrical system is critical, so stability depends on good coordination and supervision between generation and dispatch.

One of the measures currently in operation integrates conventional and nonconventional renewable energies to create an electrical network in order to mitigate and carry out an adequate energy balance quickly, accurately, and reliably. One of the main challenges is short-term forecasting that helps make allocation and dispatch decisions in the day-ahead market. The medium- and long-term estimates allow

estimating the demand in the range of months and years for the planning and operation of the energy system.

There is a large amount of research to solve problems related to the optimal dispatch of electrical energy [1–3], where different methods have been used, including linear programming, nonlinear programming, and quadratic programming [4–7]. These methods require too much computational resource and execution time, which is why their application is complex. However, they can be overcome using metaheuristic algorithms or evolutionary algorithms [8–11].

The objective of this article is to establish a platform capable of applying different techniques implemented through a customizable, dynamic, and interactive interface in order to coordinate and maintain different types of power

generation plants in operation under conditions of efficiency, quality, safety, reliability, and sustainability for various regions of the country considering different time horizons [12].

The actions carried out by the staff frequently provide a priori preparation by accumulating experiences and sharpening common sense. But novice specialists may not have enough skill to take the best action, or experienced specialists may not be available when needed, drop out, or forget how to deal with some scenarios. So, in these cases, a platform that helps reduce response and support times to make the best decision is appropriate.

A platform built with free, open-source software [13, 14], which is modular and scalable, is proposed. One of the components that integrates it contains the extraction, cleaning, transformation, and loading of information in a centralized database (ETL), available to build different models, which avoids investing time and effort in these actions again and is usable for all areas that require the information, reducing rework times and focusing on priority activities, as well as avoiding duplication and ensuring the reliability and completeness of the data [15–18].

The platform provides three different user profiles: operator, boss, and administrator. The boss must configure the five parameters available in the dashboard, although there is a preestablished option to run the forecast models to generate the results. The administrator must register the users, configure their profiles, and have the same permissions as the head of operations. The operator only has the display, without the option of modifying the parameters or execution.

The dashboard has six performance indicators (KPIs) related to energy generation and the results of the forecasts can be displayed in a linear or interval graphical representation.

Tests were performed on this platform using a centralized database of the energy generated from January 29, 2016, to April 30, 2021, from the Eastern region of Mexico, since 25% is generated in this area and there is also a great potential for wind energy. From this information, knowledge is extracted for the construction of 7 forecast models implemented with the following techniques: autoregressive (AR), moving average (MA), autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA), support vector machine (SVM), machine learning with Bayesian networks (BN), and deep learning using neural networks (NN) applying data mining techniques [19–22].

Electricity is generated in Mexico through various sources. 68.5% of energy is generated through fossil fuels and 3.74% using nuclear power [23–26] which helps to avoid problems of reliability, security, and stability in the electrical network, unlike the 27.82% that is generated from renewable and clean energies such as hydroelectric, geothermal, wind, biomass, and photovoltaic energies.

The dashboard uses this knowledge in a dynamic and interactive way that supports various users, where the results show the reduction of interoperability times at different time horizons, compared to learning models in a standard

separate way, availability of a centralized base to organize, facilitating the distribution and creation of new knowledge, which allows you to quickly answer hypotheses and consult results and metrics in a visual and personalized way to get the most out of the information.

This document is organized as follows: Section 2 presents the theoretical basis, proposed architecture, the implemented models, and the applied methodology. In Section 3, it addresses the results obtained for the problem that occurs in the Eastern region of Mexico applied to different periods of time using different models implemented. The experiments and results obtained for each time horizon generated by the platform are described. Section 4 shows the advantages, results, limitations, and conclusions for the energy forecast estimated by the platform.

2. Methods

The proposed platform is implemented by modules independently to achieve scalability, portability, and reusability.

For scalability, when implementing components, new behaviors can be modified, discarded, or added to existing ones without affecting others. Portability is allowing it to run seamlessly on multiple platforms, which means that it is not dependent on a single operating system in particular and maintainable using the principle called “Don’t Repeat Yourself” (DRY) which applies to design patterns that encourage the creation of maintainable and reusable code in order not to repeat the code so that there is no duplication [27–29].

It also allows you to focus at the appropriate levels with the details that are required in specific areas, as in data processing and forecast models, in views or communication between them.

Figure 1 shows the proposed general architecture that offers a set of decoupled components.

The first bottom-up component of the image considers the exploration of data related to energy generation extracted from 2016 to 2021 from the Eastern region of the country from the public and reliable source of information from Mexico by CENACE, 2021 [30]. This set of data to be used is defined as

$$C = \{x_1, x_2, x_3, \dots, x_n\}, \quad (1)$$

where C is data set and x_n is value for data n .

Subsequently, the data is *extracted, cleaned, transformed, and loaded* into a centralized repository, so that it is available for use by the following processes. The D_i functions represent the sets of 1,790 daily records categorized and stored by the type to which they correspond: generation, day, month, year, and season.

The *application server* implements different forecast models δ_j built by applying the AR, MA, ARMA, ARIMA, SVM, BN, and NN algorithms.

The controller responds to events generated by the interface, usually triggered by the user, or invoked by requests managed by the model caused by some data update or execution from the dashboard, in such a way that it works as

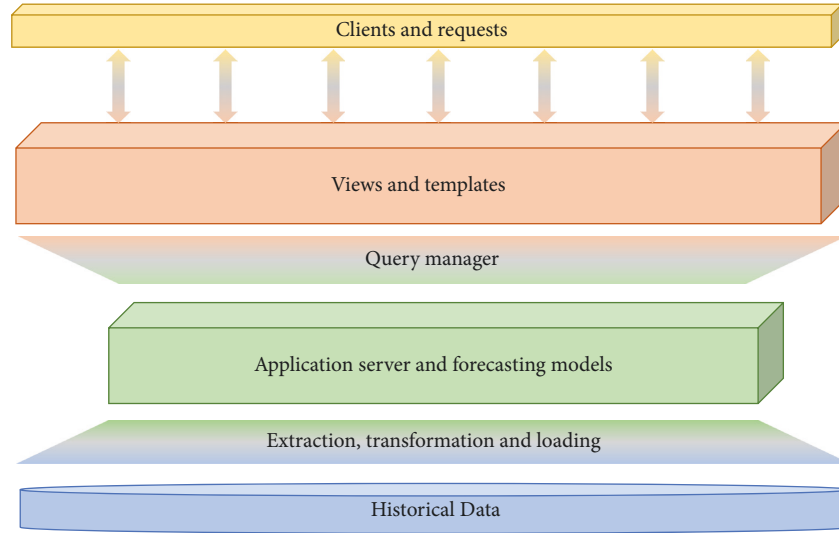


FIGURE 1: Proposed platform architecture.

a *query manager* or intermediary between the views and the application server.

