



HAL
open science

SmartGrid Improvements on Photovoltaic Systems by Analyzing Energy and Weather Measures

Vagner Henrique De Souza, Antonio Wendell de Oliveira Rodrigues, Matheus Lima, Cláudio Marques de Sá Medeiros, Edilson Mineiro Sá Júnior

► **To cite this version:**

Vagner Henrique De Souza, Antonio Wendell de Oliveira Rodrigues, Matheus Lima, Cláudio Marques de Sá Medeiros, Edilson Mineiro Sá Júnior. SmartGrid Improvements on Photovoltaic Systems by Analyzing Energy and Weather Measures. 10th International Workshop on ADVANCEs in ICT Infrastructures and Services (ADVANCE 2023), Federal University of Ceara, University of Evry, Feb 2023, Fortaleza-Jericoacoara, Brazil. 8p, 10.48545/advance2023-fullpapers-1_2 . hal-04077336

HAL Id: hal-04077336

<https://hal.science/hal-04077336v1>

Submitted on 21 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SmartGrid Improvements on Photovoltaic Systems by Analyzing Energy and Weather Measures*

Vagner Henrique de Souza¹, Wendell Rodrigues², Matheus Lima³, Cláudio Marques de Sá Medeiros⁴, and Edilson Mineiro Sá Júnior⁵

¹ PPGCC - Federal Institute of Ceara, Fortaleza, Brazil
vagner@lit.ifce.edu.br

<https://orcid.org/0000-0002-0928-1609>

² PPGCC - Federal Institute of Ceara, Fortaleza, Brazil
wendell@ifce.edu.br

<https://orcid.org/0000-0001-6140-3610>

³ PPGCC - Federal Institute of Ceara, Fortaleza, Brazil
matheuslimaufc@gmail.com

⁴ PPGCC - Federal Institute of Ceara, Fortaleza, Brazil
claudiosa@ifce.edu.br

⁵ IFCE - Federal Institute of Ceara, Fortaleza, Brazil
edilson.mineiro@gmail.com

Abstract

This paper presents an approach for power generation analysis in photovoltaic systems. The objective is to identify the qualitative impact on the energy produced according to fault detection and prediction based on the electrical and weather measurement dataset. Currently, machine and deep learning have been used effectively to recognize patterns in solar panel measurements. Dirt, breakage and shading can greatly reduce the designed efficiency of solar power plants. Thus, increasing the quantity and quality of the failure measurement dataset can significantly improve reliability and its performance.

Contents

1	Introduction	2
2	Photovoltaic Systems	2
3	Fault Detection and Prediction Methods	3
3.1	Light Gradient Boosting Machine	3
4	PV Measurement Dataset	4
4.1	Measures	4
5	Results	4
5.1	Comparing Methods	4
5.2	Training Score for lightGBM	5
5.3	Feature Importance	5
5.4	Prediction Error for lightGBM Regressor	5
6	Conclusion and Perspectives	7

*This is a full research paper submitted to ADVANCE 2023 in the area of interest: *Verticales Industries, Usage and Societal Impacts*.

1 Introduction

Smart grid [5][6] is a system that consists of, by using digital solutions, monitoring, controlling and analyzing in bidirectional power flow among all types of generating units like Thermal power station, hydropower generation, solar power plant, nuclear power station, etc. and end-use consumers. It is used to monitor and control the power flow automatically, through a master station with the help of technology. Smart grid technology enables the controlling of components of the power system in generation, transmission, distribution system. It enables monitoring and controlling the power consumption at consumers' premises by the automation system and provides proper control in the distribution system. It also analyzes the power consumption so that it provides proper control in the generation system and transmission automatically. Smart grid can be applied on all layers of energy handling. Regarding energy producing, solar panel systems provide an interesting power source.

Energy produced by solar panel systems is proliferating fast in recent years. This is due, mainly, to the reduced cost and the benefits relied to environment. However, the performance of those systems is impacted by many issues. Ancuta and Cepisca in [1] present fault analysis for photovoltaic (PV) panels that decrease their result according to plant design.

