



# Selection of Characteristics by Hybrid Method: RFE, Ridge, Lasso, and Bayesian for the Power Forecast for a Photovoltaic System

Jose Cruz<sup>(✉)</sup>, Wilson Mamani, Christian Romero,  
and Ferdinand Pineda

Universidad Nacional del Altiplano, Puno, Peru

{josecruz,cromero,ferpineda}@unap.edu.pe, wmamani@estudiante.unap.edu.pe

**Abstract.** Currently, the generation of alternative energy from solar radiation with photovoltaic systems is growing, its efficiency depends on internal variables such as powers, voltages, currents; as well as external variables such as temperatures, irradiance, and load. To maximize performance, this research focused on the application of regularization techniques in a multiparametric linear regression model to predict the active power levels of a photovoltaic system from 14 variables that model the system under study. These variables affect the prediction to some degree, but some of them do not have so much preponderance in the final forecast, so it is convenient to eliminate them so that the processing cost and time are reduced. For this, we propose a hybrid selection method: first we apply the elimination of Recursive Feature Elimination (RFE) within the selection of subsets and then to the obtained results we apply the following contraction regularization methods: Lasso, Ridge and Bayesian Ridge; then the results were validated demonstrating linearity, normality of the error terms, without autocorrelation and homoscedasticity. All four prediction models had an accuracy greater than 99.97%. Training time was reduced by 71% and 36% for RFE-Ridge and RFE-OLS respectively. The variables eliminated with RFE were “Energia total”, “Energia diaria” e “Irradiancia”, while the variable eliminated by Lasso was: “Frecuencia”. In all cases we see that the root mean square errors were reduced for RFE.Lasso by 0.15% while for RFE-Bayesian Ridge by 0.06%.

**Keywords:** Regularization Shrinkage · Lasso · Ridge · Bayesian Ridge · RFE · Linear regression · Homoscedasticity

## 1 Introduction

Due to the increasing use of photovoltaic systems in the generation of alternative energy, it is difficult to obtain mathematical or physical models that result in the efficient use of such systems. Likewise, the tools and algorithms provided

by Machine Learning in the use and treatment of data, result in useful tools to model photovoltaic generation systems. Within the field of data-based forecasts we have multiparametric linear regression, which allows forecasts taking into account a set of independent variables that affect the target or dependent variable. These variables affect the prediction to some degree, but some of them do not have so much preponderance in the final forecast, so it is convenient to eliminate them so that the processing cost and time are reduced. Among some of the techniques to exclude irrelevant variables or predictors we have: Subset Selection, Shrinkage Regularization, and Dimension Reduction. Within the first group that identifies and selects among all the available predictors that are most related to the target variable, we have: Bestsubset selection and Stepwise selection. Within this last group we have: forward, backward, and hybrid. In the backward method we have the elimination of recursive functions (RFE), which is the algorithm used in this paper to model the multiparameter photovoltaic system. RFE is used in various studies such as the selection of attributes in classifiers based on artificial neural networks in the detection of cyberbullying [1]. In conjunction with other techniques such as SVR for feature selection based on twin support vector regression [2]; with SVM and Bayes for categorical classifications [3]. For the modeling of emotions and affective states from EEG, combining RFE with Random Forest (RF), Support Vector Regression (SVR), Tree-based bagging [4]. In identifying features for football game earnings forecast, combining it with Gradient Boosting and Random Forest [5]. In the prediction of boiler system failures, using the RFE algorithm in combination with the elimination of recursive functions by vector machine (SVM-RFE) [6]. In the phenotyping of high-yield plants [7], to eliminate spectral characteristics, the elimination of vector-machine recursive characteristics (SVM-RFE), LASSO logistic regression and random forest are used. To perform the short-term electricity price and charge forecast using KNN, [8] uses RFE to eliminate redundancy of functions. To perform heart transplant tests, [9] in pig tests use a combination of RFE-SVM to select the parameters for the estimation of  $V_0$ . In the present work, we perform the combination of RFE with Shrinkage regularization algorithms: Ridge, Lasso, and Bayesian Ridge, establishing a hybrid algorithm for modeling the multiparameter photovoltaic system.