An application based on the model view controller (MVC) waits for requests from the browser of the website. When it receives a request from the interface, the model generates the representation of the information managing the access permissions to the controller, given the configuration of the user, whether operator, operation manager, or administrator, and the model sends the information to the *views and templates* to present it as output.

An administrator user profile or boss can establish the configuration of the parameters date, season, models, type of graph, and time horizon to use (P_k) which invoke the necessary data for the selected algorithms and execute the models.

The *clients and requests* module allows interaction with users from any access point where the application is available according to the permissions assigned to each user profile.

This strategy allows the adaptation of a particular model or new models to be implemented on the platform in an agile and efficiently integrated manner, if required by the acquisition of new knowledge [31–34].

2.1. Model Building. The functions ϕ_j contain the implemented algorithms available to estimate the generation and where it relates the necessary data sets D_i to execute the operation.

The general data analysis of D_i is shown in Table 1.

The annual data can be visualized, since it is a daily time series; in Figure 2, each year is plotted as a separate line in the same plot. This allows the patterns to be compared side by side.

The time series has a pattern with lows at the beginning and end of the year affected by the temperatures recorded in this geographical area of the country, which reduces power generation.

For this, a sample of the population of the data representative of the universe of 80% is considered, saving 20% for model validation.

TABLE 1: Analysis of energy generation in Mexico from the Eastern region.

	Day	Month	Year	Generation
Count	1919.0	1919.0	1919.0	1919.0
Mean	15.739969	6.350182	2018.216780	7471.461079
Std	8.807205	3.456770	1.537813	912.990937
Min	1.0	1.0	2016.0	4506.442532
25%	8.0	3.0	2017.0	6857.759842
50%	16.0	6.0	2018.0	7409.817362
75%	23.0	9.0	2020.0	8077.294021
Max	31.0	12.0	2021.0	10210.865800

Source: Estimate of Real Demand in Mexico by CENACE (2021).

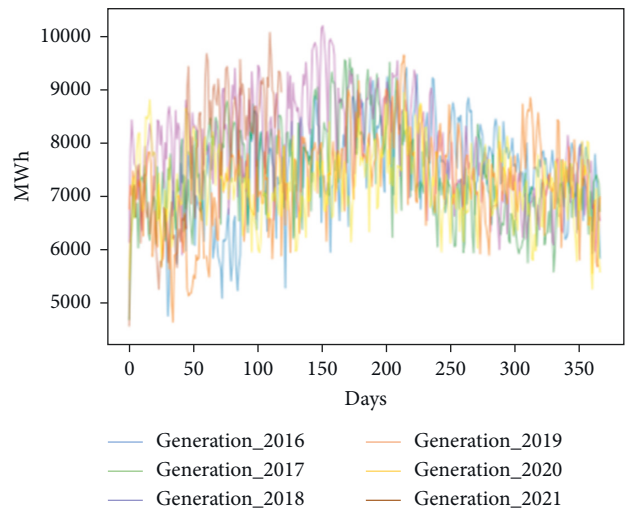


FIGURE 2: Graph of energy generation data by year.

$$D_i = \{x_1, x_2, x_3, \dots, x_i\}, \quad (2)$$

where D_i is the data set to use for model building and x_i is the value for data i .

Default execution parameters are established using all available data with a short-term time horizon in a multi-linear graph by considering all available algorithms simulating a Supplementary Allocation of Power Plant Units for Reliability (Asignación Suplementaria de Unidades de Central Eléctrica para Confiabilidad AUGC) forecast for the following seven days of operation, considering the time series $\{X_n\}$; the values observed are

$$x_1, x_2, \dots, x_{t-1}, x_t, x_{t+1}, \dots, x_n, \quad (3)$$

where x_1 represents the first value of the series, x_2 represents the second value of the series, x_{t-1} represents the value for a period $t-1$ of the series, x_t represents the value for a period t of the series, and x_{t+1} represents the value for a following period $t+1$ of the series.

If $x_1 \rightarrow x_2$ is represented like x_1 influences x_2 , the simple autocorrelation function aims to study the influence of the various observations:

$$x_1 \rightarrow x_2 \rightarrow \dots x_{t-1} \rightarrow x_t \rightarrow x_{t+1} \rightarrow \dots x_{n-1} \rightarrow x_n. \quad (4)$$

With these data μ_j , models are trained and built, applying the following techniques.

2.1.1. Autoregressive (AR). The autoregression method models the behavior pattern of past observations as a linear function to make time series forecasts of future trends [35, 36].

The notation for the model involves specifying the order of model p as a parameter of the AR function written as

$$x_t = c + \varnothing_1 x_{t-1} + \varnothing_2 x_{t-2} + \dots + \varnothing_p x_{t-p} + \varepsilon_t, \quad (5)$$

$$x_t = c + \sum_{i=1}^p \varnothing_i x_{t-i} + \varepsilon_t,$$

where x_t is the estimated value of the time series in period t , c is a constant; $\varnothing_1, \dots, \varnothing_p$ are the model parameters; x_{t-i} are the values in period $t-i$ of the series; ε_t is a white noise error term.

In this regression model, the response variable at the previous time period becomes the new predictor, and the errors are assumed on errors in any simple linear regression model.

That is, for an estimate of time t , it is based on the data up to $t-1$ and so on.

2.1.2. Moving Average (MA). Unlike the AR model, which uses past data to predict trends, the moving average method models the sequence as a linear function of past residual errors in a regression model to construct an averaged trend across the data [37, 38].

It can be defined as the weighted sum of current random errors and past errors, as shown in the following equation:

$$x_t = c + \theta \varepsilon_{t-1} + \theta \varepsilon_{t-2} + \dots + \theta \varepsilon_{t-q} + \varepsilon_t, \quad (6)$$

$$x_t = c + \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i},$$

where x_t is the estimated value of the time series in period t ; c refers to noise; ε_t is error term; θ_i is data point coefficient; ε_{t-i} represents prior period errors.

2.1.3. Autoregressive Moving Average (ARMA). The ARMA method combines autoregression (AR) and moving average (MA) models [39, 40].

Therefore, the ARMA(p, q) notation defines order p of the autoregressive part and order q of the moving average part:

$$x_t = c + \sum_{i=1}^p \varnothing_i x_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t, \quad (7)$$

where x_t is the estimated value of the time series in period t ; c is a constant; p is the order of the autoregressive model; $\varnothing_1, \dots, \varnothing_p$ are the parameters of the model; x_{t-i} are the values in period $t-i$ of the series; q is the order of the moving average; θ_i is the coefficient of the data point; ε_{t-i} is the error of the previous period; ε_t is the current error.

The autoregressive model extracts the pattern from the trend and the moving average captures the white noise.