Thus, this work addresses the external factors that directly impact the power loss. For this, we analyze some industries that use solar plants and enumerate the main issues regarding failures. Most of them are related to vandalism and dust, even bird dust. As the dataset for this problem is almost non-existent, we experimentally collected measurement data from a small-scale solar panel plant from our institution and applied some pattern recognition methods to provide a real-time solution to improve the system performance.

2 Photovoltaic Systems

Krauter[2] defines Photovoltaics(PV) as a solar electric power generated via the direct conversion of solar radiation into electricity. Currently, this conversion can achieve around 26% of a theoretical 33% of the incoming solar irradiance.

According to Sumathi et al [3], in 2010, among the world's total power, the contribution of solar PV was a very small fraction. However a rapid growth is observed in the installation of solar PV and would continue to increase in the upcoming years. In the New Policies Scenario, power generated from solar PV in 2035 is over 26-times that of 2010, expanding from 32 Twh to 846 Twh. The energy from installed solar PV would increase from 67 GW in 2011 to 600 GW in 2035, on account of decreasing expenses and government aids.

Thus, the industry and other productive sectors of society have high competitive gains adopting this technology.

Briefly, a single PV device is known as a cell. An individual PV cell is usually small, typically producing about 1 or 2 watts of power. These cells are made of different semiconductor materials and are often less than the thickness of four human hairs. In order to withstand the outdoors for many years, cells are sandwiched between protective materials in a combination of glass and/or plastics.

PV modules and arrays are just one part of a PV system. Systems also include support structures, along with the components, such as inverters, that take the direct-current (DC) electricity produced by modules and convert it to the alternating-current (AC) electricity used to power electrical devices.

3 Fault Detection and Prediction Methods

An effective way to detect pattern changes is comparing a large number of measures in different days in different weather conditions. Usually, voltage and current have small changes if we have same solar irradiance. Hence, an experimental survey was implemented in order to acquire electrical measures from PV in several time and weather conditions. Figure 1 shows the elements of this additional testbed for small-scale analysis adding external interference such as dirties or shadows. During energy production to supply a small set of loads (lamps and sockets), three measures are made: voltage, current (so power), and solar irradiance from an external metering sensor. Current and voltage come from the installed inverter and all these measures are stored as a time-series data. Data are processed using some python scripts in order to be used as input information to pattern recognizing methods. However, in order to validate small-scale data on larger power plants, we have used data acquired from commercial inverters and meteorological station.

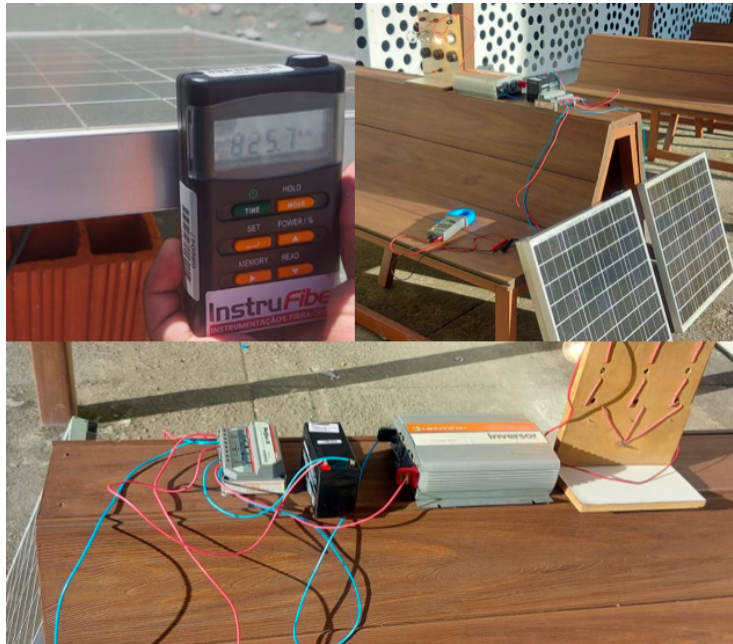


Figure 1: Experimental Dataset Capture. Source: authors

This problem fits typically in regression methods. Regression analysis is a set of statistical processes for estimating the relationships between a dependent variable (often called the *outcome variable*, or *target*) and one or more independent variables (often called *features*, *predictors*, or *covariates*). The objective of regression in machine learning is to predict continuous values such as generated power.

