Monographic Section

DOI: http://dx.doi.org/10.25115/eea.v39i3.4723

Studies of Applied Economics
Estudios de Economía Aplicada

Volume 39-3, May 2021 // ISSN: 1133-3197

Financial Viability of Households in the Long-Term Care System in Spain: Regional Evidence

ROMÁN MÍNGUEZ SALIDO[1], RAÚL DEL POZO RUBIO[2], MARÍA DEL CARMEN GARCÍA-CENTENO[3]

[1]Department of Public Economy, Statistics and Economic Policy, UNIVERSIDAD DE CASTILLA-LA MANCHA. ESPAÑA. E-mail: roman.minguez@uclm.es

[2]Department of Economics Analysis and Finance, UNIVERSIDAD DE CASTILLA-LA MANCHA. ESPAÑA
E-mail: Raul.delPozo@uclm.es

[3]Department of Applied Mathematics and Statistics, UNIVERSIDAD SAN PABLO-CEU, ESPAÑA
E-mail: garcen@ceu.es

ABSTRACT

One of the most extensively analysed issues in recent decades has been financial catastrophe due to out-of-pocket payments (OOP) made by households to access and use health systems. This paper has two main objectives. The first is to predict the rates of financial catastrophe and determine the importance of the chosen variables for predicting the rates of catastrophe for high, medium and low income levels in the different Spanish regions. To this end, a comparison will be made between two machine learning algorithms, one based on elastic-net regressions to estimate generalised linear models; and another based on *random forest* algorithms, which makes it possible to capture the possible non-linearities and interactions that may occur in the data. The results show that the random forest is more appropriate. Based on these results, the second objective is to establish a ranking of the different regions by income level for the different categories of financial catastrophic expenditure rates, using a discrete multi-criteria decision model (PROMETHEE method).

Keywords: Catastrophic expenditure rate, Co-payment, Dependency, Machine learning algorithms, Discrete multi-criteria decision.

JEL classification: I14, N30, R10, P46

Received: 19 March 2021
Accepted: 24 May 2021

Sección Monográfica

Studies of Applied Economics
Estudios de Economía Aplicada

DOI: http://dx.doi.org/10.25115/eea.v39i3.4723

Volumen 39-3, Mayo 2021 // ISSN: 1133-3197

Viabilidad Financiera de los Hogares en el Sistema de Atención a la Dependencia en España: Evidencia Regional

ROMÁN MÍNGUEZ SALIDO[1], RAÚL DEL POZO RUBIO[2], MARÍA DEL CARMEN GARCÍA-CENTENO[3]
[1]Department of Public Economy, Statistics and Economic Policy, UNIVERSIDAD DE CASTILLA-LA MANCHA, ESPAÑA. E-mail: roman.minguez@uclm.es
[2]Department of Economics Analysis and Finance, UNIVERSIDAD DE CASTILLA-LA MANCHA. ESPAÑA
E-mail: Raul.delPozo@uclm.es
[3]Department of Applied Mathematics and Statistics, UNIVERSIDAD SAN PABLO-CEU, ESPAÑA
E-mail: garcen@ceu.es

RESUMEN

Uno de los temas más analizados en las últimas décadas ha sido el catastrofismo financiero debido a los Pagos de Bolsillo (PDB) que realizan los hogares por el acceso y utilización de los sistemas de salud. En este trabajo se persiguen fundamentalmente dos objetivos. El primero, se centra en predecir la tasa de catastrofismo financiero y obtener la importancia de las variables para predecir la tasa de catastrofismo para un nivel de renta alto, medio o bajo de las diferentes Comunidades Autónomas. Para ello, se establecerá una comparativa entre dos algoritmos *machine learning*, uno basado en regresiones *elastic-net* para estimar modelos lineales generalizados; y, otro basado en algoritmos *random forest*, que permite captar las posibles no linealidades e interacciones que se pueden producir en los datos. Los resultados muestran que es más adecuado el *random forest*. A partir de estos resultados, el segundo objetivo, se centra en establecer un ordenamiento entre las diferentes Comunidades Autónomas según su nivel de renta para las diferentes categorías de las tasas de catastrofismo mediante la utilización de un modelo de decisión multicriterio discreto (método PROMETHEE).

Palabras clave: Tasa de catastrofismo, Co-pago, Dependencia, Algoritmos machine learning, Decisión multicriterio discreta.

Clasificación JEL: I14, N30, R10, P46

Recibido: 19 de marzo de 2021
Aceptado: 24 de mayo de 2021

1. Introduction

The population's access to and use of health systems in different countries around the world often requires monetary outlays financed by families, whether through fees, co-payments or taxes (World Health Organization, 2016), known as out-of-pocket payments (OOP) (Ke et al., 2011). The amounts can be so high as to cause significant financial stress (Altice, Banegas *et al*., 2017; Yabroff, Zhao et al., 2019), thus limiting and even restricting access to and utilisation of these health services (Kolasa and Kowalczyk, 2016). Measures of financial catastrophe are used to assess the magnitude of these financial hardships (Wagstaff and van Doorslaer, 2003; Xu et al., 2003), according to which a household is defined as catastrophic when the amount of financial resources spent on health payments exceeds a certain threshold of its equivalent income (Wyszewianski, 1986). The established thresholds may vary by disease or health system, country, or point in time. The thresholds commonly used in the literature are 10%, 20%, 30%, and 40% (Wagstaff and van Doorslaer, 2003; Casado 2008; Wang et al., 2015).

Among the variables found in the literature to be associated with a higher risk of financial catastrophe due to the use of health or long-term care services are living in low- and middle-income countries (Alam and Mahal, 2014) or in regions with lower per capita income (Wagstaff and van Doorslaer, 2003; Buigut et al., 2015). The family environment strongly influences the risk of financial catastrophe: living in low-income households markedly increases said risk (Xu et al., 2003) as does living in households where the head of the household is unemployed (Del Pozo-Rubio and Jiménez-Rubio, 2019).

On the demographic side, the risk is increased when household members include older people (Scheil Adlung and Bonan, 2013; Wang et al., 2015), the chronically ill (Choi et al., 2015), elderly people with chronic diseases (Wang et al., 2015; Arsenijevic et al. 2016), and disabled (Mitra et al., 2009; Lee et al., 2016) or dependent persons (Del Pozo-Rubio et al., 2019). In Spain, substantial financial catastrophe is associated with making the co-payment for long-term care: simply having to make such a co-payment increases the probability of suffering financial stress by 18.9% (Del Pozo-Rubio et al., 2019).