2.1.4. Autoregressive Integrated Moving Average (ARIMA). This model combines autoregression differencing and a moving average model for data series. It has three components: autoregressive (AR), which refers to a model that shows a variable that changes its own lagged or previous values; integrated (I), which represents the differentiation of raw observations to allow the time series to become stationary; that is, the data values are replaced by the difference between the data values and the previous values; and the moving average of the data set (MA), which incorporates the dependence between an observation and a residual error of a moving average model applied to lagged observations [41, 42].

The general model of ARIMA is (p, d, q), where the parameters p, d , and q are nonnegative integers, and it can be represented as

$$x_t = c + \sum_{i=1}^p \varnothing_i \Delta^d x_{t-i} - \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t, \quad (8)$$

where x_t is the estimated value of the time series in period t ; c is a constant; p is the order of the autoregressive model; \varnothing_p is order of the autoregressive part of the stationary series; x_{t-i} are the values in period $t-i$ of the series; q is the order of the moving average; θ_q is order representing the moving average of the stationary series; ε_t is error term; d represents the number of differences needed to make the original series stationary.

2.1.5. Support Vector Machines (SVM). As an example of supervised machine learning models, the support vector machine focused on regression is called SVR for its acronym in English for Support Vector Machine for Regression [43, 44].

It consists of mapping a data set $\{(x_1, y_1), \dots, (x_m, y_m)\}$, such that $x \in X$, with a linear function given by

$$f(x) = wx + b, \quad (9)$$

where w is the magnitude of the vector of hyperplane; x represents pairs of values in the plane; b is a scalar threshold; and $|b|/|w|$ represents the perpendicular distance from the separating hyperplane to the origin.

In the case of regression, a tolerance margin ξ is established near the vector in order to minimize the error, taking into account the fact that part of this error is tolerated, so it is used by restrictions on Lagrange multipliers that are denoted by α , described as

$$|y_i - f(x_i, \alpha)| = \begin{cases} \xi si y_i - f(x_i, \alpha) \leq \xi \\ |y_i - f(x_i, \alpha)| \text{ other case} \end{cases}. \quad (10)$$

The objective is to find a function $f(x_i, \alpha)$, which maximizes the deviation of ξ with respect to the objectives y_i for the entire data set, at the same time being as flat as possible; it refers to looking for the smallest possible parameter for w . One way to ensure this is through the minimization of the norm $\|w\|^2 = w \blacksquare w$; this problem can be rewritten as a convex optimization problem:

$$\begin{aligned} \text{Minimize: } & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \psi_i, \\ \text{Subject to restrictions: } & \begin{cases} y_i - wx_i + b \leq \varepsilon + \psi_i, \\ y_i - wx_i + b \geq -\varepsilon - \psi_i, \end{cases} \end{aligned} \quad (11)$$

where $w: w \in \mathbb{R}^n$ is a normal vector that defines the boundary; C is a constant and it must be greater than 0; ξ and ξ^* are the variables that control the error made by the regression function when approximating the bands; b is distance to origin.

2.1.6. Machine Learning with Bayesian Networks (BN). It is a method based on the theory of probability [45–47], a branch of mathematics that studies random and stochastic phenomena in a rigorous way, expressing it through a set of axioms.

These axioms formalize the probability in terms of a value in a model determined by the conclusions of the observations between 0 and 1, called a probability measure that can occur in an event, in order to quantify and know if an event is more likely than not.

Complex models, where not all the parameters involved, are known or what their relationships between them are are modeled using probability distributions.

To describe experiments with random outcomes mathematically, the notion of space of elementary events or outputs corresponding to the experiment under consideration is needed. Let Ω denote any set such that each outcome of the experiment of interest can be uniquely specified by the elements of Ω .

Consider finite or infinite spaces of elementary events Ω . These are the so-called discrete spaces. The elements of a

space Ω are denoted by the letter w and we will call them elementary events. The probability of an event A is defined as follows:

$$P(A) = \sum_{w \in A} P(w). \quad (12)$$

In 1763, Bayes' theorem, proposed by the English mathematician Thomas Bayes, was published, which expresses the conditional probability of a random event, based on prior knowledge of the conditions that could be related to said event and where one learns about the world through the approach.

One gets closer to the truth as more evidence is collected. This argument can be expressed mathematically through Bayes' theorem.

Let A be any event and let $\{B_1, B_2, \dots, B_n\}$ be a set of mutually exclusive and exhaustive events, such that the probability of each of them is different from zero, with positive probabilities such that the sequence of events B_1, B_2, \dots, B_n can be infinite. Then the following probability formula for A is given:

$$\begin{aligned} P(B_i | A) &= \frac{P(A | B_i)P(B_i)}{P(A)}, \\ P(A | B_i) &= \frac{P(B_i | A)P(A)}{P(B_i)}, \end{aligned} \quad (13)$$

where $P(B_i)$ are the observed probabilities in event B_i , $P(A | B_i)$ is the conditional probability of event A in hypothesis B_i , and $P(B_i | A)$ is the probability of any of events B_i given by A .

2.1.7. Deep Learning Using Neural Networks (NN). Neural networks try to mimic the way a human brain approaches problems and uses interconnections between interconnected units to learn and infer relationships based on observed data.

A neural network is made up of a network of neurons that are organized into at least two layers: an input layer with the predictor variables and an output layer made up of the forecasts. There may also be one or more intermediate layers that contain hidden neurons, which is why it is called the hidden layer. In each of them, there are a set of units called artificial neurons, which are connected to each other to transmit signals through links [48–50].

More complex systems have more connected layers, and when there is more than one hidden layer, it gives greater depth to the model; hence the adjective deep learning, where different heterogeneous layers are used, different types of signals in time and interpolate greater complexity than a set of Boolean variables. As can be seen in Figure 3, neural networks are often used when data is unlabeled or unstructured.

Deep learning represents an approach that is more similar to the functioning of the human nervous system. Since the brain has a highly complex microarchitecture, in which nuclei and different areas whose networks of neurons are specialized to perform specific tasks have been

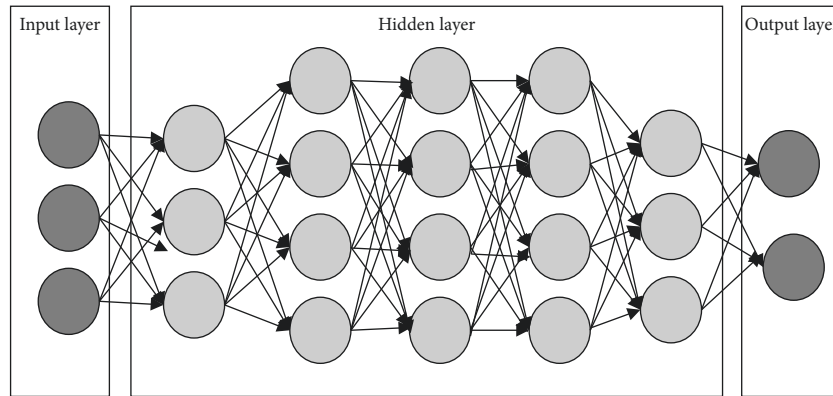


FIGURE 3: Structure of a multilayer neural network.