3.1 Light Gradient Boosting Machine

Next sections show how interesting is Light Gradient Boosting Machine (LightGBM) in results. LightGBM [7] extends the Gradient Boosting Regressor (GBR) [9] algorithm by adding a type of automatic feature selection as well as focusing on boosting examples with larger gradients.

This can result in a dramatic speedup of training and improved predictive performance. Later is preferable for our approach.

4 PV Measurement Dataset

The data were collected over 6 months in northeastern Brazil during 2021, where the sun is available most of the time. For training and validation procedures, we have acquired data from a commercial inverter connected to some solar panel strings and, also, a weather station. Data were matched in time in order to link measures. Inverters can handle multiple MPPT modules and, thus, allows better layout scenarios.

4.1 Measures

There are many output measures from data source. However, many of them can be calculated from the other columns. Pre-analysis shows non-relevant features and enlightens others such as solar irradiance and temperature. In fact, as the panels transform sunlight into energy, irradiance plays a key role in production. However, it cannot be disregarded that the temperature and even the wind speed (which helps to cool the panels) directly contribute to the best performance in power generation. Thus, dataset is provided as shown in Table 1 with 35336 rows and 6 columns.

DateTime	Irrad	StTemperature	WindSpeed	PnTemperature	Power
2021-05-01 05:22:00	0	31.3	0.0	29	0
2021-05-01 05:23:00	0	31.3	0.0	31	0
...

Table 1: Cross Data Measures from Inverters and Weather Station

5 Results

Dataset was split in a twofold subset. First one has 24735 rows for training and a second one has 10601 rows for validation test. The prediction is based on the regression in search of the generated power according to the characteristics obtained in the field.

5.1 Comparing Methods

Table 2 presents the regression methods comparison ordered by R^2 , the coefficient of determination indicating the proportion of variation between generated power and its prediction.

Although the training time for the Light Gradient Boosting Machine (lightgbm) method is the longest, we prefer to use it as the best result knowing that the training time does not have much impact on the final application. Moreover, it is always possible to decrease that time by computing on GPUs.

Still observing the results, it is important to note that other methods such as Random Forests (rf) [8] show expressive gains. The low training time has a good R^2 value and presents a better result for the accuracy indicated in the Mean Absolute Percentage Error (MAPE). This method is an excellent candidate when training time is a critical factor in implementing the application.

Model	MAE	MSE	RMSE	R^2	RMSLE	MAPE	TT (Sec)
lightgbm	2953.3167	3.523394e+07	5933.3738	0.8705	0.5873	0.1750	300
gbr	3076.8711	3.620912e+07	6014.6031	0.8670	0.6267	0.2085	2.186
rf	2821.3143	3.858068e+07	6209.9684	0.8582	0.4255	0.1496	9.808
knn	3200.9721	4.239244e+07	6508.1509	0.8442	0.4262	0.1769	231
lr	4094.1461	4.443212e+07	6662.7248	0.8367	0.8606	0.4469	403
lasso	4092.0608	4.443848e+07	6663.1764	0.8367	0.8591	0.4479	62
ridge	4093.5858	4.443253e+07	6662.7510	0.8367	0.8601	0.4469	36
lar	4094.5139	4.443441e+07	6662.8956	0.8367	0.8635	0.4468	41
br	4092.0380	4.443597e+07	6662.9925	0.8367	0.8584	0.4474	77
llar	4120.2396	4.489454e+07	6696.9883	0.8351	0.8407	0.4629	40
et	2879.3797	4.660053e+07	6825.7398	0.8288	0.4145	0.1501	9.893
en	4542.0858	5.224133e+07	7223.4427	0.8081	0.9578	0.7955	84
omp	4672.4580	5.351067e+07	7310.9689	0.8034	0.8998	0.5791	27
huber	3771.4987	5.446469e+07	7374.9502	0.7999	0.6986	0.2994	1.293
ada	5317.3509	5.663545e+07	7524.5133	0.7919	0.7790	0.4537	635
dt	3354.4411	6.235039e+07	7894.1402	0.7708	0.4479	0.1737	186
par	4750.7526	7.142720e+07	8354.1221	0.7377	0.7240	0.3689	179
dummy	14337.4690	2.723538e+08	16502.6821	-0.0008	1.6260	4.1296	29