The explanatory variables included in the analyses of this paper have been selected from the literature review, and are essentially socio-demographic characteristics (Xu et al., 2003; van Doorslaerl et al., 2007; Xu et al., 2007; Wang et al., 2015, Arsenijevic et al., 2016; Del Pozo-Rubio et al., 2019). These socio-demographic characteristics are gender (male, female); age; marital status (married, single, widowed, separated/divorced); educational level (very low level: no studies or incomplete primary education; low level: primary education or equivalent; medium level: secondary education, baccalaureate, vocational training or equivalent; high level: university degree or similar); economic activity (pensioner or retired; employed; unemployed; other situations (housewife, student, etc.)); monthly household income (<€500; €500-1000; €1000-1500; €1500-2000; >€2000); equivalent household members; degree of dependency (Grade I (25-49 points); Grade II (50-74 points); Grade III (75-100 points)); number of hours of informal care received; and members with mental illness.

After this brief introduction, the rest of the paper is structured as follows: In section two, a comparison is made between the predictive power of the different categories of financial catastrophic expenditure rates in the different Spanish regions using machine learning algorithms from elastic-net regressions (which enable the estimation of generalised linear models) and random forest algorithms (non-parametric algorithms to capture possible non-linearities and interactions that may be present in the data). In addition, the importance of the different variables in predicting catastrophic expenditure rates is determined. In section 3, based on the importance identified with the random forest algorithms, a discrete multi-criteria decision model is proposed, which makes it possible to establish a ranking of the different regions by income level for the different catastrophic expenditure rate categories. Finally, the main conclusions are drawn from the results obtained.

2. Machine learning algorithms for catastrophic expenditure rates

This study groups the catastrophic expenditure rates according to the corresponding region, dividing them into three groups defined by their per capita income (low, medium or high). The regions included in each group are as follows:

1. Low income: Andalusia, Castile-La Mancha, Extremadura, Murcia, the Canary Islands and Ceuta-Melilla.
2. Medium income: Valencia, Galicia, Asturias, Castile-Leon, Cantabria, and the Balearic Islands.
3. High income: La Rioja, Aragon, Catalonia, Navarra, the Basque Country, and Madrid.

The aim of this part of the study is to identify which factors best help to predict catastrophic expenditure rates within each type of region. Once these have been identified, it can be established which are the common or idiosyncratic factors that characterise the predictability of catastrophic expenditure rates within each type of region.

Firstly, looking at the frequency table of catastrophic expenditure rates, it should be noted that for rates of <10% and 10-20% there is a preponderance of individuals from low-income regions, while for the higher catastrophic expenditure rates (20-30%, 30-40%, and >40%) this proportion is reversed and there is a majority of individuals from middle- and high-income regions (see Table 1).

Table 1. Catastrophic expenditure rate frequencies

	Catastrophic expenditure rates by region				
Region income	< 10%	10-20%	20-30%	30-40%	> 40%
Low income	856	1025	351	242	280
Medium income	298	531	612	387	477
High income	235	484	313	113	319

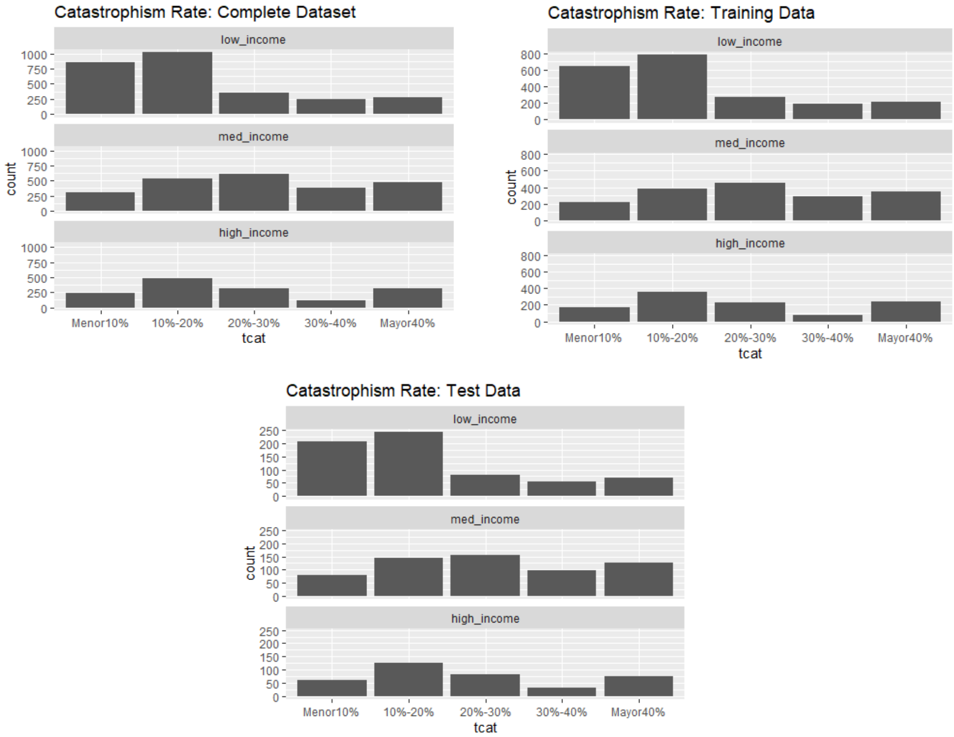
2.1. Partition of the database into training and test data

The aim of dividing the database into *training* and *test* groups is to carry out a prediction exercise with various types of algorithms to evaluate the performance in predicting the individual catastrophic expenditure rate. This prediction exercise will be repeated for the different types of region (low, medium or high income) in an attempt to identify common patterns or differences in the detected profiles.

To carry out this prediction exercise, first of all, the available database is divided into training and test sets. The training set is used to train the algorithms while the test set is used to evaluate the out-of-sample predictions. Given the availability of data for each type of region, the decision was made to take 75% of the data for the training set (25% for the test set), and to apply stratified sampling by catastrophic expenditure rate to ensure that each set is similar to the available population in the different region types (James et al., 2017; Johnson and Kuhn, 2018; Boehmke, 2020).

The following charts (see Figure 1) show that the catastrophic expenditure rate profile is indeed similar in the training and global sets, as well as in the full database, for each type of region.

Figure 1. Catastrophic expenditure rate profile



2.2. Machine learning algorithms

Once the training and test databases have been chosen, some algorithms commonly used in machine learning are selected as representative of the wide range of algorithms available for supervised learning (see James et al., 2017; Dinov, 2019; Boehmke, 2020 for recent reviews). These algorithms are used in cases where a response variable is to be predicted as a function of predictor variables (or *features*). To focus the comparison, two types of algorithms have been chosen as representative of the bias-variance trade-off associated with the different types of models according to their associated complexity and non-linearity (Hastie et al., 2009; Efron and Hastie, 2016; James et al., 2017).

