Models for the optimization and evaluation of photovoltaic self-consumption facilities

Llanos Mora-López[1] and Mariano Sidrach-de-Cardona[2]

[1] Dpto. Lenguajes y Ciencias de la Computación/Universidad de Málaga, Málaga (España)

[2] Dpto Física Aplicada II/ Universidad de Málaga, Málaga (España)

Abstract

The results obtained for the modeling and optimization of photovoltaic self-consumption facilities are presented. The study has been carried out for three Spanish cities with different climatic conditions. The self-consumption and self-sufficiency curves for different hourly consumption profiles have been obtained based on the installed peak power and the size of the battery. Different models of machine learning are proposed to predict these parameters. The input variables of these models are related to the configuration of the installation, its location and the type of consumption profile. The model with best predictions of self-sufficiency is Random Forest, which in cross-validation has a relative error of 5%. For the prediction of self-consumption, the model that performs best is the multilayer perceptron, with an absolute relative error of 3% and a mean absolute error of 0.55 (%, as the self-consumption is a percentage), that is, the mean difference between the estimated and actual self-consumption values is 0.55.

Keywords: photovoltaic self-consumption, photovoltaic self-sufficiency, evaluation, modelling

1. Introduction

Photovoltaic energy is called to play an increasingly important role within the current energy mix. Within photovoltaic systems, self-consumption facilities allow empowering citizens and hold them responsible for the production and use of energy. In recent years, and thanks to the growth that is taking place in this type of facilities, the concept of self-sufficient housing and zero energy housing has emerged. A house or a building of zero energy (known as ZEB, Zero Energy Building) is a concept that is used in buildings with an energy balance between generation and consumption of energy close to zero or even positive in a year typical (NREL, 2006), (NREL, 2010). The massive development of this type of housing could mitigate economic and environmental problems, such as CO2 emissions and dependence on fossil energy sources.

The analysis and modeling of self-consumption photovoltaic systems allows us to determine the different scenarios that define the conditions of optimal design and operation of this type of systems, the energy that can be used directly and the energy that is going to be exchanged with the electrical power grid.

In this paper, we present the results obtained for the optimization of self-consumption facilities. The study has been carried out for different Spanish cities with different climatic conditions. The self-consumption and self-sufficiency curves have been obtained based on the installed peak power and the size of the battery, for different hourly consumption profiles. Different models of machine learning are proposed to predict these parameters.

2. Materials and methods

2.1 Parameters for evaluating a self-consumption facility

The energy generated by a photovoltaic system depends, on the one hand, on climatological parameters, such as the solar radiation it receives and the ambient temperature and, on the other, on the technology of the modules, such as type of modules, performance, etc. In addition, the different losses that occur in the system must be

doi:10.18086/swc.2019.30.01 Available at http://proceedings.ises.org

taken into account.

The method proposed in (Osterwald, 1986) is used to estimate the power generated by the photovoltaic generator (P_m). It uses the incident irradiance and the working temperature of the modules as inputs; the following expression:

$$P_m(\mathrm{W}) = P_{m,ref} \cdot \frac{G}{G_{ref}} \cdot [1 + \gamma \cdot (T_m - T_{m,ref}) \qquad \text{(eq. 1)}$$

were: $P_{m,ref}$ is the peak power of the PV generator in standard conditions (STC), G is the global irradiance incident on the modules (W/m^2), G_{ref} is the global irradiance in STC (1000 W/m^2), T_m is the module temperature, $T_{m,ref}$ is the module temperature in STC (25º) and γ is the power losses temperature coefficient (%/ºC).