Table 2: Regression Methods Comparison :: gbr(gradient boosting), rf(random forests), knn(k-nearest neighbour), lr(logistic regression),lasso(least absolute shrinkage and selection operator),lar(least-angle regression),br(bayesian ridge),llar(lasso Least angle regression),et(extra tree),en(elastic net),omp(orthogonal machine pursuit),ada(adaboost regressor),dt(decision tree),par(passive aggressive regressor)

5.2 Training Score for lightGBM

A learning curve shows the validation and training score of the estimator for varying numbers of training samples. It allows us to find out how much we benefit from adding more training data and whether the estimator suffers more from a variance error or a bias error. In Figure 2, both the cross validation score and the training score converge to a value that is quite low with increasing size of the training instances.

5.3 Feature Importance

The *feature importance* will return an array where each element of it is a feature of our model. It will tell you, in proportions, how important that feature is to the model, where the higher the value, the more important the feature is to the model.

As expected, irradiance is the most important feature in the model (see Figura 3). However, it is not a direct function of the generation and, in cases where there are external interferences such as dirt on the panels, this feature may indicate an anomalous pattern in the generation.

5.4 Prediction Error for lightGBM Regressor

A prediction error plots the actual targets (generated power on panels) from the dataset against the predicted values generated by our model. This allows us to see how much variance is in the model. The regression model have a identity line well fitting when comparing against the 45 degree line, where the prediction exactly matches the model (Figure 4).