1. Algorithms based on regularised parametric regression models for automatic variable selection. The elastic-net type algorithms have been chosen as representative of this group (Zou and Hastie, 2005; Friedman et al., 2010); they are based on generalised linear regression models with a penalty term on the coefficients which is a combination of LASSO (Tibshirani, 1996) and Ridge type penalties (Hoerl and Kennard, 1970). These algorithms are implemented in R language (R Core Team, 2020) through the *glmnet* package (Friedman et al., 2010).
2. Algorithms based on non-parametric models that capture possible non-linear relationships and interactions between variables without specifying a priori predetermined functional forms. As is well known, these algorithms can be appropriate for modelling complex relationships although they present high instability. In other words, they have low bias but can have high variance and are prone to overfitting, whereas parametric algorithms are more stable but more prone to specification bias (Hastie et al., 2009; KjellJohnson, 2018). There is a large number of algorithms in this group although random forest type algorithms (Breiman, 2001; Hastie et al., 2009; Boehmke, 2020) and Support Vector Machines (SVM), (Hastie et al., 2009; James et al., 2017; Boehmke, 2020), tend to be used due to their predictive performance and their ability to model highly complex relationships with few a priori assumptions. In this case, since the results obtained were better with random forest, the results for SVM[1] have not been provided. These

[1] The results are available if they are of interest to the reader.

algorithms are implemented in several R packages, e.g. *ranger* (Wright and Ziegler, 2017) for random forest, and *e1071* (Meyer et al., 2020) or *svmpath* (Hastie, 2020) for SVM. In addition, the metapackage *caret* (Kuhn, 2020; Johnson and Kuhn, 2018) integrates, in a common framework, data pre-processing and coding, training/test splitting, parameter tuning and training processes for selected algorithms, and comparison of out-of-sample results.

2.3. Regularised elastic-net regressions

Elastic-net regressions allow the estimation of generalised linear models (including, among others, linear, logistic, multinomial and Poisson regression models) by including a regularisation term on the coefficients which is a combination of L_1 and L_2 norm penalties on the beta coefficients of the model. The optimisation problem solved by the algorithm is as follows (Zou and Hastie, 2005; Friedman et al., 2010):

$$\arg\min_{\beta}\left\{\frac{1}{n}\sum_{i=1}^{n}-l_i\left(\beta\right)+\lambda\left[\left(1-\alpha\right)\left\|\beta\right\|_2/2+\alpha\left\|\beta\right\|_1\right]\right\}$$

Where $l_i(\beta)$ represents the likelihood of the i-th observation while $||\beta||_1$ and $||\beta||_2$ represent, respectively, the L_1 and L_2 norms of the beta coefficients. The elastic-net penalty term is controlled by the α parameter; if $\alpha = 1$ the LASSO regression (default in the *glmnet* package) is obtained as a particular case, while the case $\alpha = 0$ corresponds to the *Ridge* regression. Any value $0 < \alpha < 1$ will correspond to a combination of both types of regression. The parameter λ controls the overall preponderance of the penalty term in the optimisation problem.

As is well known in the literature (e.g., James et al. 2017; Boehmke, 2020), the major advantage of the LASSO penalty (based on the L_1 norm on β coefficients) over the Ridge penalty (based on the L_2 norm on the same coefficients) is that LASSO allows the β_i coefficients to be 0 and, therefore, performs automatic variable selection while Ridge regression does not eliminate any coefficients and thus does not allow variable selection. On the other hand, in cases where there is strong multicollinearity among the independent variables, LASSO tends to select only one variable from the group of correlated variables while the Ridge penalty forces the β_i coefficients of the correlated variables to be close but does not eliminate them. The elastic-net penalty seeks to combine both algorithms to retain the best features of both.

In order to train the catastrophic expenditure rate prediction models corresponding to different regions using *glmnet*, a process of tuning the parameters α and λ is carried out by means of cross-validation. In this case, for each group of regions (low, medium or high income), the training set is subdivided into 10 groups, and 9 groups are used to choose the values of α and λ, but the classification error is measured on the group that has not been used in the selection of parameters. Subsequently, the process is repeated by rotating the data sets in such a way that eventually, even though all the data have been used in the training process, the classification error has always been measured on data that are not used in the estimation.

Furthermore, before carrying out the cross-validation process, imputation techniques have to be applied to some predictors as they have missing data. In this case, we have chosen to impute these data based on the values of the 7 nearest neighbours. The measure of classification error is given by the accuracy assessed by the percentage of correct classifications in the cross-validation process. The tuning process then consists of selecting values of α and λ that maximise the accuracy for each prediction model. As particular cases, if $\alpha = 1$ it corresponds to a pure LASSO and if $\alpha = 0$ it corresponds to a pure Ridge regression. The value of λ measures the degree of penalty applied to the betas of the regression.

The following table presents a summary of the α and λ parameters selected in the tuning process after the cross-validation exercise. The *Accuracy* and *AccuracySD* columns indicate, respectively, the accuracy of the predictions (percentage of correct classifications) and their standard deviation in the cross-validation process with the training set data.

Table 2. Selected parameters for the tuning process

Region income	α	λ	Accuracy	AccuracySD
Low income	0.1	0.0063	0.5889	0.0305
Medium income	0.8	0.0079	0.4493	0.0273
High income	0.2	0.0094	0.4939	0.0394

Table 3. Confusion tables for elastic net according to region income level

Confusion Table Data Test: low income					
	<10%	10%-20%	20%-30%	30%-40%	>40%
<10%	169	50	8	12	34
10%-20%	23	175	43	12	14
20%-30%	0	9	30	27	12
30%-40%	0	0	0	3	1
>40%	15	9	1	1	9
Column Percentage Confusion Table Data Test: low income					
	<10%	10%-20%	20%-30%	30%-40%	>40%
<10%	81.64	20.58	9.76	21.82	48.57
10%-20%	11.11	72.02	52.44	21.82	20.00
20%-30%	0	7.70	36.59	49.09	17.14
30%-40%	0	0	0	5.45	1.43
>40%	7.25	3.70	1.22	1.82	12.86
Confusion Table Data Test: medium income					
	<10%	10%-20%	20%-30%	30%-40%	>40%
<10%	14	32	9	0	20
10%-20%	27	61	24	19	9
20%-30%	16	45	93	40	31
30%-40%	1	5	15	25	19
>40%	3	18	13	15	46
Column Percentage Confusion Table Data Test: medium income					
	<10%	10%-20%	20%-30%	30%-40%	>40%
<10%	40.51	9.79	5.84	0	16.00
10%-20%	34.18	42.66	15.58	19.19	7.20
20%-30%	20.25	31.47	60.39	40.40	24.80
30%-40%	1.27	3.50	9.74	25.25	15.20
>40%	3.80	12.59	8.44	15.15	36.80
Confusion Table Data Test: high income					
	<10%	10%-20%	20%-30%	30%-40%	>40%
<10%	38	23	3	0	12
10%-20%	16	70	36	10	7
20%-30%	0	11	30	18	15
30%-40%	0	0	0	0	0
>40%	7	20	14	3	40
Column Percentage Confusion Table Data Test: high income					
	<10%	10%-20%	20%-30%	30%-40%	>40%
<10%	62.30	18.55	3.61	0	16.22
10%-20%	26.23	56.45	43.37	32.26	9.46
20%-30%	0	8.87	36.14	58.06	20.27
30%-40%	0	0	0	0	0
>40%	11.48	16.13	16.87	9.68	54.05

