

An Ensemble Model Based in AI Using Past Output Data for Forecast of Operating Variables of Distributed Resources

JUAN HERNÁNDEZ ARRIETA¹, IDI ISAAC MILLÁN², JAVIER SIERRA CARRILLO³

^{1,2} Department of Electrical and Electronic Engineering, Universidad Pontificia Bolivariana UPB, Medellín, COLOMBIA

³ Department of Electronic Engineering, Universidad de Sucre, Sincelejo, COLOMBIA

*This work is part of the project: “Estrategia de transformación del sector energético colombiano en el horizonte de 2030” funded by the research call 778 of MinCiencias Ecosistema Científico. Contract FP44842-210-2018

Abstract: - Non-conventional renewable energies, such as solar, have great potential for generation, however, there is no accurate and timely information for making decisions that lead to the installation of microgrids in interest zones. For this, an ensemble model has been developed to forecast the main operating variables associated with the distributed resources of potential microgrids. To develop the model, a comprehensive technological surveillance of the strategies, models and techniques used for the valuation and forecast of photovoltaic generation and energy storage resources was carried out, then the model K Nearest Neighbors (KNN) was referenced, selected and programmed for the treatment (purification and imputation) of the data; The energy models for these systems were referenced and selected, using the HOMER PRO software model, where the main operating variables that are the object of forecasting were determined for each of the selected distributed resources. The graphical interface of the ensemble model was programmed for the short-term forecast of these variables, using the selected artificial intelligence forecasting techniques (decision trees, k Nearest Neighbor and Random Forests), and finally the results obtained were validated with the model, versus the commercial software HOMER PRO and another artificial intelligence model (Artificial Neural Network - Multilayer Perceptron).

Key-Words: - Generation forecasting, forecasting techniques, artificial intelligence, microgrid, statistical models, assembly model.

Received: June 15, 2021. Revised: August 16 2022. Accepted: September 22, 2022. Published: October 26, 2022.

1 Introduction

Due to their climatology, many areas can be regions that potentially generate high and/or medium powers as a result of DER (Distributed Energy Resources), so based on this research, which emerges from the macroproject “Estrategia de transformación del sector energético colombiano en el horizonte de 2030 (Energética 2030) – Proyecto 08 (58855): Pruebas de concepto de Microrredes en Colombia”, a model was developed to forecast the main operational variables associated with the distributed resources of potential microgrids in the department of Sucre.

Today, for data consultation, the Institute of Hydrology, Meteorology and Environmental Studies - IDEAM, in its Atlas of Solar Radiation of Colombia (2006), has available a series of measurements resulting from the installation and monitoring of meteorological stations in Colombia, among these is the horizontal global irradiation [1]. For the department of Sucre, IDEAM only has 3 automatic stations, located in the municipalities of Sincelejo, Morroa, San Benito, Coveñas and San Marcos, with pyranometers (so-

lar radiation meter), as can be seen in Figure 1.

Since these maps are the most complete in the country, and as a result of the limited existence of meteorological stations installed in the Department, there is no accurate and timely information on the variables associated with the resources distributed in the Department. This shortcoming prevents decision-making related to the installation and implementation of microgrids based on renewable energies and the subsequent forecasting of parameters to facilitate their operational management. Therefore, the model developed and programmed in an easy-to-use graphical interface will help this decision-making, supporting information on output power forecasts in solar modules, among others variables of consideration.

2 Data cleaning and imputation techniques

Data cleaning is a technique that allows you to filter out unwanted values that can affect the fidelity of the original data and lead to errors in subsequent forecasting techniques or any other data processing.

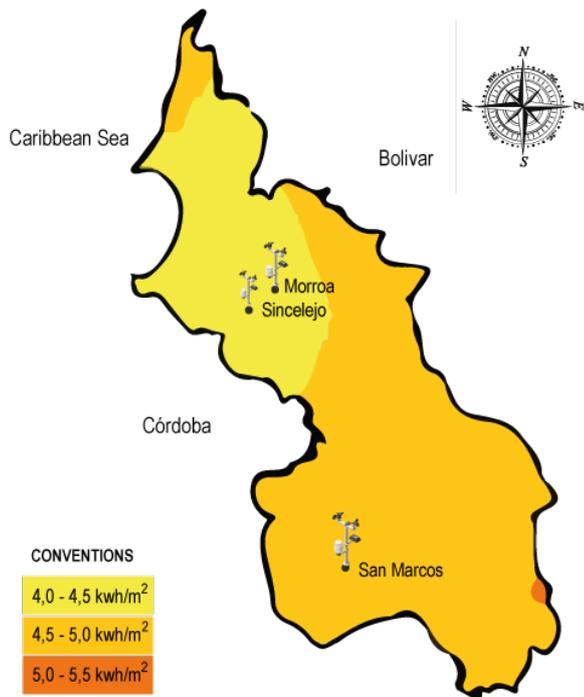


Figure 1: Annual Average Global Horizontal Irradiation in the department of Sucre. Source: Author elaboration based on IDEAM.

In the case of Solar Radiation Databases, it is important to recognize and calculate the number of Missing Values (MV) in these per day, month, hour, etc., depending on the user’s transcendence interval. This is why there are data ranges where the user only ignores or removes the MVs based on knowledge of their behavior.

It is important to categorize the mechanisms that lead to the introduction of Missing Values in data mining [2]. This separate research is based on studies by Luengo et al in [3].

Assumptions that are made about the absence mechanism and PV pattern may affect which imputation method might be applied, if any. There are three different mechanisms for MV induction:

- Missing Completely At Random (MCAR), when the distribution of an example that has a missing value for an attribute does not depend on either the observed data or the MVs.
- Missing At Random (MAR), when the distribution of an example that has a missing value for an attribute depends on the observed data, but does not depend on the MVs.
- Not Missing At Random (NMAR), when the distribution of an example that has a missing value

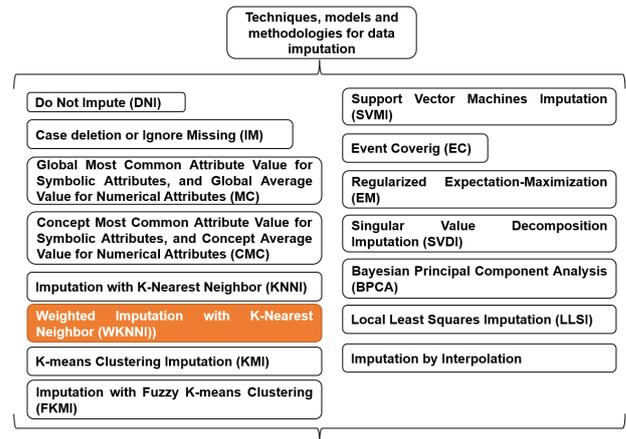


Figure 2: Summary of the most used techniques, models and methodologies for data imputation. Source: Author elaboration based on [3].

for an attribute depends on the MV.

In the case of the MCAR mode, the underlying distributions of the missing and complete data are assumed to be the same, while for the MAR mode they are different, and the MVs can be predicted using the complete data. These two mechanisms are assumed by the imputation methods that were studied in [3] [17] [18].

Below are the imputation methods, which are the most representative and used in the literature presented in [3], as shown in Figure 2, where the highlighted methods correspond to those used in the proposed research model.

Weighted Imputation with K-Nearest Neighbor [4]: The Weighted K Nearest Neighbors method selects the instances with similar values (in terms of distance) to the considered one, so it can impute as KNNI does. However, the estimated value now takes into account the different distances of the neighbors, using a weighted average or the most repeated value according to the distance.

3 Distributed Energy Resources: Primary resource assessment techniques and forecast

Various modeling approaches have been used for the assessment and forecasting of Photovoltaic generation: physical, statistical, artificial intelligence (including deep neural network), ensemble-based prediction models, hybrids, among others [19]. Some of these techniques are listed below, and are shown in Figure 3 with the models and techniques used in the research highlighted.