The King model has been used to calculate the operating temperature, T_m, of the different modules in outdoor condition. This model is widely used in the literature (Mora-Segado, 2015). The model proposes a relationship between module temperature, ambient temperature (T_{amb} , ºC), incident irradiance (G, W/m^2) and wind speed (W, m/s) according to the expression:

$$T_m = T_{amb} + G \cdot e^{(m+n \cdot W)} \qquad \text{(eq. 2)}$$

where m is an empirical dimensionless coefficient that describes the impact of the irradiance in module temperature and n is an empirical coefficient that describes the effect of the wind in the module temperature. For the most common technologies, these empirical coefficients of this model are described in (Mora-Segado, 2014).

The losses produced in a photovoltaic system in direct current are due to the following factors:

- Losses due to angular reflectance and variation of the incident radiation spectrum. In this work, they will be assumed to be 3%.
- Ohmic losses due to wiring (L_o). In this study, they will be assumed to be 2%, that mean L_o=0.98
- Losses due to the dispersion of parameters in the generator (L_p). In this study, they will be assumed to be 2%. L_p=0.98
- Losses due to errors in the tracking of the maximum power point of the inverter (L_t). In this study, they will be assumed to be 1%. L_t=0.99

The power at the input of the inverter is estimated using the expression:

$$P_{in} = L_o L_p L_t P_m \qquad \text{(eq. 3)}$$

Finally, to determine the power generated in alternating current at the output of the inverter (P_{out}), the performance of the inverter must be taken into account. This yield will be calculated according to the equation proposed by Jantsch (M. Jantsch el., 1992) from the input power to the inverter normalized to its nominal power (η_{in}):

$$\eta_{inv} = \frac{P_{in} - (b_0 + b_1 \cdot P_{in} + b_2 \cdot P_{in}^2)}{P_{in}} \qquad \text{(eq. 4)}$$

$$P_{out} = \eta_{inv} P_{in} \qquad \text{(eq. 5)}$$

where b_i are the fitting coefficients empirically estimated (b_0=0.04, b_1=0.002, b_2=0.03).

The parameters to be estimated and analyzed to evaluate the operation of a self-consumption photovoltaic installation are similar to those proposed in (Sartori et al., 2012):

- Energy generated by the photovoltaic installation
- Energy injected into the power grid
- Energy imported from the power grid
- Energy consumed

In addition, in the evaluation of photovoltaic self-consumption facilities, two parameters are particularly useful according to (Luthander et al., 2015):

- Percentage of self-consumption (SC_t): defined as the part of energy produced by the photovoltaic system that is directly consumed in the home where the installation is, with respect to the total production of the photovoltaic system.

$$SC_t\ (\%) = \frac{\sum E_{t,FV,self}}{\sum E_{t,FV}} \times 100 \qquad \text{(eq. 6)}$$

- Percentage of self-sufficiency (SS_t): defined as the part of energy produced by the photovoltaic system that is directly consumed in the home where the installation is, with respect to the total consumption of the home.

$$SS_t\ (\%) = \frac{\sum E_{t,FV,sc}}{\sum C_t} \times 100\ \ (\%) \qquad \text{(eq. 7)}$$

2.2 Data mining models

Data mining has been previously used for the prediction of the production of photovoltaic systems; for example, in (Alfadda et al., 2017) and (Sharma et al., 2011) the use of vector support models is proposed and in (Nageem and Jayabarathi, 2017) correction factors that are included in this type of models are included. They take into account different weather conditions. Hybrid models have also been proposed that include electrical and statistical models (Filipe et al., 2015) and models based on neural networks and fuzzy logic (Sivaneasan et al., 2017). In all of them what is done is the prediction of the production of a system, but the values of self-consumption and self-sufficiency are not modeled in any case.