Once the tuning parameters have been chosen, in order to assess the predictive quality of each model, the data available in the test set are used to make a confusion table for each type of region. In this way, the predictive evaluation corresponds to new data different from those used in the previous tuning process. This prevents the measures of prediction error obtained from being biased due to the overfitting that can arise from using only the data from the training set in the training process.

The following can be seen from the tables above:

- In the low income group, catastrophic expenditure rates are better predicted in the segments below 20%, while the accuracy decreases drastically in the higher groups. This can be explained by the larger number of observations in the first two segments, which enhances the models' ability to capture the dependencies in the data.
- The middle and high income groups have a similar accuracy across all segments, showing less heterogeneity than the low income group.
- The 30-40% segment shows poor predictive performance in all groups (although the upper-middle group is the one that yields the best result). This is due to a shortage of observations in this segment, and probably to the fact that the characteristics of the data corresponding to the predictor variables in this segment are similar to those of the neighbouring segments, which may be an obstacle to correct classification.
- Thus, classification errors in each segment usually occur because neighbouring segments are predicted. The exception to this rule is in the low income group for segments higher than 20%.

In general, the above results indicate that this type of algorithm can be useful for predicting new data (such as those used in the test set), although in some catastrophic expenditure rate segments (especially 30-40% and >40%) they achieve lower accuracy in the predictions. However, for catastrophe thresholds below 20% the results improve substantially.

2.4. Random forest

Having evaluated the predictive power of algorithms based on generalised linear models, the objective now is to evaluate the capacity of non-parametric algorithms to capture the possible non-linearities and interactions that may be present in the data. Among the wide range available in the machine learning world, we have chosen the random forest[2] (Breiman, 2001) and SVM type algorithms (Vapnik, 2000; Hastie et al., 2009) as representatives. The reason for choosing these algorithms is that, in both cases, they tend to yield a good predictive result for problems with little a priori information (Efron and Hastie, 2016; James et al., 2017; Boehmke 2020). Moreover, they are able to capture potential non-linear relationships and interactions between predictor variables without the need to specify a priori complex functional forms. In both cases, although it is true that we are dealing with black-box algorithms, which prevents us from having exact knowledge of the functional form obtained and the effects of each predictor variable, it is possible to obtain quantitative measures of importance associated with each predictor variable. This feature will be especially useful in the variable selection section of the multi-criteria method. Moreover, the random forest algorithm allows the researcher to work directly with both quantitative and qualitative variables without being affected by the corresponding recoding of each type of variable.

Focusing on the random forest algorithm, it is based on the use of classification (or regression) trees that partition the space of predictor variables to optimise a certain criterion (in our case, to reduce impurities by minimising the Gini index in each of the partitions in such a way that the terminal nodes are as homogeneous as possible). However, individual trees tend to be prone to

[2] As mentioned above, the results of the random forest have been compared with those of SVM, with the former being better in this case.

overfitting, as they have a very high variance; moreover, in many cases they do not show better predictive accuracy than other algorithms (e.g., James et al. 2017; García Centeno et al. 2021). To improve the capacity of the individual trees, the random forest algorithm has two additional features:

1. To reduce the high variance associated with each individual tree, bootstrapping is applied to the original data and a different tree is trained for each of the samples obtained. By so doing, the high variance is reduced by averaging the trees obtained for all the samples used. On the other hand, as in each sample there will be observations that have not been used in the training (out-of-bag observations), these observations can be used in the process of tuning the algorithm parameters.
2. If the predictor variables are highly correlated, the trees tend to reproduce similar results. In an effort to avoid this, the random forest algorithm performs, in each partition, a sampling among the potential predictor variables, randomly selecting only a subset of them from which it chooses the one that partitions best. In this way, the trees obtained are decorrelated, thus mitigating the effects of multicollinearity when averaging the trees.

In our case, two parameters have been chosen in the training process, corresponding to the number of predictor variables considered in each partition (*mtry*) and the minimum number of observations allowed in each of the final nodes of the tree (*min.nod.size*). Logically, a lower value of *min.nod.size* allows a greater depth to each individual tree although, in that case, they are more prone to overfitting. However, when training many trees in the different bootstrap samples, the propensity is mitigated, so this value is usually low in the training process. The error value measures the percentage of incorrect classifications obtained by the random forest algorithm in the out-of-bag samples that approximate the expected results in the test set. The following table summarises the 10 best models for each type of region (errors are measured as the rate of misclassifications in the unused out-of-bag observations in each bootstrap sample).

Table 4. Confusion tables for each region type

Low income			**Medium income**			**High income**		
mtry	*min.node.size*	error	*mtry*	*min.node.size*	error	*mtry*	*min.node.size*	error
2	1	0.3938	2	1	0.5155	1	1	0.4894
2	3	0.3938	2	3	0.5155	1	3	0.4894
2	5	0.3938	2	5	0.5155	1	5	0.4894
2	10	0.3938	2	10	0.5155	1	10	0.4894
1	1	0.4167	1	1	0.5255	4	1	0.4958

Once the training process has been carried out for each region, the best models obtained are applied to the test set data in order to carry out an assessment of the predictive quality of the algorithm on new data.

The results of the above confusion tables show that the predictive performance of the random forest is similar to that of the elastic-net model, although there is a slight improvement in some of the segments (especially in the >40% segment). This may indicate a greater complexity in the relationships in the data of this group with respect to the other segments.

Table 5. Confusion table for Random Forest according to income level by region

Confusion Table Data Test: low income					
	<10%	10%-20%	20%-30%	30%-40%	>40%
<10%	178	54	9	10	31
10%-20%	20	168	30	12	15
20%-30%	0	13	33	9	11
30%-40%	1	2	9	20	5
>40%	8	6	1	4	8

Column Percentage Confusion Table Data Test: low income					
	<10%	10%-20%	20%-30%	30%-40%	>40%
<10%	85.99	22.22	10.98	18.18	44.29
10%-20%	9.66	69.14	36.59	21.82	21.43
20%-30%	0	5.35	40.24	16.36	15.71
30%-40%	0.48	0.82	10.98	36.36	7.14
>40%	3.86	2.47	1.22	7.27	11.43
Confusion Table Data Test: medium income					
	<10%	10%-20%	20%-30%	30%-40%	>40%
<10%	33	18	11	2	11
10%-20%	14	62	21	7	13
20%-30%	16	38	98	35	22
30%-40%	2	5	11	38	26
>40%	14	20	13	17	53
Column Percentage Confusion Table Data Test: medium income					
	<10%	10%-20%	20%-30%	30%-40%	>40%
<10%	41.77	12.59	7.14	2.02	8.80
10%-20%	17.72	43.36	13.64	7.07	10.40
20%-30%	20.25	26.57	63.64	35.35	17.60
30%-40%	2.53	3.50	7.14	38.38	20.08
>40%	17.72	13.99	8.44	17.17	42.40
Confusion Table Data Test: high income					
	<10%	10%-20%	20%-30%	30%-40%	>40%
<10%	30	17	2	0	13
10%-20%	16	79	39	15	2
20%-30%	0	7	15	4	0
30%-40%	0	0	1	0	0
>40%	15	21	26	12	59
Column Percentage Confusion Table Data Test: high income					
	<10%	10%-20%	20%-30%	30%-40%	>40%
<10%	49.18	13.71	2.41	0	17.57
10%-20%	26.23	73.71	46.99	48.39	2.70
20%-30%	0	5.65	18.07	12.90	0
30%-40%	0	0	1.20	0	0
>40%	24.59	16.94	31.33	38.71	79.73

2.5. Measuring the importance of the variables

Once the above algorithms have been trained and their predictive power has been evaluated, the importance of each predictor variable in the resulting models is measured (Brandon and Bradley, 2020). In the random forest algorithm, the variables are ordered according to their contribution to the improvement in the Gini index (decrease in impurity at each node) using the training data. The following table shows the resulting ranking of the importance of the predictor variables.

As can be seen, the variables selected within each region are similar, with *Age*, *Dependency score*, *Monthly household income*, *Informal care hours*, *Degree of dependency*, *Education level* and *Number of equivalent household members* appearing in all regions, albeit not in the same order.

These predictor variables will be used in the classification criteria for PROMETHEE decision methods and the weights used will correspond to the measure of importance obtained.

Table 6. Importance of variables by regional income level based on Gini

Variable	Importance Low income	Importance Medium income	Importance High income
Dependency score	277.96544	239.83709	73.37118
Monthly household income	235.24561	234.59103	83.17819
Age	228.12613	194.18867	57.62070
Informal care hours	145.06378	156.71018	41.74362
Degree of dependency	119.86045	74.05475	35.71228
Number of equivalent household members	66.97857	70.38221	24.06899
Education level	55.39017	56.30598	23.78526

3. Analysis using discrete multi-criteria decision methods

After identifying the most appropriate model to determine which variables are most important in explaining the different categories of catastrophic expenditure rate in the different regions, it makes sense to conduct a comparison between the different regions for each of the established levels of GDP per capita (high, medium or low). To this end, Visual PROMETHEE, one of the most widely-used multi-criteria decision methods in practice (Brans and Mareschal, 2000; Goumans and Lygerou, 2000; Mareschal, 2013; Fernández, 2006), will be used. The aim of these methods is to conduct a pairwise comparison of different alternatives, simultaneously evaluating them using different criteria. Through this comparison, priorities among them can be established, determining which are the best alternatives and which are the worst under the criteria analysed.

A key element for the use of these methods is the decision matrix. This matrix consists of the following elements:

1. *The alternatives.* These are the elements among which a ranking is to be established. In this case, the different Spanish regions plus Ceuta and Melilla will be ranked according to their level of GDP pc (high, medium or low) and then the similarities and differences between the five catastrophic expenditure rate categories for these levels of GDP pc (less than 10%; between 10% and 20%; between 20% and 30%; between 30% and 40%; and greater than 40%) will be analysed.
2. *The criteria.* These are the variables under which each of the regions is studied. These can be maximised (i.e. the higher the value of the criterion, the better the alternative) or minimised (in which case, the lower the value of the criterion, the more preferable the alternative). In order to determine the importance of each of the criteria in establishing the ranking between regions, each of these criteria has been assigned a standardised weighting. This weighting has been calculated on the basis of the importance determined for them through the machine learning random forest algorithm.

   In this case, the variables to be considered are age (the later a person needs long-term care, the better); income (the higher the level of income, the more resources households can devote to long-term care); level of education (the higher the level of education, the better able people are to provide long-term care). The variables to be minimised are hours of informal care (the lower the average monthly hours of informal care the dependent person has to receive, the better) and the number of equivalent household members (the fewer people in the household who have to be involved in the care of a dependent person, the better).

   In order to calculate the preference indices, it is necessary to assign a generalised criterion to each of these criteria. In this case, the linear criterion without thresholds has been used, which implies that one region will be preferred to another when its results are better.
3. *The results.* These are the values obtained by evaluating each of the regions for each of the criteria in the different categories of the catastrophic expenditure rate.

With the above elements, the decision matrices are calculated. In this case, there are 15, since for each of the five catastrophic expenditure rate categories, decision matrices have been calculated for each of the three levels of GDP pc. From these decision matrices, the corresponding preference index matrices have been calculated for the different regions. The indices of these matrices are calculated as follows:

$$I\left(a_i, a_j\right) = \sum_i w_i H_i(d)$$

where a_i, a_j represent any two regions; w_i are the standardised weights corresponding to each criterion, and $H_i(d)$ is the result corresponding to each preference function.

3.1. Partial ranking (PROMETHEE I)

The next step is to obtain a partial ranking (PROMETHEE I) from these preference index matrices. To do this, it is necessary to simultaneously take into account the positive flows (i.e. the degree of preference of a region in average terms with respect to the rest of the regions when evaluated under the different criteria) and the negative flows (the purpose of which is to determine the degree to which a region is dominated by the rest, that is, the opposite effect of the positive flows).

In many cases, when comparing the two flows, incomparabilities may arise, since one region may be preferred over another on the basis of the positive flows (*Phi+*) but not on the basis of the negative flows (*Phi-*). If this occurs, it is necessary to resort to PROMETHEE II, or full ranking, which allows the calculation of net flows (*Phi*) as the difference between positive and negative flows. The results obtained for the different flows for each category of catastrophic expenditure rate can be seen in tables 7, 8, and 9 below:

Table 7. Positive, negative and net flows for a high income level in the different catastrophic expenditure rate categories

	Catastrophic expenditure rate lower than 10%			**Catastrophic expenditure rate between 10% and 20%**		
Region	**Phi+**	**Phi-**	**Phi Net**	**Phi+**	**Phi-**	**Phi Net**
Aragon	0.5680	0.3700	0.1980	0.8120	0.1880	0.6240
Basque Country	0.7000	0.3000	0.4000	0.3260	0.6560	-0.3300
Catalonia	0.8280	0.1720	0.6560	0.7460	0.2360	0.5100
La Rioja	0.1220	0.8780	-0.7560	0.0900	0.8920	-0.8020
Madrid	0.4180	0.5820	-0.1640	0.5500	0.4500	0.1000
Navarra	-0.3020	0.6360	-0.3340	0.4400	0.5420	-0.1020

	Catastrophic expenditure rate between 20% and 30%			**Catastrophic expenditure rate between 30% and 40%**		
Region	**Phi+**	**Phi-**	**Phi Net**	**Phi+**	**Phi-**	**Phi Net**
Aragon	0.7980	0.1580	0.6400	0.6840	0.2640	0.4200
Basque Country	0.4040	0.5780	-0.1740	03560	0.4960	-0.1400
Catalonia	0.7340	0.2400	0.4940	0.7520	0.1960	0.5560
La Rioja	0.1540	0.8280	-0.6740	0.1860	0.7700	-0.5840
Madrid	0.4400	0.5600	-0.1200	0.3780	0.5960	-0.2180
Navarra	0.4080	0.5740	-0.1660	0.4700	0.5040	-0.0340

	Catastrophic expenditure rate greater than 40%		
Region	**Phi+**	**Phi-**	**Phi Net**
Aragon	0.7100	0.2640	0.4460
Basque Country	0.3420	0.4980	-0.1560
Catalonia	0.5300	0.3200	0.2100
La Rioja	0.1540	0.6860	-0.5320
Madrid	0.5140	0.4680	0.0460
Navarra	0.4660	0.4800	-0.0140

Table 8. Positive, negative and net flows for a medium income level in the different catastrophic expenditure rate categories

	Catastrophic expenditure rate lower than 10%			Catastrophic expenditure rate between 10% and 20%		
Region	**Phi+**	**Phi-**	**Phi Net**	**Phi+**	**Phi-**	**Phi Net**
Asturias	0.4360	0.5640	-0.1280	0.4000	0.5640	-0.1640
Balearic Islands	0.3000	0.6680	-0.3680	0.3920	0.5720	-0.1800
Cantabria	0.1540	0.8140	-0.6600	0.0720	0.9280	-0.8560
Castile-Leon	0.7980	0.2020	0.5960	0.9280	0.0720	-0.8560
Galicia	0.7860	0.2140	0.5720	0.5280	0.4360	0.0920
Valencia	0.4940	0.5060	-0.0120	0.6260	0.3740	0.2520

	Catastrophic expenditure rate between 20% and 30%			Catastrophic expenditure rate between 30% and 40%		
Region	**Phi+**	**Phi-**	**Phi Net**	**Phi+**	**Phi-**	**Phi Net**
Asturias	0.4040	0.5280	-0.1240	0.4000	0.5060	-0.1060
Balearic Islands	0.3840	0.5480	-0.1640	0.4740	0.5080	-0.0340
Cantabria	0.0720	0.8600	-0.7880	0.1040	0.8420	-0.7380
Castile-Leon	0.8380	0.1120	0.7260	0.6580	0.3240	0.3340
Galicia	0.6080	0.3240	0.2840	0.4400	0.5060	-0.0340
Valencia	0.4080	0.3420	0.0660	0.7400	0.1300	0.6100

	Catastrophic expenditure rate greater than 40%		
Region	**Phi+**	**Phi-**	**Phi Net**
Asturias	0.3080	0.5060	-0.1980
Balearic Islands	0.3920	0.4580	-0.0660
Cantabria	0.7020	0.7420	-0.6700
Castile-Leon	0.6440	0.3560	0.2880
Galicia	0.4400	0.3740	0.0660
Valencia	0.7900	0.2100	0.5800

Table 9. Positive, negative and net flows for a low income level in the different catastrophic expenditure rate categories

	Catastrophic expenditure rate lower than 10%			Catastrophic expenditure rate between 10% and 20%		
Region	**Phi+**	**Phi-**	**Phi Net**	**Phi+**	**Phi-**	**Phi Net**
Andalusia	0.7469	0.2531	0.4939	0.7644	0.2356	0.5287
Canary Islands	0.4571	0.5429	-0.0857	0.5545	0.4455	0.1089
Castile-La Mancha	0.7653	0.2347	0.5306	0.5584	0.4416	0.1168
Ceuta and Melilla	0.1388	0.8184	-0.6796	0.1802	0.8059	-0.6257
Extremadura	0.5000	0.4551	0.0449	0.7287	0.2713	0.4574
Murcia	0.3327	0.6367	-0.3041	0.2000	0.7861	-0.5861

	Catastrophic expenditure rate between 20% and 30%			Catastrophic expenditure rate between 30% and 40%		
Region	**Phi+**	**Phi-**	**Phi Net**	**Phi+**	**Phi-**	**Phi Net**
Andalusia	0.7168	0.2277	0.4891	0.5564	0.2772	0.2792
Canary Islands	0.4515	0.4376	0.0139	0.3545	0.4238	-0.0693
Castile-La Mancha	0.2634	0.5980	-0.3347	0.4317	0.4851	-0.0535
Ceuta and Melilla	0.112	0.5267	-0.4139	0.2475	0.5030	-0.2554
Extremadura	0.1168	0.7723	0.6554	0.6376	0.3069	0.3307
Murcia	0.1129	0.5267	-0.4139	0.2733	0.5050	-0.2317

	Catastrophic expenditure rate greater than 40%		
Region	**Phi+**	**Phi-**	**Phi Net**
Andalusia	0.6099	0.2238	0.3861
Canary Islands	0.5465	0.3822	0.1644
Castile-La Mancha	0.2020	0.5327	-0.3307
Ceuta and Melilla	0.2475	0.5030	-0.2554
Extremadura	0.6178	0.3267	0.2911
Murcia	0.2614	0.5168	-0.2554

3.2. Full ranking (PROMETHEE II)

From the above flows, the rankings can be obtained for each region according to their income levels. Thus, table 10 shows the ranking of the regions with a high GDP pc level in the different catastrophic expenditure rate categories. The following findings from the analysis can be highlighted:

- La Rioja is the worst ranked for all categories.
- For all categories (except for a catastrophic expenditure rate of less than 10%), the regions in the best positions are Aragon and Catalonia.
- Madrid, Navarra and the Basque Country tend to be in an intermediate position (except for the Basque Country, which for a catastrophic expenditure rate of less than 10% is second in the ranking).

Table 11 shows, for the different catastrophic expenditure rate categories, the ranking of the regions with a medium level of GDP pc. From its analysis, we can highlight the following:

- Cantabria is the worst ranked for all catastrophic expenditure rate categories, followed by the Balearic Islands (in catastrophic expenditure rates <10%, 10-20%, and 20-30%) and Asturias (in catastrophic expenditure rates 30-40% and >40%).

Table 10. Ranking for a high GDP pc level in the different catastrophic expenditure rate categories

	Catastrophic expenditure rate categories				
Region	**<10%**	**10-20%**	**20-30%**	**30-40%**	**>40%**
Aragon	3	1	1	2	1
Basque Country	2	5	5	4	5
Catalonia	1	2	2	1	2
La Rioja	6	6	6	6	6
Madrid	4	3	3	5	3
Navarra	5	4	4	3	4

- Castile-Leon is among the best ranked for the first three catastrophic expenditure rates (<10%, 10-20%, and 20-30%), while Valencia is for the highest catastrophic expenditure rates (30-40% and >40%).
- In general, Galicia tends to be in an intermediate position for most of the catastrophic expenditure rate categories.

Table 11. Ranking for medium GDP pc level in the different catastrophic expenditure rate categories

	Catastrophic expenditure rate categories				
Region	**<10%**	**10-20%**	**20-30%**	**30-40%**	**>40%**
Asturias	4	4	4	5	5
Balearic Islands	5	5	5	3	4
Cantabria	6	6	6	6	6
Castile-Leon	1	1	1	2	2
Galicia	2	3	2	4	3
Valencia	3	2	3	1	1

Finally, table 12 shows, for the different catastrophic expenditure rate categories, the ranking of the regions with a low GDP pc level. Based on these results, we can make the following comments:

- The autonomous cities of Ceuta and Melilla are the worst ranked for all categories (except for the catastrophic expenditure rate above 40%, which corresponds to Castile-La Mancha). The next worst positioned region in the ranking for all cases is Murcia.
- Regarding the top position, there is no single one for all categories, as was the case for the high and medium levels of GDP pc. Thus, for a catastrophic expenditure rate of less than 10%, the best ranked is Castile-La Mancha; for the 10-20% and over 40% rates, it is Andalusia; while for 20-30% and 30-40% it is Extremadura.
- Finally, the Canary Islands tend to be in an intermediate position in most categories.

Table 12. Ranking for a low GDP pc level in the different catastrophic expenditure rate categories

	Catastrophic expenditure rate categories				
Region	**<10%**	**10-20%**	**20-30%**	**30-40%**	**>40%**
Andalusia	2	1	2	2	1
Canary Islands	4	4	3	4	3
Castile-La Mancha	1	3	4	3	6
Ceuta and Melilla	6	6	6	6	4
Extremadura	3	2	1	1	2
Murcia	5	5	5	5	4

4. Conclusions

The Dependency Law in Spain was designed from a state perspective. However, since it was implemented and put into effect through the Autonomous Regions, it has given rise to virtually 17 different systems of care for dependent adults. These differences can be detected in the different spheres in which said law has been constituted. First, there is the organisational-administrative sphere in terms of which ministry it is assigned to: while some regions have included the system of dependent adult care in the health department, others have chosen to manage it separately, with the relevant consequences. It should also be noted that some regions have implemented the system in full, while others have only partially replaced the existing system of social services.

A second sphere refers to the administrative waiting times for management and response; in this regard, it is worth noting the assessment times (for the score awarded to the situation of dependency and its intensity), the allocation of benefits (a range of benefit options to be granted, including financial benefits, services or a combination of the two depending on the socio-demographic characteristics and the availability of public resources in the applicant's place of residence), and finally, the effective implementation of the benefits (date on which the beneficiary begins to receive their benefit). For example, the average waiting time for the applicant until the benefit is granted was 219 days in April 2013 (Spanish Court of Auditors, 2015): a time that has increased on average to 426 days in December 2019, with significant inter-regional differences. For example, Ceuta (70 days) and Melilla (170), the Basque Country (137) and Navarre (155) are the regions with the shortest waiting times, while the Canary Islands (785 days), Extremadura (675), and Andalusia (621) have very long waiting times. The consequence of these waiting times is that in 2019 some 31,000 people died without actually receiving a dependency benefit that had been recognised and assigned to them.

Perhaps it is the third aspect that generates the greatest inter-regional differences, which is the nature of the distribution of benefits—and as an immediate consequence, the co-payments they generate can also be diametrically opposed. For example, the provision of services accounts for 47% of the benefits granted in 2019. The region with the greatest share in that year is La Rioja (69.70%), followed by Galicia (63.80%) and Andalusia (62%), while Navarra, Valencia and the Balearic Islands are the ones that allocate the smallest share of service benefits, with 21%, 22.26% and 23.56%, respectively. User co-payments are known to be substantially higher in the case of service benefits than in the case of economic benefits. This leads to the first conclusion of this paper, which takes as its basis the ranking of the different regions; generally speaking, it can be said that the positions change according to the different rates of catastrophe and levels of income.

However, within the same income level, some similarities can be highlighted. Thus, among the regions with a high income level, La Rioja is always the worst ranked, while Aragon and Catalonia, on the other hand, tend to be among the best ranked for most of the catastrophic expenditure rate categories. For those with a medium income level, Cantabria is the worst ranked, while Castile-Leon is among the best ranked for rates below 30% and Asturias for rates above 30%. Finally, among those with low income, Ceuta and Melilla is the worst ranked for all categories of the catastrophic expenditure rate, while there is no single generalisable ranking for the best positioned.

Future studies are needed with new, updated databases incorporating the socio-demographic effects of the global pandemic caused by SaRS-COV-2, in order to apply the set of methodologies presented here to assess the financial risk associated with dependent persons. This will provide policymakers with relevant information to evaluate and design social policies aimed at the financial protection of those households sensitive to financial catastrophe.

References

1. Alam, K. and Mahal, A. (2014). Economic impacts of health shocks on households in low and middle income countries: a review of the literature. *Global Health*, 10(1), 21.
2. Altice, C. K.; Banegas, M. P.; Tucker-Seely, R. D. and Yabroff, K. R. (2017). Financial hardships experienced by cancer survivors: a systematic review. *Journal of the National Cancer Institute*, 109(2), 242-249.
3. Arsenijevic, J.; Pavlova, M.; Rechel, B. and Groot, W. (2016). Catastrophic Health Care Expenditure among Older People with Chronic Diseases in 15 European Countries. *PloS One*, 11(7), e0157765.
4. Boehmke, B. (2020). Hands-on Machine Learning with R. FL: CRC Press, Boca Raton, USA.
5. Brandon, M.; Greenwell, H. and Bradley, C. (2020). Variable Importance Plots. An Introduction to the Vip Package. *The R Journal,* 12(1), 343-366. https://doi.org/10.32614/rj-2020-013.
6. Brans, J. P. and Mareschal, B. (2000). Multicriteria Decision Aid. The PROMETHEE GAIA Solution. *Journal of Decision Systems*, 12, 297-310.
7. Breiman, L. (2001). Random Forest. *Machine Learning,* 45(1), 5–32. https://doi.org/10.1023/a:1010933404324.
8. Buigut, S.; Ettarh, R. and Amendah, D. D. (2015). Catastrophic health expenditure and its determinants in Kenya slum communities. *International Journal for Equity in Health*, 14, 46.
9. Casado, D. (2008). Políticas públicas alternativas en el ámbito de la dependencia: un ejercicio de simulación para el caso español. *Hacienda Pública Española*, 186, 61-90.
10. Choi, J. W.; Choi, J. W.; Kim, J. H.; Yoo, K. B. and Park, E. C. (2015). Association between chronic disease and catastrophic health expenditure in Korea. *BMC Health Service Research*, 15:26.
11. Del Pozo Rubio, R. and Jiménez Rubio, D. (2019). Catastrophic risk associated with out-of-pocket payments for long term care in Spain. *Health Policy*, 123(6), 582-589.
12. Del Pozo Rubio, R.; Mínguez Salido, R.; Pardo-García, I. and Escribano-Sotos, F. (2019). Catastrophic long-term care expenditure: associated socio-demographic and economic factors. The *European Journal of Health Economics*, 20, 691-701.
13. Dinov, I. D. (2019). *Data Science and Predictive Analytics: Biomedical and Health Applications Using R*. Springer, Heildeberg, Berlin.
14. Efron, B. and Hastie, T. (2016). *Computer Age Statistical Inference*. Cambridge University Press, Cambridge, Mass., USA.

   https://www.ebook.de/de/product/26311758/bradley_efron_trevor_hastie_computer_age_statistical_inference.html
15. Fernandez, G. (2006). Robustness Analysis: A powerful tool in the Multiple Criteria Decision Making Field. *Newsletter of the European Working Group Multicriteria Aid of Decision*, 3(13), 3-9.
16. Friedman, J.; Hastie, T. and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software,* 33(1), 232-240. https://doi.org/10.18637/jss.v033.i01.
17. García Centeno, M. C.; Mínguez Salido R. and Del Pozo Rubio, R. (2021). The classification of profiles of financial catastrophe caused by out-of-pocket payments: A methodological approach. *Mathematics*, 9(11), 1-20. https://doi.org/10.3390/math9111170
18. Goumans, M. and Lygerou, V. (2000). An extension of the PROMETHEE method for decision making in fuzzy environment: Ranking of alternative energy exploitation projects. *European Journal of Operational Research*, 123(3), 606-613.
19. Hastie, T. (2020). *Svmpath: The SVM Path Algorithm*. https://CRAN.R-project.org/package=svmpath.

20. Hastie, T.; Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, USA.
21. Hoerl, A. E. and Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics* 12(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634.
22. James, G.; Witten, D.; Hastie, T. and Tibshirani, R. (2017). *An Introduction to Statistical Learning*. Springer-Verlag, Berlin, Germany.
    https://www.ebook.de/de/product/20292548/gareth_james_daniela_witten_trevor_hastie_robert_tibshirani_an_introduction_to_statistical_learning.html.
23. Johnson, K. and Kuhn, M. (2018). *Applied Predictive Modeling*. Springer, New York, USA. https://www.ebook.de/de/product/20211095/kjell_johnson_max_kuhn_applied_predictive_modeling.html.
24. Ke, X.; Saksena, P. and Holly, A. (2011). *The determinants of health expenditure: a country-level panel data analysis*. World Health Organization, Geneva, Switzerland.
25. Kolasa, K. and Kowalczyk, M. (2016). Does cost sharing do more harm or more good?-a systematic literature review. *BMC Public Health*, 16(1), pp. 992.
26. Kuhn, M. (2020). Caret: Classification and Regression Training. https://CRAN.R-project.org/package=caret.
27. Lee, J.-E.; Shin, H.-I.; Do, Y.K. and Yang, E.J. (2016). Catastrophic Health Expenditures for Households with Disabled Members: Evidence from the Korean Health Panel. *Journal of Korean Medical Science*, 31(3), pp. 336-344.
28. Mareschal, B. (2013). Promethee Methods. Visual PROMETHEE- versión 1.4.
29. Meyer, D.; Dimitriadou, E.; Hornik, K.; Weingessel, A. and Friedrich, L. (2020). e1071: Miscelanea Functions of the Department of Statistics, Probability Theory Group (formerly: E1071), TU Wien. https://CRAN.R-project.org/package=e1071.
30. Mitra, S.; Findley, P.A. and Sambamoorthi, U. (2009). Health Care Expenditures of Living With a Disability: Total Expenditures, Out-of-Pocket Expenses, and Burden, 1996 to 2004. *Archives of Physical Medicine and Rehabilitation*, 90(9), 1532-1540.
31. R Core Team. (2020). R: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
32. Scheil-Adlung, X. and J. Bonan (2013). Gaps in social protection for health care and long-term care in Europe: Are the elderly faced with financial ruin? *International Social Security Review*, 66(1), 25-48.
33. Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58(1), 267–88. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.
34. Tribunal De Cuentas De España. (2015). *Informe de Fiscalización 977 de la gestión económico-financiera y de la aplicación de la Ley 39/2006, de 14 de diciembre, de promoción de la Autonomía Personal y Atención a las personas en situación de dependencia*. Tribunal de Cuentas de España, Madrid, Spain. (in Spanish)
35. Van Doorslaer, E.; O'Donnell, O.; Rannan-Eliya, R. P.; Somanathan, A.; Adhikari, S. R.; Garg, C. C.; Harbianto, D.; Herrin, A. N.; Huq, M. N.; Ibragimova, S.; Karan, A.; Lee, T. J. et al. (2007). Catastrophic payments for health care in Asia. *Health Economics*, 16, 1159-1184.
36. Vapnik, V. N. (2000). *The Nature of Statistical Learning Theory*. Springer-Verlag GmbH, Berlin, Germany.
    https://www.ebook.de/de/product/2082300/v_n_vapnik_the_nature_of_statistical_learning_theory.html.

37. Wagstaff, A. and Van Doorslaer, E. (2003). