Title: Comparison of artificial intelligence algorithms to estimate sustainability indicators

Abstract: The monitoring of sustainability indicators allows behavioural tendencies of a region to be controlled, so that adequate policies could be established in advance for a sustainable development. However, some data could be missed in the monitoring of these indicators, thus making the establishment of sustainability policies difficult. This paper therefore analyses the possibility to forecast the sustainability indicators of a region by using four different artificial intelligent algorithms: linear regression, multilayer perceptron, random forest, and M5P. The study area selected was the Algarve region in Portugal, and 180 monitored indicators were analysed between 2011 and 2017. The results showed that M5P is the most appropriate algorithm to estimate sustainability indicators. M5P was the algorithm obtaining the best estimations in a greater number of indicators. Nevertheless, the results showed that MP5 was not the best option for all indicators, since in some of them, the use of other algorithms obtained better results, thus reflecting the need of an individual previous study of each indicator. With these algorithms, it is possible for public bodies and institutions to evaluate the sustainable development of the region and to have reliable information to take corrective measures when needed, thus contributing to a more sustainable future.

Keywords:

Artificial intelligence; sustainability indicators; OBSERVE platform; data mining; monitoring process

1. Introduction

The development of regions is an important aspect of the political strategy and nowadays the main aim is to achieve a more sustainable development (Akande et al., 2019). In this regard, one of the activities influencing a good sustainability level is the tourist activity (Asmelash & Kumar, 2019; Liao & Chern, 2015; Moussiopoulos et al., 2010). Such activity is constantly increasing (Hatipoglu et al., 2016), which is reflected by job creation (Rylance, 2012) and the increase of the Gross Domestic Product (GDP) (Pérez-Rodríguez et al., 2015). In addition, the tourist activity has contributed to the restoration of the historical heritage (Almeida et al., 2018) as it is a great attraction of tourists. The tourist activity, however, implies a huge consumption of energy and production of greenhouse gas emissions which strongly affect the ecological footprint (Castellani & Sala, 2012; W. Lin et al., 2018) and climate changes (Robaina-Alves et al., 2016; S. Wang et al., 2019). Tourism represents 4.9% of carbon dioxide emissions, with an annual increase of 2.5% (Whittlesea & Owen, 2012). Other aspects, such as the consumption of raw materials to make tourist products with a short useful life (He et al., 2018), also generate a significant impact, thus implying that tourism has a direct repercussion on some sectors, e.g., the energy (Rizzo, 2017) or the food sector (Pérez Gálvez et al., 2017), and generating a pressure on the environment of the region (Feleki et al., 2018; Michailidou et al., 2015).

The tourist activity should therefore be managed in a broader sustainable tourism context (Higgins-Desbiolles, 2018). The term sustainable tourism has different meanings according to the specific characteristics of each region (Lu & Nepal, 2009). The definition by the World Tourism Organization (UNWTO) and the United Nations Environment Programme (UNEP) is usually considered the most representative (UNWTO and UNEP, 2005): "tourism that takes full account of its current and future economic, social and environmental impacts, addressing the needs of visitors, the industry, and the environment and host communities". Sustainable tourism is consequently used to balance environmental, economic, and social dimensions (Liu et al., 2013) to clearly improve the life quality of people (Lozano, 2012), the economic advancement of the activities related to the sector (Lane, 2018) and the improvement of competitiveness (Crouch, 2011; Pulido-Fernández et al., 2019). Users recognise the improvement of sustainability thanks to the great deal of information being updated in the internet (F. Wang et al., 2020).

The tourist development of a region should be improved by local governments to ensure the achievement of the United Nations Sustainable Development Goals (United Nations General Assembly, 2015). However, there are some cases in which local governments have detected conflicts with the tourist sector (Kapera, 2018). For this reason, the monitoring of sustainability indicators allows behavioural tendencies of a region to be controlled, so that adequate policies could be established for a sustainable development (Hermans et al., 2011; Verma & Raghubanshi, 2018). Sustainability indicators are variables used to know the sustainability degree of a region (Manning, 1996), and such variables should be quantitative to carry out objective assessments (Michael et al., 2014). There are many typologies of indicators varying according to the region (Kristjánsdóttir et al., 2018) and the goals to achieve.

For the Algarve region (Figure 1), sustainability indicators can be monitored by OBSERVE - Observatory of Sustainability of the Algarve Region for Tourism (https://observe.ualg.pt/) (Farinha et al., 2019). Algarve is the region of Portugal located further south of the Iberian Peninsula, with approximately 200 km of coastline. Despite its population represents 5% of the total population in the country (Instituto Nacional de Estatística, 2019), such population is tripled in the hottest periods, so that tourism is among the main activities of the region (Coelho et al., 2006; Ramos, 2009). However, this tendency is not the same in the 16 municipalities of the region (Figure 1) as there are differences between coastal areas (with a greater tourist activity) and the interior (less population and a predominant activity in primary sectors) (Mascarenhas et al., 2014). Regardless of these differences among municipalities, the tourist activity generates a high pressure on the area of conservation of biodiversity, considering that 33% of the territory is included in the network Natura 2000 (Mascarenhas et al., 2010). Consequently, development plans of the Algarve region should be focused on a greater sustainability (Mascarenhas et al., 2015).

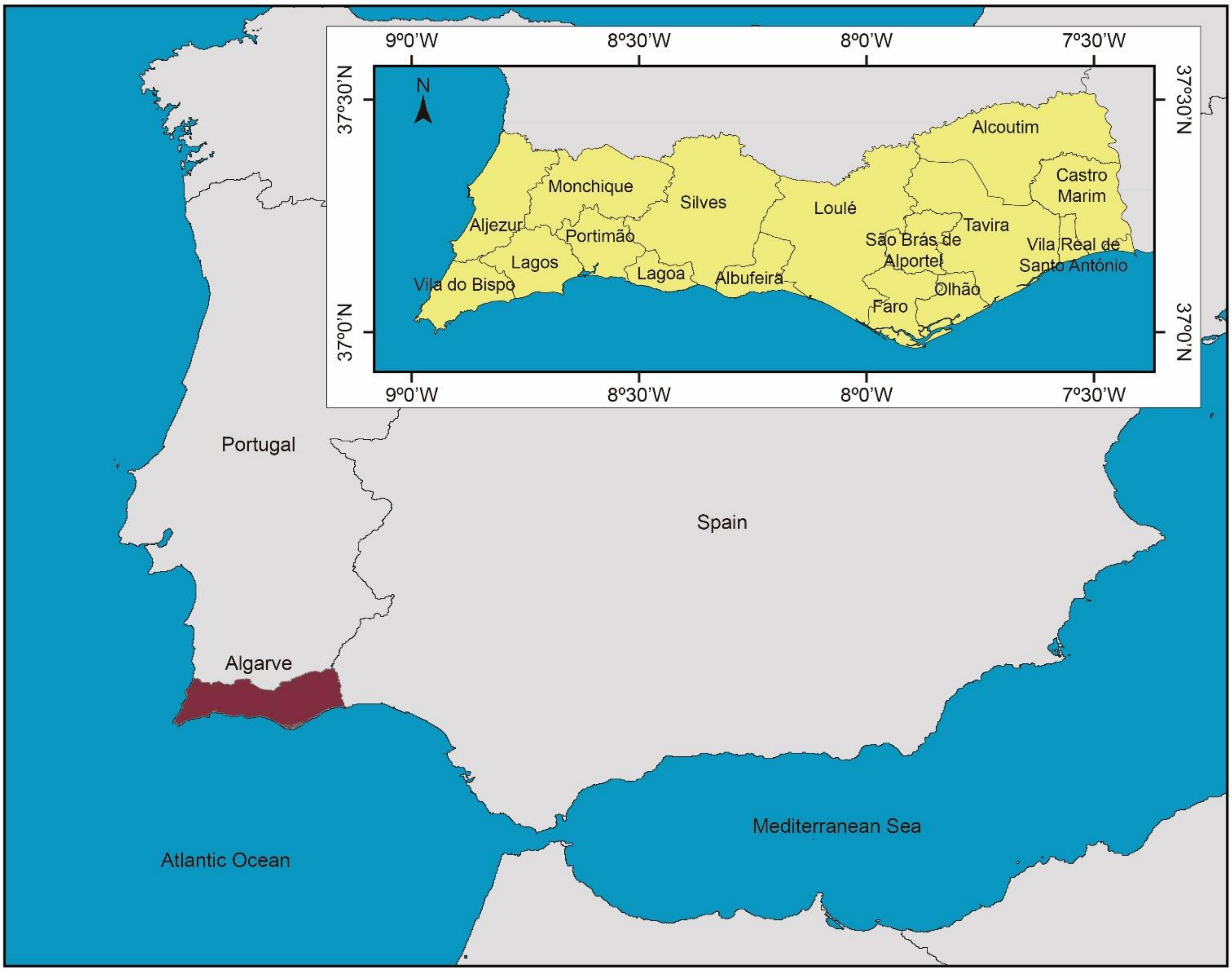

Figure 1. The Algarve region and the sixteen municipalities.

The OBSERVE platform has 65 indicators divided into 4 dimensions (Figure 2): environmental, institutional, economic, and sociocultural. The indicators, for each dimension, were chosen after meeting and surveying different stakeholders from the region, such as the Associação dos Industriais Hoteleiros, Restauração e Bebidas (in English, Tourism Industrial Association) or the Agência Portuguesa do Ambiente (in English, Environment Portuguese Agency) (Farinha et al., 2019). consequently, the indicators which were extremely important to be monitored by OBSERVE were selected. Therefore, all the indicators monitored by OBSERVE are important indicators for the stakeholders in the region.

OBSERVE constitutes an important opportunity for local organisations to assess the sustainability of the region. However, one of the limitations of this monitoring project (which could be extrapolated to other regions) is the lack of values recorded in some of the indicators. In this regard, full data from some years is sometimes not available in the monitoring processes. So, having statistical techniques and methodologies to estimate the values of the indicators which were not monitored (i.e., to fill the gaps) guarantees a greater effectiveness of the decision-making.

Here is where artificial intelligence could be useful to estimate missing values. There are various research studies assessing the possibilities of using artificial intelligence for the statistical treatment of sustainability indicators. Some relevant examples are: (i) Zhang et al. (2015) assessed the tourism sustainability in the Tibet Autonomous Region by using neural networks; (ii) Wu et al. (2019) used neural networks to forecast the ecological footprint and the ecological capacity of the urban development in Tianjin (China); (iii) D'Amico et al. (2019) used neural networks to forecast the energy and environmental behaviour of Italian buildings; and (iv) Antanasijević et al. (2013) developed artificial neural networks to forecast missing data of municipal waste generation in developing countries.

Nevertheless, none of these studies used various regression algorithms to assess the possibilities (most studies used artificial neural networks without analysing other algorithms). This research therefore suggest using 180 sustainability indicators monitored by the OBSERVE platform between 2011 and 2017: models were trained with the yearly data compiled between 2011 and 2016, and these models estimated the values of indicators in the year 2017. For this purpose, training datasets were designed based on an algorithm of input attribute selection in WEKA, and prediction models were trained for each output indicator by using four different algorithms: multiple linear regression, multilayer perceptron, random forest, and M5P. The results showed the possibilities to estimate sustainability indicators and the most appropriate methodology to be used. In addition, this approach could be extrapolated to other indicators or regions where sustainability indicators are monitored.

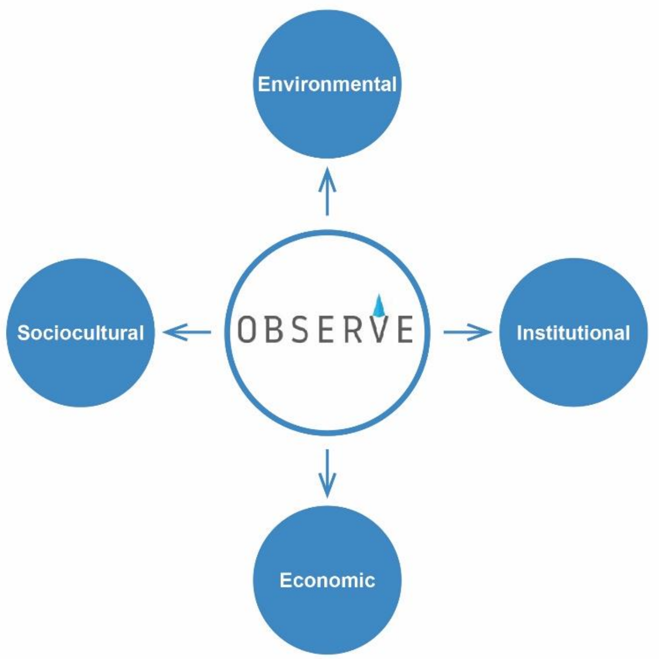

Figure 2. The 4 dimensions of the sustainability indicators of the OBSERVE platform.

2. Methodology

The methodological framework of this research has five main phases: (i) definition of the characteristics of the regression algorithms used in the research; (ii) definition of the output variables intended to be estimated; (iii) definition of the procedure to select attributes; (iv) training of the networks by using four regression machine learning algorithms; and (v) test of the models developed.

2.1. Regression machine learning algorithms

This section describes briefly the four regression machine learning algorithms considered in this study: multiple linear regression, multilayer perceptron, random forest, and M5P.

2.1.1. Multiple linear regression

The multiple linear regression (MLR) is a classical regression algorithm, which consists in connecting independent variables through regression coefficients to obtain the value of the output variable by their sum (Eq. (1)). The MLR algorithm has several advantages (Pino-Mejías et al., 2017), namely: possibility of being adjusted over the transformations of the variables, interpretability, simplicity, supposing the hypothesis of normality, homoscedasticity and intercorrelation between the error ε and the predictor variables.

$$\hat{Y}_{MLR} = \beta_0 + \sum_{i=1}^{v}(\beta_i x_i) + \varepsilon \quad (1)$$

where β_0 is the independent term, β_i are the regression coefficients, x_i are the predictor variables, and ε is the error.