We propose the use of the following models:

- Linear regression: it is used with numerical variables. In the event that there are independent variables of nominal type, these must be previously transformed into dummy variables. Linear regression is a model that attempts to adjust independent variables using a linear equation.
- Multilayer perceptron: it is a useful mathematical model for modeling nonlinear relationships between input and output data. It is a type of neural network, which implies a generalization of the simple perceptron in which several simple perceptrons are combined and allows treating some nonlinear problems (Minsky and Papert, 1969). From this proposal, (Rumelhart et al., 1986) presented a modification that allowed the back-propagation of the errors measured at the network exit to the hidden neurons. From the point of view of its architecture it is characterized because it has its neurons grouped in layers of different levels: an input layer, an output layer and hidden layers; each layer has a group of neurons.
- Decision trees: a decision tree is a logical structure constructed from a set of rules. The first proposals to use decision trees from data sets were proposed in (Hunt et al., 1966), (Quinlan, 1979), (Quinlan, 1983) and (Quinlan, 1986).
- M5’: it is an algorithm derived from the M5 method defined by Quinlan (Quinlan, 1992). It was proposed by Wang and Witten (Wang and Witten, 1997). It was defined to predict numerical values.
- REPTree: it is a fast learning model of decision trees. For the classification of numerical variables, the algorithm first orders the values of those variables and starts the execution. Then, use the ordered lists to calculate the best way to divide into each node of the tree. This way of dividing minimizes the variance. The measure that is used is entropy.
- Random Forest: it is part of what is known as multiclassifier systems. These methods are usually very accurate (Sardá-Espinosa, 2017) and robust in the event that there is noise in the data, they also do not produce overfitting. However, they are more difficult to interpret when compared to models based on simple regression trees. It is an algorithm that induces a series of individual trees (Breiman, 2001). To do this, the number of variables that are used to generate each of those trees that will be created from each node is limited. This number is usually less than the number of variables available.

For the linear regression model, the independent variable is the self-sufficiency (or the self-consumption) and the dependent variables are the those enumerated in Data section. The estimation is done by using ordinary least square method. For the rest of the models, these independent variables are used as input data and the output self-sufficiency (or self-consumption) is obtained by the classification model that can be a decision tree or a forest or trees.

3. Data

The study has been carried out for three different Spanish cities with different climatic conditions (Table 1).

Using the meteorological parameters of hourly global incident radiation and hourly temperature, the energy balances of a self-consumption facility has been obtained using different configurations for peak power photovoltaic modules and battery capacities.

The hourly consumption profiles used are based on the profiles proposed by the UKERK center (UKERK). From these profiles, the percentage of consumption produced every hour with respect to total daily consumption (in annual average values) has been considered. To build the hourly profiles for the 365 days of the year, the periods proposed by this center have been considered, except for the period they call "high summer". The profiles corresponding to this period have been generated for those days when the average daily temperature is over 25 degrees.

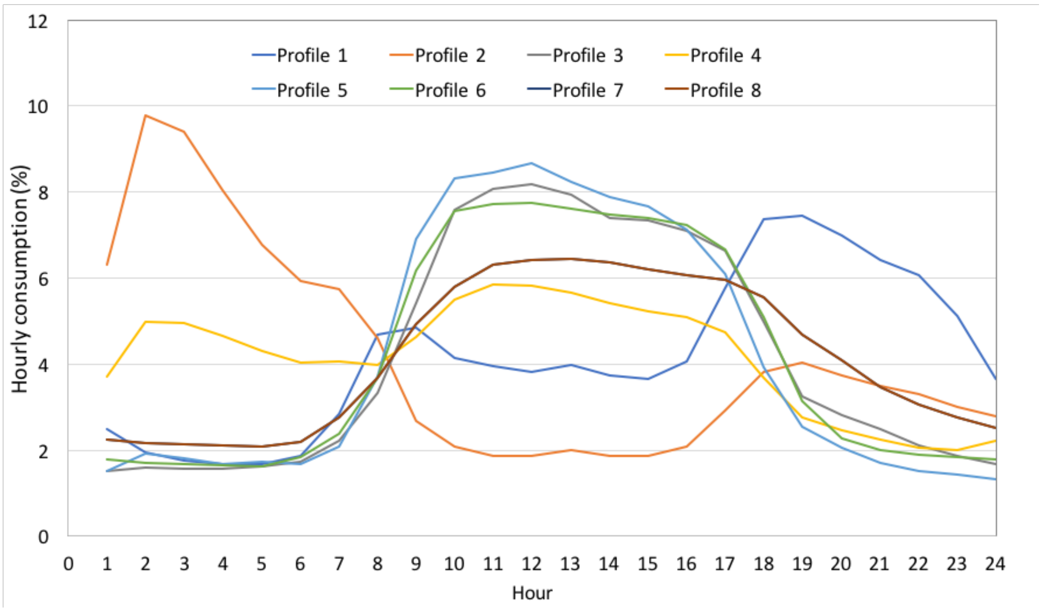

Figure 1. Load profiles analyzed

The peak powers of the installation considered are related to the peak power that causes the installation to generate the same energy as that consumed throughout the year; this is known as the peak power that makes a zero energy building (or house), ZEB. From this value to which the value of one is assigned, the other sizes relative to it are estimated from a size of 0.25 to a size 2 times. The estimated consumption, the photovoltaic energy produced by 1 kWp photovoltaic system and the peak power ZEB values are shown in the table Table 1. For the battery capacity the sizes considered vary from 0 to 25 kWh, with an increase of 2.5 kWh.

Table 1. Meteorological and photovoltaic facility parameters for each location

Location	Gdm (kWh/m^2)	STD Gdm (kWh/m^2)	Temp (°C)	STD T (°C)	Load (kWh)	FV 1kWp (kWh)	Wp (ZEB)
Málaga	5.98	2.04	18.1	0.8	11	4.81	2.3
Madrid	5.09	2.46	14.2	1.6	14.3	4.07	3.5
Santander	4.61	2.39	14.9	0.9	11.8	3.71	3.2

The self-consumption and self-sufficiency coefficients have been obtained based on the installed peak power and the size of the battery using the described hourly consumption profiles.

So, the input variables of the models are related to the configuration of the installation, its location and the type of consumption profile. Specifically, the independent proposed variables are:

- Type of consumption profile
- Average annual value of daily global radiation
- Standard deviation of the daily average values of global radiation
- Average annual value of daily temperature
- Standard deviation of daily global temperature values
- Peak power of the installation
- Battery capacity

Using all these independent variables, the values of self-sufficiency and self-consumption have been calculated for each of the possible combinations. In total, 1056 different values were obtained for each of these two parameters.

4. Results

In order to estimate the self-consumption and self-sufficiency, different models of machine learning have been checked to predict these parameters. The models that have been used are the following:

- Linear Regression
- Multilayer Perceptron
- M5'
- RepTree
- Random Forest

The estimation of the parameters of each of them, as well as of the different error metrics has been done using the Weka tool (Hall et al., 2009). The errors obtained have been evaluated both when using cross correlation and when using test set. The results obtained for estimating the self-sufficiency and self-consumption using cross-validation are shown in Table 2 and Table 3, respectively, and the results obtained when using test set are shown in Table 4 and Table 5 respectively.

Table 2. Metrics obtained for each of the analyzed models when estimating self-sufficiency using cross-validation

Metrics	RL	MP	RF	M5P	REPTree
R	0.798	0.9835	0.9986	0.9927	0.9925
MAE	11.1385	3.2984	0.9364	1.8181	1.6386
RMSE	14.0129	4.2624	1.5568	2.8995	2.8507
RAE	60.03%	17.78%	5.05%	9.80%	8.83%
rRSE	60.23%	18.32%	6.69%	12.46%	12.25%
N	1056	1056	1056	1056	1056

Table 3. Metrics obtained for each of the analyzed models when estimating self-sufficiency for test set

Metrics	RL	MP	RF	M5P	REPTree
R	0.8291	0.9889	0.9983	0.9933	0.9914
MAE	10.4183	2.6083	0.985	1.7628	1.6595
RMSE	12.9941	3.4691	1.6182	2.8199	3.0409
RAE	55.50%	13.90%	5.25%	9.39%	8.84%
rRSE	56.13%	14.99%	6.99%	12.18%	13.14%
N	211	211	211	211	211

Table 4. Metrics obtained for each of the analyzed models when estimating self-consumption using cross-validation

Metrics	RL	MP	RF	M5P	REPTree
R	0.8599	0.9992	0.9963	0.9963	0.9912
MAE	10.2565	0.8015	1.1807	1.4163	1.7489
RMSE	12.9934	1.0402	2.2602	2.2642	3.3616
RAE	51.73%	4.04%	5.95%	7.14%	8.82%
rRSE	50.94%	4.08%	8.86%	8.88%	13.18%
N	1056	1056	1056	1056	1056

Table 5. Metrics obtained for each of the analyzed models when estimating self-consumptio for test set

Metrics	RL	MP	RF	M5P	REPTree
R	0.8472	0.9993	0.9966	0.9947	0.9879
MAE	9.2919	0.5511	1.0817	1.2948	1.9273
RMSE	11.578	0.7945	1.8556	2.1982	3.3243
RAE	51.37%	3.05%	5.98%	7.16%	10.66%
rRSE	53.06%	3.64%	8.50%	10.07%	15.23%
N	211	21	211	211	211

As can be seen, all the proposed models, except linear regression, have correlation coefficients greater than 0.98. Among them, the model with which better predictions of self-sufficiency are obtained is Random Forest, which in cross-validation has an average absolute error of less than 1 and a relative error of 5%. Very good results are also obtained for the M5' and REPTree models. In all cases the relative absolute error is less than 10%, which means for a wide range of self-sufficiency values errors in percentage points of the order of 1%. Figure 2 shows the self-sufficiency values obtained by simulating the behavior of the installation against the values obtained by the Random Forest and REPTree models.

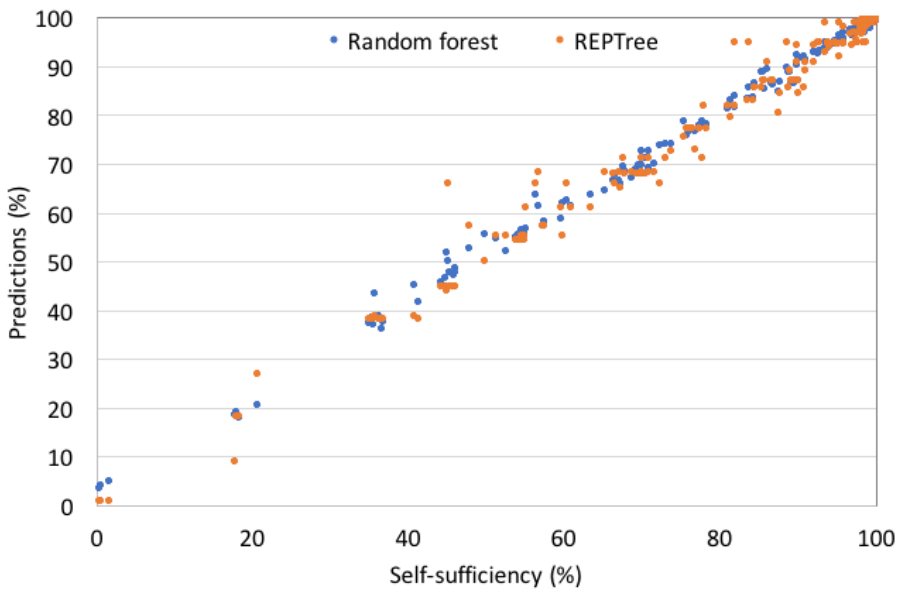

Figure 2. Self-sufficiency values versus the predictions of these values for the Random Forest and REPTree models.

For the estimation of self-consumption, the model that works best is the multilayer perceptron, with an mean absolute error of 0.55 and an relative absolute error of 3% for test set. The decision tree models also have small errors, especially Random Forest and M5P. In Figure 3 self-consumption values versus the predictions of these values for the multilayer perceptron models and Random Forest are shown.

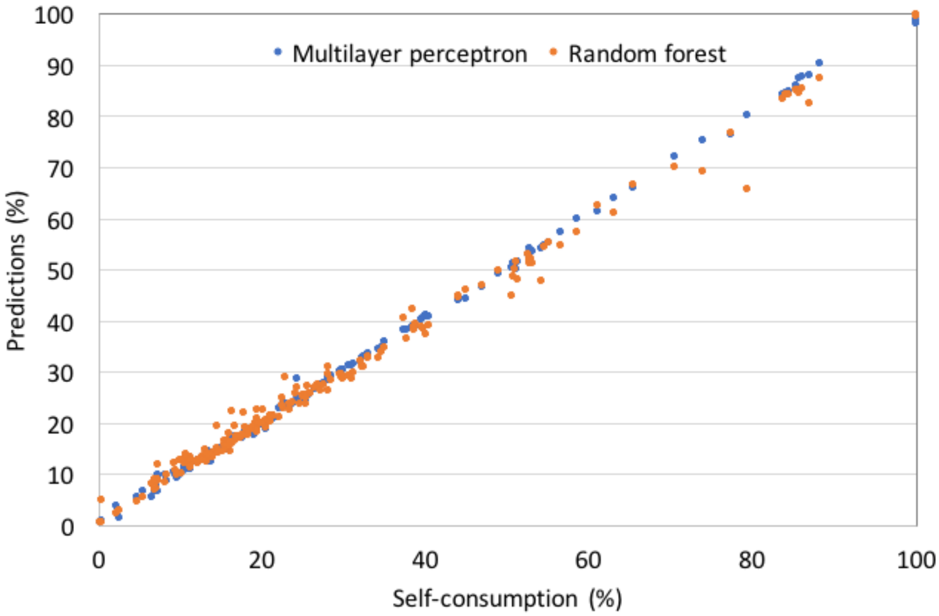

Figure 3. Self-consumption values versus the predictions of these values for the Multilayer Perceptron and Random Forest models.

5. Conclusions

Different data mining models have been evaluated for the modeling of the parameters of self-consumption and self-sufficiency. Specifically, a linear regression, a multilayer perceptron and several types of decision trees have been used. In all cases except linear regression, the results obtained allow us to affirm the validity of the different models. For self-consumption, the model with the best results is the multilayer perceptron, while for self-sufficiency the one that obtains the most accurate results is Random Forest. The correlation coefficient is higher in both cases than 0.99, while the average absolute error is less than 1 and the relative average error is 3% for self-consumption and 7% for self-sufficiency. The best results obtained in the estimation of self-consumption with respect to self-sufficiency are explained because in the first case the self-sufficiency values have been used as an independent variable; this is possible in real situations since the estimation of these parameters can be done in two phases, since for each one of them a different model has been estimated.

6. Acknowledgments

This work has been supported by the project RTI2018-095097-B-I00 at the 2018 call for I+D+i Project of the Ministerio de Ciencia, Innovación y Universidades, Spain.

7. References

Alfadda, R. Adhikari, M. Kuzlu and S. Rahman, Hour-ahead solar PV power forecasting using SVR based approach. IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), Washington, DC, pp. 1-5, 2017.

Breiman, L. Random forests. *Machine Learning*, 45 (1), pp. 5–32, 2001.

Filipe, J.M. Bessa, R.J. Sumaili, J. Tomé, R. and Sousa, J.N. A hybrid short-term solar power forecasting tool. *18th International Conference on Intelligent System Application to Power Systems (ISAP)*, Porto, pp. 1-6, 2015.

Hall, M. Frank, E. Holmes, G. Pfahringer, B. Reutemann, P. I. H. Wit- ten, The weka data mining software: an update, SIGKDD Explorations Newsletter 11, 10–18, 2009.

Hunt, E.B. Marin, J. Stone, P.J. Experiments in Induction, Academic Press, New York, 1966.

Jantsch, M. Schmidt, H. Schmid, J. Results on the concerted action on power conditioning and control. *11th European Photovoltaic Solar Energy Conference*. Montreux, pp. 1589-1592, 1992.

Luthander, R. Widén, J. Nilsson, D. Palm, J. Photovoltaic self- consumption in buildings: A review. Applied Energy, 142, pp. 80-94. 2015.

Minsky, M. Papert, S. Perceptrons. An Introduction to Computational Geometry. M. Minsky and S. Papert. M.I.T. Press, Cambridge, Mass., 1969.

Mora Segado, P. Carretero, J. Sidrach-de-Cardona, M. Models to predict the operating temperature of different photovoltaic modules in outdoor conditions. *Progress in Photovoltaics: Research and Applications*, 23 (10), pp. 1267-1282, 2014.

Mora Segado, P. Contribución al estudio de la temperatura de módulos FV de diferentes tecnologías en condiciones de sol real. Tesis Doctoral. Universidad de Málaga, 2015.

Nageem, R. Jayabarathi, R. "Predicting the Power Output of a Grid- Connected Solar Panel Using Multi-Input Support Vector Regression," *Procedia Comput. Sci.*, vol. 115, pp. 723–730, January 2017.

NREL. National Renewable Energy Laboratory. Zero energy buildings: a critical look at the definition. Colorado. 2006.

NREL. National Renewable Energy Laboratory. Net-Zero Energy Buildings: a classification system based on renewable energy supply options. Colorado. 2010.

Osterwald, C. R. Translation of Device Performance Measurements to Reference Conditions. *Solar Cells*, 18 (3-4), pp. 269-279, 1986.

Quinlan, J. Discovering rules by induction from large collections of examples, in: D. Michie (Ed.), *Expert Systems in the Micro Electronic Age*, 1979.

Quinlan, J. Learning efficient classification procedures, in: R.S. Michlaski, J.G. Carbonell, T.M. Mitchell (Eds.), Machine Learning: An Artificial Intelligence Approach, Tioga Press, Palo Alto, CA, 1983.

Quinlan, J. Induction of decision trees, Mach. Learn. 1, 81–106, 1986.

Quinlan, J. Learning with continuous classes. In: *5th Australian Joint Conference on Artificial Intelligence*. World Scientific, Singapore, pp. 343–348, 1992.

Rumelhart, D.E. Hinton, G.E. Williams, R.J. Learning representations by back-propagating errors. Nature, 323, 533-536, 1986.

Sardá-Espinosa, A. Subbiah, S. Bartz-Beielstein, T. Conditional inference trees for knowledge extraction from motor health condition data. *Engineering Applications of Artificial Intelligence*, 62 (Supplement C), pp.26-37, 2017.

Sartori, I. Naplitano, A. and Voss, K. Net zero energy buildings: A consistent definition framework. *Energy and Buildings*, 48, pp.220–232. 2012.

Sharma, N. Sharma, P. Irwin, D. and Shenoy, P. "Predicting solar generation from weather forecasts using machine learning," 2011 *IEEE International Conference on Smart Grid Communications*, Brussels, 2011, pp. 528-533.

Sivaneasan, B. Yu, C.Y. Goh, K.P. Solar Forecasting using ANN with Fuzzy Logic Pre-processing, *Energy Procedia*, vol. 143, pp. 727– 732, Dec. 2017.

Wang, Y. Witten, I.H. Induction of model trees for predicting con- tinuous classes. In: *9th European Conference on Machine Learning* (poster papers). Springer, pp. 128 – 137, 1997.