Renewable Energy Prediction through Machine Learning Algorithms

Luisa Fernanda Jiménez Alvarez
Facultad de Ingeniería
Universidad Panamericana
Aguascalientes, México
0204923@up.edu.mx

Sebastián Ramos González
Facultad de Ingeniería
Universidad Panamericana
Aguascalientes, México
0204700@up.edu.mx

Antonio Delgado López
Facultad de Ingeniería
Universidad Panamericana
Aguascalientes, México
0204856@up.edu.mx

Diego Alonso Hernández Delgado
Facultad de Ingeniería
Universidad Panamericana
Aguascalientes, México
0205007@up.edu.mx

Ricardo Espinosa
Facultad de Ingeniería
Universidad Panamericana
Aguascalientes, México
respinosa@up.edu.mx

Sebastián Gutiérrez
Facultad de Ingeniería
Universidad Panamericana
Aguascalientes, México
jsgutierrez@up.edu.mx

***Abstract*—This paper aims to implement an efficient renewable energy selection (either solar or wind) based on the chosen geographic location of Aguascalientes, Mexico through a Machine Learning (ML) method. Likewise, the information listed below will provide both a critical analysis and review of the state-of-the-art applications for ML Algorithms such as Support Vector Machines (SVM), Linear Regression (LR) and Neural Network Models (NNM). Rigorous data measurements taken over a period of six months, including those of solar irradiance, temperature, wind speed and wind direction to name a few, were the inputs used in different algorithms in order to find the one that could most accurately predict future weather conditions. Based on the obtained results, the best ML Algorithm ended up being Random Forest; an approach that is capable of building an accurate prediction through the calculation of two crucial parameters; Mean Square Error (MSE) and Mean Absolute Error (MAE).**

***Keywords*—Forecasting, Machine Learning, MAE, MSE, NNM, Random Forest, Renewable Energy, Solar Energy, SVP, Wind Energy**

I. Introduction

Along with the increasing and urgent abandonment of fossil fuels and the development of sustainable methods for energy obtainment, the need has risen for new ways of measuring, predicting and choosing the best renewable alternatives available to generate electricity. According to the World Resources Institute in 2017, 81% of the global energy consumed was accounted to the use of oil, natural gas, and coal. Annually, 15 billion metric tons of fossil fuels are consumed, being the three biggest fossil-fuel consuming countries China, the United States, and India respectively [1].

The presence of renewable energies in electrical systems of second and first-world countries has had a steady and promising growth over the years. This could be due to the concern for climate change and CO2 emissions, and the constant increase of cost per kWh. These factors are essential in order to develop an energy production forecast , since solar and wind energy are making their way through the energy mix. This depends on various meteorological factors that affect most of that production, which will allow the user to reduce costs, while adding a greater demand and value to the energy market. For solar thermal and photovoltaic (PV) energies,the main variable that affects their energy production is the global solar irradiance. This refers to the amount of solar radiation per unit area (W/m^2) [2]. Mahmud Wasfi talked about two different methods in order to convert the energy from the sun: thermal and PV processes; the first one consists on accumulating heat from the sun (typically by reflection) in order to generate vapor to produce mechanical energy with a turbine. The PV process uses the photoelectric effect of some elements like Silicon to convert the solar power into electricity [3].

There are many factors that affect solar power generation, like solar irradiance, humidity, atmospheric pressure, temperature, wind speed, cloud coverage, etc. [4]. These factors, along with transmission lines and the proximity of electrical substations are critical aspects to take into account for any type of solar project. To analyze solar irradiance calculations, satellite imagery has sometimes used, since the Direct Normal Irradiance (DNI) is recorded here. For this, the minimum measurement period of time for some softwares like Meteonorm for example,is approximately 11 years. Using radiometers, the solar radiation can be calculated, while only needing a period of time of one year for data collection and processing [5].

Wind power generation depends on the wind speed and on the aerodynamic and physical characteristics of the wind turbine. Several approaches have been previously proposed for predicting wind speeds, which include physical models for numerical weather prediction, statistical and soft computing approaches; including Linear Regression (LN), Multi-Layer Perceptron (MLP) and Support Vector Machine (SVM) models. MLP's have been used for wind speed energy predictions because of their flexibility and its adaptability of learning abilities [6].

Data collection for wind farms varies depending on the

978-1-7281-9365-6/20/$31.00 ©2020 IEEE

period established for the measurement part of the process. This is a very important step that has to be taken into consideration in order to calculate the viability of the project. In some cases, the minimum period of time needed to acquire solid and trustful information is about three years. Through the application of ML algorithms, this time can be reduced, while also predicting more precisely future data [7].

The National Technical University of Athens is one of the many institutes around the globe that is using ML algorithms as a powerful tool to predict the availability of the resource that is being leveraged [8]. ML, a branch of artificial intelligence is a relatively new sort of modeling approach that has shown really accurate results in system modeling and analysis in various branches of science, including energy generation.

In this work we propose an efficient method for wind and solar energy prediction, forecasting two very important metrics into renewable energy in order to assist early selection of removable energy depending on geographic location. The data was obtained from "Planta Fotovoltaica Aguascalientes Sur 1" located in Aguascalientes, Mexico. The data obtained was temperature, pressure, humidity, wind direction, wind speed, time sunrise, and time sunset. This information allowed us to forecast solar radiation and wind speed with accurate results measured by MSE and MAE obtaining the best results with Random Forest algorithm as a ML method.

The rest of the document is structured as follows. Section II presents the state-of-the-art of the present work. Section III describes the methodology of the proposal description. Section IV presents the experimental results of the study. Section V illustrated the discussion and Section VI concludes the paper.

II. State of the Art

The use Artificial intelligence methods for future Weather Prediction (WP) forecasts is not new and there has been plenty of investigation on this subject. LN is most commonly used in solar irradiance measurements matched with available forecast of Global Horizontal solar Radiation (GHR) [9]. Likewise, Neural Networks have been applied as a tool to reduce the relative Root Mean Square Error (RMSE) of daily average weather forecasting [10] and Gennetics Algorithms (GA's) [24].

Papacharalampous et al recently worked on precipitation and temperature forecasting using classical and ML algorithms in Greece. They focused on two ML algorithms: Support Vector Machine (SVM) and MLP. Likewise, four of the classical methods for precipitation and temperature forecast were used. According to the study, both ML and classical seemed to be competitive, showing similar results [8].

The prediction of PV power generation is based on solar and weather conditions. The data needs to be captured throughout a long period of time in order to be useful for computer models such as the Autoregressive (AR), Moving Average (MA) and Autoregressive Moving Average (ARMA). These algorithms are used in linear weather forecasting to define the linear relationships between the input and output parameters. This results in an Autoregressive Integrated Moving Average (ARIMA) [11].

The first attempts to successfully predict irradiance casting included the Model Output Statistics (MOS). This is a post-processing technique used to objectively interpret numerical model outputs. It determines the statistical relationship between observed weather elements and variable forecast through a numerical model [12]. Later studies involving Neural Networks (NN) for predicting the global and diffuse solar radiation incident on the earth's surface are found in [13].

In [14] smart baseline models for solar irradiation forecasting are presented, comparing different ML approaches such as SVR, NN and Random Forest RF. The results of these ML techniques were compared with genetic algorithms (GA's) approaches, achieving a comparable performance between SVR and GA´s in MSE metric.

The ML methods have been proven to be successful tools for wind speed and wind power prediction. Techniques like k-nearest neighbors (KNN) algorithms have been used worldwide for quite some years, in 2016 [15], O. Kramer shown that KNN and MLP have successfully been applied. In later investigations, Treiber et al [16], proved that Support Vector Regression (SVR) is more accurate for weather predictions in a shorter-term forecast. For post-processed meteorological models, ML techniques were proven to be more accurate. In the future, a hybridization with meteorological methods could be beneficial for different forecast horizons.

The application for ML ensembles to the field of wind power prediction has also been shown to work successfully by Kusiak et al, in 2009, apply different methods for short-term wind power prediction [17]. Various algorithms were compared in 2008 by Fugon, for wind power forecasting and show that the RF algorithm with or without a random input selection, yielded a prediction performance similar to SVR, but in cases where the computation time grows it is recommend to use a linear model [18]. Heinermann and Kramer implemented two methods to evaluate ML for wind energy prediction, comparing the particular performance of the Decision Tree (DT) algorithms, KNN and SVR and then mixing them to test multiple databases with them. The combination of SVR and DT presented a 37% improvement rate compared to SVR [19]. Perera et al. proposed the use of ML algorithms applied to three actual aspects present in the planning of renewable energy plants; first is the forecast of energy generation from wind energy, the solar resource (either for PV systems or solar concentration systems) and hydro power generation. The second aspect is the plant's size and organization selection (this is completely linked to the availability of the resource that will be used for the generation of electrical energy, which is calculated in the energy generation section forecast). Lastly, the integration to the smart grid and its availability take part, since it is important to control the intermittence of solar and wind energy in order to avoid grid destabilization [20].

III. METHODOLOGY

A. Data collection

Many variables were taken into consideration but the ones that were chosen, given the previous analysis, were temperature, pressure, humidity, wind direction, wind speed, time of sunrise, and time of sunset. Most of the already mentioned parameters were obtained from “Planta Fotovoltaica Aguascalientes Sur 1” (PSFV AGS 1) which gave us the data collected from three different weather stations distributed strategically throughout the PV plant, located in Aguascalientes, Mexico (21.8° N, –102.2° E). Other factors such as pressure, humidity, temperature and wind direction were collected in an average-value form during a measurement period of six months.

Automatic professional weather stations are those used by official meteorological services to record and transmit data from remote or unattended areas. They have very high quality; so much so that they are certified (meaning their data is official) and their installation is carried out by specialists. If they can’t transmit data, they can store it until you have the availability to do so automatically. Normally the variables that are measured are temperature, relative humidity, atmospheric pressure, wind, precipitation, and radiation. They have autonomous energy systems either with solar panels or with long-lasting batteries.

The irradiance was measured with 6 digital pyranometers ISO 9060 /second class/ RS-485 SR20-D2 distributed in a close range nearby the weather stations to transmit the data that they collect from different points throughout the installation. The distribution of these weather stations and the pyranometers are shown in Figure 1.

The data encompasses the readings obtained from the first of January of 2020 to the fifth of June of this same year, collecting information every hour during this period of time, in which the plant has been in full production. Throughout the sorting of all the data recollected, the average values of these readings were obtained.

B. Pre-processing

To make the database reading more agile, the same format was given to each of the columns that stored the variables involved in the study. For this, the average measurements per hour of irradiance, wind speed and direction, temperature, and humidity were used, derived from the following formula:

$$x = \frac{\sum_{i=1}^{N} X_i}{N}$$

where each value (for irradiance) of Xi represents the measurement from the previously exposed pyranometers . In the same way, we opted to convert the time for sunrise and sunset to decimal form (based on a 24-hour format). The measurement units for each parameter are shown in Table I.

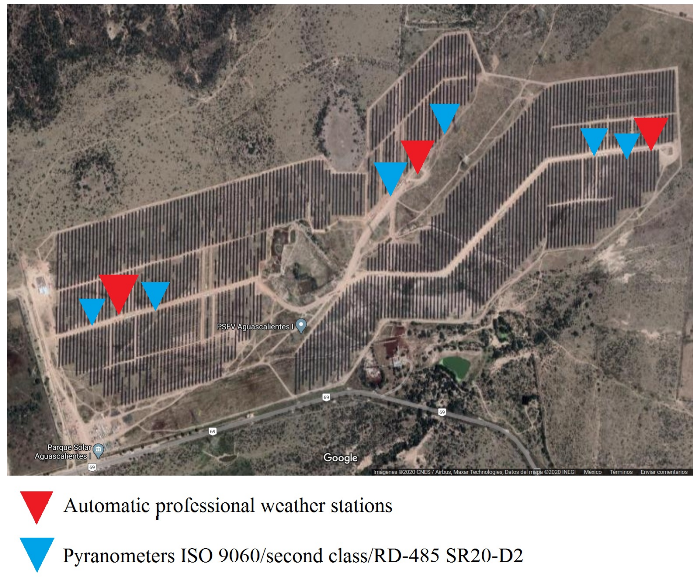

Fig. 1. Distribution of the weather station and pyranometers along PSFV AGS 1.

TABLE I
UNITS USED IN THE STUDY

Values	*Units*
Wind Velocity	m/s
Wind Direction	°(Degrees)
Temperature	°C
Pressure	mmHg
Irradiance	W/m2
Humidity	%

C. Data analysis

1) Irradiance: Laguarda et al propose that the empirical model for irradiance evaluation is based on statistical relationships between horizontal solar irradiance and other variables that can be measured, calculated and obtained by satellite. Empirical models can be accurate if sufficient technology is available in the area to make accurate measurements.The physical model takes into account variables such as the atmospheric state, clouds and water vapor; variables that we are not going to use [21]. This empirical model fits our way of working with variables.

The variables taken from the Data Collection are measured in order to know how they affect the irradiation available. After analyzing the correlation matrix for irradiance, shown in the Figure 2, it is observed that the time, month, and time of sunset have altogether, the majority of influence over the wind direction, whose impact is very considerable. Other variables such as humidity, day, pressure and time of sunrise will have the least direct influence on the solar resource.

2) Relative Wind Energy: There is literature available that analyzes the main characteristics of wind resource, both theoretically and experimentally, from where it is obtained that there are two very important variables for wind energy production; the direction of the wind and its speed (amongst others not stated in this document). These two wind characteristics are carefully analyzed [22]. In this way, the information obtained from the data collection was used to create a correlation matrix, shown in Figure 3, which clearly illustrates that the wind speed has most of the influence on the relative wind energy, with wind direction following up in the correlation, alongside Time Sunrise and Hour, Month, Day, Time Sunset, Humidity and Pressure in that specific order for the wind resource.

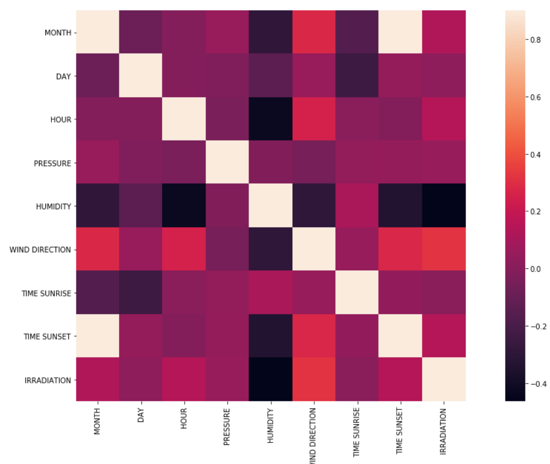

Fig. 2. Correlation Matrix for Solar Irradiance data.

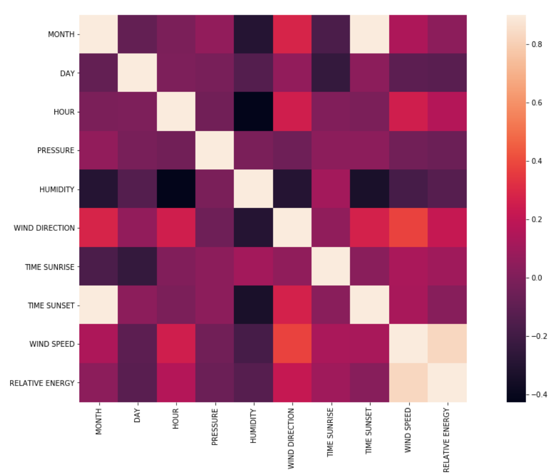

Fig. 3. Correlation Matrix for Relative Wind Energy data.

D. Learning and Prediction

One of the ML algorithms used in this study is the SVM, a method that is capable of solving binary classification problems by formulating them as convex optimization problems [23]. This has already been used to predict the availability of natural resources such as rain in Greece [8], in the same way it has been used for wind power forecasting on its own and in combination with other algorithms [18] [19]. For this case, SVM is applied for relative wind energy and irradiance regression, practically SVR, as one of the most complex and efficient ML algorithms, used in weather forecasting in order to reduce the MSE and MAE [10].

IV. Results

The ML algorithms used for this study were programmed in the 3.7 version of Python. For reading and manipulating purposes of the obtained datasets, the Pandas library was used; for the graphic analysis of information like the Correlation Matrices, Heat-Map and plotting in general, Pyplot and Color libraries from matplotlib were used. In addition, the sklearn library was used for the implementation of all the ML algorithms.

For each of the algorithms, 25% of both datasets, Wind and Solar, was used for testing. To analyze our proposal, the experiments were carried out with four classic ML algorithms: (i) Multi-layer Perceptron, (ii) Support Vector Machine, (iii) Random Forest and (iv) Linear Regression Their individual performance was based on two accuracy-indicators; MSE as shown in equation (1) and MAE as shown in equation (2).

$$MSE = \frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{1}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i|^2 \tag{2}$$

A. Multi-layer Perceptron

The MLP algorithm consists of 3 essential parts; the first one being the input layer from which the data is obtained. In this specific case, the input layer are the datasets mentioned before (Solar and Wind data). The second part consists of a hidden layer where all the input layers are processed, with the last part being the output layer. In this part we have data processed by an N quantity of layers. This algorithm can be compared to a human neuron since it follows the same sequence and structure (receiving data, processing data and final data). For this specific case there were 100 hidden layers, one random state and max iterations of 1000 since it was concluded that with that specific quantity of layers and iterations the algorithm was able to portray the most accuracy in MSE and MAE. For this specific situation, the MLP accuracy resulted insufficient, hence unable to formulate the optimal decision intended. The results from the the testing are shown in Tables II and III. The MLP formula is presented in equation (3).

$$Sum = \sum_{i=1}^{N} I_i W_i \tag{3}$$

B. Support Vector Machines

For the SVM code, after the import of all the necessary libraries and the .csv file reading for each dataset, creating the Correlation Matrix was necessary in order to know which variables had the greatest relation and impact on both types of energy. The Numpy library was used to read and separate, irradiance (for the solar analysis) and relative wind energy (for the wind analysis), since these variables are the (y)outputs, and the all the other ones are the (x)inputs. For the acquisition of better results, another variable pre-processing task had to be carried out within the code, especially for the ones with an hour format. The training was modeled with 75% of the dataset and the regression was made with SVR, non-linear kernel Radial Basis Function (RBF) with a Constant (C) = 100, Epsilon = 0.01 for Wind Data and 0.9 for Solar Data. The results from the the testing are shown in Tables II and III. RBF kernel formula is presented in equation (4).

$$K(x, x') = \exp\left(-\gamma||x - x'||\right) \tag{4}$$

C. Random Forest

RF is an meta-estimator algorithm that fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. One of the multiple reasons why RF is mainly used in plenty of other investigations, is the accurate compile efficiency with large databases to predict in different settings, providing the minimum error possible.

Cardiovascular disease (CVD) is the leading cause of death worldwide. CVD prediction is one of the most effective measures for CVD control by using an electronic health record system collected from 29930 subjects with high-risk in 2014. Almost 30 indicators related to CVD were tested in different regression and classification methods, showing that RF was superior to the others, providing a prediction of a 3-year risk assessment of CVD with an Area Under the Curve (AUC) value of 0.7143. This analysis proved the reliability of CVD treatment predictions in China [24].

Another RF application is on the Internet of Things Technology field [25], where there are always possible risks, one of the most common ones being the classified as unauthorized access to the user's devices and technologies, since authentication schemes for protecting smart devices are vulnerable to many attacks. The solution proposed with the use of ML algorithms is to identify the behavioral traits that a user possesses with patterns like biometrics and the interaction of the users with his/her device to avoid any unauthorized access. Different methods were tested, with the best average accuracy was the one achieved through RF with a 74.97% , showing a importance of recognizing the user's interaction with their smartphones.The examples previously mentioned give us a background of RF's accuracy in the prediction field with multiple uses.

In this project multiple RF iterations were performed using a cross-branching method, going from the second maximum depth value to the sixth, using the same values from the LR and SVM for training and testing, applying identical conditions to all of them. This approach was taken in order to achieve a minimum prediction error. The results from the testing are given in Table IV. When making the regression, the RF algorithm uses the same equation for MSE.

D. Linear Regression

The LR algorithm performs a regression task. This approach is used to find out the linear relationship between an input and the output, in this case being the recorded variables and the forecasting, respectively. LR's purpose is to predict a dependent value (y) and if there is enough correlation between the independent data (x) the algorithm will predict future dependent values with acceptable accuracy. 25% of the data from each energy dataset was used for testing and the remaining 75% for training. An important aspect to take into consideration is that in this case, there is no random state. An LR algorithm is considered successful if the error rate is almost non-existent, which was achieved since MSE and MAE from each specific energy were very close to zero. The results from the the testing are given in Tables II and III. Linear Regression equation is shown in equation (5).

$$y = \theta_1 + \theta_2 * x \tag{5}$$

TABLE II
PERFORMANCE OBTAINED BY MACHINE LEARNING ALGORITHMS (MLP, SVM, RF, LR) WITH SOLAR IRRADIATION FORECASTING

Algorithm	*MAE*	*MSE*
MLP (Multi-layer preceptron)	0.377550	0.2222
SVM (Support Vector Machines)	43.7748	6351.9697
RF (Random Forest)	0	0
LR (Linear Regression)	4.23100e-13	2.8026e-25

TABLE III
PERFORMANCE OBTAINED BY MACHINE LEARNING ALGORITHMS (MLP, SVM, RF, LR) WITH RELATIVE WIND ENERGY FORECASTING

Algorithm	*MAE*	*MSE*
MLP (Multi-layer preceptron)	0.2969	0.1336
SVM (Support Vector Machines)	2.7396	68.9663
RF (Random Forest)	0	0
LR (Linear Regression)	4.1717e-13	2.7176e-25

TABLE IV
CROSS VALIDATION PROCESS IN RANDOM FOREST ALGORITHM

Max Depth	*MSE*	*MAE*
2	1.8173	4.3835
3	0.8455	0.9091
4	0.3499	0.1392
5	0.0024	0.0001
6	0.0	0.0

V. DISCUSSION

As seen in the State-of-Art section, SVM was not the best option to work with, so much so, that it was better to integrate it with another ML algorithm such as KNN or DT, for both wind and solar regressions. However, SVM performance for relative wind energy was much better than with irradiance, so it can be stated that an improvement in accuracy can be looked upon to calculate the lowest MAE and MSE possible in future predictions.

RF had the best performance as to the accurate prediction of relative wind speed and solar irradiance. It is important to mention that the scope of this project contemplated a specific location hence, testing this trained model in different locations will deauthenticate the reliability of its accuracy.

VI. Conclusions

While comparing all the different error rates calculated by the different ML approaches previously mentioned, the best outcome for the optimal renewable energy selection in Aguascalientes, was throughout the RF algorithm, achieving MAE and MSE of zero as shown in experimental results. Through its implementation, it can be stated with certainty that this algorithm can effectively and precisely predict solar irradiance and relative wind energy, for solar and wind energy production respectively. Likewise, this is a very helpful alternative that can significantly reduce the data recollection period, which is part of the first stage of any project; its analysis, needed in order to predict its viability.

From the experimental results, we concluded that our proposal is completely reliable in forecasting solar irradiance and relative wind speed in order to assist the early selection of renewable energy depending on a specific geographic location.

Lastly, a long-term objective of these findings is the implementation of this approach in a global scale, therefore combining data sets from different geographic locations in order to improve the flexibility and availability of future models will be needed.

Acknowledgments

This research was supported with datasets obtained from the Photovoltaic Solar Plant Aguascalientes Sur 1 of the OBRASCÓN HUARTE LAIN, S.A. (OHL) Company.

References

[1] Emily Cassidy. (May 2, 2019). Which Countries Use The Most Fossil Fuels?. June 6, 2020, World Resources Institute, [Online]. Available:https://blog.resourcewatch.org/2019/05/02/which-countries-use-the-most-fossil-fuels/.

[2] Y. Gala, Á. Fernández, J. Díaz, and J. R. Dorronsoro, "Hybrid machine learning forecasting of solar radiation values," Neurocomputing, 2016, doi: 10.1016/j.neucom.2015.02.078.

[3] M. Wasfi, Solar Energy and Photovoltaic Systems, Cyber Journals, Journal of Selected Areas in Renewable and Sustainable Energy (JRSE), Vol. 1, No. 2, February 2011, pp. 1-8

[4] M. Gao, J. Li, F. Hong, and D. Long, "Day-ahead power forecasting in a large-scale photovoltaic plant based on weather classification using LSTM," Energy, 2019, doi: 10.1016/j.energy.2019.07.168.

[5] W. Derouich, M. Besbes, and J. D. Olivencia, "Prefeasibility study of a solar power plant project and optimization of a meteorological station performance," J. Appl. Res. Technol., 2014, doi: 10.1016/S1665-6423(14)71607-4.

[6] R. Ak, O. Fink, and E. Zio, "Two Machine Learning Approaches for Short-Term Wind Speed Time-Series Prediction," IEEE Trans. Neural Networks Learn. Syst., 2016, doi: 10.1109/TNNLS.2015.2418739.

[7] M. Abbes and J. Belhadj, "Development of a methodology for wind energy estimation and wind park design," J. Renew. Sustain. Energy, 2014, doi: 10.1063/1.4895919.

[8] G. Papacharalampous, H. Tyralis, and D. Koutsoyiannis, "Univariate Time Series Forecasting of Temperature and Precipitation with a Focus on Machine Learning Algorithms: a Multiple-Case Study from Greece," Water Resour. Manag., 2018, doi: 10.1007/s11269-018-2155-6.

[9] J.Jensenius, G.Cotton,The development and testing of automated solar energy forecasts based on model output statistics (MOS) technique, in:1st Workshop on Terrestrial Solar Resource Forecasting and on Use of Satellites for Terrestrial Solar Resource Assessment, 1981, pp.22–29.

[10] L. Benali, G. Notton, A. Fouilloy, C. Voyant, and R. Dizene, "Solar radiation forecasting using artificial neural network and random forest methods: Application to normal beam, horizontal diffuse and global components," Renew. Energy, 2019, doi: 10.1016/j.renene.2018.08.044.

[11] G. Reikard, "Predicting solar radiation at high resolutions: A comparison of time series forecasts," Sol. Energy, 2009, doi: 10.1016/j.solener.2008.08.007.

[12] [1] H. R. Glahn and D. A. Lowry, "The Use of Model Output Statistics (MOS) in Objective Weather Forecasting," J. Appl. Meteorol., 1972, doi: 10.1175/1520-0450(1972)011¡1203:tuomos¿2.0.co;2.

[13] M. Santamouris, G. Mihalakakou, B. Psiloglou, G. Eftaxias, and D. N. Asimakopoulos, "Modeling the global solar radiation on the earth's surface using atmospheric deterministic and intelligent data-driven techniques," J. Clim., 1999, doi: 10.1175/1520-0442(1999)012¡3105:MTGSRO¿2.0.CO;2.

[14] Urraca, R., Antonanzas, J., Alia-Martinez, M., Martinez-De-Pison, F., & Antonanzas-Torres, F. (2016). Smart baseline models for solar irradiation forecasting. Energy Conversion and Management, 108, 539-548. doi:10.1016/j.enconman.2015.11.033

[15] J. Heinermann and O. Kramer, "Machine learning ensembles for wind power prediction," Renew. Energy, 2016, doi: 10.1016/j.renene.2015.11.073.

[16] N. A. Treiber, S. Späth, J. Heinermann, L. Von Bremen, and O. Kramer, "Comparison of numerical models and statistical learning for wind speed prediction," in 23rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2015 - Proceedings, 2015.

[17] A. Kusiak, H. Zheng, and Z. Song, "Short-term prediction of wind farm power: A data mining approach," IEEE Trans. Energy Convers., 2009, doi: 10.1109/TEC.2008.2006552.

[18] L. Fugon, J. Juban, and G. Kariniotakis, "Data mining for wind power forecasting," in European Wind Energy Conference and Exhibition 2008, 2008.

[19] J. Heinermann and O. Kramer, "Precise wind power prediction with SVM ensemble regression," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2014, doi: 10.1007/978-3-319-11179-7_100.

[20] K. S. Perera, Z. Aung, and W. L. Woon, "Machine Learning Techniques for Supporting Renewable Energy Generation and Integration: A Survey," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 2014, doi: 10.1007/978-3-319-13290-7_7.

[21] N. Chrysochoidis-Antsos, A. V. Amoros, G. J. W. van Bussel, S. M. Mertens, and A. J. M. van Wijk, "Wind resource characteristics and energy yield for micro wind turbines integrated on noise barriers – An experimental study," J. Wind Eng. Ind. Aerodyn., 2020, doi: 10.1016/j.jweia.2020.104206.

[22] M. Awad and R. Khanna, "Support Vector Regression," in Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers, Berkeley, CA: Apress, 2015, pp. 67–80.

[23] A. Laguarda, R. Alonso-Suárez, and R. Terra, "Solar irradiation regionalization in Uruguay: Understanding the interannual variability and its relation to El Niño climatic phenomena," Renew. Energy, 2020, doi: 10.1016/j.renene.2020.05.083.

[24] L. Yang et al., "Study of cardiovascular disease prediction model based on random forest in eastern China," Sci. Rep., 2020, doi: 10.1038/s41598-020-62133-5.

[25] Alqarni, M.A., Chauhdary, S.H., Malik, M.N., Ehatisham-ul-Haq, M., Azam, M.A., "Identifying smartphone users based on how they interact with their phones", (2020) Human-centric Computing and Information Sciences, doi=10.1186/s13673-020-0212-7.