TABLE 2: Analysis of algorithms for forecast models.

Algorithm	RMSE	MAE	MSE	MAPE	Time (seconds)
AR	523.73	398.21	274300.49	5.55	0.03
MA	611.08	470.37	373424.97	6.5	0.03
ARMA	573.27	448.07	328640.59	6.29	85.58
ARIMA	559.73	436.08	313302.53	6.08	120.19
SVM	863.24	695.08	745192.52	9.97	0.19
BN	865.19	691.37	748566.32	9.86	4.8
NN	879.77	723.71	773998.31	10.31	87.98

discovered, this allows for networks of process units within the global system that specialize in discovering certain hidden features in the data.

Each model μ_j provides a forecast F_j , which estimates power generation. These values are validated and adjusted by the test data.

2.2. Evaluation of Optimal Models. The models were reviewed again and again until obtaining an algorithm (O_j) with the best results in the statistical tests, with the lowest RMSE, MAE, MSE, and MAPE, sufficiently valid, without overfitting the function, considering its restrictions and characteristics, detected for each of them, which is shown in Table 2.

Once these techniques have been optimized, δ_j are obtained, which refer to the possible solutions adjusted and implemented in the platform. Although adjustments to the models can be made in a simple or new way, more precise algorithms can be integrated to those presented previously in order to achieve the established objectives.

By considering several forecast models integrated in the platform, it allows obtaining different estimated values for the same scenario, which allows evaluating different possible solutions for the same problem.

The advantages and disadvantages of each of the methods can be analyzed in relation to the results obtained, together with the techniques used. Each approach has strengths and weaknesses, but the positive is the complementary use to increase the possibilities so that they are used effectively and provide the enormous wealth of information

that all methodologies can provide and thus make better decisions based on evidence.

3. Results

The platform and mechanism for generating forecast models for electricity generation in the eastern sector of the country have been implemented. First, different algorithms are evaluated within the platform, and then the results are presented in the user interface for each of them, so that users can make decisions based on the information presented.

The main menu houses the interactive dashboard, shown in Figure 4, which is an information management tool that visually shows the six key performance indicators (KPI), metrics, and fundamental data of the business, which allows analyzing and monitoring its status.

The information is arranged in the following way for the administrator user or head of operation:

- (i) In the *orange* box on the left, it contains a menu to configure the five parameters (date, season, forecast methods, type of graph, and horizon).
- (ii) A data panel appears on the right side of the interface, in the area marked *yellow*.

For all users, the different types of descriptive analysis of generation and demand data are available on the upper central *screen* in the green area.

It uses historical data from 2016 to 2021 of generation and demand, which are presented in box diagrams

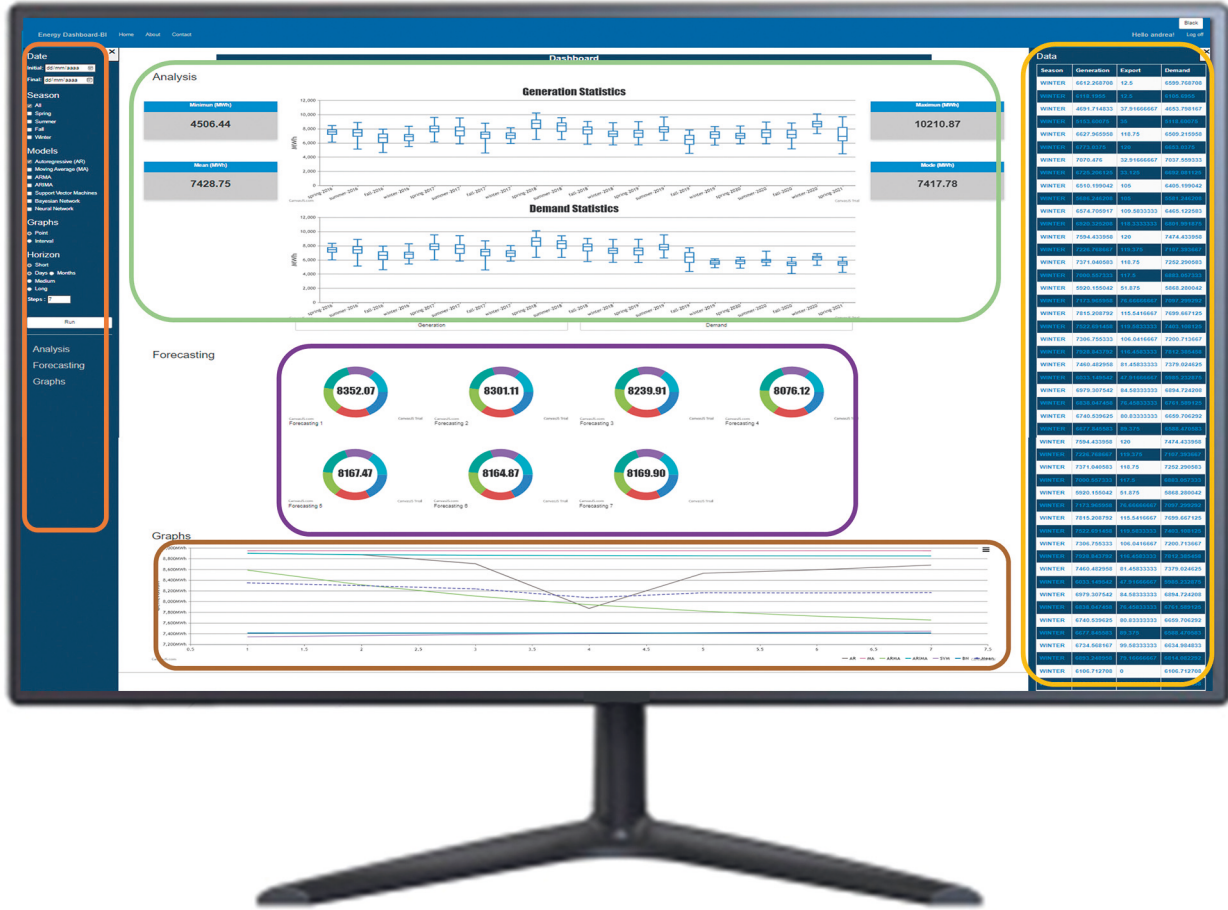


FIGURE 4: Interactive dashboard of the proposed platform.

grouped by seasons of the year, as well as obtaining quantitative metrics. The indicators shown, in this case, are the minimum, maximum, mean, and mode values which are obtained from the generation data.