## 2 Methodology

In regression models, a compromise must be made between the bias and the variance provided by the data to be predicted and the model performed. For this, the theory provides us with the following variable selection methods (feature selection): Subset selection, Shrinkage, and Dimension reduction. The first identifies and selects among all the available predictors those that are most related to the response variable. Shrinkage or Shrinkage fits the model, including all predictors, but including a method that forces the regression coefficient estimates to zero. While Dimension Reduction creates a small number of new variables from combinations of the original variables. Each of them has a subset of techniques

such as for subset selection: best subselection and stepwise selection (forward, backward and hybrid). For Shrinkage: Ridge, Lasso and ElasticNet. For Dimension Reduction we have Principal components, Partial Least Square and tSNE. Subset selection is the task of finding a small subset of the most informative elements in a basic set. In addition to helping reduce computational time and algorithm memory, due to working on a much smaller representative set, he has found numerous applications, including image and video summary, voice and document summary, grouping, feature selection and models, sensor location, social media marketing and product recommendation [10]. The recursive feature removal method (RFE) used works by recursively removing attributes and building a model on the remaining attributes. Use precision metrics to rank the feature based on importance. The RFE method takes the model to be used and the number of characteristics required as input. Then it gives the classification of all the variables, 1 being the most important. It also provides support, True if it is a relevant feature and False if it is an irrelevant feature.

The data was pre-processed by eliminating the null values. Next, the non-multicollinearity between the predictors was determined using a heat diagram. Three hybrid methods of variable selection were performed: RFE-Lasso, RFE-Ridge, RFE-Bayesian Ridge, comparing them with RFE-OLS, it was used as a baseline for our work. Finally, the results were validated under conditions of linearity, normality, no autocorrelation of error terms, and homoscedasticity.

### 3 Methods

#### 3.1 Recursive Feature Elimination

For RFE we will use the following algorithm:

- 1 Refine/Train the model in the training group using all predictors
- 2 Calculate model performance
- 3 Calculate the importance of variables or classifications
- 4 For (for) each subset size  $S_i$ ,  $i = 1 \dots S$  do (do
  - 4.1 Keep the most important variables of  $S_i$
  - 4.2 Optional: Pre-process the data
  - 4.3 Refine/Train the model in the training group using  $S_i$  predictors
  - 4.4 Calculate model performance
  - 4.5 Optional: Recalculate rankings for each predictor
  - 4.6 End (end)
- 5 Calculate the performance profile on  $S_i$
- 6 Determine the appropriate number of predictors
- 7 Use the model corresponding to the optimal  $S_i$

The algorithm fits the model to all predictors, each predictor is classified using its importance for the model. Let  $S$  be a sequence of ordered numbers that are candidate values for the number of predictors to retain ( $S_1, S_2, \dots$ ). At each iteration of the feature selection, the highest ranked  $S_i$  predictors are retained,

the model is readjusted, and performance is evaluated. The best performing Si value is determined and the main Si predictors are used to fit the final model. The algorithm has an optional step just at the end of its sequence (8) where the predictor ratings recalculate into the reduced feature set model. For the random forest models, there was a decrease in performance when the rankings were recalculated at each step. However, in other cases when the initial classifications are not good (for example, linear models with highly collinear predictors), the recalculation may slightly improve performance [11].

### 3.2 Ridge

For Ridge the sum of squared errors for linear regression is defined by Eq. 1:

$$E = \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (1)$$

Just as the data set we want to use to make machine learning models must follow the Gaussian distribution defined by its mean,  $\mu$  and variance  $\sigma^2$  and is represented by  $N(\mu, \sigma^2)$ , *i.e.*,  $X \sim N(\mu, \sigma^2)$  where X is the input matrix.

For any point  $x_i$ , the probability of  $x_i$  is given by Eq. 2.

$$P(x_i) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2} \frac{(x_i - \mu)^2}{\sigma^2}} \quad (2)$$

The occurrence of each  $x_i$  is independent of the occurrence of another, the joint probability of each of them is given by Eq. 3:

$$p(x_1, x_2, \dots, x_N) = \prod_{i=1}^N \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2} \frac{(x_i - \mu)^2}{\sigma^2}} \quad (3)$$

Furthermore, linear regression is the solution that gives the maximum likelihood to the line of best fit by Eq. 4:

$$P(X | \mu) = p(x_1, x_2, \dots, x_N) = \prod_{i=1}^N \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2} \frac{(x_i - \mu)^2}{\sigma^2}} \quad (4)$$