2.1.2. Multilayer perceptron

Neural networks are bioinspired statistical models simulating the neurological brain structure to solve regression and classification problems (Haykin et al., 2009). Multilayer perceptrons (MLPs) are the artificial neural networks offering the best features due to their capacities of universal approximation (Barron, 1993; Cybenko, 1989; Hornik et al., 1989). MLPs are characterised by having an architecture of three or more layers, with a series of nodes or neurons in each (Figure 3 (a)):

(i) an input layer whose nodes correspond to the different input variables considered for the model; (ii) one or several intermediate layers with interconnected nodes; (iii) an output layer corresponding to the output variable (or dependent variable) whose value is obtained by summing the values of the input neurons weighted by synaptic weights and applying an activation function:

$$\hat{Y}_{MLP} = \sigma\left(\sum_{k=1}^{M} w_{lk}^{(2)} \sigma\left(\sum_{j=0}^{d} w_{kj}^{(1)} x_j\right) + w_{l0}^{(2)} y_0\right) \tag{2}$$

where $\hat{Y}_{MLP}$ is the estimation conducted by the MLP, x_j are the values of the input layer, $w_{k0}^{(1)}$ and x_0 are the weight and the input value of the bias neuron of the input layer, $w_{kj}^{(1)}$ are the weights of the hidden layer, $w_{l0}^{(2)}$ and y_0 are the weight and the input value of the bias neuron of the hidden layer, $w_{lk}^{(2)}$ are the weights of the output layer, y_k is the output of a neuron of the hidden layer, and σ is the activation function.

For this research, models with a hidden layer were considered and a sigmoidal activation function was used in the hidden layer and in the output layer (Eq. (3)), similarly to other studies in which such models were applied (Bienvenido-Huertas et al., 2019), since their performance is better than that of more complex structures (Kumar et al., 2013). The correct number of nodes from the hidden layer was assessed by analysing the error associated with the training and testing of the models. To do this, the number of neurons ranged from 2 to 16.

$$\sigma = \frac{1}{1 + e^{-x}} \tag{3}$$

As mentioned above, the output value is obtained from the weighted propagation of the input signs. One of the most important aspects of MLPs is therefore the adjustment of the synaptic weights reducing the error between estimations and actual values. For this purpose, the models were trained through backpropagation (Rumelhart et al., 1986; Y. N. Wang, 1994; Werbos, 1974), using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) (Fletcher, 1980) algorithm (which belongs to quasi-Newton methods), due to the high accuracy achieved in the results of other studies (Ahmad et al., 2017; Golbabai & Seifollahi, 2007; Pino-Mejías et al., 2018).

2.1.3. Random forest

Random forest (RF) is an evolution of the classification and regression trees (CART) algorithm (Breiman, 1996, 2001), so understanding how CART work is crucial.

The CART algorithm develops reverse tree models whose internal nodes correspond to the input variables, arches correspond to the values of the root node and are connected to other nodes or leaves, and leaves correspond to the value of the model. These models develop a series of if-then rules which, following the rules indicated in each node, lead to the output value. They are, therefore, characterised by dividing the input space into subregions, simplifying complex problems with simple models (Sun, 2018). It is important to stress that the output value included in each leaf is a unique numeric value, so no equation is included to obtain the response of the CART model.

Thus, CART models are easy to understand the solution adopted for the problem (Xu et al., 2005), so many research studies have applied them (Mousa et al., 2017; Tso & Yau, 2007; Williams & Gomez, 2016).

However, the use of this algorithm is limited to different applications (Dudoit et al., 2002; Larivière & Van Den Poel, 2005). Due to this circumstance, RF allows a more robust application than CART models as RF develops a set of CART models (i.e., a forest of tree models) (Figure 3 (c)), which reduce the variance (Breiman, 1996, 2001) and the influence of atypical values (Assouline et al., 2018). It is an ensemble learning algorithm, so a better behaviour is obtained than that with an individual model (Dietterich, 2000).

To train RF, N bootstrapped sample sets are obtained from the training dataset (Breiman, 2001). Each bootstrapped sample generates a CART model. Also, each node of each tree is divided by using a subset of m predictors randomly selected, thus reducing the influence of the strongest predictors (Rodriguez-Galiano et al., 2015). The model is estimated by the average of the output value of the CART set (Eq. (4):

$$\hat{Y}_{RF} = \frac{1}{T}\sum_{t=1}^{T} \hat{Y}_t \tag{4}$$

where $\hat{Y}_{RF}$ is the estimation of the RF model, T is the number of trees, and $\hat{Y}_t$ is the estimation of the t-th CART model. As can be seen, the number of trees affects the result obtained. In general, when a certain number of trees is overcome, the model always obtains the same output. Determining the limit number of trees is fundamental to reduce the time required to train RF models. For this reason, the optimal number of trees was assessed in all the RF models developed in this research.

2.1.4. M5P

The M5P algorithm is another evolution of the CART algorithm (Quinlan & others, 1992; Y. Wang & Witten, 1997). The main difference with respect to CART is that M5P develops tree models whose leaves are MLR models (Figure 3 (b)). Unlike RF models, a unique tree model is developed. The algorithm, therefore, works by developing optimal MLR models in the various subregions or divisions made by the dataset. It is also an algorithm from which the rules established between the different variables of the dataset are known. Another advantage of the M5P is that it effectively uses big datasets, which are robust due to the lack of values in the observations of the dataset analysed (Behnood et al., 2017; L. Lin et al., 2016).

In the development process of the M5P model, instead of maximising the information gain, the internal variation of the subsets for the class values of each branch is minimised. After building the model, the pruning (i.e., the removal of inefficient nodes) reduces the overfitting (Rodriguez-Galiano et al., 2015).

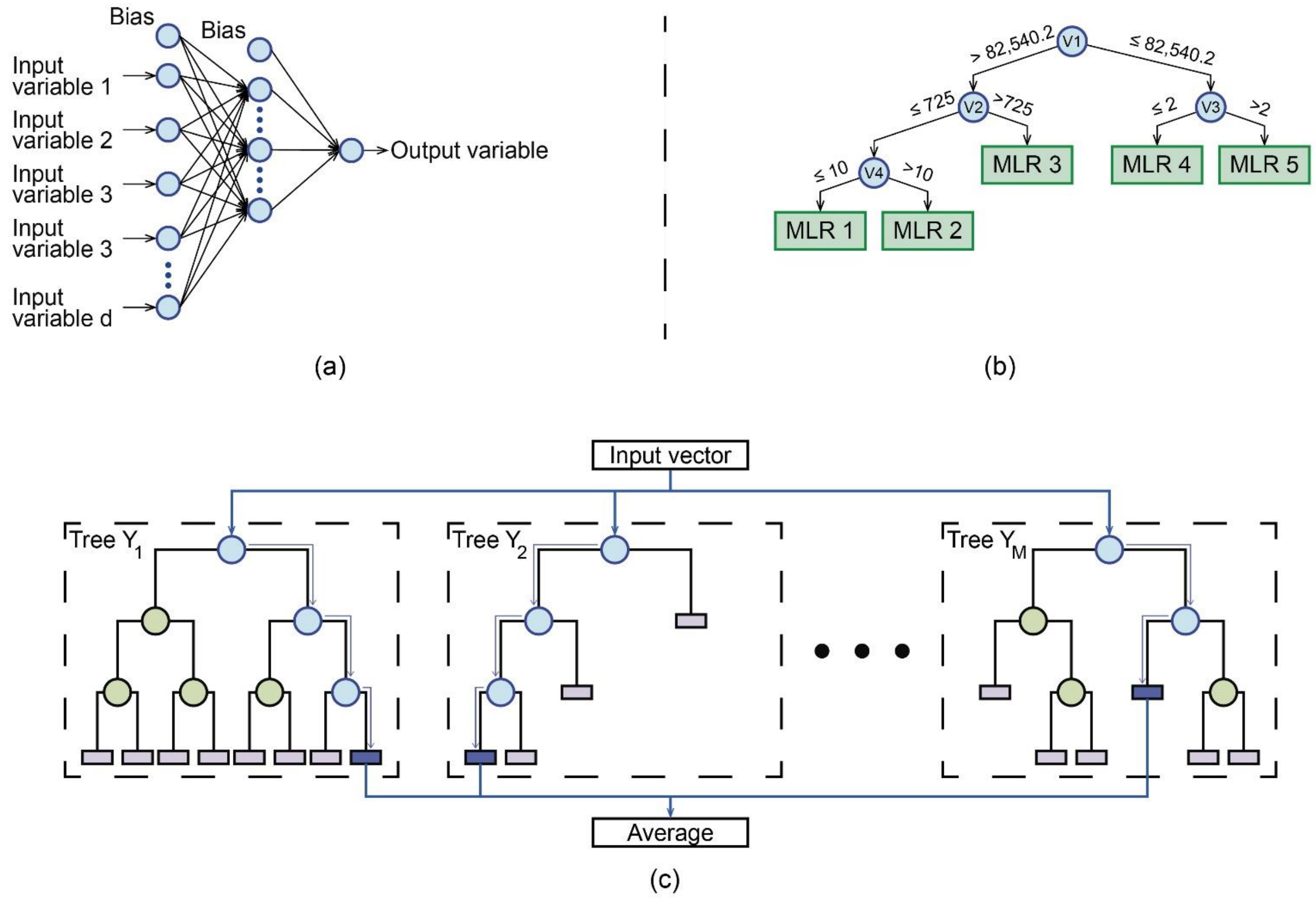

Figure 3. Algorithms´ schemes: (a) MLP model, (b) M5P model, and (c) RF model.

2.2. Training and validation procedures

For this research, the database of sustainability indicators compiled by the OBSERVE platform was analysed. A total of 43 sustainability indicators had yearly data between 2011 and 2017 (Table 1). So, this set of indicators were used. It is also important to emphasise that most of indicators had various subcategories, thus increasing the number of values to be estimated. Table 2 shows that the actual number of indicators to be estimated was 180. It is important to stress that the number of indicators monitored by OBSERVE is greater than 43 (OBSERVE monitors 65 indicators) and, therefore, greater than the 180 sub-indicators analysed in the research. Thus, not all sustainability indicators were used for the purpose of this research. The reason was that the remaining 22 indicators lacked some annual data or began to be monitored after 2011. So, they were not used to assess the suitability of applying the regression algorithms to estimate missing data, since actual values are required to evaluate the error associated with the estimations. Likewise, this aspect reflects the need to analyse the objective of this study as sustainability indicators data may be lacking in monitoring.

By using these data, a total of 180 regression models were developed for each indicator. As the 180 models were developed by each type of algorithm, the total number of models developed in this research was 720. Figure 4 sums up the workflow of the training and test processes.

Input variables were defined in the development of the dataset used for each sustainability indicator (output variable). For this purpose, the selection process of input variables included in WEKA was used (Yadav et al., 2014). In particular, the attribute evaluator CorrelationAttributeEval was used as it evaluates the most suitable input variables by measuring the Pearson correlation between them and the output variable. A total of 10 input variables were defined by this process for each output variable. After defining the structure of the 180 datasets, the training and testing subsets were defined: the training subset was made up of the data compiled between 2011 and 2016, and the testing subset of the data compiled in 2017. As actual data of the year intended to be estimated was available, the error associated to estimations could be assessed. For this purpose, the mean absolute percentage error (MAPE) was used in this study as a statistical parameter to assess the error (Eq. (5)). By using the MAPE's assessment, the accuracy of the estimations conducted by the algorithms and the most appropriate approach were obtained.

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{a_i - e_i}{a_i}\right| \qquad (5)$$

where n is the number of instances in the testing subset (in this study, it is a unique instance per each testing subset), a_i is the actual value of the indicator, and e_i is the value predicted by the model.

Table 1. Sustainability indicators of the OBSERVE platform used in the research.

Nº	Dimension	Subject area	Description
01	Environment	Environmental Management	Environmental expenditure of municipalities by 1000 inhabitants
02	Environment	Mobility	Number of embarked and disembarked passengers in Faro Airport
03	Environment	Mobility	Number of passengers-kilometre carried by enterprises exploring inland transportation
04	Environment	Mobility	Movement of passengers in inland waterways
05	Environment	Energy Management	Consumption of electric energy by inhabitant
06	Environment	Energy Management	Consumption of motor fuel by inhabitant
07	Environment	Water Cycle Management	Percentage of safe water
08	Environment	Environmental management	Number of blue flags beaches
09	Environment	Water Cycle Management	Fresh water supplied per inhabitant
10	Environment	Water Cycle Management	Wastewater sewerage per capita
11	Environment	Materials and Waste Management	Urban waste selectively collected per inhabitant
12	Environment	Natural Capital Management	Burnt area
13	Environment	Environmental Management	Number of bathing waters and quality classes
14	Environment	Natural Capital Management	Investments on protection of biodiversity and landscapes of municipalities
15	Environment	Territory Management	Percentage of reconstructed total area
16	Environment	Mobility	Number of embarked and disembarked passengers of cruise ships in Portimão port
17	Institutional	Governance and Citizenship	Percentage of capital expenditure
18	Institutional	Governance and Citizenship	Broadband internet accesses per 100 inhabitants
19	Institutional	Innovation and Knowledge	Gross expenditure on research and development of institutions and enterprises
20	Economic	Economic Impact	Gross value added (GVA) of enterprises
21	Economic	Tourist Occupation	Nights in hotel establishments
22	Economic	Tourist Occupation	Revenue per available room (RevPAR) of hotel establishments
23	Economic	Tourist Occupation	Average stay in hotel establishments
24	Economic	Economic Impact	Apparent labour productivity in establishments, food and beverage service activities
25	Economic	Economic Impact	Inflation
26	Economic	Economic Impact	Number of establishments and economic activity
27	Economic	Economic Impact	Persons employed of establishments and economic activity
28	Economic	Economic Impact	Turnover of establishments and economic activity
29	Economic	Job	Employment by gender and economic sector
30	Economic	Economic Impact	Relative contribution of establishments, food and beverage service activities to the Algarve economy (GVA per Enterprises)
31	Sociocultural	Demography	Annual population balances: natural and migratory
32	Sociocultural	Culture	Number of cultural properties
33	Sociocultural	Health Care	Health care
34	Sociocultural	Safety	Crime rate
35	Sociocultural	Safety	Number of registered crimes
36	Sociocultural	Social Cohesion	Regional development composite index
37	Sociocultural	Social Cohesion	Beneficiaries of social integration income per 1000 inhabitants in working age
38	Sociocultural	Demography	Resident population
39	Sociocultural	Demography	Foreign population with status of resident
40	Sociocultural	Pressure	Lodging capacity in hotel establishments by 1000 inhabitants
41	Sociocultural	Education	Population education level with 15 and more years
42	Sociocultural	Pressure	Regional tourist density
43	Sociocultural	Pressure	Municipal tourist density

Table 2. Sustainability sub-indicators used in the research.

Indicator	Sub-indicator	Nº.
01	Total; Waste management; Noise and vibration abatement; Protection of biodiversity and landscape; Research and development; Others	6
02	Total; Embarked; Disembarked	3
03	Total; Rail; Road	3
04	Total; Ria Formosa; Rio Guadiana	3
05	Total	1
06	Total	1
07	Total	1
08	Total	1
09	Total	1
10	Total	1
11	Total	1
12	Total; Forest stands; Shrub land; Agricultural area	4
13	Total; Inland; Coastal/transition	3
14	Total; Prevention against forest fires; Others	3
15	Total	1
16	Total; Embarked; Disembarked; Transition	4
17	Total	1
18	Residential; Non residential	2
19	State; Enterprises; Higher education; Private non-profit institutions	4
20	Total; Agriculture, farming of animals, hunting and forestry; Mining and quarrying; Manufacturing; Electricity, gas, steam, cold and hot water and cold air; Water collection, treatment and distribution, sewerage, waste management and remediation activities; Construction; Wholesale and retail trade, repair of motor vehicles and motorcycles; Transportation and storage; Accommodation and food service activities; Information and communication activities; Real estate activities; Consultancy, scientific and technical activities; Administrative and support service activities; Education; Human health and social work activities; Arts, entertainment, sports and recreation activities; Others	18
21	Hotels; Apartment hotels; Tourist villages; Tourist apartments	4
22	Total; Hotels; Guest houses; Lodging houses; Hotel apartments; Tourist villages; Tourist apartments	7
23	Total	1
24	Total; Accommodation and food service activities	2
25	Total; Total excluding housing; Total excluding unprocessed food and energy; Total excluding unprocessed food; Total excluding energy; Unprocessed food; Energy	7
26	Agriculture, farming of animals, hunting and forestry; Mining and quarrying; Manufacturing; Electricity, gas, steam, cold and hot water and cold air; Water collection, treatment and distribution, sewerage, waste management and remediation activities; Construction; Wholesale and retail trade, repair of motor vehicles and motorcycles; Transportation and storage; Accommodation and food service activities; Information and communication activities; Real estate activities; Consultancy scientific and technical activities; Administrative and support service activities; Education; Human health and social work activities; Arts, entertainment, sports and recreation activities; Others	17
27	Agriculture, farming of animals, hunting and forestry; Manufacturing; Electricity, gas, steam, cold and hot water and cold air; Water collection, treatment and distribution, sewerage, waste management and remediation activities; Construction; Wholesale and retail trade, repair of motor vehicles and motorcycles; Transportation and storage; Accommodation and food service activities; Information and communication activities; Real estate activities; Consultancy scientific and technical activities; Administrative and support service activities; Education; Human health and social work activities; Others	15
28	Total; Agriculture, farming of animals, hunting and forestry; Manufacturing; Electricity, gas, steam, cold and hot water and cold air; Water collection, treatment and distribution, sewerage, waste management and remediation activities; Construction; Wholesale and retail trade, repair of motor vehicles and motorcycles; Transportation and storage; Accommodation and food service activities; Information and communication activities; Real estate activities; Consultancy scientific and technical activities; Administrative and support service activities; Education; Human health and social work activities; Others	16
29	Total-Gender; Men-Gender; Women-Gender; Total-Sector; Agriculture, forestry and fishing-Sector; Industry-Sector; Services-Sector	7
30	Accommodation and food service activities; Others	2
31	Natural increase; Net migration	2
32	Monuments; Sets; Sites	3
33	Beds; Doctors; Nurses	3
34	Crimes of assault; Theft/purse snatching; Theft of and from motor vehicles; Driving a motor vehicle with a blood alcohol equal or above; Driving without legal documentation; Crimes against patrimony	6
35	Crimes against persons [except voluntary manslaughter]; Crimes of voluntary manslaughter; Crimes against patrimony; Crimes against life in society; Crimes against the State; Crimes set out in sundry legislation	6
36	Total	1
37	Total	1
38	Total; Men; Women	3
39	Men; Women	2
40	Total	1
41	Total; Men; Women	3
42	Hotels; Boarding houses; Inns; Lodging houses; Apartment hotels; Tourist villages; Tourist apartments	7
43	Portugal; Other countries	2

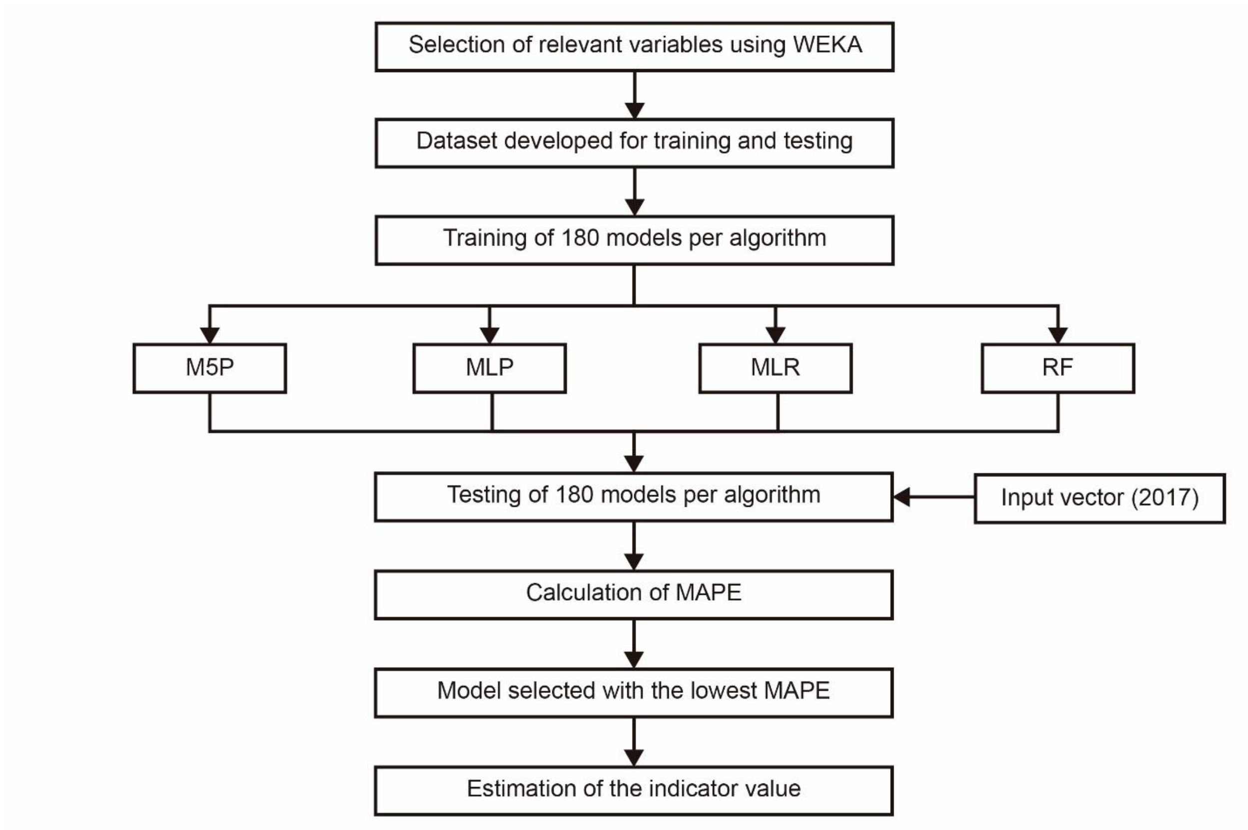

Figure 4. Workflow of the training and test processes.

3. Results and discussion

First, the 180 models of each algorithm were trained. For this purpose, the WEKA algorithm was applied to optimally select the 10 input attributes of each model. By selecting these 10 most suitable input variables to estimate the 180 sustainability indicators, the training datasets were designed to train the models. After training the models, the model was tested by using values from the 180 indicators in the year 2017. The assessment of MAPE determined the robustness of the estimations conducted. It is important to remember that, in the case of MLPs, the optimal number of nodes was determined. The analysis was performed by evaluating the optimal number of nodes between 2 and 16, and that with the lowest error in both training and testing phases was considered as the optimal number. Figure 5 includes the optimal case numbers obtained for each number of nodes in the hidden layer. As it is shown, the optimal number of nodes varied between 2 and 13 depending on the sustainability indicator analysed. In this regard, many indicators obtained the best performance with simple 2-node architectures in the hidden layer, although there were a high number of indicators in which more complex architectures (between 8 and 13 nodes) were the most appropriate. These results therefore showed the need for an individual analysis of the most appropriate MLP architecture to estimate each sustainability indicator.

After determining the most appropriate MLP architectures, the results of the estimations obtained with both the MLP and the other algorithms were analysed. To provide readers with a summarised information of the quality of the estimations, Figure 6 includes the histogram of the MAPEs obtained in the estimations of M5P, MLP, MLR, and RF. Likewise, Annex A includes the MAPE values obtained in each indicator. The models developed with each algorithm made adjusted estimations in many indicators, although different patterns were found in the quality of the estimations. For high MAPE values (in this case, values greater than 20% were considered), RF, MLP, and M5P obtained a similar number of indicators (25, 18, and 22, respectively), whereas MLR obtained a greater number of indicators (69 indicators). In addition, MLR was the algorithm obtaining an estimation with greater MAPE in the indicator I12 (Agricultural area), with a value of 1,316.29%. The other algorithms also made estimations with high maximum percentage deviations (276.85, 474.40, and 827.36% for RF, MLP, and M5P, respectively). M5P was the second algorithm with the highest MAPE value. However, this was the case of the same indicator with which a greater MAPE was obtained with MLR (I12 (Agricultural area)), which indicates the possible estimation limitation of this indicator. Also, M5P was the algorithm characterised by obtaining a larger number of estimations with a lower MAPE (Figure 7). In this regard, M5P was the best option in 75 out of the indicators analysed and was the algorithm obtaining the lowest number indicators with the worst values. These aspects show the greatest robustness of M5P to make adjusted estimations, which is reflected in Figure 6 by analysing the density of indicators per ranges of 1%. In this regard, M5P obtained a total of 129 indicators with a MAPE lower than 5%.

Despite the robustness of most estimations carried out with M5P, more adjusted estimations were obtained in 105 indicators by using the other algorithms. In this regard, the differences between M5P and MLP were insignificant. The MLP was the second algorithm obtaining a better performance (it was the best option in 61 sustainability indicators) and with a

MAPE of less than 5% in 120 indicators. The MLP also showed great potential to estimate sustainability indicators. Regarding the other two algorithms (MLR and RF), a greater error was found in the estimations, especially in the MLR with the worst estimation in 121 of the sustainability indicators analysed. RF obtained a high number of indicators with the worst estimations (42 indicators). However, bad estimations were not always obtained with MLR and RF. In this regard, MLR was the best option in 26 indicators, while RF was the best option in 18 indicators. Likewise, the number of cases with a MAPE of less than 5% was 52 and 62 in RF and MLR, respectively. This aspect shows that, despite M5P is a quite efficient algorithm to estimate sustainability indicators, some indicators shows a better behaviour for other types of algorithms. For this reason, to analyse previously the most appropriate algorithm for each indicator would determine the most appropriate approach for each sustainability indicator. To assess this, the estimations obtained by the most appropriate combination for each indicator were studied. Figure 8 shows the histogram of MAPE with the optimal combination, whereas Annex B provides the results obtained between the actual values and the best values predicted in each indicator. As can be seen, the distribution of MAPE presents a greater density of indicators in values close to 0 with respect to M5P. The number of indicators with MAPE values lower than 5% was 147, overcoming 14 and 23% of the number of indicators obtained with M5P and MLP, respectively. Also, the number of cases with MAPE values greater than 20% in the optimal combination was 10, whereas in the other algorithms it was 25, 69, 18, and 22 with RF, MLR, MLP, and M5P, respectively.

Therefore, this optimal combination of algorithms for each indicator allows appropriate estimations to be carried out in 94.44% of indicators. However, it is important to highlight the need to carry out a preliminary study on the most suitable algorithm and architecture for each sustainability indicator. In the case of the MLPs designed for the study, the optimal number of nodes was determined. However, a quick MLP design without determining the optimal number of nodes could generate variations in the most suitable algorithm for each sustainability indicator To assess this aspect, the results obtained by MLPs with an optimal number of nodes were compared with those obtained by MLPs designed with the rule of number of nodes of Eq. (6). Figure 9 shows the effect of the optimization of the number of nodes of the MLPs on the number of cases with the best and worst estimations obtained by each algorithm. The determination of the optimal number of nodes can imply that the MLP is the second best algorithm instead of the third algorithm with the best results (24 indicators, behind RF and M5P). There was also a decrease in the number of cases in which the worst estimation was made, from 23 cases (when the optimal number of nodes was not determined) to 11 cases. Furthermore, the optimization of the number of nodes decreased the number of optimal cases of the other algorithms, with special emphasis on M5P. In this regard, the optimization of the MLPs represented a decrease of 23 cases in which M5P was the best option, while in MLR and RF that decrease was of 7 cases in each algorithm. It is important to note that the optimization of the number of nodes usually involved small variations in the MAPE value. As can be seen in Figure 10, the highest concentration of MAPE variations obtained with the node optimization was less than 2%. A total of 135 sustainability indicators were concentrated in the range of MAPE variations between 0 and 3%. However, these small percentage variations may imply that the best option is MLP. Table 3 includes the MAPE variations found in the 37 indicators in which the MLP was the best algorithm by optimizing the number of nodes. A total of 23 indicators obtained a decrease in MAPE of less than 3%. Even though this variation was low, MLPs were the best option. Thus, this process of optimizing the number of nodes can lead to obtaining more adjusted results in the estimations of some sustainability indicators, although the other algorithms analysed (especially M5P) may be suitable algorithms to make precise estimations without the need to carry out an optimization process, as in the case of the MLPs. Likewise, its white-box model structure allows decision makers to know how the model works.

$$Number\ of\ nodes = \frac{NI + NO}{2} \tag{6}$$

where NI is the number of nodes of the input layer (i.e., the input variables of the dataset), and NO is the number of nodes of the output layer (i.e., the output variables of the dataset).

In some cases, as can be seen in the Tables included in Annex A and B, estimations were not appropriate in some indicators since they have random behavior, i.e., not predictable, as is the case of I12 - Burnt area, so the estimations conducted in these indicators were individually analysed. Table 4 includes the predicted and actual MAPE values of indicators with a MAPE greater than 20%. It is important to highlight that the typology of these indicators is quite different, and there are sustainability indicators of the 4 dimensions (environmental, institutional, economic, and sociocultural). Likewise, except indicator I15 (% Reconstructed total area), indicators correspond to a subcategory within the indicator (e.g., indicator I12 (Agricultural area) belongs to the indicator category of burn area with other subcategories, such as forest stands or shrub land). There are also some similarities in the estimation limitations by similarity of the event monitored in the indicator. In this regard, I12 (Agricultural burnt area), I14 (Prevention against forest fires), and I15 are related to fires. This type of phenomena could strongly vary over the years (Figure 11), as the year 2016 reflects. In this year, there was an increase in the burnt surface area of forests and bushes, whereas agricultural areas were not affected. Therefore, it is supposed that in other years there were limitations in the estimations of the prediction models with the remaining indicators related to such phenomena (e.g., the remaining subcategories of I12).

By analysing Table 4, it is seen that the MAPEs obtained correspond to the highest values, except in 4 indicators. The remaining indicators obtained MAPEs greater than 50% with all the algorithms analysed, thus leading to the fact that the values obtained in the different estimations could not be considered valid. These indicators could therefore be limited to be correctly estimated. So, the methodology used is limited, since some types of sustainability indicators could be incorrectly estimated. Nevertheless, the number of indicators where this aspect was detected was low with respect to the size of the sample of indicators. Likewise, as there is a greater training sample, the estimation carried out by regression models could solve such limitations. In this regard, the results of this study are based on models trained with data from the 6 years before the year assessed.

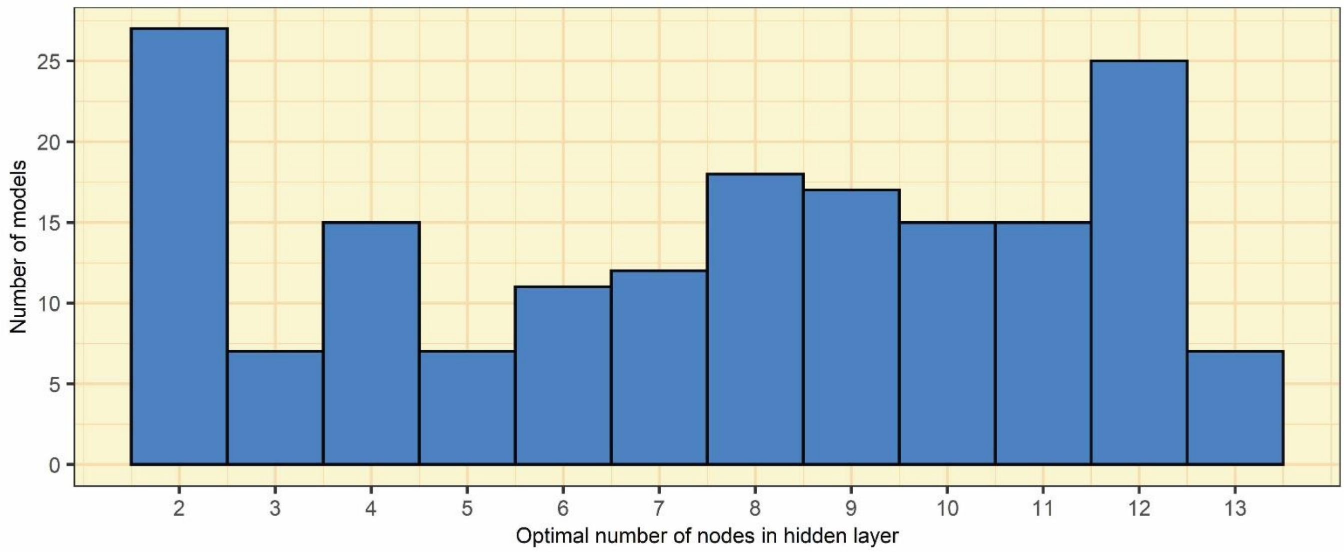

Figure 5. Histogram with the optimal number of nodes obtained in the MLPs. The histogram is represented by a bin width of 1 node.

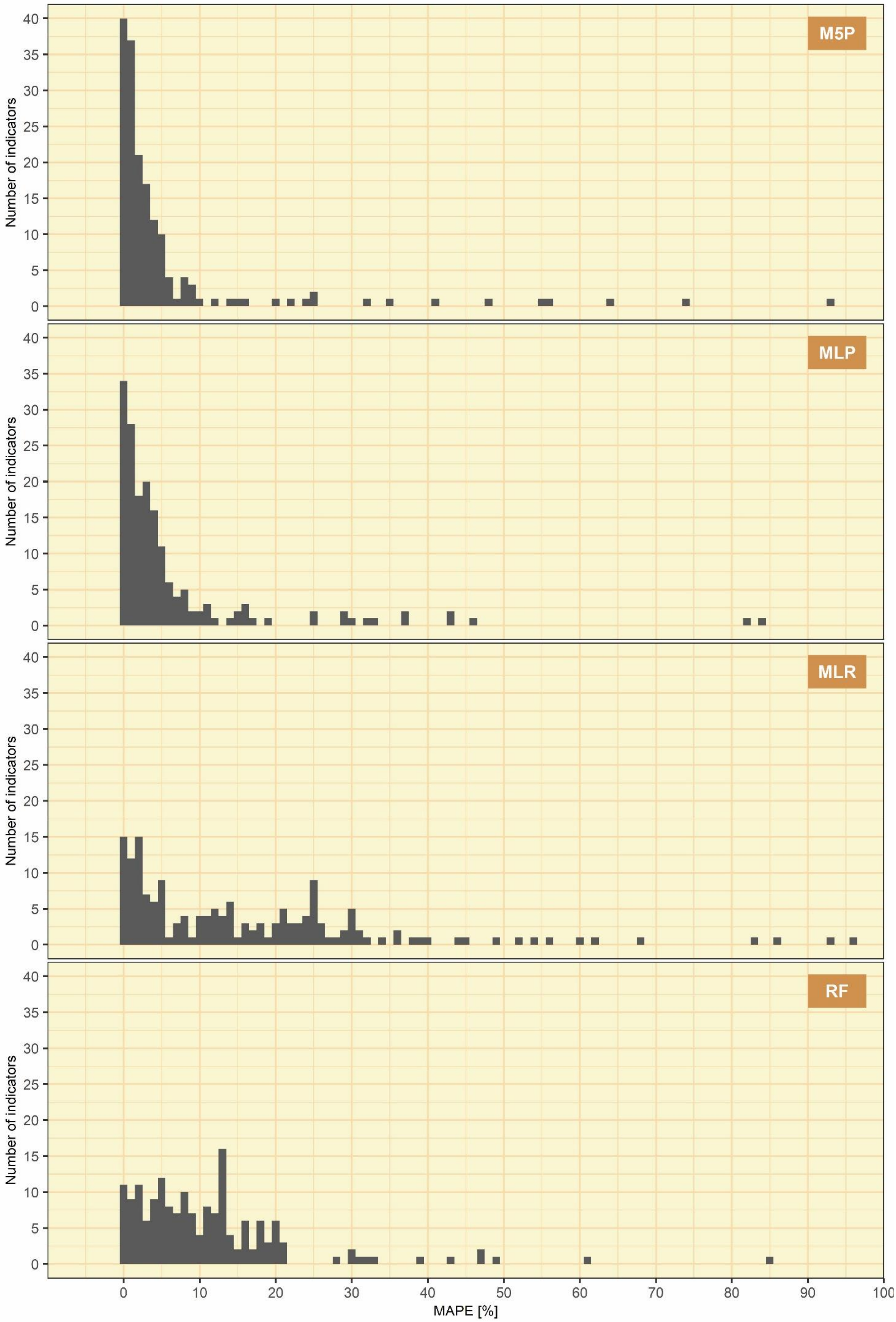

Figure 6. Comparison of the MAPE obtained in the estimation of the sustainability indicators from 2017 with the 4 algorithms. The histogram is represented by a bin width of 1%.

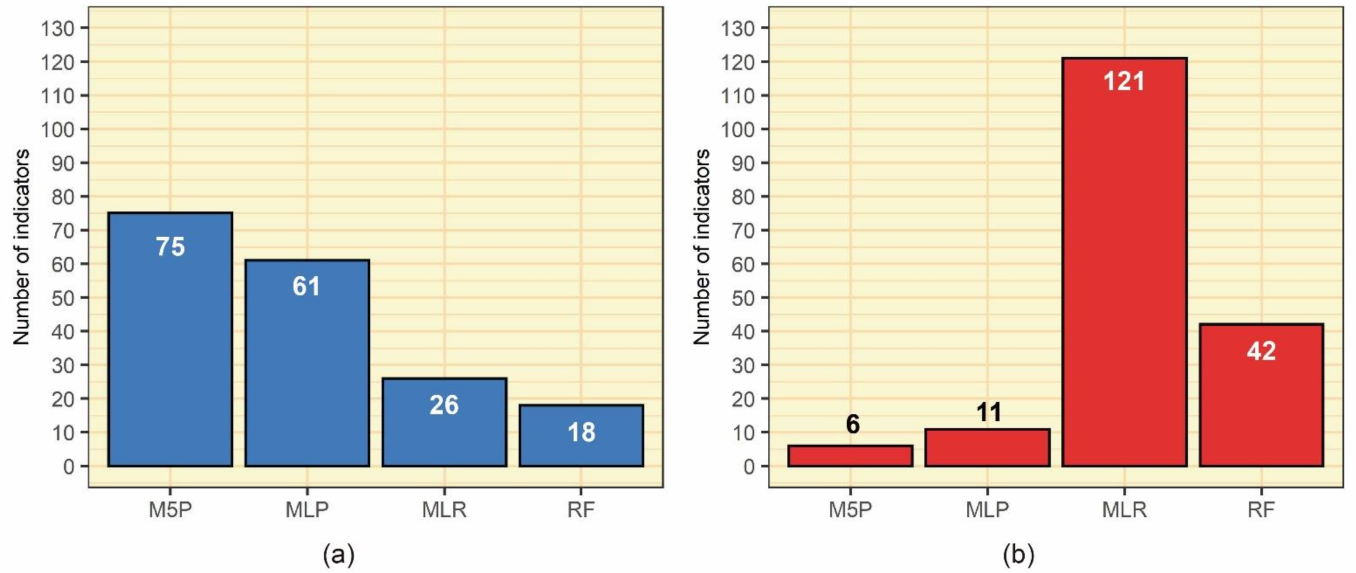

Figure 7. Representation of the estimations obtained by the algorithms: (a) number of cases per algorithm with the best estimations; and (b) number of cases per algorithm with the worst estimations.

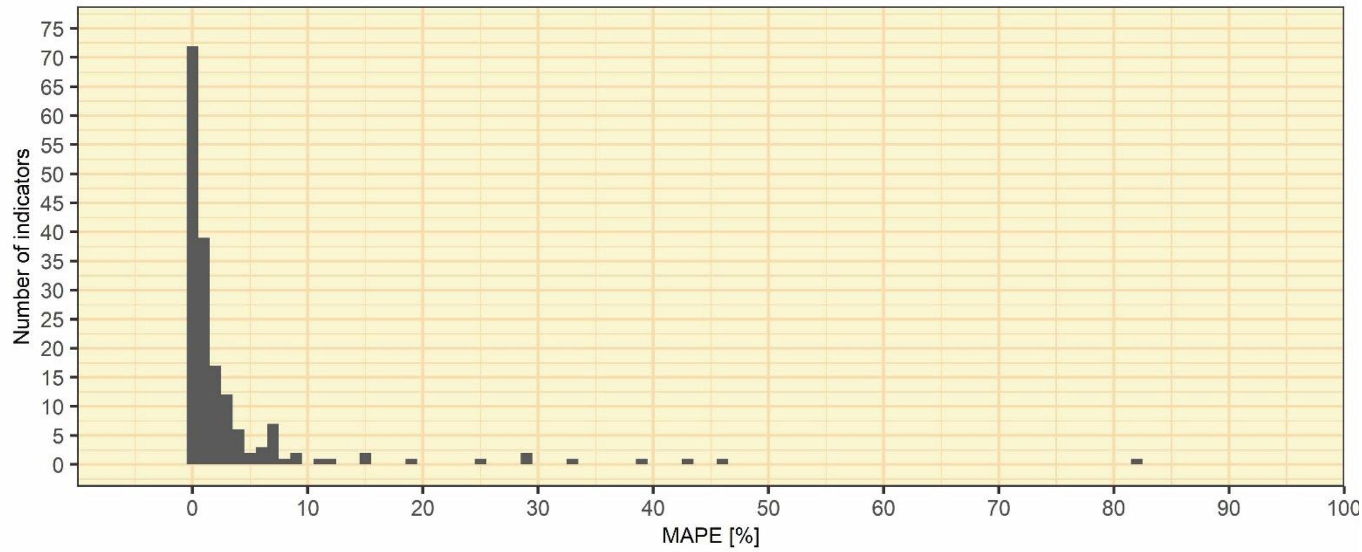

Figure 8. Histogram with the combination of the best estimations of the sustainability indicators from 2017. The histogram is represented by a bin width of 1%.

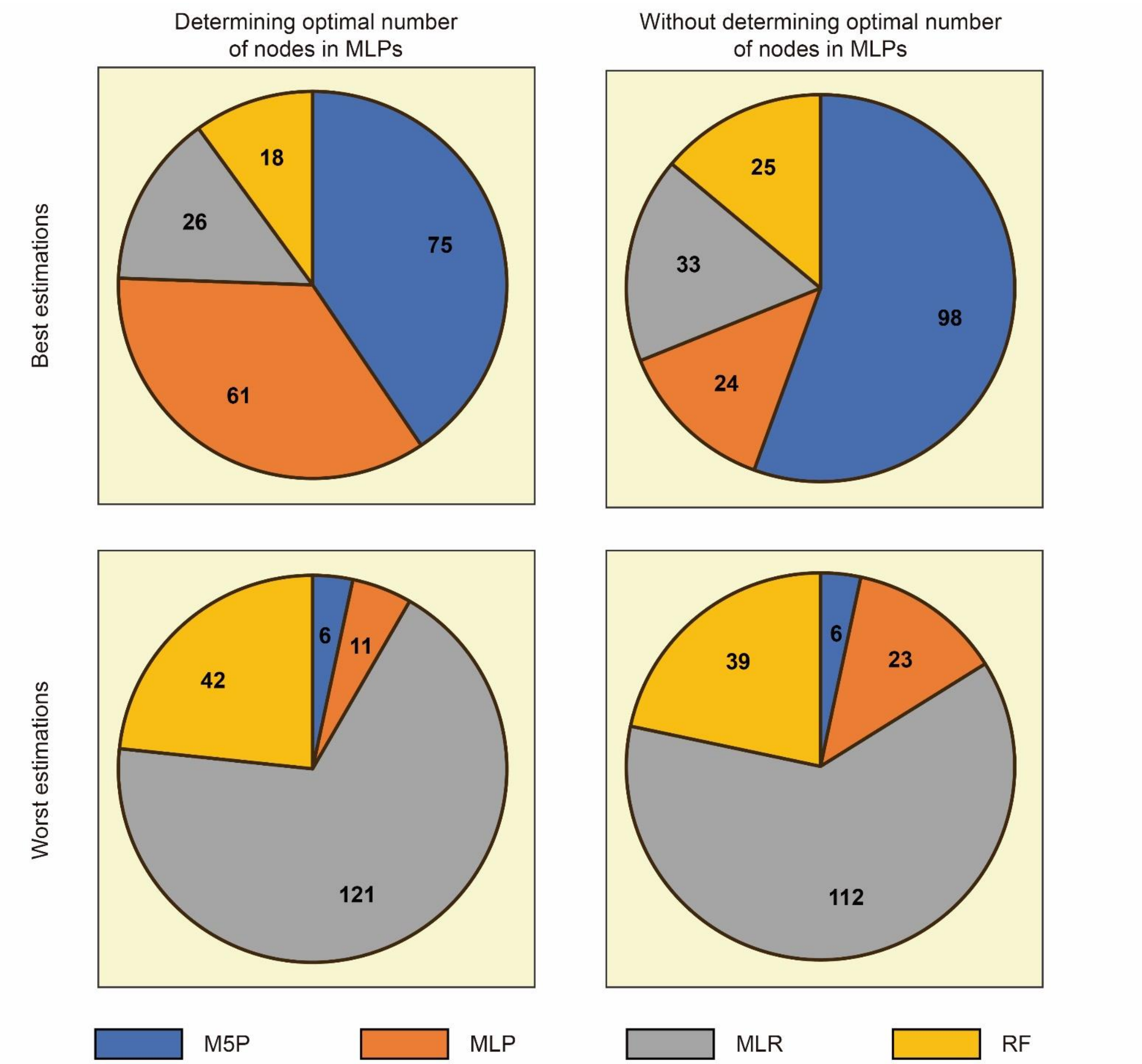

Figure 9. Effect of determining the optimal number of nodes of the MLPs on the best and worst estimations made in each sustainability indicator.

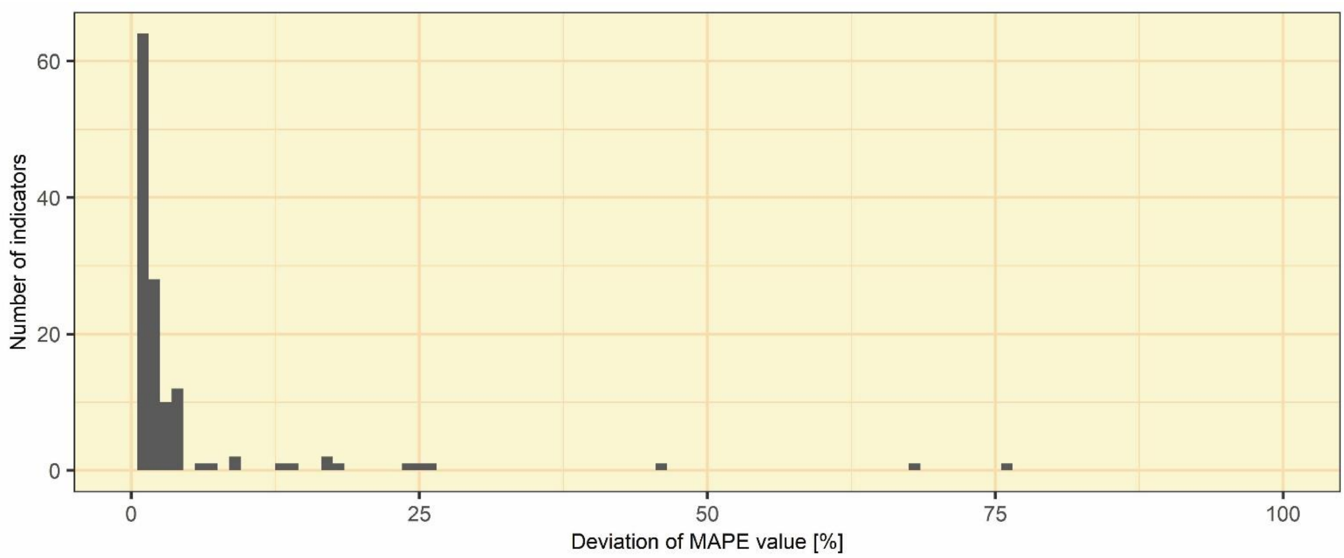

Figure 10. Deviation in the MAPE value between the MLPs that determined the optimal number of nodes and those that did not.

Table 3. Analysis of the MAPE variation in the 37 indicators in which the determination of the optimal number of nodes allows the MLP to be selected as the best algorithm.

Indicator	MAPE obtained with the optimal number of nodes (MLP) [%]	MAPE without determining the optimal number of nodes (MLP) [%]	Better MAPE obtained with the other algorithms (M5P, MLR or RF) [%]
I01 (Total)	2.26	7.85	4.85
I03 (Road)	0.04	3.43	3.05
I08	1.44	2.44	2.28
I09	3.38	5.67	5.47
I13 (Inland)	7.19	75.33	20.16
I15	15.40	229.62	196.61
I16 (Embarked)	42.60	88.37	47.57
I16 (Disembarked)	7.17	155.83	11.76
I16 (Total)	0.11	2.43	0.36
I18 (Residential)	1.25	3.39	2.87
I20 (Total)	0.96	3.31	1.64
I20 (Accommodation and food service activities)	2.69	6.41	4.13
I25 (Total excluding unprocessed food and energy)	45.80	245.80	148.28
I25 (Total excluding unprocessed food)	0.81	7.44	5.51
I25 (Total excluding energy)	10.61	28.20	15.02
I25 (Energy)	28.59	239.10	143.58
I26 (Electricity, gas, steam, cold and hot water and cold air)	0.14	2.63	1.88
I26 (Accommodation and food service activities)	3.58	5.19	3.85
I27 (Electricity, gas, steam, cold and hot water and cold air)	0.01	3.90	1.04
I27 (Water collection, treatment and distribution, sewerage, waste management and remediation activities)	5.15	5.84	5.49
I27 (Transportation and storage)	1.04	2.24	2.07
I27 (Real estate activities)	0.02	1.93	1.11
I27 (Human health and social work activities)	1.21	2.30	1.83
I27 (Others)	0.30	2.24	1.88
I28 (Manufacturing)	0.13	1.23	0.68
I28 (Human health and social work activities)	1.24	2.42	1.44
I29 (Agriculture, forestry and fishing-Sector)	0.35	1.92	0.39
I33 (Doctors)	0.78	1.47	1.12
I34 (Crimes of assault)	0.23	0.56	0.47
I34 (Theft/purse snatching)	2.28	3.75	3.40
I34 (Driving without legal documentation)	4.65	8.44	4.76
I35 (Crimes against persons [except voluntary manslaughter])	1.60	2.61	1.93
I35 (Crimes of voluntary manslaughter)	81.81	157.68	85.20
I37	7.35	8.78	7.99
I39 (Men)	2.19	3.11	2.79
I41 (Men)	0.27	1.38	0.89
I43 (Other countries)	0.12	1.73	1.31

Table 4. Indicators with a MAPE greater than 20% in the best estimation.

Indicator	MAPE [%]				Predicted value [a]				Actual value [a]
	M5P	MLP	MLR	RF	M5P	MLP	MLR	RF	
I12 (Agricultural area)	827.36	424.13	1316.29	33.29	-320.04	-142.62	-535.17	29.35	44.00
I14 (Prevention against forest fires)	258.91	620.22	545.00	247.07	179.45	360.11	322.50	173.54	50.00
I16 (Embarked)	47.57	88.37	59.58	61.04	277.42	354.13	300.01	302.76	188.00
I19 (State)	203.65	97.79	95.58	39.38	2,886.82	1,880.43	1,859.33	1,325.10	950.70
I20 (Real estate activities)	24.58	40.76	54.20	42.77	126,593,076.00	99,446,034.35	76,881,706.17	96,074,502.85	167,861,472.00
I25 (Total excluding unprocessed food and energy)	148.28	245.80	489.52	194.31	0.44	0.61	1.05	0.52	0.18
I25 (Energy)	209.27	239.10	222.09	143.58	-2.24	-2.85	-2.50	-0.89	2.05
I31 (Net migration)	198.46	291.70	353.14	208.85	-2,390.70	-3,137.55	-3,629.68	-2,473.92	-801.00
I34 (Theft of and from motor vehicles)	40.83	29.57	61.84	31.53	2.25	2.68	1.45	2.60	3.80
I35 (Crimes of voluntary manslaughter)	93.33	157.68	93.33	85.20	0.33	-2.88	0.33	0.74	5.00

[a] Units of output values are different. For more information about the units of each indicator see Annex A.

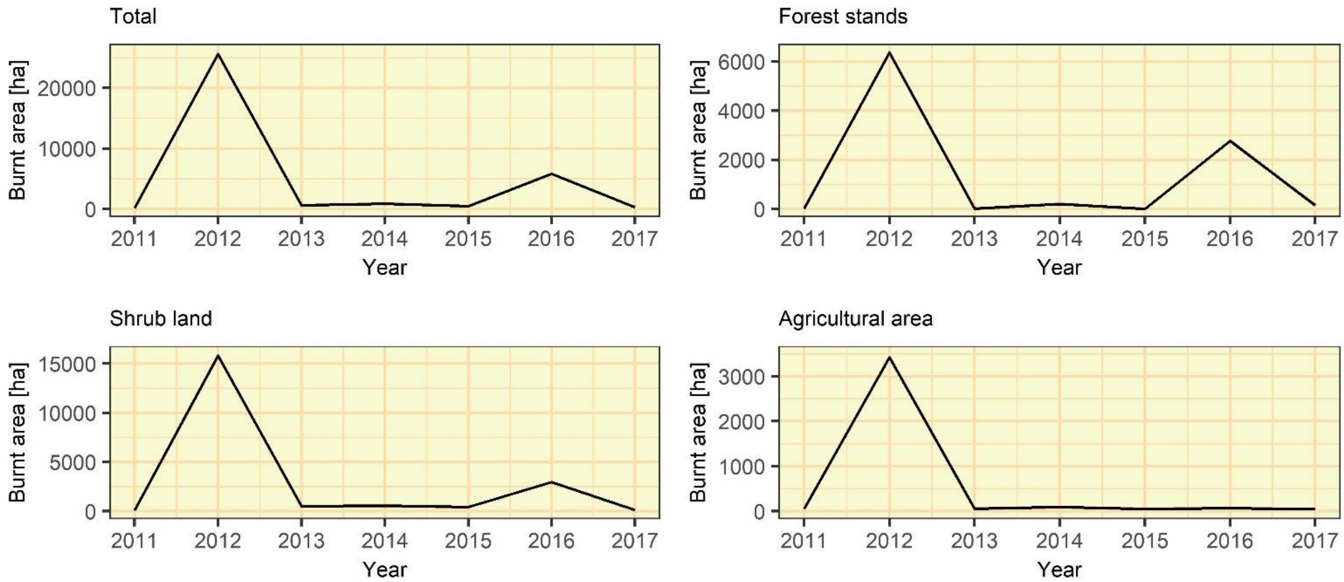

Figure 11. Monitored time series of the sub-indicators of burnt area.

4. Conclusions

This research analysed the possibility to estimate sustainability indicators of a region using four regression algorithms: multiple linear regression, multilayer perceptron, random forest, and M5P. The study area selected was the Algarve region (in the South of Portugal), and the data collected were from 180 indicators of the Observatory of Sustainability of the Algarve Region for Tourism (OBSERVE) platform.

Based on the results obtained with 720 models trained in the study, it was possible to determine that M5P and multilayer perceptrons were the algorithms which obtained the best estimations. In this regard, the number of cases with the best estimations was 75 with M5P and 61 with multilayer perceptrons. However, the use of multiple linear regression or random forest allowed the best estimation to be obtained in some indicators. This aspect suggests the need to carry out a previous study of each indicator to determine the most appropriate regression algorithm. In some cases, the use of M5P was an appropriate algorithm for most indicators. A high percentage of the indicators analysed in the study obtained low errors with the use of M5P. Although some of these indicators were not the best option, it was stressed the potential of using M5P when most appropriate algorithm could not be previously studied. For example, the analysis to determine the optimal number of nodes in the hidden layer allowed the performance of the multilayer perceptrons to be improved (obtaining better results than the M5P in some indicators), but the differences between the estimations were minor and the estimations obtained with M5P can be considered valid. In addition, the M5P models have an advantage over the multilayer perceptrons that allow the potential of their use to be influenced: stakeholders could extract a knowledge of the rules established by the model. The reason is that they are white box models with a tree structure which is easy to interpret. This would allow public bodies and institutions to apply these models without having advanced knowledge of these techniques.

In some cases, there were some limitations with the methodology analysed. In this regard, some types of indicators presented limitations in the estimation of their actual value, as a prediction process could not be applied to them – there are sustainable indicators whose occurrence is not predictable as was evident in the burnt area indicator. However, the low number of cases happening this (10 out of 180 indicators) guarantees the effectiveness of the methodology to estimate indicators in which data is not available through monitoring.

In conclusion, the results of this research could be very important: public bodies and institutions responsible for taking corrective measures may have complete information to take decisions by using the methodology used in this research. Also, this methodology could be extrapolated to other regions having monitoring databases of sustainability indicators, thus contributing to a more sustainable future and a better world.

References

Ahmad, M. W., Mourshed, M., & Rezgui, Y. (2017). Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. *Energy and Buildings*, *147*, 77–89. https://doi.org/10.1016/j.enbuild.2017.04.038

Akande, A., Cabral, P., Gomes, P., & Casteleyn, S. (2019). The Lisbon ranking for smart sustainable cities in Europe. *Sustainable Cities and Society*, *44*(October 2018), 475–487. https://doi.org/10.1016/j.scs.2018.10.009

Almeida, C. P., Ramos, A. F., & Silva, J. M. (2018). Sustainability assessment of building rehabilitation actions in old urban centres. *Sustainable Cities and Society*, *36*(May 2017), 378–385. https://doi.org/10.1016/j.scs.2017.10.014

Antanasijević, D., Pocajt, V., Popović, I., Redžić, N., & Ristić, M. (2013). The forecasting of municipal waste generation using artificial neural networks and sustainability indicators. *Sustainability Science*, *8*(1), 37–46.

Asmelash, A. G., & Kumar, S. (2019). Assessing progress of tourism sustainability: Developing and validating sustainability indicators. *Tourism Management*, *71*(October 2018), 67–83. https://doi.org/10.1016/j.tourman.2018.09.020

Assouline, D., Mohajeri, N., & Scartezzini, J. L. (2018). Large-scale rooftop solar photovoltaic technical potential estimation using Random Forests. *Applied Energy*, *217*(February), 189–211. https://doi.org/10.1016/j.apenergy.2018.02.118

Barron, A. R. (1993). Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information Theory*, *39*(3), 930–945.

Behnood, A., Behnood, V., Modiri Gharehveran, M., & Alyamac, K. E. (2017). Prediction of the compressive strength of normal and high-performance concretes using M5P model tree algorithm. *Construction and Building Materials*, *142*, 199–207. https://doi.org/10.1016/j.conbuildmat.2017.03.061

Bienvenido-Huertas, D., Rubio-Bellido, C., Pérez-Ordóñez, J. L., & Moyano, J. (2019). Optimizing the evaluation of thermal transmittance with the thermometric method using multilayer perceptrons. *Energy & Buildings*, *198*, 395–411. https://doi.org/10.1016/j.enbuild.2019.06.040

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Castellani, V., & Sala, S. (2012). Ecological Footprint and Life Cycle Assessment in the sustainability assessment of tourism activities. *Ecological Indicators*, *16*, 135–147. https://doi.org/10.1016/j.ecolind.2011.08.002

Coelho, P., Mascarenhas, A., Vaz, P., Beja, I., Dores, A., & Ramos, T. B. (2006). A methodological framework for indicators of sustainable development at the regional scale: the case of a Portuguese region (Algarve). *Proceedings of 12th Annual International Sustainable Development Research Conference*.

Crouch, G. I. (2011). Destination competitiveness: An analysis of determinant attributes. *Journal of Travel Research*, *50*(1), 27–45.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, *2*(4), 303–314.

D'Amico, A., Ciulla, G., Traverso, M., Lo Brano, V., & Palumbo, E. (2019). Artificial Neural Networks to assess energy and environmental performance of buildings: An Italian case study. *Journal of Cleaner Production*, *239*, 117993. https://doi.org/10.1016/j.jclepro.2019.117993

Dietterich, T. G. (2000). Experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization. *Machine Learning*, *40*(2), 139–157. https://doi.org/10.1023/A:1007607513941

Dudoit, S., Fridlyand, J., & Speed, T. P. (2002). Comparison of discrimination methods for the classification of tumors using gene expression data. *Journal of the American Statistical Association*, *97*(457), 77–87.

Farinha, F., Oliveira, M. J., Silva, E. M. J., Lança, R., Pinheiro, M. D., & Miguel, C. (2019). Selection process of sustainable indicators for the Algarve region-OBSERVE project. *Sustainability (Switzerland)*, *11*(2), 1–24. https://doi.org/10.3390/su11020444

Feleki, E., Vlachokostas, C., & Moussiopoulos, N. (2018). Characterisation of sustainability in urban areas: An analysis of assessment tools with emphasis on European cities. *Sustainable Cities and Society*, *43*(August), 563–577. https://doi.org/10.1016/j.scs.2018.08.025

Fletcher, R. (1980). *Practical methods of optimization*. John Wiley&Sons.

Golbabai, A., & Seifollahi, S. (2007). Radial basis function networks in the numerical solution of linear integro-differential equations. *Applied Mathematics and Computation*, *188*(1), 427–432. https://doi.org/10.1016/j.amc.2006.10.004

Hatipoglu, B., Alvarez, M. D., & Ertuna, B. (2016). Barriers to stakeholder involvement in the planning of sustainable tourism: The case of the Thrace region in Turkey. *Journal of Cleaner Production*, *111*, 306–317. https://doi.org/10.1016/j.jclepro.2014.11.059

Haykin, S. S., Haykin, S. S., Haykin, S. S., & Haykin, S. S. (2009). *Neural networks and learning machines* (Vol. 3). Pearson Upper Saddle River.

He, P., He, Y., & Xu, F. (2018). Evolutionary analysis of sustainable tourism. *Annals of Tourism Research*, *69*(January), 76–89. https://doi.org/10.1016/j.annals.2018.02.002

Hermans, F. L. P., Haarmann, W. M. F., & Dagevos, J. F. (2011). Evaluation of stakeholder participation in monitoring regional sustainable development. *Regional Environmental Change*, *11*(4), 805–815.

Higgins-Desbiolles, F. (2018). Sustainable tourism: Sustaining tourism or something more? *Tourism Management Perspectives*, *25*(November 2017), 157–160. https://doi.org/10.1016/j.tmp.2017.11.017

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*(5), 359–366. https://doi.org/10.1016/0893-6080(89)90020-8

Instituto Nacional de Estatística. (2019). *Estatísticas*. https://ine.pt

Kapera, I. (2018). Sustainable tourism development efforts by local governments in Poland. *Sustainable Cities and Society*, *40*(May), 581–588. https://doi.org/10.1016/j.scs.2018.05.001

Kristjánsdóttir, K. R., Ólafsdóttir, R., & Ragnarsdóttir, K. V. (2018). Reviewing integrated sustainability indicators for tourism. *Journal of Sustainable Tourism*, *26*(4), 583–599. https://doi.org/10.1080/09669582.2017.1364741

Kumar, R., Aggarwal, R. K., & Sharma, J. D. (2013). Energy analysis of a building using artificial neural network: A review. *Energy and Buildings*, *65*(2), 352–358. https://doi.org/10.1016/j.enbuild.2013.06.007

Lane, B. (2018). Will sustainable tourism research be sustainable in the future? An opinion piece. *Tourism Management Perspectives*, *25*(December 2017), 161–164. https://doi.org/10.1016/j.tmp.2017.12.001

Larivière, B., & Van Den Poel, D. (2005). Predicting customer retention and profitability by using random forests and regression forests techniques. *Expert Systems with Applications*, *29*(2), 472–484. https://doi.org/10.1016/j.eswa.2005.04.043

Liao, Y. T., & Chern, S. G. (2015). Strategic ecocity development in urban-rural fringes: Analyzing Wulai District. *Sustainable Cities and Society*, *19*(116), 98–108. https://doi.org/10.1016/j.scs.2015.07.014

Lin, L., Wang, Q., & Sadek, A. W. (2016). A combined M5P tree and hazard-based duration model for predicting urban freeway traffic accident durations. *Accident Analysis and Prevention*, *91*, 114–126. https://doi.org/10.1016/j.aap.2016.03.001

Lin, W., Li, Y., Li, X., & Xu, D. (2018). The dynamic analysis and evaluation on tourist ecological footprint of city: Take Shanghai as an instance. *Sustainable Cities and Society*, *37*(100), 541–549. https://doi.org/10.1016/j.scs.2017.12.003

Liu, C. H., Tzeng, G. H., Lee, M. H., & Lee, P. Y. (2013). Improving metro-airport connection service for tourism development: Using hybrid MCDM models. *Tourism Management Perspectives*, *6*, 95–107. https://doi.org/10.1016/j.tmp.2012.09.004

Lozano, R. (2012). Towards better embedding sustainability into companies' systems: An analysis of voluntary corporate initiatives. *Journal of Cleaner Production*, *25*, 14–26. https://doi.org/10.1016/j.jclepro.2011.11.060

Lu, J., & Nepal, S. K. (2009). Sustainable tourism research: an analysis of papers published in the Journal of Sustainable Tourism. *Journal of Sustainable Tourism*, *17*(1), 5–16. https://doi.org/10.1080/09669580802582480

Manning, T. (1996). *What tourism managers need to know: A practical guide to the development and use of indicators of sustainable tourism*. World Tourism Organization Pubns.

Mascarenhas, A., Coelho, P., Subtil, E., & Ramos, T. B. (2010). The role of common local indicators in regional sustainability assessment. *Ecological Indicators*, *10*(3), 646–656. https://doi.org/10.1016/j.ecolind.2009.11.003

Mascarenhas, A., Nunes, L. M., & Ramos, T. B. (2014). Exploring the self-assessment of sustainability indicators by different stakeholders. *Ecological Indicators*, *39*, 75–83. https://doi.org/10.1016/j.ecolind.2013.12.001

Mascarenhas, A., Nunes, L. M., & Ramos, T. B. (2015). Selection of sustainability indicators for planning: Combining stakeholders' participation and data reduction techniques. *Journal of Cleaner Production*, *92*, 295–307. https://doi.org/10.1016/j.jclepro.2015.01.005

Michael, F. L., Noor, Z. Z., & Figueroa, M. J. (2014). Review of urban sustainability indicators assessment - Case study between Asian countries. *Habitat International*, *44*, 491–500. https://doi.org/10.1016/j.habitatint.2014.09.006

Michailidou, A. V., Vlachokostas, C., & Moussiopoulos, N. (2015). A methodology to assess the overall environmental pressure attributed to tourism areas: A combined approach for typical all-sized hotels in Chalkidiki, Greece. *Ecological Indicators*, *50*, 108–119. https://doi.org/10.1016/j.ecolind.2014.11.006

Mousa, W. A. Y., Lang, W., Auer, T., & Yousef, W. A. (2017). A pattern recognition approach for modeling the air change rates in naturally ventilated buildings from limited steady-state CFD simulations. *Energy and Buildings*, *155*, 54–65. https://doi.org/10.1016/j.enbuild.2017.09.016

Moussiopoulos, N., Achillas, C., Vlachokostas, C., Spyridi, D., & Nikolaou, K. (2010). Environmental, social and economic information management for the evaluation of sustainability in urban areas: A system of indicators for Thessaloniki, Greece. *Cities*, *27*(5), 377–384. https://doi.org/10.1016/j.cities.2010.06.001

Pérez-Rodríguez, J. V., Ledesma-Rodríguez, F., & Santana-Gallego, M. (2015). Testing dependence between GDP and tourism's growth rates. *Tourism Management*, *48*, 268–282. https://doi.org/10.1016/j.tourman.2014.11.007

Pérez Gálvez, J. C., Granda, M. J., López-Guzmán, T., & Coronel, J. R. (2017). Local gastronomy, culture and tourism sustainable cities: The behavior of the American tourist. *Sustainable Cities and Society*, *32*(May), 604–612. https://doi.org/10.1016/j.scs.2017.04.021

Pino-Mejías, R., Pérez-Fargallo, A., Rubio-Bellido, C., & Pulido-Arcas, J. A. (2017). Comparison of linear regression and artificial neural networks models to predict heating and cooling energy demand, energy consumption and CO2 emissions. *Energy*, *118*, 24–36. https://doi.org/10.1016/j.energy.2016.12.022

Pino-Mejías, R., Pérez-Fargallo, A., Rubio-Bellido, C., & Pulido-Arcas, J. A. (2018). Artificial neural networks and linear regression prediction models for social housing allocation: Fuel Poverty Potential Risk Index. *Energy*, *164*, 627–641. https://doi.org/10.1016/j.energy.2018.09.056

Pulido-Fernández, J. I., Cárdenas-García, P. J., & Espinosa-Pulido, J. A. (2019). Does environmental sustainability contribute to tourism growth? An analysis at the country level. *Journal of Cleaner Production*, *213*, 309–319. https://doi.org/10.1016/j.jclepro.2018.12.151

Quinlan, J. R., & others. (1992). Learning with continuous classes. *5th Australian Joint Conference on Artificial Intelligence*, *92*, 343–348.

Ramos, T. B. (2009). Development of regional sustainability indicators and the role of academia in this process: the Portuguese practice. *Journal of Cleaner Production*, *17*(12), 1101–1115. https://doi.org/10.1016/j.jclepro.2009.02.024

Rizzo, A. (2017). Managing the energy transition in a tourism-driven economy: The case of Malta. *Sustainable Cities and Society*, *33*(2017), 126–133. https://doi.org/10.1016/j.scs.2016.12.005

Robaina-Alves, M., Moutinho, V., & Costa, R. (2016). Change in energy-related CO2 (carbon dioxide) emissions in Portuguese tourism: A decomposition analysis from 2000 to 2008. *Journal of Cleaner Production*, *111*, 520–528. https://doi.org/10.1016/j.jclepro.2015.03.023

Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews*, *71*, 804–818. https://doi.org/10.1016/j.oregeorev.2015.01.001

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536. https://doi.org/10.1038/323533a0

Rylance, A. (2012). Local economic development in Mozambique: an assessment of the implementation of tourism policy as a means to promote local economies. In *Responsible Tourism* (pp. 55–68). Routledge.

Sun, W. (2018). River ice breakup timing prediction through stacking multi-type model trees. *Science of the Total Environment*, *644*, 1190–1200. https://doi.org/10.1016/j.scitotenv.2018.07.001

Tso, G. K. F., & Yau, K. K. W. (2007). Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy*, *32*(9), 1761–1768. https://doi.org/10.1016/j.energy.2006.11.010

United Nations General Assembly. (2015). *Resolution Adopted by the General Assembly on 25 September 2015: Transforming Our World: the 2030 Agenda for Sustainable Development*.

Verma, P., & Raghubanshi, A. S. (2018). Urban sustainability indicators: Challenges and opportunities. *Ecological Indicators*, *93*, 282–291.

Wang, F., Peng, X., Qin, Y., & Wang, C. (2020). What can the news tell us about the environmental performance of tourist areas? A text mining approach to China's National 5A Tourist Areas. *Sustainable Cities and Society*, *52*(September 2019), 101818. https://doi.org/10.1016/j.scs.2019.101818

Wang, S., Du, J., Li, S., He, H., & Xu, W. (2019). Impact of tourism activities on glacial changes based on the tourism heat footprint (THF) method. *Journal of Cleaner Production*, *215*, 845–853. https://doi.org/10.1016/j.jclepro.2019.01.120

Wang, Y. N. (1994). A neural network adaptive control based on rapid learning method and application. *Advance in Molding & Analysis*, *46*(3), 27–34.

Wang, Y., & Witten, I. H. (1997). *Induction of model trees for predicting continuous classes. proceedings of the poster papers of the european conference on machine learning*. Prague: University of Economics, Faculty of Informatics and Statistics.

Werbos, P. (1974). Beyond Regression: New Tools for Prediction and Analysis in the Behavior Science. In *Unpublished Doctoral Dissertation, Harvard University*. Harvard University.

Whittlesea, E. R., & Owen, A. (2012). Towards a low carbon future – the development and application of REAP Tourism, a destination footprint and scenario tool. *Journal of Sustainable Tourism*, *20*(6), 845–865. https://doi.org/10.1080/09669582.2012.680699

Williams, K. T., & Gomez, J. D. (2016). Predicting future monthly residential energy consumption using building characteristics and climate data: A statistical learning approach. *Energy and Buildings*, *128*, 1–11. https://doi.org/10.1016/j.enbuild.2016.06.076

World Tourism Organization and the United Nations Environment Programme. (2005). *Making tourism more sustainable: A guide for policy makers*. World Tourism Organization Publications.

Wu, M., Wei, Y., Lam, P. T. I., Liu, F., & Li, Y. (2019). Is urban development ecologically sustainable? Ecological footprint analysis and prediction based on a modified artificial neural network model: A case study of Tianjin in China. *Journal of Cleaner Production*, *237*, 117795. https://doi.org/10.1016/j.jclepro.2019.117795

Xu, M., Watanachaturaporn, P., Varshney, P. K., & Arora, M. K. (2005). Decision tree regression for soft classification of remote sensing data. *Remote Sensing of Environment*, *97*(3), 322–336. https://doi.org/10.1016/j.rse.2005.05.008

Yadav, A. K., Malik, H., & Chandel, S. S. (2014). Selection of most relevant input parameters using WEKA for artificial neural network based solar radiation prediction models. *Renewable and Sustainable Energy Reviews*, *31*, 509–519. https://doi.org/10.1016/j.rser.2013.12.008

Zhang, J., Ji, M., & Zhang, Y. (2015). Tourism sustainability in Tibet - Forward planning using a systems approach. *Ecological Indicators*, *56*, 218–228. https://doi.org/10.1016/j.ecolind.2015.04.006

Annex A. MAPE values obtained in each estimation

Table A1. MAPE values obtained in the estimations of the environmental indicators in 2017.

Indicator	MAPE [%]			
	MLR	MLP	RF	M5P
I01 (Total)	5.13	2.26	4.85	10.76
I01 (Waste management)	5.13	0.48	6.19	6.13
I01 (Noise and vibration abatement)	34.80	7.41	55.79	7.41
I01 (Protection of biodiversity and landscape)	11.56	9.14	17.31	9.14
I01 (Research and development)	-	-	-	-
I01 (Others)	1.91	1.91	45.05	3.40
I02 (Total)	0.10	0.00	0.00	18.43
I02 (Embarked)	0.09	0.01	0.01	18.43
I02 (Disembarked)	0.09	0.00	0.00	18.42
I03 (Total)	1.19	1.19	25.09	15.73
I03 (Rail)	3.07	3.02	3.02	17.06
I03 (Road)	3.05	0.04	25.25	13.25
I04 (Total)	0.30	0.30	1.55	13.02
I04 (Ria Formosa)	0.30	0.30	20.42	13.25
I04 (Rio Guadiana)	1.53	1.53	11.54	8.62
I05	3.13	0.02	4.70	1.65
I06	1.18	1.18	14.18	8.17
I07	0.19	0.04	0.30	0.27
I08	3.11	1.44	11.55	2.28
I09	5.60	3.38	14.08	5.47
I10	0.33	0.33	12.53	3.51
I11	8.89	2.66	23.46	12.10
I12 (Total)	16.12	0.01	0.01	49.32
I12 (Forest stands)	63.64	6.66	387.94	6.66
I12 (Shrub land)	32.09	0.00	0.00	276.85
I12 (Agricultural area)	827.36	33.29	1,316.29	33.29
I13 (Total)	0.64	0.15	1.48	0.15
I13 (Inland)	20.16	7.19	83.33	47.00
I13 (Coastal/transition)	0.09	0.01	2.45	0.38
I14 (Total)	24.14	12.30	24.79	12.30
I14 (Prevention against forest fires)	258.91	247.07	545.00	247.07
I14 (Others)	24.80	19.45	48.75	19.45
I15	244.08	15.40	196.61	207.82
I16 (Embarked)	47.57	42.60	59.58	61.04
I16 (Disembarked)	209.28	7.17	11.76	46.80
I16 (Transition)	0.34	1.20	24.92	29.90
I16 (Total)	0.36	0.11	24.90	30.97

Table A2. MAPE values obtained in the estimations of the institutional indicators in 2017.

Indicator	MAPE [%]			
	MLR	MLP	RF	M5P
I17	15.23	15.23	21.12	16.26
I18 (Residential)	2.87	1.25	27.79	11.05
I18 (Non-residential)	7.17	7.17	35.91	18.89
I19 (State)	203.65	39.38	95.58	39.38
I19 (Enterprises)	4.02	4.02	12.67	5.93
I19 (Higher education)	14.16	6.78	6.78	9.29
I19 (Private non-profit institutions)	-	-	-	-

Table A3. MAPE values obtained in the estimations of the economic indicators in 2017.

Indicator	MAPE [%]			
	MLR	MLP	RF	M5P
I20 (Total)	1.64	0.96	3.48	18.16
I20 (Agriculture, farming of animals, hunting and forestry)	3.52	0.03	33.60	13.45
I20 (Mining and quarrying)	7.71	7.70	7.70	11.87
I20 (Manufacturing)	1.76	1.76	19.59	11.35
I20 (Electricity, gas, steam, cold and hot water and cold air)	8.46	1.35	23.84	1.35
I20 (Water collection, treatment and distribution, sewerage, waste management and remediation activities)	5.11	2.16	2.16	3.15
I20 (Construction)	10.23	5.50	30.44	12.05
I20 (Wholesale and retail trade, repair of motor vehicles and motorcycles)	4.66	0.08	26.47	14.14
I20 (Transportation and storage)	0.48	0.48	22.91	12.56
I20 (Accommodation and food service activities)	4.13	2.69	39.17	20.92
I20 (Information and communication activities)	3.15	3.15	38.33	19.90
I20 (Real estate activities)	24.58	24.58	54.20	42.77
I20 (Consultancy, scientific and technical activities)	2.21	2.21	28.89	19.76
I20 (Administrative and support service activities)	3.22	1.16	1.16	13.60
I20 (Education)	2.64	0.24	0.24	11.20
I20 (Human health and social work activities)	3.64	3.64	5.05	12.66
I20 (Arts, entertainment, sports and recreation activities)	8.40	0.04	32.43	12.65

I20 (Others)	2.49	2.49	26.92	19.01
I21 (Hotels)	0.00	0.00	25.22	12.53
I21 (Apartment hotels)	0.00	0.00	15.50	7.53
I21 (Tourist villages)	0.00	0.00	0.00	7.98
I21 (Tourist apartments)	0.00	0.00	6.86	0.66
I22 (Total)	0.54	0.54	30.22	15.60
I22 (Hotels)	0.46	0.26	0.26	15.86
I22 (Guest houses)	0.58	0.58	39.89	15.11
I22 (Lodging houses)	3.12	2.19	2.19	19.74
I22 (Hotel apartments)	4.23	3.49	30.24	16.03
I22 (Tourist villages)	2.28	2.28	25.63	14.39
I22 (Tourist apartments)	1.96	1.96	30.78	16.81
I23	0.84	0.00	2.22	2.09
I24 (Total)	2.31	0.44	2.82	8.68
I24 (Accommodation and food service activities)	2.10	0.07	20.62	10.43
I25 (Total)	3.25	3.25	18.19	20.12
I25 (Total excluding housing)	3.30	3.30	145.32	20.23
I25 (Total excluding unprocessed food and energy)	148.28	45.80	489.52	194.31
I25 (Total excluding unprocessed food)	56.28	0.81	86.31	5.51
I25 (Total excluding energy)	55.21	10.61	109.46	15.02
I25 (Unprocessed food)	21.97	0.08	44.35	12.50
I25 (Energy)	209.27	28.59	222.09	143.58
I26 (Agriculture, farming of animals, hunting and forestry)	5.47	0.03	22.38	2.21
I26 (Mining and quarrying)	1.34	0.54	14.24	0.54
I26 (Manufacturing)	0.92	0.92	1.48	2.93
I26 (Electricity, gas, steam, cold and hot water and cold air)	1.88	0.14	68.00	29.78
I26 (Water collection, treatment and distribution, sewerage, waste management and remediation activities)	6.20	3.46	3.46	4.66
I26 (Construction)	1.64	0.07	5.47	4.13
I26 (Wholesale and retail trade, repair of motor vehicles and motorcycles)	0.02	0.02	0.24	0.08
I26 (Transportation and storage)	1.29	1.29	7.93	5.83
I26 (Accommodation and food service activities)	3.85	3.58	31.07	11.68
I26 (Information and communication activities)	2.34	2.32	2.32	6.73
I26 (Real estate activities)	0.72	0.72	22.23	14.01
I26 (Consultancy scientific and technical activities)	0.05	0.05	11.46	6.99
I26 (Administrative and support service activities)	1.02	1.02	24.51	13.34
I26 (Education)	2.10	0.01	4.62	0.01
I26 (Human health and social work activities)	0.20	0.20	8.90	4.58
I26 (Arts, entertainment, sports and recreation activities)	1.19	1.19	20.96	12.56
I26 (Others)	1.04	1.04	15.56	10.32
I27 (Agriculture, farming of animals, hunting and forestry)	5.17	0.01	23.84	5.38
I27 (Manufacturing)	0.86	0.86	2.06	1.93
I27 (Electricity, gas, steam, cold and hot water and cold air)	3.85	0.01	1.04	10.12
I27 (Water collection, treatment and distribution, sewerage, waste management and remediation activities)	5.49	5.15	9.71	8.31
I27 (Construction)	0.97	0.97	7.77	6.11
I27 (Wholesale and retail trade, repair of motor vehicles and motorcycles)	1.30	0.83	0.83	3.96
I27 (Transportation and storage)	2.07	1.04	9.86	4.40
I27 (Accommodation and food service activities)	0.03	0.03	3.36	13.28
I27 (Information and communication activities)	6.17	6.17	11.34	9.04
I27 (Real estate activities)	1.11	0.02	22.21	12.23
I27 (Consultancy scientific and technical activities)	0.35	0.35	14.83	9.09
I27 (Administrative and support service activities)	0.15	0.15	29.54	18.50
I27 (Education)	0.77	0.00	4.41	0.00
I27 (Human health and social work activities)	1.83	1.21	14.14	6.56
I27 (Others)	2.11	0.30	1.88	7.70
I28 (Total)	0.20	0.20	21.09	12.78
I28 (Agriculture, farming of animals, hunting and forestry)	3.64	0.59	25.58	12.56
I28 (Manufacturing)	5.11	0.13	0.68	6.80
I28 (Electricity, gas, steam, cold and hot water and cold air)	74.28	4.41	4.41	4.66
I28 (Water collection, treatment and distribution, sewerage, waste. management and remediation activities)	0.59	0.59	17.77	10.40
I28 (Construction)	4.84	1.44	23.60	7.99
I28 (Wholesale and retail trade, repair of motor vehicles and motorcycles)	1.17	1.17	17.74	11.53
I28 (Transportation and storage)	0.05	0.05	20.85	12.71
I28 (Accommodation and food service activities)	3.58	3.58	35.87	20.85
I28 (Information and communication activities)	0.17	0.17	25.37	5.80
I28 (Real estate activities)	5.31	0.77	0.77	28.25
I28 (Consultancy scientific and technical activities)	0.43	0.43	30.36	18.35
I28 (Administrative and support service activities)	2.60	2.60	2.77	16.12
I28 (Education)	2.76	2.76	14.45	8.59
I28 (Human health and social work activities)	1.44	1.24	20.09	11.17
I28 (Others)	3.85	3.85	24.07	11.17
I29 (Total-Gender)	0.00	0.00	0.00	5.27
I29 (Men-Gender)	0.24	0.24	4.02	1.22
I29 (Women-Gender)	1.28	1.28	12.52	7.96
I29 (Total-Sector)	0.00	0.00	0.00	5.27
I29 (Agriculture, forestry and fishing-Sector)	0.39	0.35	13.60	4.54
I29 (Industry-Sector)	3.87	0.49	13.29	0.49
I29 (Services-Sector)	0.83	0.83	10.08	6.18
I30 (Accommodation and food service activities)	2.37	2.37	11.66	3.85
I30 (Others)	0.16	0.16	4.91	1.66

Table A4. MAPE values obtained in the estimations of the sociocultural indicators in 2017.

Indicator	MAPE [%]			
	MLR	MLP	RF	M5P
I31 (Natural increase)	8.80	8.80	28.69	21.17
I31 (Net migration)	198.46	198.46	353.14	208.85
I32 (Monuments)	3.83	2.47	3.61	4.17
I32 (Sets)	0.21	0.21	1.84	1.52
I32 (Sites)	1.34	0.09	1.34	6.31
I33 (Beds)	1.01	1.01	4.49	3.46
I33 (Doctors)	1.12	0.78	10.68	3.62
I33 (Nurses)	0.72	0.72	10.21	5.21
I34 (Crimes of assault)	0.47	0.23	4.66	1.13
I34 (Theft/purse snatching)	3.40	2.28	19.23	3.92
I34 (Theft of and from motor vehicles)	40.83	29.03	61.84	31.53
I34 (Driving a motor vehicle with a blood alcohol equal or above 1)	1.15	1.15	4.29	2.20
I34 (Driving without legal documentation)	9.11	4.65	4.76	20.36
I34 (Crimes against patrimony)	0.40	0.40	3.19	1.70
I35 (Crimes against persons [except voluntary manslaughter])	2.59	1.60	4.92	1.93
I35 (Crimes of voluntary manslaughter)	93.33	81.81	93.33	85.20
I35 (Crimes against patrimony)	0.42	0.42	16.95	1.97
I35 (Crimes against life in society)	6.44	6.44	7.58	6.81
I35 (Crimes against the State)	1.50	1.11	1.11	2.63
I35 (Crimes set out in sundry legislation)	2.45	0.07	2.45	3.10
I36	0.81	0.81	1.51	0.99
I37	7.99	7.35	51.94	11.31
I38 (Total)	0.51	0.43	0.76	0.49
I38 (Men)	0.66	0.56	0.56	0.81
I38 (Women)	0.31	0.14	0.40	0.14
I39 (Men)	2.79	2.19	11.60	4.63
I39 (Women)	2.03	1.71	1.71	4.00
I40	1.80	1.14	1.14	4.70
I41 (Total)	0.67	0.37	2.14	0.37
I41 (Men)	0.89	0.27	2.02	1.08
I41 (Women)	0.76	0.00	0.00	0.24
I42 (Hotels)	0.00	0.00	25.22	12.53
I42 (Boarding houses)	-	-	-	-
I42 (Inns)	-	-	-	-
I42 (Lodging houses)	1.34	1.34	2.21	8.07
I42 (Apartment hotels)	0.00	0.00	15.50	7.80
I42 (Tourist villages)	0.00	0.00	0.00	7.98
I42 (Tourist apartments)	0.00	0.00	6.86	1.35
I43 (Portugal)	2.68	2.68	10.89	5.35
I43 (Other countries)	1.31	0.12	23.24	10.72

Annex B. Results obtained in the best estimations

Table B1. Results obtained in the best estimations of the environmental indicators in 2017.

Indicator	Unit	Model	Predicted value	Actual value
I01 (Total)	€/1000inh.	MLP	92,219.47	94,350.00
I01 (Waste management)	€/1000inh.	MLP	61,440.99	61,736.00
I01 (Noise and vibration abatement)	€/1000inh.	RF	59.08	55.00
I01 (Protection of biodiversity and landscape)	€/1000inh.	RF	28,206.31	31,044.00
I01 (Research and development)	€/1000inh.	RF	0.00	0.00
I01 (Others)	€/1000inh.	M5P	1,486.14	1,515.00
I02 (Total)	No.	MLR	8,682,119.95	8,682,120.00
I02 (Embarked)	No.	MLR	4,345,641.82	4,346,157.00
I02 (Disembarked)	No.	MLR	4,335,962.65	4,335,963.00
I03 (Total)	No.	M5P	473,363.93	479,050.00
I03 (Rail)	No.	M5P	197,301.13	203,559.00
I03 (Road)	No.	M5P	267,099.72	275,491.00
I04 (Total)	No.	M5P	5,228,302.20	5,243,998.00
I04 (Ria Formosa)	No.	M5P	4,948,049.36	4,962,940.00
I04 (Rio Guadiana)	No.	M5P	276,767.47	281,058.00
I05	kWh/inh.	MLP	5,110.86	5,111.90
I06	toe/inh.	M5P	0.64	0.65
I07	%	MLP	99.19	99.23
I08	No.	MLP	86.73	88.00
I09	m^3/inh.	MLP	125.61	130.00
I10	m^3/inh.	M5P	101.73	101.40
I11	kg/inh.	MLP	248.21	255.00
I12 (Total)	ha	MLR	300.03	300.00
I12 (Forest stands)	ha	RF	132.55	142.00
I12 (Shrub land)	ha	MLR	114.01	114.00
I12 (Agricultural area)	ha	RF	29.35	44.00

I13 (Total)	No.	RF	109.84	110.00
I13 (Inland)	No.	MLP	0.93	1.00
I13 (Coastal/transition)	No.	MLP	108.99	109.00
I14 (Total)	1000 x €	RF	1,573.33	1,401.00
I14 (Prevention against forest fires)	1000 x €	RF	173.54	50.00
I14 (Others)	1000 x €	RF	1,613.73	1,351.00
I15	%	MLP	3.06	2.65
I16 (Transition)	No.	MLP	107.92	188.00
I16 (Embarked)	No.	MLP	302.63	326.00
I16 (Disembarked)	No.	MLR	29,287.53	29,188.00
I16 (Total)	No.	MLP	29,670.22	29,702.00

Table B2. Results obtained in the best estimations of the institutional indicators in 2017.

Indicator	Unit	Model	Predicted value	Actual value
I17	%	M5P	17.55	20.70
I18 (Residential)	%	MLP	37.13	37.60
I18 (Non-residential)	%	M5P	7.80	8.40
I19 (State)	1000 x €	RF	1,325.10	950.70
I19 (Enterprises)	1000 x €	M5P	5,064.23	4,868.70
I19 (Higher education)	1000 x €	MLR	22,960.42	21,502.10
I19 (Private non-profit institutions)	1000 x €	RF	1.80	0.00

Table B3. Results obtained in the best estimations of the economic indicators in 2017.

Indicator	Unit	Model	Predicted value	Actual value
I20 (Total)	€	MLP	2,919,211,744.09	2,947,518,306.00
I20 (Agriculture, farming of animals, hunting and forestry)	€	MLP	113,504,315.99	113,540,927.00
I20 (Mining and quarrying)	€	M5P	4,326,569.83	4,687,854.00
I20 (Manufacturing)	€	M5P	97,038,920.60	98,776,987.00
I20 (Electricity, gas, steam, cold and hot water and cold air)	€	RF	8,446,653.13	8,333,760.00
I20 (Water collection, treatment and distribution, sewerage, waste management and remediation activities)	€	MLR	89,810,392.00	87,913,082.00
I20 (Construction)	€	MLP	257,073,874.87	272,039,670.00
I20 (Wholesale and retail trade, repair of motor vehicles and motorcycles)	€	MLP	501,397,668.12	501,789,676.00
I20 (Transportation and storage)	€	M5P	92,530,017.80	92,980,254.00
I20 (Accommodation and food service activities)	€	MLP	849,984,072.57	873,508,333.00
I20 (Information and communication activities)	€	M5P	22,493,938.81	23,224,834.00
I20 (Real estate activities)	€	M5P	126,593,076.00	167,861,472.00
I20 (Consultancy, scientific and technical activities)	€	M5P	154,317,657.92	157,799,995.00
I20 (Administrative and support service activities)	€	MLR	251,505,754.51	254,462,932.00
I20 (Education)	€	MLR	34,300,101.17	34,381,423.00
I20 (Human health and social work activities)	€	M5P	123,270,055.08	127,928,784.00
I20 (Arts, entertainment, sports and recreation activities)	€	MLP	97,118,325.60	97,153,510.00
I20 (Others)	€	M5P	30,360,259.96	31,134,813.00
I21 (Hotels)	No.	M5P	7,981,812.67	7,981,933.00
I21 (Apartment hotels)	No.	M5P	4,579,216.73	4,579,264.00
I21 (Tourist villages)	No.	MLR	2,088,181.20	2,088,189.00
I21 (Tourist apartments)	No.	M5P	4,105,435.64	4,105,505.00
I22 (Total)	€	M5P	51.62	51.90
I22 (Hotels)	€	MLR	69.62	69.80
I22 (Guest houses)	€	M5P	33.60	33.80
I22 (Lodging houses)	€	MLR	82.06	83.90
I22 (Hotel apartments)	€	MLP	49.41	51.20
I22 (Tourist villages)	€	M5P	39.77	40.70
I22 (Tourist apartments)	€	M5P	33.33	34.00
I23	No.	MLP	4.50	4.50
I24 (Total)	€	MLP	17,227.13	17,303.36
I24 (Accommodation and food service activities)	€	MLP	19,797.21	19,810.48
I25 (Total)	%	M5P	0.58	0.60
I25 (Total excluding housing)	%	M5P	0.58	0.60
I25 (Total excluding unprocessed food and energy)	%	MLP	0.26	0.18
I25 (Total excluding unprocessed food)	%	MLP	0.34	0.35
I25 (Total excluding energy)	%	MLP	0.42	0.47
I25 (Unprocessed food)	%	MLP	2.75	2.75
I25 (Energy)	%	MLP	2.64	2.05
I26 (Agriculture, farming of animals, hunting and forestry)	No.	MLP	6,530.14	6,532.00
I26 (Mining and quarrying)	No.	RF	48.26	48.00
I26 (Manufacturing)	No.	M5P	1,990.46	2,009.00
I26 (Electricity, gas, steam, cold and hot water and cold air)	No.	MLP	186.74	187.00
I26 (Water collection, treatment and distribution, sewerage, waste management and remediation activities)	No.	MLR	102.33	106.00
I26 (Construction)	No.	MLP	5,719.12	5,723.00
I26 (Wholesale and retail trade, repair of motor vehicles and motorcycles)	No.	M5P	12,728.55	12,731.00
I26 (Transportation and storage)	No.	M5P	1,155.92	1,171.00
I26 (Accommodation and food service activities)	No.	MLP	13,913.82	14,431.00
I26 (Information and communication activities)	No.	MLR	593.88	608.00

I26 (Real estate activities)	No.	M5P	3,287.05	3,311.00
I26 (Consultancy scientific and technical activities)	No.	M5P	5,267.26	5,270.00
I26 (Administrative and support service activities)	No.	M5P	10,338.62	10,445.00
I26 (Education)	No.	RF	2,363.32	2,363.00
I26 (Human health and social work activities)	No.	M5P	3,726.70	3,734.00
I26 (Arts, entertainment, sports and recreation activities)	No.	M5P	2,088.76	2,114.00
I26 (Others)	No.	M5P	3,607.00	3,645.00
I27 (Agriculture, farming of animals, hunting and forestry)	No.	MLP	11,067.21	11,068.00
I27 (Manufacturing)	No.	M5P	6,415.65	6,471.00
I27 (Electricity, gas, steam, cold and hot water and cold air)	No.	MLP	337.96	338.00
I27 (Water collection, treatment and distribution, sewerage, waste management and remediation activities)	No.	M5P	2,368.43	2,497.00
I27 (Construction)	No.	M5P	16,127.36	15,972.00
I27 (Wholesale and retail trade, repair of motor vehicles and motorcycles)	No.	MLR	39,717.47	40,050.00
I27 (Transportation and storage)	No.	MLP	5,768.40	5,829.00
I27 (Accommodation and food service activities)	No.	M5P	48,244.46	48,258.00
I27 (Information and communication activities)	No.	M5P	1,472.12	1,569.00
I27 (Real estate activities)	No.	MLP	6,614.61	6,616.00
I27 (Consultancy scientific and technical activities)	No.	M5P	9,121.19	9,153.00
I27 (Administrative and support service activities)	No.	M5P	24,907.77	24,946.00
I27 (Education)	No.	RF	3,594.08	3,594.00
I27 (Human health and social work activities)	No.	MLP	7,124.49	7,212.00
I27 (Others)	No.	MLP	4,996.88	5,012.00
I28 (Total)	€	M5P	11,795,968,197.73	11,820,106,678.00
I28 (Agriculture, farming of animals, hunting and forestry)	€	MLP	291,257,593.25	292,983,079.00
I28 (Manufacturing)	€	MLP	409,278,264.55	409,813,951.00
I28 (Electricity, gas, steam, cold and hot water and cold air)	€	MLR	37,941,477.43	36,337,473.00
I28 (Water collection, treatment and distribution, sewerage, waste management and remediation activities)	€	M5P	212,864,553.09	214,124,514.00
I28 (Construction)	€	MLP	944,820,234.80	958,577,257.00
I28 (Wholesale and retail trade, repair of motor vehicles and motorcycles)	€	M5P	5,147,551,516.70	5,208,628,882.00
I28 (Transportation and storage)	€	M5P	394,481,497.23	394,661,774.00
I28 (Accommodation and food service activities)	€	M5P	2,204,591,146.47	2,286,374,704.00
I28 (Information and communication activities)	€	M5P	147,686,827.87	147,941,850.00
I28 (Real estate activities)	€	MLR	482,337,590.71	486,061,545.00
I28 (Consultancy scientific and technical activities)	€	M5P	299,706,013.66	301,007,282.00
I28 (Administrative and support service activities)	€	M5P	629,405,543.59	646,209,771.00
I28 (Education)	€	M5P	52,480,258.51	53,967,523.00
I28 (Human health and social work activities)	€	MLP	297,467,107.97	301,192,291.00
I28 (Others)	€	M5P	79,061,251.60	82,224,782.00
I29 (Total-Gender)	1000 x No.	M5P	204.80	204.80
I29 (Men-Gender)	1000 x No.	M5P	103.28	103.53
I29 (Women-Gender)	1000 x No.	M5P	99.97	101.28
I29 (Total-Sector)	1000 x No.	M5P	204.80	204.80
I29 (Agriculture, forestry and fishing-Sector)	1000 x No.	MLP	9.22	9.25
I29 (Industry-Sector)	1000 x No.	RF	19.48	19.58
I29 (Services-Sector)	1000 x No.	M5P	174.51	175.98
I30 (Accommodation and food service activities)	%	M5P	28.94	29.64
I30 (Others)	%	M5P	70.25	70.36

Table B4. Results obtained in the best estimations of the sociocultural indicators in 2017.

Indicator	Unit	Model	Predicted value	Actual value
I31 (Natural increase)	No.	M5P	-1,143.45	-1,051.00
I31 (Net migration)	No.	M5P	-2,390.70	-801.00
I32 (Monuments)	No.	MLP	122.89	126.00
I32 (Sets)	No.	M5P	20.96	21.00
I32 (Sites)	No.	MLP	25.98	26.00
I33 (Beds)	No.	M5P	2.57	2.60
I33 (Doctors)	No.	MLP	3.87	3.90
I33 (Nurses)	No.	M5P	6.16	6.20
I34 (Crimes of assault)	‰	MLP	6.78	6.80
I34 (Theft/purse snatching)	‰	MLP	1.27	1.30
I34 (Theft of and from motor vehicles)	‰	MLP	2.70	3.80
I34 (Driving a motor vehicle with a blood alcohol equal or above 1)	‰	M5P	3.46	3.50
I34 (Driving without legal documentation)	‰	MLP	1.33	1.40
I34 (Crimes against patrimony)	‰	M5P	26.69	26.80
I35 (Crimes against persons [except voluntary manslaughter])	No.	MLP	4,710.25	4,787.00
I35 (Crimes of voluntary manslaughter)	No.	MLP	0.91	5.00
I35 (Crimes against patrimony)	No.	M5P	11,722.89	11,772.00
I35 (Crimes against life in society)	No.	M5P	2,659.85	2,843.00
I35 (Crimes against the State)	No.	MLR	399.50	404.00

I35 (Crimes set out in sundry legislation)	No.	MLP	1,272.07	1,273.00
I36	-	M5P	95.06	95.83
I37	‰	MLP	20.50	22.13
I38 (Total)	No.	MLP	437,721.38	439,617.00
I38 (Men)	No.	MLR	208,723.96	209,898.00
I38 (Women)	No.	RF	229,388.63	229,719.00
I39 (Men)	No.	MLP	34,104.25	34,867.00
I39 (Women)	No.	MLR	33,372.64	33,953.00
I40	No.	MLR	286.30	289.60
I41 (Total)	1000 x No.	RF	295.31	296.40
I41 (Men)	1000 x No.	MLP	136.33	136.70
I41 (Women)	1000 x No.	MLR	159.70	159.70
I42 (Hotels)	No./km^2	M5P	133.11	133.12
I42 (Boarding houses)	No./km^2	M5P	-1.55	0.00
I42 (Inns)	No./km^2	MLR	-0.31	0.00
I42 (Lodging houses)	No./km^2	M5P	1.19	1.21
I42 (Apartment hotels)	No./km^2	M5P	76.37	76.37
I42 (Tourist villages)	No./km^2	MLR	34.82	34.82
I42 (Tourist apartments)	No./km^2	M5P	68.47	68.47
I43 (Portugal)	No./km^2	M5P	843.78	866.97
I43 (Other countries)	No./km^2	MLP	3,173.22	3,177.04