3.1 Statistical models

Statistical models do not need any internal system information to perform a model. It is a data-driven approach that can extract relationships on past data to predict future plant behavior. Therefore, the quality of historical data is essential for forecast accuracy [5] [20].

There are several types of statistical models, the ones used in the investigation are the following:

3.1.1 Nonlinear models [6]

For about a decade, there has been a great deal of research interest in artificial intelligence techniques, not only for prediction but also for a wide range of applications, including control, data compression, optimization, pattern recognition, and classification. *Machine Learning Forecasting Techniques* [7] are other statistical models that take advantage of advances in machine learning, a computer-based approach or artificial intelligence. The method is based on the ability of Artificial Intelligence to learn from experience with historical data and to further hone its predictive abilities through training runs. Powerful computers are required to run numerous iterations before a final prediction can be achieved. You can perceive impossible representations without predetermined formulas or equations. Its applications abound: pattern recognition, data mining, classification problems, filtering, and forecasting. The main machine learning techniques are ANN [21], Multi-layer Perceptron Neural Network (MLPNN) [22] [23], Recurrent Neural Network (RNN) [24], Feed-Forward Neural Network (FFNN) [25] and Feedback Neural Network (FBNN) [7], but others machine learning techniques are used as well, such as, Random Forest, Decision Trees, K Nearest Neighbors, etc.

- *Random Forests (RF)*: Although SVMs and ANNs are popular for short-term solar forecasting, random forest models are also used. A random forest is a Collection of Single Classification and Regression Trees (CART) in which each CART is trained by a bagging algorithm that avoids overfitting the RF models. To train each CART, the training set is partitioned using the bootstrap example method. The robustness of a CART can be improved by combining CARTs according to their performance. Although each CART may be biased due to its structure and the specific subset of features selected, aggregation of all decision trees can significantly reduce the error bias of the final result. The RFs do this by averaging all the CARTs in the set [8].
- *k-Nearest Neighbors (k-NN)*: It is one of the simplest machine learning methods. It is based on

an algorithm for pattern recognition, which compares the current state with training samples in a feature space. Euclidean distances are calculated and the first k nearest neighbors are selected for predictions [9].

- *Decision Trees (DT)*: this is a non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from data features. A tree can be viewed as a piecewise constant approximation [10].
- *Ridge Regression with Built-in Cross Validation (RidgeCV)*: Ridge Regression (ridges) addresses some of the ordinary least squares problems by imposing a penalty on the size of the coefficients. Crest coefficients minimize a penalized residual sum of squares. This estimator has built-in cross-validation capabilities to automatically select the best hyperparameters. The advantage of using a cross-validation estimator is that you can take advantage of hot start by reusing the results precomputed in the previous steps of the cross-validation process, this generally leads to speed improvements [11].

In addition to the models chosen above, another model is defined below for comparison purposes:

- *Multilayer Perceptron Neural Network (MLPNN) [7]*: Many researchers treat the MLPNN model as a benchmark. It is a technique for elementary and effective ANN approach for design and prediction. It is so powerful that this network is used in universal approximation, and in nonlinear models and complex problems that cannot be solved by an ordinary single-layer neural network. In general, MLP is a composite of three or more layers of nodes of incoherent activation. These nodes in any layer are connected through a certain amount of weight to other nodes in the next layer. Therefore, it has the ability to correlate input and output through proper learning. The correlation between the number of nodes and the hidden layer is essential.

3.2 Hybrid Models

According to [9], individual models may omit certain information due to the way each technique transforms the data. Therefore, it is also common to combine techniques to build on their strengths in order to improve accuracy, denoted as hybrid, blended, combined, or ensemble models. Models can be mixed in various ways, such as bagging, boosting, voting, or

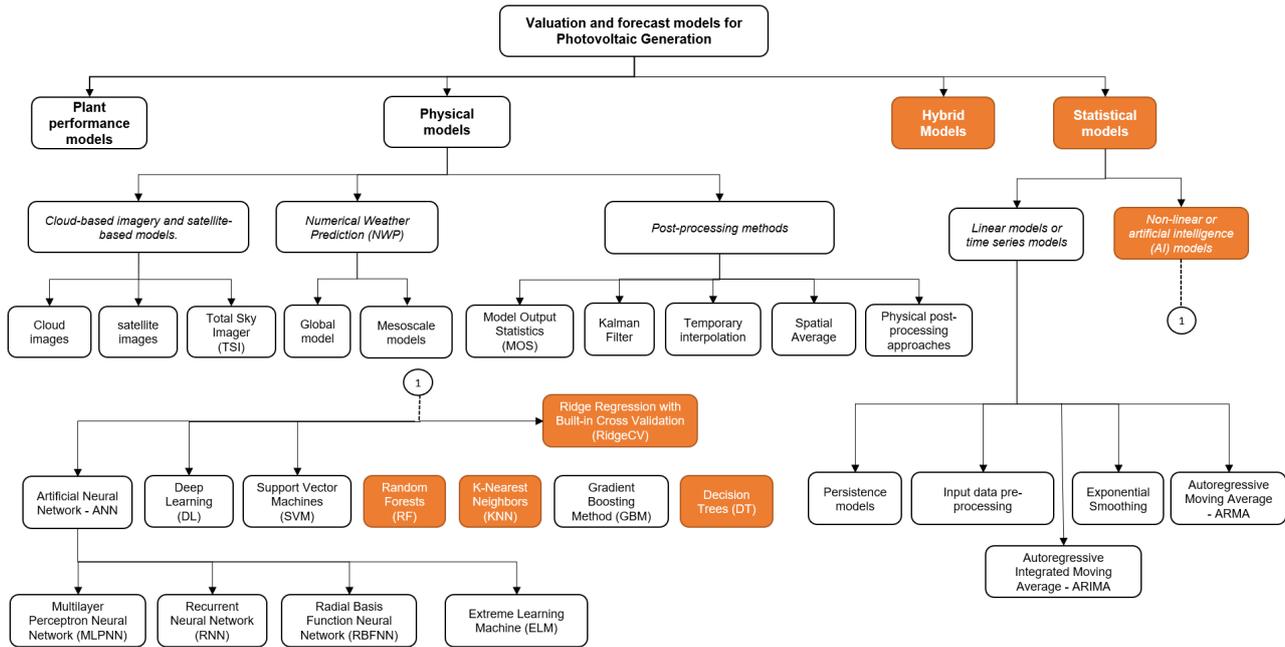


Figure 3: Models and techniques for assessing and forecasting Photovoltaic generation. Source: Author elaboration.

stacking. Two approaches can be followed, either by combining two or more statistical techniques (hybrid statistics) or by joining a statistical technique to a PV performance model (physical hybrid). Hybrid methods can be implemented in three different ways; linear models, nonlinear models and linear and nonlinear models [6].

3.3 Forecast model performance

Performance estimation is essential to assess the forecast accuracy of a model. Common tools include: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE) [26] [27]. In this investigation, the RMSE was used, which estimates the mean value of the error using the square root of the average of the squared differences between the predicted values and the actual observations. Therefore, it is more robust in dealing with large deviations that are especially undesirable, giving the researcher the ability to identify and remove outliers [7].

In addition to the above, the performance can be calculated using the coefficient of determination R^2 , which consists of the proportion of the total variance of the variable explained by the regression. This parameter reflects the goodness of fit of a model to the variable to be explained, it ranges between 0 and 1, the closer it is to the value 1, the greater the fit to the model, but it can also be negative, because the model

Table 1: Solar irradiance forecast categorization based on approach. Adapted from [12].

Approach	Input	Examples
Physical	Meteorological data	NWP, TSI, GOES (Geostationary Operational Environmental Satellite)
Statistical	Historical data	ANN, MLP (Multilayer Perceptron), ARIMA, RNN.

can be arbitrarily worse.

3.4 Solar Radiation Forecast

Solar irradiance can be divided into three categories: direct, diffuse and global. Direct irradiance is solar radiation that travels directly to the earth’s surface. Diffuse irradiance is the solar radiation scattered from the direct beam. The Global Horizontal Irradiance (GHI) is the sum of the previous two [12].