Linear regression maximizes this function for the sake of finding the line of best fit. For this, we take the natural logarithm of the probability function (likelihood) (L), then differentiate and equal zero by Eq. 5.

$$\ln(P(X | \mu)) = \ln(p(x_1, x_2, \dots, x_N)) = \quad (5)$$

$$\ln \prod_{i=1}^N \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2} \frac{(x_i - \mu)^2}{\sigma^2}} = \sum_{i=1}^N \ln \left( \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2} \frac{(x_i - \mu)^2}{\sigma^2}} \right) = \quad (6)$$

$$\sum_{i=1}^N \ln\left(\frac{1}{2\pi\sigma^2}\right) - \sum_{i=1}^N \left| \frac{1}{2} \frac{(x_i - \mu)^2}{\sigma^2} \right| \tag{7}$$

$$\frac{\partial \ln(P(X | \mu))}{\partial \mu} = \frac{\partial \sum_{i=1}^N \ln\left(\frac{1}{2\pi\sigma^2}\right)}{\partial \mu} - \frac{\partial \sum_{i=1}^N \frac{1}{2} \frac{(x_i - \mu)^2}{\sigma^2}}{\partial \mu} \tag{8}$$

$$= 0 + \sum_{i=1}^N \frac{(x_i - \mu)}{\sigma^2} = \sum_{i=1}^N \frac{(x_i - \mu)}{\sigma^2} \tag{9}$$

$$\frac{\partial \ln(P(X | \mu))}{\partial \mu} = \sum_{i=1}^N \frac{(x_i - \mu)}{\sigma^2} = 0 \implies \mu = \frac{\sum_{i=1}^N x_i}{N} \tag{10}$$

We take into account here is that maximizing the probability function (likelihood)  $L$  is equivalent to minimizing the error function  $E$ . Furthermore, and it is Gaussian distributed with mean transposition ( $w$ ) \*  $X$  and variance  $\sigma^2$  is show in Eq. 11.

$$y \sim N(\omega^T X, \sigma^2) \quad o \quad y = \omega^T X + \varepsilon \tag{11}$$

Where  $\varepsilon \sim N(0, \sigma^2)$   $\varepsilon$  is Gaussian distributed noise with zero mean and variance  $\sigma^2$ . This is equivalent to saying that in linear regression, the errors are Gaussian and the trend is linear. For new or outliers, the prediction would be less accurate for least squares, so we would use the L2 regularization method or Ridge regression. To do this, we modify the cost function and penalize large weights as follows by Eq. 12:

$$J_{RIDGE} = \sum_{i=1}^N (y_i - \hat{y}_i)^2 + \lambda |w|^2 \tag{12}$$

Where:  $|w|^2 = w^T w = w_1^2 + w_2^2 + \dots + w_D^2$

We, now have two probabilities:

Posterior:

$$P(Y|X, w) = \prod_{i=1}^N \frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2\sigma^2} (y_n - w^T x_n)^2\right) \tag{13}$$

A priori:

$$P(w) = \frac{\lambda}{\sqrt{2\pi}} \exp\left(-\frac{\lambda}{2} w^T w\right) \tag{14}$$

### 3.3 Ridge-Bayesian

So, applying Bayes

$$\exp(J) = \prod_{n=1}^N \exp(-(y_n - w^T x_n)^2) \exp(\lambda w^T w) \tag{15}$$

$$\begin{aligned} \text{Applying Bayes: } J &= (Y - Xw)(Y - Xw)^T + \lambda w^T w \\ &= Y^T Y - 2Y^T Xw + w^T X^T Xw + \lambda w^T w \end{aligned} \quad (16)$$

To minimize J, we use  $\frac{\partial J}{\partial w}$  and set its value to 0. Therefore,  $-2X^T Y + 2X^T Xw + 2\lambda w = 0$

$$\text{So } (X^T X + \lambda I)w = X^T Y \text{ or } w = (X^T X + \lambda I)^{-1} X^T Y$$

This method encourages weights to be small since P (w) is a Gaussian centered around 0. The anterior value of w is called the MAP (maximize posterior) estimate of w.

### 3.4 Lasso

In the same way for Lasso

$$J_{LASSO} = \sum_{n=1}^N (y_i - \hat{y}_i)^2 + \lambda ||w|| \quad (17)$$

Maximizing the likelihood

$$P(Y|X, w) = \prod_{n=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_n - w^T x_n)^2\right) \quad (18)$$

and prior (previous) is given by:

$$P(w) = \frac{\lambda}{2} \exp(-\lambda|w|) \quad (19)$$

So that  $J = (Y - Xw)^T (Y - Xw) + \lambda|w|$   
 $\frac{\partial J}{\partial w} = -2X^T Y + 2X^T Xw + \lambda \text{sign}(w) = 0$   
 Where  $\text{sign}(w) = 1$  If  $x > 0$  and  $-1$  if  $x < 0$  and  $0$  if  $x = 0$

## 4 Data Set

### 4.1 Data Acquisition

The data was collected in the department of Puno whose coordinates are: 15° 29' 27" S and 70° 07' 37" O. The time period was April and August 2019.

The data to be analyzed were: DC Voltage, AC Voltage, AC Current, Active Power, Apparent Power, Reactive Power, Frequency, Power Factor, Total Energy, Daily Energy, DC Voltage, DC Current, and DC Power. Those that were obtained through the StecaGrid 3010 Inverter. The temperature of the environment and the photovoltaic panel were obtained by the PT1000 sensors that are suitable for temperature-sensitive elements given their special sensitivity, precision and reliability. Irradiance was obtained through a calibrated Atersa

brand cell, whose output signal depends exclusively on solar irradiance and not on temperature. The amount of data is reduced from 331157 to 123120 because many of the values obtained are null, for example, the values obtained at night time. Characteristics such as mean, standard deviation, minimum value, maximum value and percentages of the pre-processed data are presented in Table 1 and Table 2. The statistics of the data obtained are shown as median, standard deviation, values: maximums, minimums, and interquartile ranges.

**Table 1.** Statistic table

Statistic	Tension AC	Corriente AC	Potencia activa	Potencia aparente	Potencia reactiva	Frecuencia	Factor de potencia	Energia total
mean	235.45	6.965	1,621.97	1,643.26	219.883	60.003	0.951	5,233.50
std	2.944	2.931	708.23	696.367	66.593	0.046	0.189	1,013.13
min	223.9	0.58	0	135	-843.9	59.5	-0.99	3,894.30
25	233.5	4.639	1,071.10	1,091.00	196.2	60	0.983	4,184.70
50%	235.4	7.564	1,764.00	1,779.55	228.4	60	0.991	5,910.30
75%	237.6	9.43	2,219.30	2,232.70	256.2	60	0.994	6,175.40
max	247.9	12.416	2,879.20	2,898.00	485.1	60.5	0.998	6,427.60

**Table 2.** Statistic table

Statistic	Energia diaria	Tension DC	Corriente DC	Potencia DC	Irradiancia	Temp modulo	Temp ambiente
mean	127.683	334.805	5.558	1,831.38	669.007	35.11	16.611
std	86.399	17.334	2.389	737.254	291.94	11.274	3.773
min	0	220.8	0	0	0	2.4	-2
25	56.666	321.9	3.62	1,261.44	432	27.6	14.5
50%	113.423	332.7	5.89	1,973.63	706	37	17.4
75%	190.369	346	7.65	2,450.61	926	44.2	19.4
max	342.906	420.8	10.78	3,142.27	1,522.00	60.3	27.7

## 5 Results

### 5.1 Non-multicollinearity Between Predictors - Correlation

The independent variables (predictors) should not be correlated with each other, as they would cause problems in the interpretation of the coefficients, as well as the error provided by each one. To determine this, a correlation heat map was used. Correlation is the basis to eliminate or minimize some variables, this is done by a variable selection algorithm or by the researcher’s criteria, of course, advanced methods use an algorithm as will be done later, however, Fig. 1 displays the matrix to validate subsequent results.

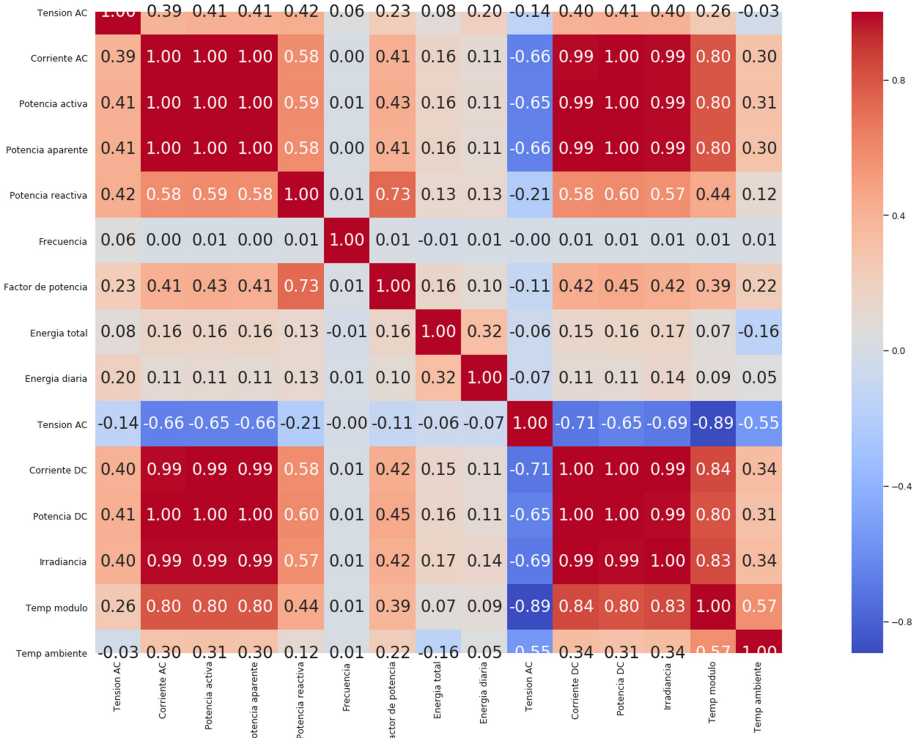


Fig. 1. Correlation matrix.

### 5.2 Prediction

First the RFE method was applied for the selection of variables, to the obtained results we applied the following Shrinkage regularization methods: Lasso, Ridge and Bayesian Ridge The data set is divided into training data 98496 (80%) and test data set 24624 (20%), for better performance seeds are also used. The best seed is also 8849. The RFE algorithm is applied, the following result is obtained:

[‘Tension AC’, ‘Corriente AC’, ‘Potencia aparente’, ‘Potencia reactiva’, ‘Frecuencia’, ‘Factor de potencia’, ‘Energia total’, ‘Energia diaria’, ‘Tension DC’, ‘Corriente DC’, ‘Potencia DC’, ‘Irradiancia’, ‘Temp modulo’, ‘Temp ambiente’]

[True, True, True, True, True, True, False, False, True, True, True, False, True, True]

[‘Tension AC’, ‘Corriente AC’, ‘Potencia aparente’, ‘Potencia reactiva’, ‘Frecuencia’, ‘Factor de potencia’, ‘Tension DC’, ‘Corriente DC’, ‘Potencia DC’, ‘Temp modulo’, ‘Temp ambiente’]

Of the 14 variables evaluated, for RFE the optimal number of characteristic variables was 11 with a score of 0.999768. It is important to mention that RFE discards: “Energia total”, “Energia diaria” e “Irradiancia”. The hyperparameters are then determined for Ridge an alpha value = 1,538 and for Lasso an alpha



value = 0.01. For the models found, we determined  $R^2$  and adjusted  $R^2$ , the mean absolute error of R (MAE), the mean square error of R (RMSE) and Score.

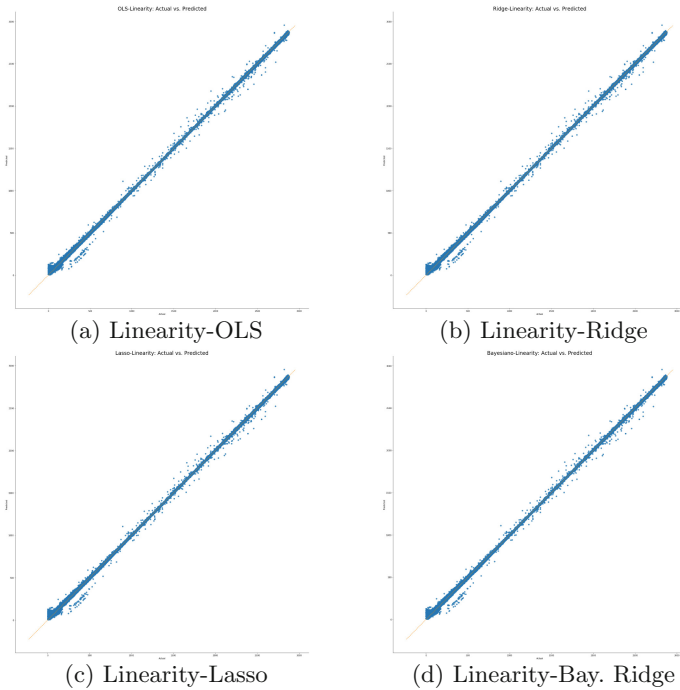
Table 3 and Table 4 shows the values obtained for the proposed groups, where the RFE method with OLS is not part of the research proposal, this result is also used to compare the research results. The following RFE methods with Lasso, RFE with Ridge and RFE with Bayesian Ridge; form the proposal of this research.

## 6 Validation of the Results

To check the results provided by the model, we must check certain assumptions about linear regression. If they are not fulfilled, the interpretation of results will not be valid.

### 6.1 Linearity

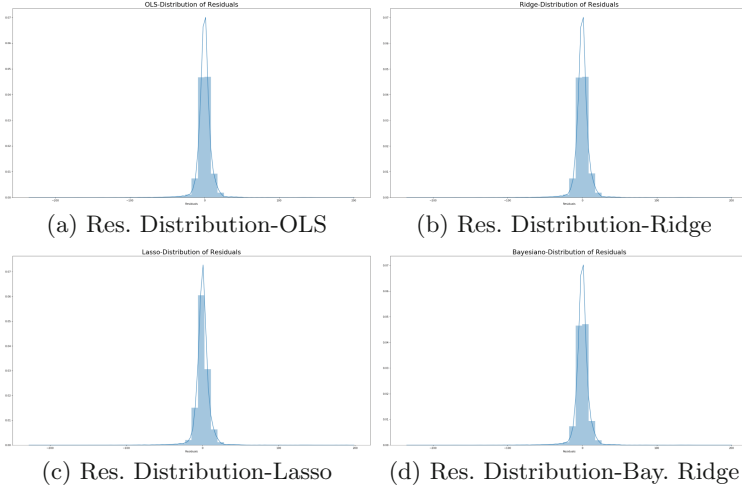
There must be a linear relationship between the actual data and the prediction so that the model does not provide inaccurate predictions. It is checked using a scatter diagram in which the values or points must be on or around the diagonal line of the diagram Fig. 2 shows the linear relationship.



**Fig. 2.** RFE-Linearity, (a) Correspond to model OLS, (b) Correspond to model Ridge (c) Correspond to model Lasso, (d) Correspond to model Bayesian Ridge

### 6.2 Normality of Error Terms

The error terms should be distributed normally. The histogram and the probability graph are shown in Fig. 3.



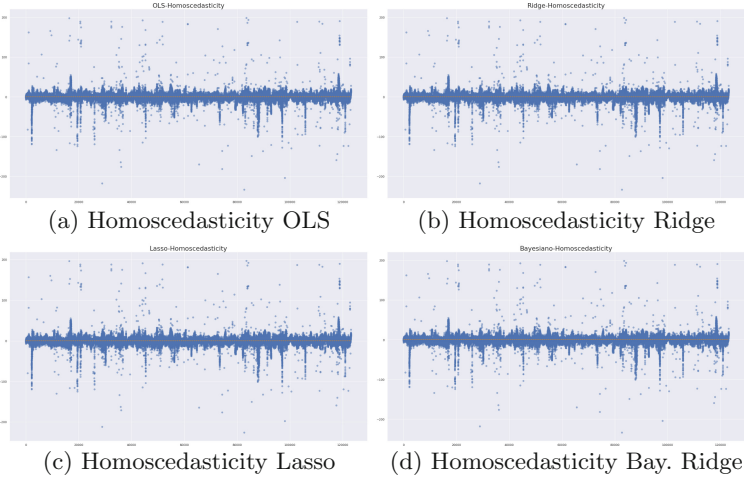
**Fig. 3.** RFE-Residual Distribution, (a) Correspond to model OLS, (b) Correspond to model Ridge (c) Correspond to model Lasso, (d) Correspond to model Bayesian Ridge

### 6.3 No Autocorrelation of the Error Terms

Autocorrelation indicates that some information is missing that the model should capture. It would be represented by a systematic bias below or above the prediction. For this we will use the Durbin-Watson test. Value from 0 to 2 is positive autocorrelation and value from 2 to 4 is negative autocorrelation. For RFE - OLS there is no autocorrelation. Durbin-Watson Test is 2.0037021333412754, little to no autocorrelation. For RFE - Bayesian Ridge there is no autocorrelation. Durbin-Watson Test is 2.0037008807358965, little to no autocorrelation. For RFE - Lasso there is no autocorrelation. Durbin-Watson is 2.0037472224605053, little to no autocorrelation. Have a For RFE - Ridge there is no autocorrelation. Durbin-Watson is 2.0037017639830537, little to no autocorrelation.

### 6.4 Homocedasticity

It must be fulfilled that the error made by the model always has the same variance. It is presented when the model gives too much weight to a subset of data, particularly where the variance of the error was the greatest: to detect it, residuals are plotted to see if the variance is uniform (Fig. 4).



**Fig. 4.** RFE-Homoscedasticity, (a) Correspond to model OLS, (b) Correspond to model Ridge (c) Correspond to model Lasso, (d) Correspond to model Bayesian Ridge

## 7 Description and Analysis of the Results

In this article we present three hybrid methods for the selection of variables in the multiparameter regression of photovoltaic systems to predict the levels of the active power of the photovoltaic system with 14 independent variables, these methods are RFE - Lasso, RFE - Ridge and RFE - Bayesian Ridge.

Table 3 and Table 4 shows the method comparison, RFE-OLS, which is not part of our proposal, was compared with OLS to have a benchmark for the following comparisons that are part of the proposal. RFE-Lasso: it has an absolute error of approximately 0.035% greater than Lasso, which is taken as a disadvantage of the proposal, it has a mean squared error of approximately 0.057% less than Lasso, which is a significant result considered as an advantage, it has a coefficient of determination of approximately 0.0000309% higher than Lasso, this is considered greater but almost the same, so it is not considered very advanta-

**Table 3.** RFE with OLS and Shrinkage

Comparison of methods (%)	RFE-OLS vs OLS	RFE-Lasso vs Lasso
Mean absolute error R	0.003062198	0.034722952
Square root error $R^2$	-0.068102080	-0.056497772
Determination coefficient	0.000037258	0.000030848
Adj coef of determination	0.000037928	0.000031516
Training time	-37.027209963	-30.903451764
Test time	13.394018205	4.161073826

**Table 4.** RFE with OLS and Shrinkage

Comparison of methods (%)	RFE-Ridge vs Ridge	RFE-Bayesian Ridge vs Bayesian Ridge
Mean absolute error R	0.003071248	0.003841610
Square root error $R^2$	-0.068083868	-0.067919409
Determination coefficient	0.000037248	0.000037158
Adj coef of determination	0.000037918	0.000037828
Training time	-24.017312331	-32.443632788
Test time	2.165087957	16.901408451

geous but in no way a disadvantage of the proposal, it has an adjusted coefficient of determination of approximately 0.0000315% greater than Lasso, this is considered an advantage as the previous case, the training time is approximately 30,904% less, which is considered a great contribution of this hybrid method, the test time is approximately 4,161% greater than Lasso, which is considered a disadvantage of the proposed model. For the following two hybrid RFE-Ridge methods compared to Ridge and RFE-Bayesian Ridge compared to Bayesian Ridge, and shown in Table 3 and Table 4. The description and analysis is similar to RFE-Lasso compared to Lasso.

## 8 Conclusions

The selection of independent variables of the multi-parameter photovoltaic system allowed us to develop four prediction models with an accuracy greater than 99.97% in all cases. Three RFE proposals are presented, RFE-Ridge, RFE-Lasso, and RFE-Bayesian Ridge; training time was reduced by 71% for RFE-Ridge over Ridge and and 36% RFE-OLS over OLS. The variables eliminated with RFE-Ridge and RFE-Bayesian Ridge were: “Energia total”, “Energia diaria” e “Irradiancia”, and additionally the variable eliminated by RFE-Lasso was: “Frecuencia”. In all cases we see that the root mean square errors were reduced for RFE-Lasso by 0.15% over Lasso while for RFE-Bayesian Ridge by 0.06% over Bayesian Ridge. From all that has been done, we note that the proposed hybrid method, by eliminating variables that are not significant for the system, achieves a decrease in training times, without losing accuracy in predictions. The results can be improved by implementing algorithms in pre-processing stages such as imputation of values; or perform techniques as linear regression such, neural networks or XGBoost.

## References

1. Çürük, E., Acı, Ç., Saraç Eşsiz, E.: The effects of attribute selection in artificial neural network based classifiers on cyberbullying detection. In: 2018 3rd International Conference on Computer Science and Engineering (UBMK), Sarajevo, pp. 6–11 (2018). <https://doi.org/10.1109/UBMK.2018.8566312>

2. Wu, Q., Zhang, H., Jing, R., Li, Y.: Feature selection based on twin support vector regression. In: 2019 IEEE Symposium Series on Computational Intelligence (SSCI), Xiamen, China, pp. 2903–2907 (2019). <https://doi.org/10.1109/SSCI44817.2019.9003001>
3. Zheng, Z., Cai, Y., Yang, Y., Li, Y.: Sparse weighted Naive Bayes classifier for efficient classification of categorical data. In: IEEE Third International Conference on Data Science in Cyberspace (DSC), Guangzhou, pp. 691–696 (2018). <https://doi.org/10.1109/DSC.2018.00110>
4. Al-Fahad, R., Yeasin, M., Anam, A.I., Elahian, B.: Selection of stable features for modeling 4-D affective space from EEG recording. In: International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, pp. 1202–1209 (2017). <https://doi.org/10.1109/IJCNN.2017.7965989>
5. Tanizaka Filho, M.O., Marujo, E.C., dos Santos, T.C.: Identification of features for profit forecasting of soccer matches. In: 8th Brazilian Conference on Intelligent Systems (BRACIS), Salvador, Brazil, pp. 18–23 (2019). <https://doi.org/10.1109/BRACIS.2019.00013>
6. Qin, H., Yin, S., Gao, T., Luo, H.: A data-driven fault prediction integrated design scheme based on ensemble learning for thermal boiler process. In: 2020 IEEE International Conference on Industrial Technology (ICIT), Buenos Aires, Argentina, pp. 639–644 (2020). <https://doi.org/10.1109/ICIT45562.2020.9067216>
7. Moghimi, A., Yang, C., Marchetto, P.M.: Ensemble feature selection for plant phenotyping: a journey from hyperspectral to multispectral imaging. *IEEE Access* **6**, 56870–56884 (2018). <https://doi.org/10.1109/ACCESS.2018.2872801>
8. Ashfaq, T., Javaid, N.: Short-term electricity load and price forecasting using enhanced KNN. In: 2019 International Conference on Frontiers of Information Technology (FIT), Islamabad, Pakistan, pp. 266–2665 (2019). <https://doi.org/10.1109/FIT47737.2019.00057>
9. Xiao, W., et al.: Single-beat measurement of left ventricular contractility in normothermic ex situ perfused porcine hearts. *IEEE Trans. Biomed. Eng.* (2020). <https://doi.org/10.1109/TBME.2020.2982655>
10. Elhamifar, E., De Paolis Kaluza, M.C.: Subset selection and summarization in sequential. In: Guyon, I., et al. (eds.) *Advances in Neural Information Processing Systems* 30, pp. 1035–1045. Curran Associates Inc. (2017). <http://papers.nips.cc/paper/6704-subset-selection-and-summarization-in-sequential-data.pdf>
11. Svetnik, V., Liaw, A., Tong, C., Wang, T.: Application of Breiman’s random forest to modeling structure-activity relationships of pharmaceutical molecules. In: Roli, F., Kittler, J., Windeatt, T. (eds.) *MCS 2004*. LNCS, vol. 3077, pp. 334–343. Springer, Heidelberg (2004). [https://doi.org/10.1007/978-3-540-25966-4\\_33](https://doi.org/10.1007/978-3-540-25966-4_33)