The following charts focus on predictive analytics, where statistical techniques, vector support machines, machine learning, and deep learning are used, as well as models trained with historical data. This analysis looks at past events to estimate future events, which gives an answer to what is likely to happen under certain conditions.

The donut graphs in the purple box represents the results for each of the times $t + i$, corresponding to the period of the horizon and around it you can see the values estimated by each of the selected algorithms and in the central part the calculation of the mean considering all forecasts.

In the lower part, framed in the brown area, the graph of forecasts of the selected type is shown and also the calculated mean of the estimates generated from the selected techniques is graphed; this is to have another point of comparison using a dotted line.

The selection of a method depends on the context of the forecast and the degree of accuracy desired.

Once the parameters are set within the allowed ranges, the platform can be run and the corresponding algorithms are called to generate the forecasts.

TABLE 3: Execution time of each of the implemented algorithms.

ID	Algorithms	Time (seconds)
1	AR	0.04
2	MA	0.02
3	ARMA	1.24
4	ARIMA	1.89
5	SVM	0.13
6	BN	0.09
7	NN	3.48

The time it takes to execute each of the algorithms considering all days, seasons, and selecting a type of linear graph for a short-term time horizon simulating a forecast of Supplementary Allocation of Power Plant Units for Reliability (AUGC) for the following seven days of operation of the Wholesale Electricity Market is represented in Table 3.

The time it takes to execute all the algorithms on the platform is 6.89seconds, because it is cumulative, since parallelism is not being applied.

The platform offers various parameters that produce different combinations and thus different results, according to the values selected in each of the variables. Table 4 shows the possible choices for each of the implemented models.

TABLE 4: Selectable parameters for each model.

Parameters	AR	MA	ARMA	ARIMA	SVM	BN	NN
Date	X	X	X	X	X	X	X
Season	X	X	X	X	X	X	X
Linear type	X	X	X	X	X	X	X
Interval type	X	X	X	X	X	X	X
Short horizon	X	X	X	X	X	X	X
Medium horizon	X	X	X	X	—	—	—
Long horizon	X	X	X	X	—	—	—
Days period	X	X	X	X	—	—	—
Months period	X	X	X	X	—	—	—

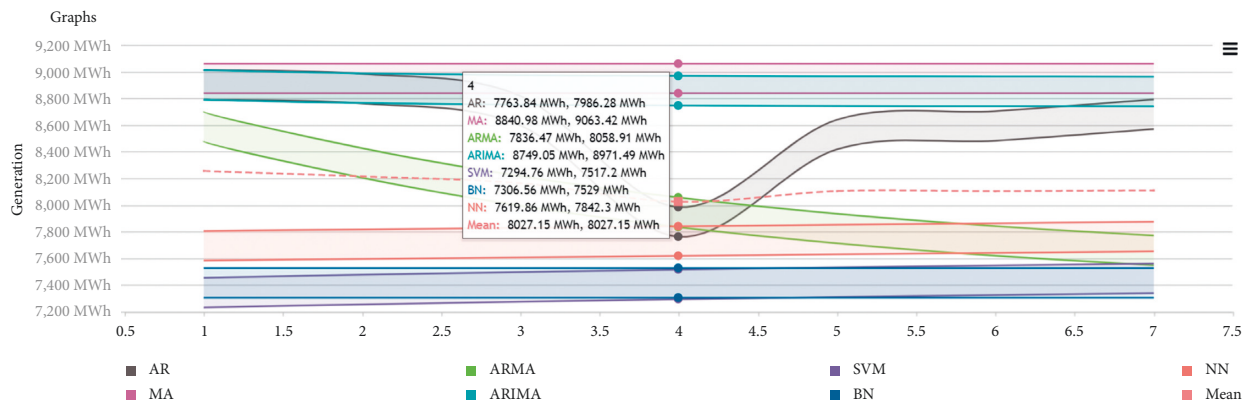


FIGURE 5: Interval plot for a short-term time horizon.

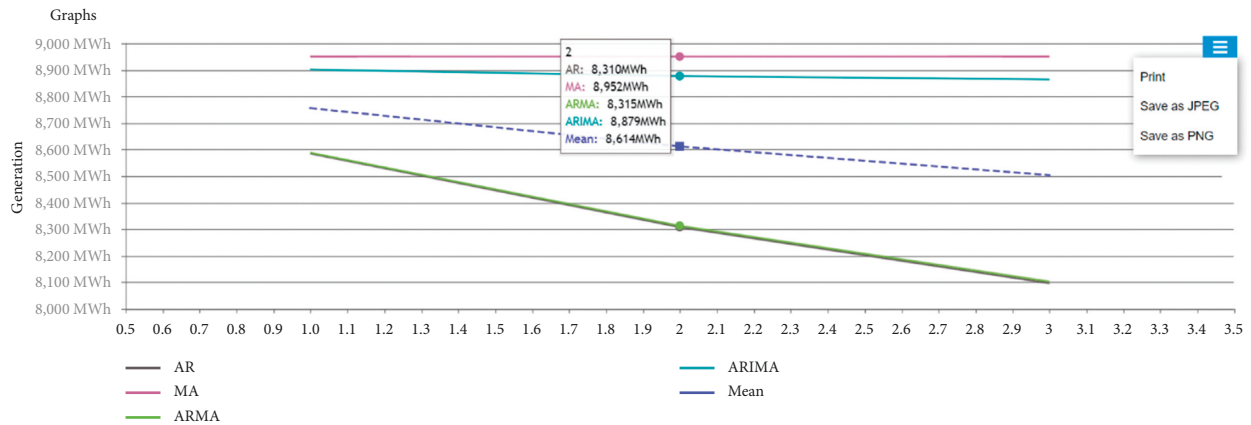


FIGURE 6: Multilinear graph for a medium-term time horizon.

The seven algorithms implemented in the platform can be executed for the short-term period of time. The use of the SVM, BN, and NN algorithms is restricted for this horizon, since in the medium-term and long-term days and months are used to estimate generation.

To generate forecasts for a short time period of months, the data is grouped according to the year to which each of the days belongs. Therefore, the models built by SVM, BN, and NN can only be executed using the time period days.

It can be plotted as an interval or with lines, and in both cases the mean is shown with a dotted line calculated from the values estimated by the algorithms.

Next, Figure 5 shows the results and graphs obtained through the dashboard of the tool, using all the data, all the seasons, selecting the interval option, by days, with the seven algorithms: AR, MA, ARMA, ARIMA, SVM, BN, and NN.

The interval graph shows the indicators of the minimum and maximum generation of the generated forecasts, as well as the estimated mean for each time point $t + i$.

In the *medium* horizon, the data is grouped by years and the available selection of steps or time period is from 1 to 3 to estimate. You can choose linear or interval graph; therefore, there are 2 different types of visual representation, and the date can be changed.

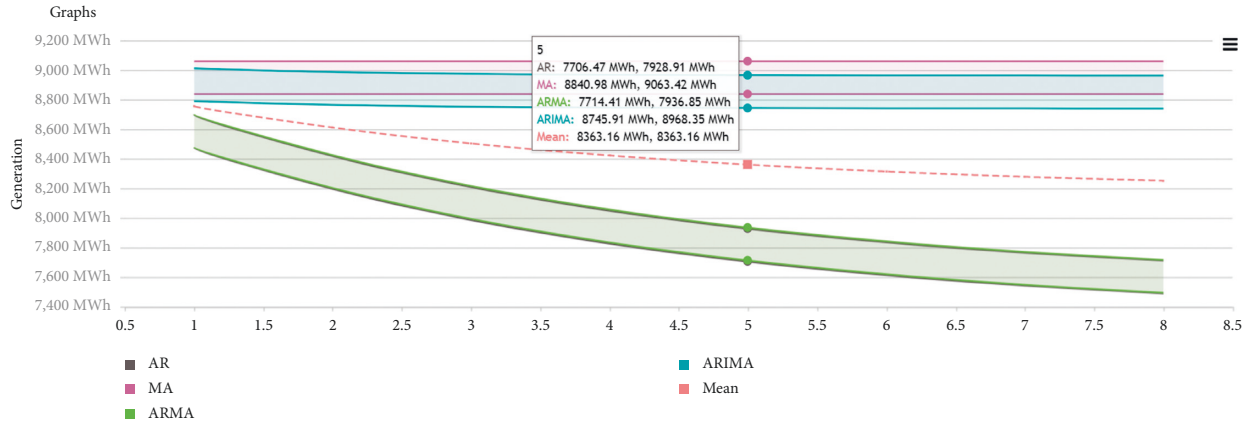


FIGURE 7: Interval plot for a long-term time horizon.

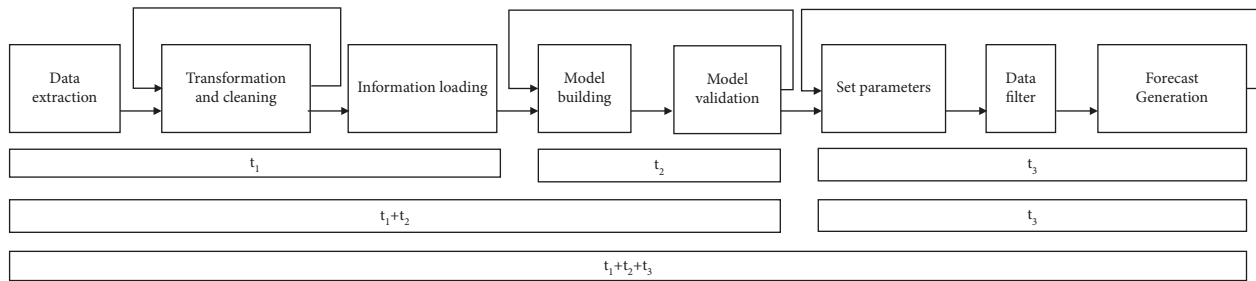


FIGURE 8: Linear execution diagram of the proposed platform.

Figure 6 shows the linear graph for a medium time horizon with 3 periods for the AR, MA, ARMA, and ARIMA algorithms. Although the AR values are very close to those estimated by the ARMA model, they can only be distinguished when the mouse is positioned over the graph at a time point $t + i$.

The figure above shows the options available on the dashboard in the upper right corner of the graph: print and save as an image in jpg or png format.

For a long-term time horizon intended for planning, the selected data are grouped by years and can be estimated from 4 to 8 years for the AR, MA, ARMA, and ARIMA algorithms.

The value of the mean calculated for an instant of time $t + i$ can be seen more clearly by positioning the mouse over the point of interest, as shown in Figure 7.

Linear graphs are recommended to represent time series; however, interval graphs show values within an estimated range between two values, a minimum and a maximum, depending on the user whose graph meets the expected results.

The process focused on the energy sector of the Eastern region within the platform which is currently linear and is described in Figure 8.

The figure describes the time that the ETL process takes, defined as t_1 , the time the construction of models and their validation takes is t_2 , and the time setting the parameters, filtering the data, and generating the forecasts take is t_3 .

The time that t_1 consumes, until the data is loaded in a central repository, is variable given the quantity and quality of the data, as well as time t_2 that is needed to generate

models and adjust them; this depends on the established considerations and the human’s expertise. However, times t_1 and t_2 , after being implemented in the platform, do not need to be considered again, only time t_3 .

Time t_3 is affected by the duration of the extraction by the query in the database, depending on the selected parameters, the number of models chosen, the generation of forecasts, and the rendering of the graphs.

4. Discussion and Conclusions

More and more companies invest in personnel, infrastructure, and equipment in the processing of information related to the business, that is, extracting key information, identifying patterns and trends, and making estimates, due to the great value and impact it generates for the strategic decision-making process, decisions in the short, medium, and long term.

There are specialized systems that generate forecasts in the market, but the prices are high. Some have payment for use, for data, for algorithms, for visitors, and in other cases for periods of time. This platform, being developed in a modular way with free and open-source software, can be maintained and scaled according to needs. Among the advantages observed given its implementation, the following can be mentioned:

- (i) It saves time and effort and focuses personnel on priority activities, since it avoids searching for information from different sources and performing

ETL procedures, over and over again, which deviates from the main activity, that is, optimizing energy generation.

- (ii) It ensures the reliability and completeness of the data by bringing together all the information available from different sources and having the same business rules applied to all the data, preventing each person from trying to standardize their own format.
- (iii) avoids the duplication of information and, therefore, reduces the number of available resources needed to store it and distribute it to all users who require it, not only to generate forecasts but also to be used for other purposes in other areas.
- (iv) high availability of the data, by having the data centralized, it helps to back it up and export it to another place, provides agility, and saves time when moving from one solution to another.
- (v) It responds to hypotheses more quickly, since times t_1 and t_2 are saved in the following executions or to generate new models.
- (vi) by allowing various profiles and different users for each of them, which facilitates access according to the assigned permissions. The established authorizations are defined on the profile and can be viewed for all people assigned to that position by strategic planning.
- (vii) dark mode for low light conditions and visual fatigue is reduced so that it can be configured depending on the user's needs.

This platform contributes to being prepared for the times that come with the acquisition of greater volumes of data (Big Data) in real time [51–54], accessing answers in a short period of time, as well as the ability to engage with the algorithm changes in a modular way, although at the moment it is necessary that the person who implements new models has programming knowledge.

This platform is limited to the energy sector, considering the data, attributes, metrics, and the business rules. Therefore, to implement it in another area, it is necessary to carry out an analysis of it to define some opportunities, determine a clear, specific, and viable objective, and review which indicators are sought according to the operation of the business, the users to those that are focused, and the permissions for each of them to prevent them from executing actions that may incur deviations.

The preprocessing of the data to clean it, adjust it, and transform it into a structure to be used is part of the ETL process within the platform, although it is not automated at the moment, which could mean saving time and money at various points for achieving greater competitiveness and other additional benefits.

Currently, only historical data from the Eastern region of the country is used; however, increasing the number of regions can be an excellent step to analyze the data and study how all the regions of the country interact.

Some improvement opportunities have been detected on the platform that is executed here which can reduce the processing time:

- (i) Connect input devices with the information directly to the central repository, to incorporate data upload in real time or where periods and frequencies can be established for data reading.
- (ii) Dynamically graph as the data arrives at the platform and represent it on the dashboard.
- (iii) Incorporate new algorithms and increase the time horizon implemented in the dashboard.
- (iv) Segment the data by similar days, special days, and vacation periods and with them generate new models.
- (v) Automatically adjust, validate, evaluate, and optimize the accuracy of forecast models. For this, it is necessary to observe the real values, compare them to measure the error (\mathcal{E}), and adjust the model that meets a satisfactory level of reliability.
- (vi) Emulate possible scenarios, modifying the values of the variables, in order to obtain personalized recommendations and anticipate actions in advance.
- (vii) Automate the models to be adaptive and evolutionary, when data arrives which is out of the expected, ask if it is acceptable, and adjust the models automatically by applying relearning.
- (viii) Add real-time visualization of geospatiality by country regions.
- (ix) Simultaneously process different models, applying parallelism on the platform.
- (x) Display automated alarms with specific expected values and unexpected values or when it falls outside of some allowed range.
- (xi) Add prescriptive analysis in the dashboard with metrics and estimates capable of proposing personalized recommendations according to each of the user profiles.

The general idea is to increase the value of the platform, turning it into a reactive system, dynamically updated in real time, with the ability to offer more accurate forecasts, with adaptive and evolutionary algorithms over time based on data.

The platform could function as a virtual laboratory, where different scenarios can be created and new hypotheses answered, thus establishing adequate strategies to deal with the expected events, as well as to evaluate forecasting methods and obtain answers as to which is the fastest method and which is the most accurate for which time horizon, putting them to compete automatically and obtain the best algorithm, assessing what corresponds and, if necessary, an adjustment can be done independently.

However, despite these limitations, the platform could constitute one of the fundamental bases for improving

energy efficiency in the world, organizations, or nations. Since it is allowed to be prepared for the operation of the operating business, project equipment maintenance, growth planning, to know when to dispatch the cheapest or cleanest energy, giving preference to renewable generation and providing benefits to the environment and the economy [16, 55–57].

An integrated system focused on the energy sector with good forecast results reduces the need for reserves, takes advantage of the diversity of energy sources, and can absorb the fluctuations that are generated to achieve a quality energy supply, making the most of the production infrastructure and distribution.

This platform takes advantage of the knowledge that is generated, its behaviors, and patterns, raising its value and potential and providing support for decision-making in the energy sector.

Data Availability

The data used to support the findings of the study were extracted from the public and available repository of the National Center for Energy Control, Mexico (CENACE), from January 29, 2016, to April 30, 2021, and were averaged to daily data corresponding to the Eastern region of the country, tagged in the files as ORI. These data are available at <https://www.cenace.gob.mx/Paginas/SIM/Reportes/EstimacionDemandaReal.aspx>. Subsequently, the methods explained in the article were carried out.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] M. Kubat, *An Introduction to Machine Learning*, Springer International Publishing, Berlin, Germany, 2017.
- [2] J. Pei, J. Han, and H. Tong, *Data Mining: Concepts and Techniques*, Elsevier Science & Technology Books, Amsterdam, Netherlands, 2021.
- [3] A. Puder, *Distributed Systems Architecture: A Middleware Approach*, Elsevier, Amsterdam, Netherlands, 2005.
- [4] C. Chase, *Demand-Driven Forecasting*, John Wiley & Sons, Hoboken, NJ, USA, 2009.
- [5] R. Nahle, “En México se tiene capacidad instalada para generar electricidad a través de energía renovable en 31 por ciento,” 2020, <https://twitter.com/rocionahle/status/1262054906594123776>.
- [6] J. Wakefield, *Bayesian and Frequentist Regression Methods*, Springer, Berlin, Germany, 2013.
- [7] R. Weron, *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*, John Wiley & Sons, Hoboken, NJ, USA, 2007.
- [8] I. Azmira, A. Razak, S. Majid Mohd, and H. A. Rahman, “Electricity load forecasting using data mining technique,” in *Data Mining Knowledge Discovery and Applications*, Intechopen, London, UK, 2012.
- [9] T. W. S. Chow and C. T. Leung, “Neural network based short-term load forecasting using weather compensation,” *IEEE Transactions on Power Systems*, vol. 11, no. 4, pp. 1736–1742, 1996.
- [10] J. H. Friedman, “Multivariate adaptive regression splines,” *Annals of Statistics*, vol. 19, no. 1, pp. 1–67, 1991.
- [11] A. García and H. Medina, *Predicción de demanda de energía en Colombia mediante un sistema de inferencia difusa neuronal*, Universidad Nacional De Colombia, Bogotá, Colombia, 2005, <https://www.redalyc.org/articulo.oa?id=147019387002>.
- [12] H. Danny, “Energy and the new reality 1: energy efficiency and the demand for energy services,” *Choice Reviews Online*, Routledge, London, U.K, 2010.
- [13] J. S. Canvas, “JavaScript Gallery Overview,” 2021, <https://canvasjs.com/javascript-charts/>.
- [14] S. F. Django, “The Web Framework for Perfectionists with Deadlines | Django,” 2021, <https://www.djangoproject.com>.
- [15] W. Ertel, *Introduction to Artificial Intelligence*, Springer International Publishing, Berlin, Germany, 2017.
- [16] M. K. Evans, *Practical Business Forecasting*, Blackwell Publishing Ltd, Hoboken, NJ, USA, 2003.
- [17] L. Igual and S. Seguí, *Introduction to Data Science*, Springer International Publishing, Berlin, Germany, 2017.
- [18] M. Pérez, *Business Intelligence: Técnicas, herramientas y aplicaciones. España, Alfaomega*, Alfaomega, Móstoles, Spain, Grupo RC Libros, 2015.
- [19] C. C. Aggarwal, *Data Mining*, Springer International Publishing, Berlin, Germany, 2015.
- [20] M. H. Dunham, *Data Mining: Introductory and Advanced Topics*, Prentice-Hall, Hoboken, NJ, USA, 2002.
- [21] S. S. Skiena, “The data science design manual,” *Texts in Computer Science*, Springer International Publishing, Berlin, Germany, 2017.
- [22] H. Trevor, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning Data Mining, Inference, and Prediction*, Springer, Berlin, Germany, 2001.
- [23] International Energy Agency (Iea), “Explore energy data by category, indicator, country or region,” 2021, <https://www.iea.org/data-and-statistics/data-tables/?country=WORLD&energy=Electricity>.
- [24] M. E. Ortiz, “Pronóstico de la demanda eléctrica residencial basado en el modelo de regresión adaptativa multivariante spline (MARS),” 2015, Universidad Politécnica Salesiana, Quito, Ecuador, <https://dspace.ups.edu.ec/bitstream/123456789/11290/3/UPS-KT01191.pdf>.
- [25] Secretaría de Energía (Sener), “Programa de desarrollo del sistema eléctrico nacional,” 2021, <https://www.gob.mx/sener/acciones-y-programas/programa-de-desarrollo-del-sistema-electrico-nacional-33462>.
- [26] Secretaría de Medio Ambiente y Recursos Naturales (SEMARNAT), “Capacidad instalada y generación de electricidad por sector y fuente de energía limpia,” 2021, http://dgeiawf.semarnat.gob.mx:8080/ibi_apps/WFServlet?IBIF_ex=D2_ENERGIA03_05&IBIC_user=dgeia_mce&IBIC_pass=dgeia_mce&NOMBREENTIDAD=.
- [27] I. Englander, *The Architecture of Computer Hardware, Systems Software, and Networking: An Information Technology Approach*, Wiley, Hoboken, NJ, USA, 2014.
- [28] S. Ramos, *Business Intelligence and Analytics: El arte de convertir los datos en conocimiento*, SolidQ Global S.A, España, Spain, 2016.
- [29] J. Shaffer, A. Cotgreave, and S. Wexler, *Big Book of Dashboards: Visualizing Your Data Using Real-World Business Scenarios*, Wiley & Sons, Hoboken, NJ, USA, 2017.
- [30] Centro Nacional de Control de Energía (Cenace), “Estimación de la Demanda Real,” 2021, <https://www.cenace.gob.mx/Paginas/SIM/Reportes/EstimacionDemandaReal.aspx>.

- [31] S. A. Gabriel, A. J. Conejo, J. D. Fuller, B. F. Hobbs, and C. Ruiz, "Complementarity modeling in energy markets," *International Series in Operations Research & Management Science*, Springer, New York, NY, USA, 2013.
- [32] W. Hasperué, "Extracción de conocimiento en grandes bases de datos utilizando estrategias adaptativas," 2012, <http://hdl.handle.net/10915/4215>.
- [33] O. Maimon and L. Rokach, "Introduction to knowledge discovery in databases," *Data Mining and Knowledge Discovery Handbook*, Springer, Boston, MA, USA, pp. 1–17, 2021.
- [34] F. Wolfe, "How Artificial Intelligence Will Revolutionize the Energy Industry. Special Edition on Artificial Intelligence," 2017, <https://sitn.hms.harvard.edu/flash/2017/artificial-intelligence-will-revolutionize-energy-industry/>.
- [35] L. Dannecker, *Energy Time Series Forecasting*, Springer Fachmedien Wiesbaden, Wiesbaden, Germany, 2015.
- [36] R. A. Davis and P. J. Brockwell, *Introduction to Time Series and Forecasting*, Springer, Berlin, Germany, 2016.
- [37] T. Halkjelsvik and M. Jorgensen, *Time Predictions: Understanding and Avoiding Unrealism in Project Planning and Everyday Life*, Springer, Berlin, Germany, 2018.
- [38] D. Valencia, C. Moreno, and C. Lozano, "Modelo de promedios móviles para el pronóstico horario de potencia y energía eléctrica. El Hombre y la Máquina," 2007.
- [39] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, Otexts, Australia, 2018.
- [40] J. Murillos, A. Trejos, and P. Carvajal, "Estudio del pronóstico de la demanda de energía eléctrica, utilizando modelos de series de tiempo," *Sci-Tech*, vol. 23, 2003.
- [41] M. Kuhn and K. Johnson, "Applied predictive modeling," *Applied Predictive Modeling*, Springer, New York, NY, USA, 2013.
- [42] V. M. Rueda, J. David, V. Henao, C. Jaime, and F. Cardona, "Electricidad usando modelos No lineales recent advances in load forecasting using nonlinear models," *Dyna*, vol. 78, no. 167, 2015.
- [43] A. K. Maheshwari and Business Expert Press, *Business Intelligence and Data Mining*, Business Expert Press, Bep, New York, NY, USA, 2015.
- [44] D. C. Montgomery, C. L. Jennings, and M. Kulahci, *Introduction to Time Series Analysis and Forecasting (Wiley Series in Probability and Statistics)*, Wiley-Interscience, Hoboken, NJ, USA, 2008.
- [45] A. A. Borovkov, "Probability theory," *Universitext*, Springer, London, UK, 2013.
- [46] T. J. Cleophas and A. H. Zwinderman, *Machine Learning in Medicine - a Complete Overview*, Springer, Berlin, Germany, 2015.
- [47] L. Hidalgo, *Inteligencia Artificial Y Sistemas Expertos*, Universidad de Córdoba. Servicio de Publicaciones, Córdoba, Spain, 1997.
- [48] S. Skansi, *Introduction to Deep Learning: From Logical Calculus to Artificial Intelligence*, Springer, Berlin, Germany, 2018.
- [49] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, "Dive into Deep Learning," 2021, <https://arxiv.org/abs/2106.11342>.
- [50] A. Jamwal, R. Agrawal, and M. Sharma, A. Kumar and S. Luthra, "Two decades of research trends and transformations in manufacturing sustainability: a systematic literature review and future research agenda," *Production Engineering*, vol. 16, no. 1, pp. 109–133, 2022.
- [51] M. Barlow, *Real-time Big Data Analytics: Emerging Architecture*, O'Reilly Media, Inc, Sebastopol, CA, USA, 2013.
- [52] M. Bramer, *Principles of Data Mining*, Springer, London, UK, 2a edition, 2013.
- [53] P. D. Diamantoulakis, V. M. Kapinas, and G. K. Karagiannidis, "Big data analytics for dynamic energy management in smart grids," *Big Data Research*, vol. 2, no. 3, pp. 94–101, 2015.
- [54] J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer-Verlag, New York, NY, USA, 2009.
- [55] American Petroleum Institute (Api), *Climate Change and Energy*, <https://www.api.org/%7E/media/Files/Policy/Environment/Climate-Change-and-Energy/CLIMATE-PRIMER.pdf>, 2021.
- [56] National Research Council, *Climate Change: Evidence, Impacts, and Choices*, The National Academies Press, Washington, DC, USA, 2012.
- [57] A. J. P. Moreno, "Renewable energy and energy saving: worldwide research trends," *Sustainability*, vol. 13, no. 23, Article ID 13261, 2021.