Similar to wind speed/energy forecast, solar irradiance forecast can also be classified according to approaches, as shown in Table 1.

Solar irradiance was categorized according to the time horizon for forecasting and suitable approaches for different time intervals were proposed. The cate-

Table 2: Solar irradiance forecasting categorization based on forecasting time horizon. Adapted from [12].

Time horizon	Interval	Approach
Very short time	0.5 - 6 h	TSI
Medium term	6 - 48 h	NWP with Mesoscale.

Table 3: Other Forecast Categorization of Solar Irradiance Based on the Forecasting Time Horizon. Adapted from [9].

Time horizon	Interval
Intra-hour	Few seconds – 1 hour
Intra-day	1 - 6 hours
Day ahead	6 - 48 hours

gories and approaches proposed are presented in Table 2

Many researchers to develop a forecast horizon classification approach, specifically for solar, use the following categories: intra-hour (also known as *now-casting*), intra-day, and day ahead (Table 3). These categories often overlap with the short-, medium-, and long-term categories described above [12].

4 Ensemble model for the forecast of operational variables of distributed resources in the microgrid environment

In this section, the integrated or assembly model will be structured and programmed based on the models described previously. It will begin by describing the acquisition of the data to form the databases corresponding to the Solar Photovoltaic System (SPV), consequently the process of purification and imputation of the values, continuing with the programming of the assembly forecast model, and culminating with the selection, calculation and programming of the operative variables of the DER (Power, power coefficients, Temperatures, etc.) as shown in Figure 4.

4.1 Input information

To carry out the data processing, the respective calculations and, in general, the modeling of the SPV, it is necessary to have input information that guarantees that these activities are carried out in the best way, since each input will obtain a different output. That is why the information that is reported in the different data sheets of solar modules must be taken into account when executing the application that is presented

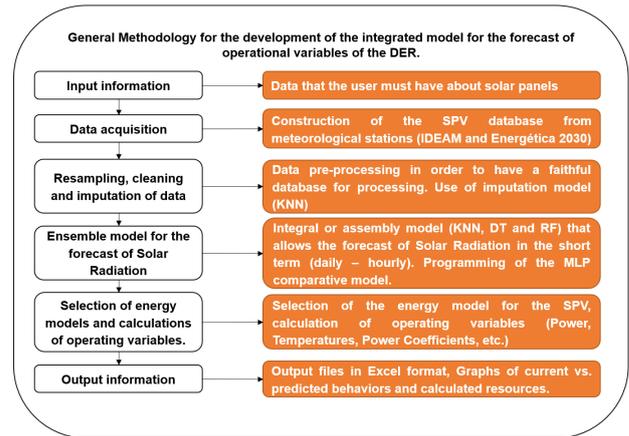


Figure 4: General methodology for the development of the integrated or assembly model. Source: Author elaboration.

as a result of the investigation. This information, for the SPV is as follows:

- The database where the hourly Global Horizontal Radiation is reported by weather station.
- Database update files (optional if you want to perform an update).
- The day, month and year to be viewed and processed.
- Capacity of the photovoltaic module(s).
- Derating factor.
- Module efficiency.
- Temperature Coefficient.
- Nominal operating temperature of the panel.
- Panel tilt.
- Panel azimuth.

With the information previously entered in the application, it is possible to carry out the pre-processing of the data (resampling, purification and imputation), but first it will be shown in detail how the acquisition of said quantities was made.

4.2 Data acquisition

5 meteorological stations are incorporated into the application and are shown in Table 4, along with their main characteristics, all for the department of Sucre (Colombia).

The database (DB) for the SPV has more than 130,000 data, therefore the execution of load and functions of these DB require machine process times ranging from 30 s to more than 2 minutes.

Table 4: Main characteristics of the Meteorological Stations.

Meteorological Station	Latitude/ Longitude	Altitude [m]	Location	Frequency	Source
Puerta Roja	9.316/-75.387	160	Sincelejo	Hourly	IDEAM
El Tesoro	9.357/-75.289	156	Morroa	Hourly	IDEAM
San Marcos	8.596/-75.142	25	San Marcos	Hourly	IDEAM
ENERGÉTICA 2030	9.315/-75.388	175	UNISUCRE (University of Sucre) - Sincelejo	5 minutes	Energética 2030
ENERGÉTICA 2030	6.240/-75.590	1486	UPB (Bolivarian Pontifical University) - Medellín	5 minutes	Energética 2030

4.2.1 IDEAM

The data from the IDEAM meteorological stations were downloaded through its Hydrometeorological data consultation and download platform: <http://dhime.ideam.gov.co>. DHIME is the development and implementation of an integrated technological solution that allows IDEAM to generate an impact on its business processes, associated with hydrological and meteorological information and for the administration and operation of the hydrometeorological network, with which the organization and information management [13].

They are non-continuous data (see Table 5), that is, they do not cover the entire time elapsed from the date of installation to the date of consultation or download and, furthermore, they do not report new data for updating.

This DB reports hourly data for the department of Sucre for solar radiation and allows downloading only for one month and for each station, so to download 10 years, 120 files had to be downloaded for each station, that is, about 600 files, to later be merged together and use only one DB for the SPV.

The data obtained through the IDEAM are: Station Code, Station Name, Latitude, Longitude, Altitude, Category, Entity, Operational Area, Department, Municipality, Installation Date, Suspension Date, Parameter Id (Solar Radiation), Label, Series Description, Frequency, Date, Value, Grade, Qualifier and Approval Level.

4.2.2 ENERGÉTICA 2030

In addition to the IDEAM data, the "Energética 2030" project provides, through its platform: <https://energetica2030.netux.com>, updated and continuous data, therefore, its data is updatable until the day of verification.

The meteorological station installed both at the University of Sucre (Sincelejo) and at the UPB (Medellín) are of the NxS-MiEstación-THVPM-RS type. It is a monitoring system that allows different

Table 5: Historical data dates where IDEAM data exists.

Meteorological Station	Existing Historical Data Dates
Puerta Roja	2005 (May), 2006 (Jun - Dec), 2007 (Jan - Dec), 2008 (Jan - Dec), 2009 (Apr), 2013 (Jul), 2014 (May-Dec), 2015 (Mar-Apr, Ago-Oct), 2016 (Apr-Jul, Sep - Dec), 2017 (Jan - Feb)
El Tesoro	2014 (Sep - Dec), 2015 (Jan - May), 2016 (May - Dec), 2017 (Jan - Feb)
San Marcos	2005 (Oct - Dec), 2006 (Jan - Dec), 2007 (Jan - Jun), 2009 (Jul - Dec), 2010 (Jan - Dec), 2011 (Jan - Jul), 2012 (Feb - May, Sep), 2013 (Oct - Dec), 2014 (Jan - Jul, Sep - Nov), 2015 (Jun - Dec), 2016 (Jan - Nov)

variables to be measured, which together allow a diagnosis and monitoring of environmental conditions, in addition to having the ability to transmit to the cloud the information that it collects with its sensors at intervals of approximately five minutes that allow remote monitoring and access to historical data.

The electrical power supply required to operate is provided by a 12 V battery rechargeable by means of the solar panel, which provides autonomy in the absence of solar radiation of 48 h.

The station allows monitoring in real time from a web and mobile platform environmental parameters such as: Temperature, Relative Humidity, Atmospheric Pressure, Wind Direction and Speed, Solar Radiation and Particulate Matter (PM 2.5 – PM 10).

Table 6: Historical data dates where "ENERGÉTICA 2030" data exists.

Meteorological Station	Existing Historical Data Dates
ENERGÉTICA 2030 - UNISU-CRE	2019 (Nov - Dec), 2020 (Jan - Dec), 2021 (Jan - Current)
ENERGÉTICA 2030 - UPB	2019 (Sep - Dec), 2020 (Jan - Dec), 2021 (Jan - Current)

Like the data obtained by IDEAM, the data presented by the ENERGÉTICA 2030 platform may be missing (without value) some hours or days of the months reported in Table 6.