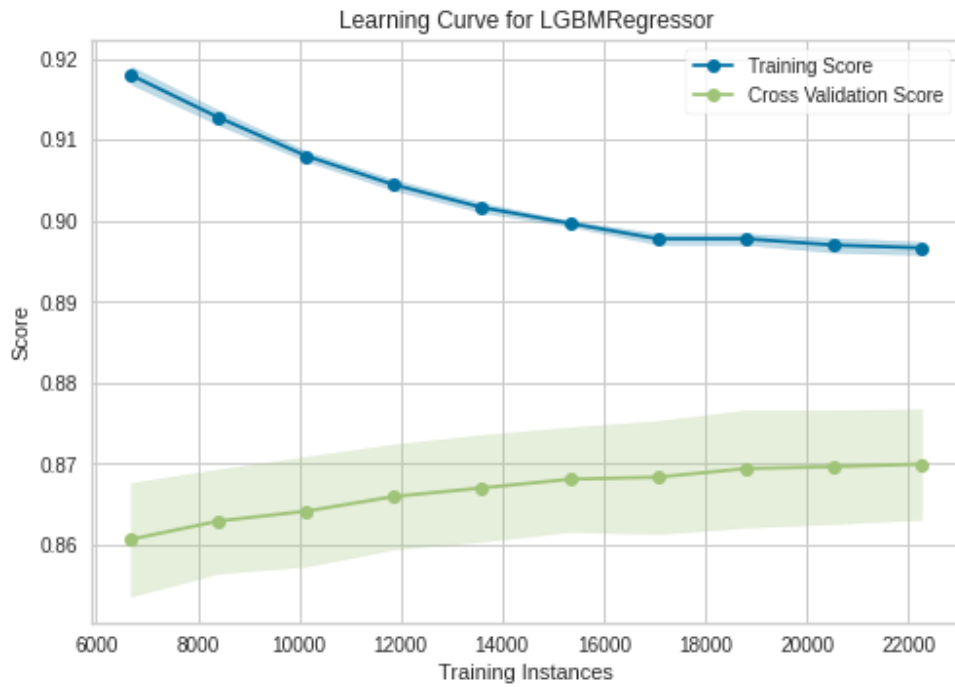


Figure 2: Learning Curve for lightGBM. Source: from experimental results

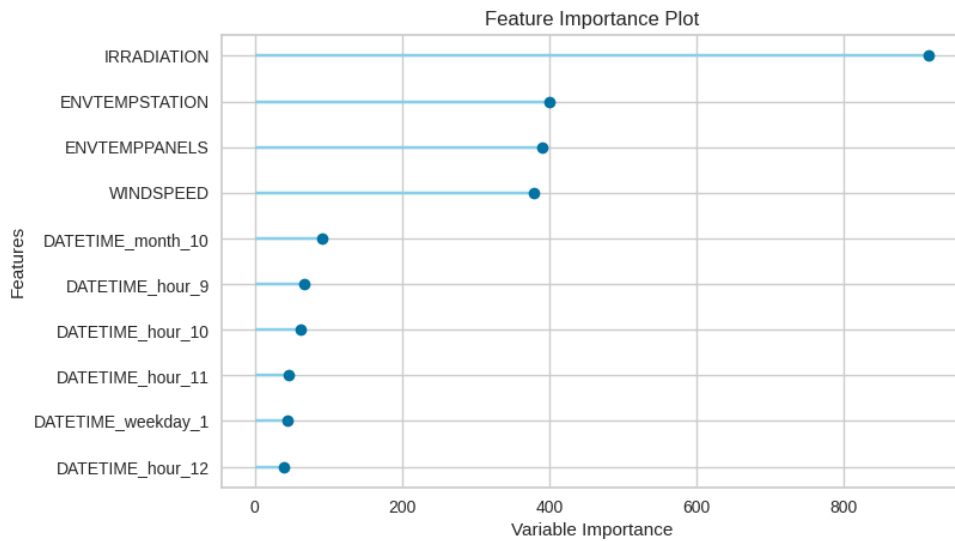


Figure 3: Feature Importance. Source: from experimental results

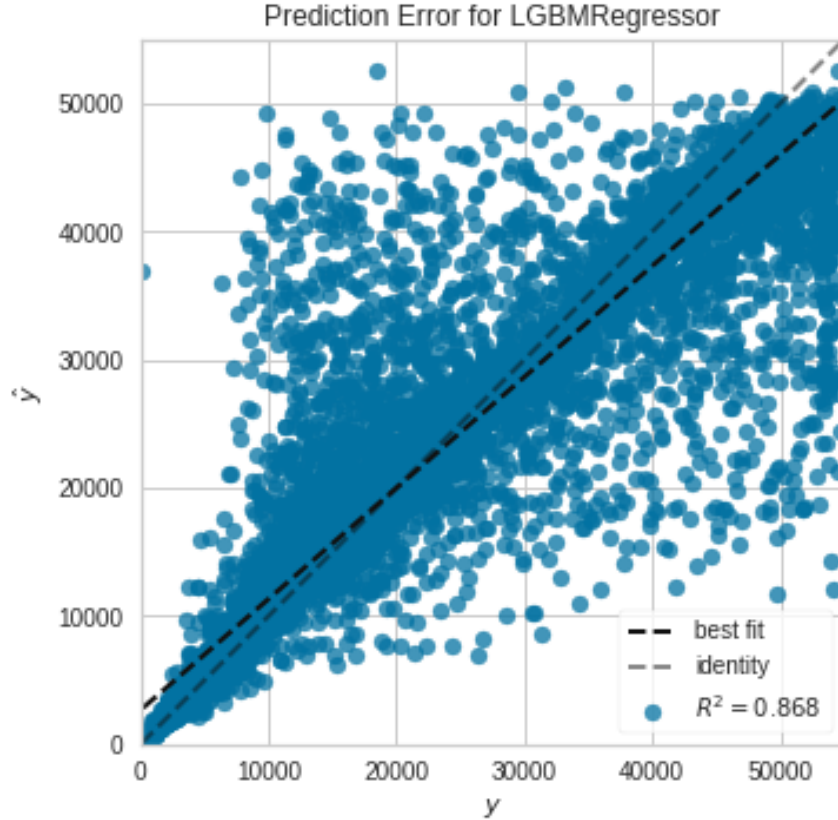


Figure 4: Prediction Error Plot. Source: from experimental results

6 Conclusion and Perspectives

This work presented a proposal to analyze a multi-source dataset to guarantee improvements in the performance of photovoltaic systems. Although the work was based on computational artificial intelligence methods of machine learning, this was not the only focus. In fact, the scarcity of datasets to train these methods has been one of the biggest problems. However, the dataset must be restructured into several data sources in addition to measurements of electrical and meteorological quantities. It is an important perspective to make an association of conventional and thermal image data to improve the assertiveness of the results. With this, the performance of photovoltaic systems is monitored in real time and it will be possible to anticipate problems. As seen in results, several machine learning methods can meet the prediction requirements. Among the methods, lightGBM stands out as it presents the best results in R^2 for our dataset. The generated model can be easily deployed on low cost hardware in order to analyze solar power plants in real-time. That allows increasing efficiency of energy generation. Nevertheless, there is still room for testing with other algorithms and when the set of features (evaluated variables such as individual MPPTs) are in a more complex definition, the use of deep learning as a tool for the identification of anomaly patterns becomes valid.

Acknowledgment

This research was supported by FUNCAP and CAPES. We thank our colleagues from Telematics and Industry Department of IFCE who provided insight and expertise that greatly assisted the research, although they may not agree with all of the interpretations of this paper.

References

- [1] F. Ancuta and C. Cepisca, "Fault analysis possibilities for PV panels," Proceedings of the 2011 3rd International Youth Conference on Energetics (IYCE), 2011, pp. 1-5.
- [2] S. C. W. Krauter, Solar Electric Power Generation, Springer, New York, NY, USA, 2006.
- [3] S. Sumathi, L. A. Kumar, and P. Surekha, Solar PV and Wind Energy Conversion Systems, Springer, Switzerland, 2015.
- [4] Alsharif, Mohammed H., Anabi Hilary, Kelechi, Khalid, Yahya, and Shehzad Ashraf, Chaudhry. "Machine Learning Algorithms for Smart Data Analysis in Internet of Things Environment: Taxonomies and Research Trends".Symmetry 12, no.1 (2020).
- [5] Qiang Sun, Xubo Ge, Lin Liu, Xin Xu, Yibin Zhang, Ruixin Niu, Yuan Zeng, "Review of Smart Grid Comprehensive Assessment Systems", Energy Procedia, Volume 12, 2011, Pages 219-229, ISSN 1876-6102, <https://doi.org/10.1016/j.egypro.2011.10.031>.
- [6] Paul, Riya and Sen, Anirban and Dey, Basudeb, "Study of Application of Smart Grid on Solar Energy" (November 25, 2019). 2nd International Conference on Non-Conventional Energy: Nanotechnology & Nanomaterials for Energy & Environment (ICNNEE) 2019, <http://dx.doi.org/10.2139/ssrn.3492860>
- [7] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. "LightGBM: a highly efficient gradient boosting decision tree." In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc., Red Hook, NY, USA, 3149–3157.
- [8] Breiman, L. Random Forests. Machine Learning 45, 5–32 (2001). <https://doi.org/10.1023/A:1010933404324>
- [9] Friedman JH. Greedy function approximation: a gradient boosting machine. Annals of statistics. 2001;1189–232.