This DB reports 5-minute values, so for the creation of the complete SSF database, a preprocessing of these values had to be carried out, calculating the hourly average. This is how it went from approximately 230,000 solar radiation data to approximately 130,000, with the consequent reduction in machine processing time.

Each file that is downloaded from the platform includes the following information: Time stamp, Date and value of the query variable (temperature, solar radiation or wind speed).

4.3 Resampling, cleaning and imputation of data

The SPV databases present gaps or time stamps without solar radiation value, which makes it necessary to pre-process the data, first with a resampling that identifies these positions and enters them as NaN data (Not a Number), this step will show the chosen programming language (Python) that it is a value that needs debugging or imputation. A NaN value will be a Missing Value (MV).

Purging is a process that is carried out mainly taking into account the number of missing values found in a day. For both systems, a purging was made of those days that presented MV greater than 70.83%, that is, they have less of 17 values of the 24 hours a day. This percentage is chosen because, empirically, in the executions of the application, when this value is exceeded, forecasts are presented with bad values of the determination coefficient R^2 and even negative ones. In addition, this percentage of MV (29.17%) is within the ranges used by [3] to perform imputation. Lower MV percentages can be used, but we wanted to cover the largest number of days with the lowest possible coefficient of determination.

After having performed the resampling and filtering, the imputation stage is accessed. This is responsi-

ble for assigning a value to the NaN data. The choice of methods for this phase starts from the behavior of solar radiation in one day. For the SPV it is logical to think that a method such as the WKNNI is viable, since it suggests the use of the closest values, weighting the distance between them and the missing value, in addition [14] found good results when using this method even in its simplest way (KNNI).

Another type of imputation used for the SPV was the direct one, since the values of "zero" are known for the night hours between 8:00 pm and 5:00 am the following day.

As a result of this stage, the complete DB for the SSF is ready for the Forecast process.

4.4 Ensemble model for the forecast of Solar Radiation

Starting from having the databases of the systems without NaN values, it is then possible to carry out the forecast of the main variables of the DER, mainly solar radiation, the rest are variables that are calculated from these.

There are integrated or assembly models for the forecast of the main variables of the investigation, and although there are models that include one or two of the chosen ones, one was not found that integrates all 3. The choice of the 3 models that were described in the section 3, part of the provision of libraries in the Python programming language but also, of the common use of these for the forecast of meteorological variables such as those studied here. This is why K nearest neighbors, decision tree, and random forests come together to form the ensemble model that predicts solar radiation and wind speed, using stacking with a final estimator called a ridge regressor. Stacked generalization consists of pooling the output of the individual estimator and using a regressor to compute the final prediction. Pooling allows you to leverage the strength of each individual estimator by using its output as input to a final estimator.

Keep in mind that the application also predicts the main variables of the model and also adds another model based on MLP-type Artificial Neural Networks, which was programmed for future comparisons.

The integrated or assembly model schedule forecasts a day with hourly resolution. It is possible to forecast the day after the last day of the database update, that is, if the database is updated until today, it is possible to forecast tomorrow, since it is very valuable to obtain results in real time for the next day. The records for the forecast were divided like this, 80% for the training data (train) and 20% for the test set (test). These percentages resulted from testing with various amounts (75% - 25%, 70% - 30%, 65% - 35%, 60% - 40%), obtaining better forecasts with (80% - 20%).

The current and forecast values are the main inputs to calculate the DER operating variables, which are presented below in the next item.

4.4.1 Calculation of the operating variables of the Photovoltaic Solar System

To make comparisons in a better way with the HOMER PRO software, the energy model used by this program to calculate the power generated by the photovoltaic modules was chosen, so this section is based on [15] and [16].

HOMER PRO uses the following equation to calculate the power output of the Solar Modules:

$$P_{pv} = Y_{pv} * f_{pv} * \left(\frac{\overline{G}_T}{\overline{G}_{T,STC}} \right) * [1 + \alpha_p * (T_c - T_{c,STC})] \quad (1)$$

where:

Y_{pv} : The rated capacity of the Photovoltaic (PV) array, meaning its power output under standard test conditions [kW].

f_{pv} : The PV derating factor [%]

\overline{G}_T : The solar radiation incident on the PV array in the current time step [kW/m²]

$\overline{G}_{T,STC}$: The incident radiation at standard test conditions [1kW/m²].

α_p : The temperature coefficient of power [%/ °C].

T_c : The PV cell temperature in the current time step [°C].

$T_{c,STC}$: the PV cell temperature under standard test conditions [25°C].

But in order to calculate this power it is necessary to calculate the temperature in the panel cell (T_c) and the incident solar radiation on the panel (\overline{G}_T), so HOMER PRO uses the following equation to calculate T_c :

$$T_c = \frac{T_a + (T_{c,NOCT} - T_a,NOCT) \left(\frac{\overline{G}_T}{\overline{G}_{T,NOCT}} \right) \left[1 - \frac{\eta_{mp,STC} (1 - \alpha_p T_{c,STC})}{\tau \alpha} \right]}{1 + (T_{c,NOCT} - T_a,NOCT) \left(\frac{\overline{G}_T}{\overline{G}_{T,NOCT}} \right) \left(\frac{\alpha_p \eta_{mp,STC}}{\tau \alpha} \right)} \quad (2)$$

Where:

T_a : The ambient temperature.

$T_{c,NOCT}$: The nominal operating cell temperature [°C].

$T_{a,NOCT}$: The ambient temperature at which the NOCT is defined [20°C].

\overline{G}_T : The solar radiation striking the PV array [kW/m²].

$\overline{G}_{T,NOCT}$: The Solar radiation at which the NOCT is defined [0.8 kW/m²]

$\eta_{mp,STC}$: The maximum power point efficiency

under standard test conditions [%].

τ : The solar transmittance of any cover over the PV array [%].

α : The solar absorptance of the PV array [%].

HOMER PRO assumes a value of 0.9 for $\tau\alpha$ and uses the following equation to calculate the solar declination:

$$\delta = 23.45^\circ \sin \left(360^\circ \frac{284 + n}{365} \right) \quad (3)$$

where:

n : the day of the year [a number 1 through 365]

The time of day affects the location of the sun in the sky, which we can describe by an hour angle. HOMER PRO uses the convention whereby the hour angle is zero at solar noon (the time of day at which the sun is at its highest point in the sky), which is negative before solar noon, and positive after solar noon. HOMER PRO uses the following equation to calculate the hour angle:

$$\omega = (t_s - 12hr) * 15^\circ/hr \quad (4)$$

Where:

t_s : The solar time [hr]

HOMER PRO assumes that all time-dependent data, such as solar radiation data and electric load data, are specified not in solar time, but in civil time (also called local standard time). HOMER PRO calculates solar time from civil time using the following equation:

$$t_s = t_c + \frac{\lambda}{15^\circ/hr} - Z_c + E \quad (5)$$

Where:

t_c : the civil time in hours corresponding to the midpoint of the time step [hr]

λ : the longitude [°]

Z_c : the time zone in hours east of GMT [hr]

E : the equation of time [hr]

The equation of time accounts for the effects of obliquity (the tilt of the earth's axis of rotation relative to the plane of the ecliptic) and the eccentricity of the earth's orbit. HOMER PRO calculates the equation of time as follows:

$$E = 3.82(0.000075 + 0.001868 \cos B - 0.032077 \sin B - 0.014615 \cos 2B - 0.04089 \sin 2B) \quad (6)$$

Where B is given by:

$$B = 360^\circ \frac{(n - 1)}{365} \quad (7)$$

Now, for a surface with any orientation, we can define the angle of incidence, meaning the angle between the sun's beam radiation and the normal to the surface, using the following equation:

$$\begin{aligned} \cos\theta = & \sin\delta \sin\phi \cos\beta - \sin\delta \cos\phi \sin\beta \cos\gamma \\ & + \cos\delta \cos\phi \cos\beta \cos\omega + \cos\delta \sin\phi \sin\beta \cos\gamma \cos\omega \\ & + \cos\delta \sin\beta \sin\gamma \sin\omega \end{aligned} \quad (8)$$

where:

- θ : The angle of incidence [°]
- β : The slope of the surface [°]
- γ : The azimuth of the surface [°]
- ϕ : The latitude [°]
- δ : The solar declination [°]
- ω : The hour angle [°]

An incidence angle of particular importance, is the zenith angle, meaning the angle between a vertical line and the line to the sun. The zenith angle is zero when the sun is directly overhead, and 90° when the sun is at the horizon. Because a horizontal surface has a slope of zero, we can find a equation for the zenith angle by setting $\beta = 0^\circ$ in the equation above, which yields:

$$\theta_z = \cos\phi \cos\delta \cos\omega + \sin\phi \sin\delta \quad (9)$$

And θ_z is the zenith angle [°]

When addressing the amount of solar radiation arriving at the top of the atmosphere over a particular point on the earth's surface, HOMER PRO assumes the output of the sun is constant in time. But the amount of sunlight striking the top of the earth's atmosphere varies over the year because the distance between the sun and the earth varies over the year due to the eccentricity of earth's orbit. To calculate the extraterrestrial normal radiation, defined as the amount of solar radiation striking a surface normal (perpendicular) to the sun's rays at the top of the earth's atmosphere, HOMER PRO uses the following equation:

$$G_{on} = G_{sc} \left(1 + 0.033 \cos \left(\frac{360n}{365} \right) \right) \quad (10)$$

Where:

- G_{on} : The extraterrestrial normal radiation [kW/m²]

G_{sc} : The solar constant [1.367 kW/m²]

n : The day of the year [a number between 1 and 365]

To calculate the extraterrestrial horizontal radiation, defined as the amount of solar radiation striking a horizontal surface at the top of the atmosphere, HOMER PRO uses the following equation:

$$G_o = G_{on} \cos\theta_z \quad (11)$$

Where:

G_o : The extraterrestrial horizontal radiation [kW/m²]

G_{on} : The extraterrestrial normal radiation [kW/m²]

θ_z : The zenith angle [°]

Because HOMER PRO simulates on a time-step-by-time-step basis, we integrate the equation above over one time step to find the average extraterrestrial horizontal radiation over the time step:

$$\bar{G}_o = \frac{12}{\pi} G_{on} \left[\cos\phi \cos\delta (\sin\omega_2 - \sin\omega_1) + \frac{\pi(\omega_2 - \omega_1)}{180^\circ} \sin\phi \sin\delta \right] \quad (12)$$

Where:

\bar{G}_o : The extraterrestrial horizontal radiation averaged over the time step [kW/m²]

G_{on} : The extraterrestrial normal radiation [kW/m²]

ω_1 : The hour angle at the beginning of the time step [°]

ω_2 : The hour angle at the end of the time step [°]

The equation above gives the average amount of solar radiation striking a horizontal surface at the top of the atmosphere in any time step. The solar resource data give the average amount of solar radiation striking a horizontal surface at the bottom of the atmosphere (the surface of the earth) in every time step. The ratio of the surface radiation to the extraterrestrial radiation is called the clearness index. The following equation defines the clearness index:

$$k_T = \frac{\bar{G}}{\bar{G}_o} \quad (13)$$

Where:

\bar{G} : The global horizontal radiation on the earth's surface averaged over the time step [kW/m²]

\bar{G}_o : The extraterrestrial horizontal radiation averaged over the time step [kW/m²]

As for the solar radiation on the earth's surface, some of the radiation is beam radiation, defined as solar radiation that travels from the sun to the earth's surface without any scattering by the atmosphere. Beam radiation (sometimes called direct radiation)

casts a shadow. The rest of the radiation is diffuse radiation, defined as solar radiation whose direction has been changed by the earth's atmosphere. Diffuse radiation comes from all parts of the sky and does not cast a shadow. The sum of beam and diffuse radiation is called global solar radiation, a relation expressed by the following equation:

$$\bar{G} = \bar{G}_b + \bar{G}_d \quad (14)$$

Where:

\bar{G}_b : The beam radiation [kW/m²]
 \bar{G}_d : The diffuse radiation [kW/m²]

The distinction between beam and diffuse radiation is important when calculating the amount of radiation incident on an inclined surface. The orientation of the surface has a stronger effect on the beam radiation, which comes from only one part of the sky, than it does on the diffuse radiation, which comes from all parts of the sky.

However, in most cases, we measure only the global horizontal radiation, not its beam and diffuse components. For that reason, you need to enter global horizontal radiation in HOMER PRO's Solar Resource Inputs. Then, in every time step, HOMER PRO must resolve the global horizontal radiation into its beam and diffuse components to find the radiation incident on the PV array. For this purpose, HOMER PRO uses correlation of Erbs et al. (1982), which gives the diffuse fraction as a function of the clearness index as follows:

$$\frac{\bar{G}_d}{\bar{G}} = \begin{cases} \frac{1.0 - 0.09k_T (*)}{0.9511 - 0.1604k_T + 4.388k_T^2 - 16.638k_T^3 + 12.336k_T^4 (**)} & 0.165 (***) \\ (*) \text{ para } k_T \leq 0.22 \\ (**) \text{ para } 0.22 \leq k_T \leq 0.80 \\ (***) \text{ para } k_T > 0.80 \end{cases} \quad (15)$$

For each time step, HOMER PRO uses the average global horizontal radiation to calculate the clearness index, then the diffuse radiation. It then calculates the beam radiation by subtracting the diffuse radiation from the global horizontal radiation.

To calculate the global radiation striking the tilted surface of the PV array, HOMER PRO uses the HDKR model, which assumes that there are three components to the diffuse solar radiation: an isotropic component that comes from all parts of the sky equally, a circumsolar component that emanates from the direction of the sun, and a horizon brightening component that emanates from the horizon. Before applying this model, we must first define three more factors.

The following equation defines R_b , the ratio of beam radiation on the tilted surface to beam radiation on the horizontal surface:

$$R_b = \frac{\cos\theta}{\cos\theta_z} \quad (16)$$

The anisotropy index, with symbol A_i , is a measure of the atmospheric transmittance of beam radiation. This factor is used to estimate the amount of circumsolar diffuse radiation, also called forward scattered radiation. The anisotropy index is given by the following equation:

$$A_i = \frac{\bar{G}_b}{\bar{G}_o} \quad (17)$$

Finally, we need to define a factor used to account for horizon brightening, or the fact that more diffuse radiation comes from the horizon than from the rest of the sky. This term is related to the cloudiness and is given by the following equation:

$$f = \sqrt{\frac{\bar{G}_b}{\bar{G}}} \quad (18)$$

The HDKR model calculates the global radiation incident on the PV array according to the following equation:

$$G_T = (\bar{G}_b + \bar{G}_d A_i) R_b + \bar{G}_d (1 - A_i) \left(\frac{1 + \cos\beta}{2} \right) \left[1 + f \sin^3 \left(\frac{\beta}{2} \right) \right] + \bar{G} \rho_g \left(\frac{1 - \cos\beta}{2} \right) \quad (19)$$

Where:

β : The slope of the surface [°]

ρ_g : The ground reflectance, which is also called the albedo [%].

5 Results

After having calculated the main operational variables of the DER, a report is then made on the results obtained with the integral or assembly model (application) and they are compared with those obtained with the HOMER PRO Software. The data to be compared consists of the calculations provided by HOMER PRO in relation to the output power, temperature, etc., of the current day and not of the

forecasts, since this Software does not make forecasts of solar radiation. Therefore, it is only possible to compare the forecast results with the data measured by the meteorological stations.

The results of the model were validated in the microgrid of the Bolivarian Pontifical University - UPB campus (within the framework of the "Energética 2030" program).

In order for the model to operate in the best way, a configuration of the parameters of each model was made, this tuning was carried out by executing the application several times and varying said parameters, obtaining the optimal data for each model, as follows: for the decision tree do not use maximum depth, for K nearest neighbors, use 7 near neighbors and for random forests, a maximum depth of 10 and 200 estimators.

The graphical interface in python language uses the following libraries in its programming:

- Wxpython (Graphical Application Programming Interface).
- Pandas (Management and analysis of data structures).
- Numpy (Numerical calculation and data analysis).
- Sklearn (Simple and efficient tools for predictive data analysis).
- Matplotlib (Specialized in creating graphs).
- Datetime (manipulation of dates and times).
- Math (Access to defined mathematical functions).

The parameters of the PV (See Table 7, Table 8, Table 9 and Table 10), correspond to those installed in UNISUCRE and UPB.

Table 7: Data of the equipment installed in UNISUCRE.

Name	Value	Unit
PV Capacity	20	kW
Derating factor	88,0	%
Module efficiency	19,88	%
Temperature Coefficient	-0,39	%/°C
Nominal operating cell temperature	45	°C
Panel tilt	9,32	°C
Panel azimuth	0	°C

Table 8: Data of the equipment installed in UPB.

Name	Value	Unit
PV Capacity	52,650	kW
Derating factor	88,0	%
Module efficiency	16,5	%
Temperature Coefficient	-0,41	%/°C
Nominal operating cell temperature	45	°C
Panel tilt	7	°C
Panel azimuth	180	°C

Table 9: Description of the equipment installed in UNISUCRE.

Name	Manufacturer	Model	Amount
PV	Jinko - Cheetah	JKM400	50
Converter	Chintpower-CPS	CPS	
		SCA6KTL- 3	
		SM	

It is only possible to validate the results of the SPV of the UPB, because the UNISUCRE microgrid still does not report data on power, temperature in the solar panels, etc, but the comparisons of horizontal global radiation reported by the weather station Vs the ensemble model will be shown.

5.1 UPB results

The evaluation of the results is mainly determined by the metrics RMSE (W/m^2) and the coefficient of determination R^2 (dimensionless) of the Global Horizontal Radiation predicted and measured in the stations, since the other variables are calculated from this parameter. operations of Distributed Resources. The errors found during the validation (20% of the data) are presented in each table, which represent the forecast information that the model has never seen. Some results obtained with the model are shown graphically.

The critical information obtained by the model, that is, minimum and maximum data of RMSE, R^2 and the dates on which this happens is shown in the following tables, for the ENERGÉTICA 2030 (UPB - UNISUCRE) stations and the one located in Puerta Roja (Sincelejo) incorporated in the application:

The RMSE value of $110 W/m^2$ downwards serves as a reference to recognize if the Global Horizontal Radiation forecast is good. Graphically it can be seen that the forecast curves and real values are very similar when they are in this range (Figure 5). The coefficient of determination R^2 shows the fit between

Table 10: Description of the equipment installed in UPB.

Name	Manufacturer	Model	Amount
PV	JA Solar	JAP6(K)	195
		60-270/4BB	
Converter	Fronius	Symo	2
		24.0-3480	

Table 11: Critical information obtained for the UPB station.

Critical information	Reported data	Date
Min_RMSE	59,62602	2021-07-16
Máx_RMSE	267,3226	2021-08-15
Min_R ²	-2,90206	2021-07-27
Max_R ²	0,96067	2021-07-08

forecast and actual value, one (1) being the best possible value (Figure 6).

For this station there are 145 days of evaluation data, obtaining the following daily results: 37.93% of data (55) of evaluation R² are in the range greater than or equal to 80%. 32.41% of the evaluation RMSE data (47) are less than or equal to 110 W/m².

It is important to clarify that the year 2021 for Colombia has been a period of time cataloged by IDEAM with the La Niña phenomenon, dominated by constant rains in the National territory, resulting in few sunny days.

The training data (80% of the total data) provided by the UPB station are in the date range from September 02, 2019 to April 3, 2021 and the validation data (20% of the total data) corresponds to the date range of April 4, 2021 to September 2, 2021.

The month with the best RMSE is June 2021, with an average value of 122.44 W/m² with an average R² of 0.6663.

Regarding the power and temperature of the modules, it is clear that the results calculated with the application (current) and HOMER PRO are consistent with what was expected (see Figure 7 and Figure 8), since the same model was used.

Tables 12 and 13 show the hourly sums of the output power calculated by the model for the days with the lowest and highest RMSE, respectively, also for the forecast and the in situ measurement. The Relative Error is calculated with respect to the power mea-

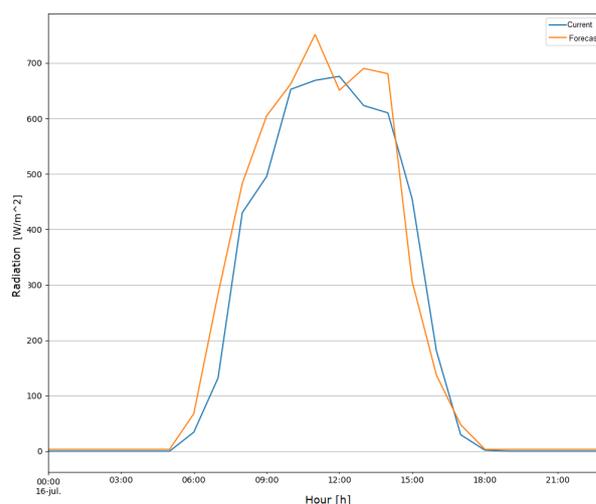


Figure 5: Forecast of Global Horizontal Radiation for the day 2021-07-16 (Lowest RMSE), UPB station. Source: Author elaboration.

sured in situ.

Table 12: Daily power output for the lowest RMSE day.

Name	Daily power [kWp]	Relative Error
Ppv current_model	215,73	11,394%
Ppv forecasted_model	232,983	4,31%
Ppv HOMER PRO	215,55	11,469%
Ppv_measured	243,47	

Table 13: Daily power output for the highest RMSE day.

Name	Daily power [kWp]	Relative Error
Ppv current_model	227,71	0,235%
Ppv forecasted_model	168,98	25,97%
Ppv HOMER PRO	227,98	0,116%
Ppv_measured	228,25	

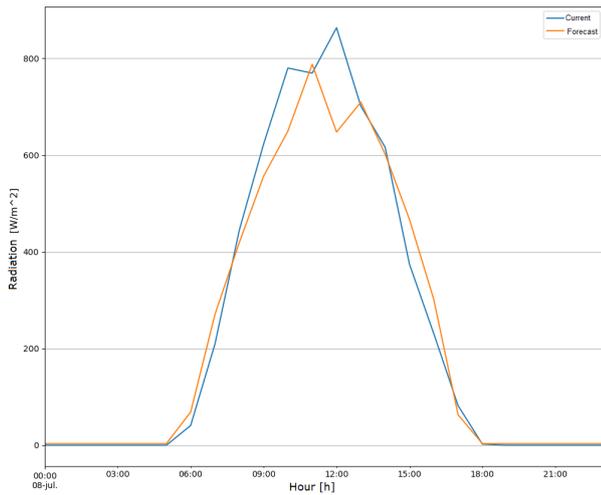


Figure 6: Forecast of Global Horizontal Radiation for the day 2021-07-08 (Major R^2), UPB station. Source: Author elaboration.

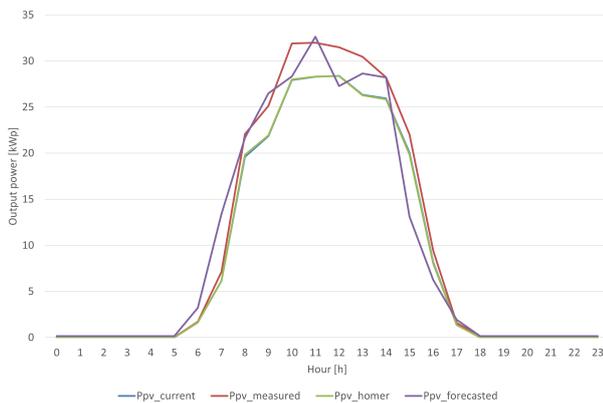


Figure 7: Output power of the modules for the day 2021-07-16 (Lowest RMSE), UPB station. Source: Author elaboration.

The results obtained by each of the artificial intelligence methods used in the integral model will be shown below, for the days with the lowest and highest RMSE reported in the assembly model for the UPB – ENERGÉTICA 2030 station:

5.1.1 AI Model: Decision Tree

The results obtained by means of this method are recorded in the table 14:

5.1.2 AI model: k Nearest Neighbors

The results obtained by means of this method are recorded in the table 15:

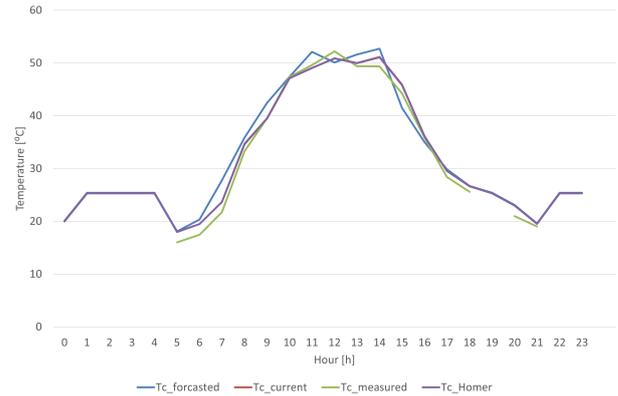


Figure 8: Module temperature for the day 2021-07-16 (Lowest RMSE), UPB station. Source: Author elaboration.

5.1.3 AI model: Random Forest

The results obtained by means of this method are recorded in the table 16:

5.1.4 AI model: Neural Network (MLP)

The results obtained by means of this method are recorded in the table 17:

5.2 UNISUCRE results

For the station located at the University of Sucre (Table 18), the evaluation data (20%) consists of 116 days, starting in March 2021 and ending in June of the same year. The following reports are available for this station: 44.83% of the data (52) of the evaluation R^2 are in the range greater than or equal to 80%. 22.41% of the evaluation RMSE data (26) are less than or equal to 110 W/m², highlighting that for these quantities the average R^2 is 0.913, which means that there is an excellent percentage of adjustment of Global Horizontal Radiation in relation to its forecast.

For the station located in Puerta Roja, very close to the ENERGÉTICA 2030 - UNISUCRE station (Table 19), the evaluation data consists of 337 days, beginning in September 2015 and ending in February 2017. For this station, the following data are available: Reports: 39.76% of data (134) of R^2 evaluation are in the range greater than or equal to 80%. 31.15% of the evaluation RMSE data (105) are less than or equal to 110 W/m², with an average R^2 of 82.26%.

6 Conclusion

The model developed in the microgrid that is being implemented on the UPB campus (in the framework of the ENERGÉTICA 2030 program) for the SPV was validated, where the forecasts obtained from the model were compared with the information measured

Table 14: Results obtained by the decision tree model.

Critical information	R^2	RMSE [W/m ²]	Measured power [kWp]	Forecast power [kWp]	Er [%]	Date
Min_RMSE	0,61367	167,512	243,47	221,72	8,93	2021-07-16
Máx_RMSE	-0.1954	340,529	228,25	207,343	9,16	2021-08-15

Table 15: Results obtained by the KNN model.

Critical information	R^2	RMSE [W/m ²]	Measured power [kWp]	Forecast power [kWp]	Er [%]	Date
Min_RMSE	0,89962	85,3876	243,47	236,98	2,67	2021-07-16
Máx_RMSE	0,08649	297,672	228,25	169,66	25,67	2021-08-15

for the variables of interest in the microgrid, using the data provided by the 52 kWp System.

From the results obtained, the following can be inferred:

- The energy model used for the SPV is adjusted to what is desired with regard to the variables studied. It is observed in the graphs of power and temperature in the modules, that the variables calculated with the model and HOMER PRO are almost the same, this small error is inferred to be due to the calculation of solar time, which HOMER PRO does not fully specify.
- The energy model is also adjusted, with regard to daily calculated power (HOMER PRO and ensemble model) and in situ measurements, with errors ranging from 0.235% to 11.394% in the ensemble model and from 0.116% to 11.47% in HOMER PRO.
- The critical information obtained for the UPB station allows us to identify that the worst RMSE result is 267.32 W/m², data that confirms that regular RMSE results are achieved, taking into account that RMSE information greater than 300 W/m² is bad, since for this station there are radiations of up to 1000 W/m². This can be verified by calculating the forecast daily power for this day and its relative error with respect to the in situ measurement, which is approximately 169 kWp and 26%, respectively.
- 31.41% of the global horizontal radiation evaluation data is in a range less than or equal to 110W/m², that is, for these data the RMSE is good, with an average R^2 of approximately 88%; obtaining very good results, such as the day with the lowest RMSE, with a forecast daily power of 233 kWp and a relative error of 4.31%.
- Of the 3 AI models that contribute to the ensemble model, it was observed that the one with the lowest relative error reported for the day with the lowest RMSE is K Near Neighbors (2.67) and the one with the highest Relative Error of RMSE presented for this same day is Decision Tree (9.93%).
- Of the 3 AI models that contribute to the assembly model, it was observed that the one with the lowest relative error reported for the day with the highest RMSE is (9.16%) and the one with the highest relative RMSE error for this same day is K Nearest Neighbors (25.67%).
- From the previous items it can be concluded that having an AI model does not mean that the least relative error for one day has to be better for all days.
- The results obtained by the MLP model show that it is in the range of errors of the assembly model, even a little lower, and that it would be an option to incorporate it into the integral model.
- Although the validation of the application was done with data provided with the UPB station, it can also be concluded that the results obtained for this station are similar to those found in UNISUCRE, which is why it is thought that when there are values for this station, such as power, temperature in the modules, etc., are close to those found in the UPB. What is clear is that the solar resource of the Department of Sucre (Colombia) is optimal for the installation of microgrids with Photovoltaic Solar Systems.
- The errors found in the SPV increase as the cloudy days increase, and there was a lot of

Table 16: Results obtained by the RF model.

Critical information	R^2	RMSE [W/m ²]	Measured power [kWp]	Forecast power [kWp]	Er [%]	Date
Min_RMSE	0,78197	125,8417	243,47	233,806	3,97	2021-07-16
Máx_RMSE	-0.0355	316,9265	228,25	189,91	16,8	2021-08-15

Table 17: Results obtained by the Neural Network MLP model.

Critical information	R^2	RMSE [W/m ²]	Measured power [kWp]	Forecast power [kWp]	Er [%]	Date
Min_RMSE	0,93283	69,84634	243,47	237,078	2,625	2021-07-16
Máx_RMSE	0,29663	261,20097	228,25	181,532	20,47	2021-08-15

Table 18: Critical information obtained for the UNISUCRE station.

Critical information	Reported data	Date
Min_RMSE	59,03201	2021-03-22
Máx_RMSE	258,78403	2021-06-12
Min_ R^2	-0,56104	2021-05-04
Max_ R^2	0,97392	2021-03-29

Table 19: Critical information obtained for the PUERTA ROJA station.

Critical information	Reported data	Date
Min_RMSE	35,73598	2015-10-16
Máx_RMSE	304,90212	2015-10-25
Min_ R^2	-11,60654	2015-10-24
Max_ R^2	0,97258	2015-10-16

information obtained, especially at the EN-
 ERGÉTICA 2030 stations, with years in which
 the La Niña phenomenon existed.

As forecasting issues are delved into, it is logical to think that while the model is more robust, it presents an improvement proportional to that dynamism, which is why the use of other AI models, such as convolutional neural networks, is recommended, precisely to improve RMSE errors.

The Multilayer Perceptron Neural Networks model - MLP, presents good results alone, for this reason it is recommended to add it to the ensemble model as a fourth AI model.

The comparison of various assembly models is

an interesting option in the application and comparison of results, including the use of satellite-type databases.

Another future work would represent making the comparison with the forecast module of the generic EMS that has been developed in the framework of ENERGÉTICA 2030.

References:

- [1] Evaluation of global horizontal irradiation in Colombia. Available online: <http://atlas.ideam.gov.co/basefiles/Evaluacion-de-la-Irradiacion-Global-Horizontal-en-Colombia.pdf> (accessed Apr. 22, 2020).
- [2] R. J. A. Little, D. B. Rubin, *Statistical analysis with missing data*, John Wiley & Sons, Inc., 2020.
- [3] J. Luengo, S. García, and F. Herrera, On the choice of the best imputation methods for missing values considering three groups of classification methods. *Knowledge and Information Systems*, Vol. 32, 2012, pp.77–108.
- [4] O. Troyanskaya et al, Missing value estimation methods for DNA microarrays, *Bioinformatics*, Vol 17, No. 6, 2001, pp. 520–525, doi: 10.1093/bioinformatics/17.6.520.
- [5] J. Antonanzas, N. Osorio, R. Escobar, R. Uraca, F. J. Martinez-de-Pison, and F. Antonanzas-Torres. Review of photovoltaic power forecasting, *Sol. Energy*, Vol 136, 2016, pp.78–111, doi: 10.1016/j.solener.2016.06.069.
- [6] M. Diagne, M. David, P. Lauret, J. Boland, and N. Schmutz, Review of solar irradiance forecasting methods and a proposition for small-scale insular grids, *Renew. Sustain. Energy Rev.*, Vol 27, 2013, pp.65–76, doi: 10.1016/j.rser.2013.06.042.

- [7] R. Ahmed, V. Sreeram, Y. Mishra, and M. D. Arif, A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization, *Renewable and Sustainable Energy Reviews*, Vol. 124, 2020, doi: 10.1016/j.rser.2020.109792.
- [8] A. Dobbs, T. Elgindy, B.-M. Hodge, and A. Florita, Short-Term Solar Forecasting Performance of Popular Machine Learning Algorithms, *Int. Work. Integr. Sol. Power into Power Syst.*, 2017, Available online: <http://www.osti.gov/scitech>.
- [9] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. Martinez-de-Pison, and F. Antonanzas-Torres, Review of photovoltaic power forecasting, *Sol. Energy*, Vol. 136, 2016, pp. 78–111, doi: 10.1016/j.solener.2016.06.069.
- [10] Decision Trees — scikit-learn 0.24.2 documentation, Available online: <https://scikit-learn.org/stable/modules/tree.html>, (accessed Jun. 02, 2021).
- [11] Sklearn linear- model RidgeCV — scikit-learn 0.24.2 documentation, Available online: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.RidgeCV.html, (accessed Jun. 02, 2021).
- [12] Y. Ren, P. N. Suganthan, and N. Srikanth, Ensemble methods for wind and solar power forecasting - A state-of-the-art review, *Renew. Sustain. Energy Rev.*, Vol. 50, 2015, 82–91, doi: 10.1016/j.rser.2015.04.081.
- [13] IDEAM, Metodología de la Operación Estadística Variables Meteorológicas, *Inst. Hidrol. Meteorol. y Estud. Ambient.*, Available Online: <http://www.ideam.gov.co/documents/11769/72085840/Documento+metodologico+variables+meteorologicas.pdf/8a71a9b4-7dd7-4af4-b98e-9b1eda3b8744>.
- [14] M. B. Mohammed, H. S. Zulkafli, M. B. Adam, N. Ali, and I. A. Baba, Comparison of five imputation methods in handling missing data in a continuous frequency table, *Proceedings of Sciemathic*, Vol. 2355, 2021, doi: 10.1063/5.0053286.
- [15] How HOMER Calculates the PV Array Power Output, Available online: https://www.homerenergy.com/products/pro/docs/latest/how_homer_calculates_the_pv_array_power_output.html (accessed May 12, 2021).
- [16] How HOMER Calculates the Radiation Incident on the PV Array. Available online: https://www.homerenergy.com/products/pro/docs/latest/how_homer_calculates_the_radiation_incident_on_the_pv_array.html, (accessed Apr. 26, 2021).
- [17] A. Farhangfar, L. Kurgan, J. Dy, Impact of imputation of missing values on classification error for discrete data, *Pattern Recognition*, Vol. 41, No. 12, 2008, pp. 3692–3705, doi:10.1016/J.PATCOG.2008.05.019.
- [18] E.T. Matsubara, R.C. Prati, G. E.A.P.A. Batista, M.C. Monard, Missing value imputation using a semi-supervised rank aggregation approach. *SBLA*, 2008, pp. 217–226.
- [19] R. Ahmed, V. Sreeram, Y. Mishra, and M. D. Arif, A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization, *Renewable and Sustainable Energy Reviews*, Vol. 124, 2020. doi: 10.1016/j.rser.2020.109792.
- [20] G. Graditi, S. Ferlito, and G. Adinolfi, Comparison of Photovoltaic plant power production prediction methods using a large measured dataset, *Renew. Energy*, Vol. 90, 2016, pp. 513–519, doi: 10.1016/J.RENENE.2016.01.027.
- [21] R.A Guarnieri, F.R Martins, and E.B Pereira, Solar radiation forecast using artificial neural networks. *National Institute for Space Research*, 2008, pp.1–34.
- [22] K.T. Bui, D. Tien Bui, J. Zou, C. Van Doan, and I. Revhaug, A novel hybrid artificial intelligent approach based on neural fuzzy inference model and particle swarm optimization for horizontal displacement modeling of hydropower dam. *Neural Comput. Appl.*, Vol. 29, No.12, 2016, pp. 1495–1506, doi: 10.1007/S00521-016-2666-0.
- [23] D. Tien Bui, K. T. T. Bui, Q. T. Bui, C. Van Doan, and N. D. Hoang, Hybrid Intelligent Model Based on Least Squares Support Vector Regression and Artificial Bee Colony Optimization for Time-Series Modeling and Forecasting Horizontal Displacement of Hydropower Dam. *Handb. Neural Comput.*, 2017, pp.279–293, doi: 10.1016/B978-0-12-811318-9.00015-6.
- [24] J. Hertz, A. Krogh, and R. G. Palmer, Introduction to the theory of neural computation. *Introd. to Theory Neural Comput.*, pp. 1–327, doi: 10.1201/9780429499661/INTRODUCTION-THEORY-NEURAL-COMPUTATION-JOHN-HERTZ-ANDERS-KROGH-RICHARD-PALMER.

- [25] S. Al-Dahidi, O. Ayadi, J. Adeb, M. Alrbai, and B. R. Qawasmeh, Extreme Learning Machines for Solar Photovoltaic Power Predictions. *Energies*, Vol. 11, No.10:2725, 2018, doi: 10.3390/EN11102725.
- [26] Q. Hu, R. Zhang, and Y. Zhou, Transfer learning for short-term wind speed prediction with deep neural networks. *Renew. Energy*, Vol.85, 2016, pp. 83–95, doi: 10.1016/J.RENENE.2015.06.034.
- [27] A. S. Qureshi, A. Khan, A. Zameer, and A. Usman, Wind power prediction using deep neural network based meta regression and transfer learning. *Appl. Soft Comput.*, Vol 58, 2017, pp. 742–755, doi: 10.1016/J.ASOC.2017.05.031.

Contribution of individual authors to the creation of a scientific article (ghostwriting policy)

Juan Andrés Hernández carried out the Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing - original draft and Writing - review & editing.

Idi Isaac Millán carried out the Conceptualization, in-

vestigation, methodology, Resources, Funding acquisition, Project administration, supervision and Writing - review & editing.

Javier Sierra Carrillo carried out Resources, Funding acquisition, Project administration, supervision and Writing - review & editing.

Sources of funding for research presented in a scientific article or scientific article itself

This work is part of the project: “Estrategia de transformación del sector energético colombiano en el horizonte de 2030” funded by the research call 778 of MinCiencias Ecosistema Científico. Contract FP44842-210-2018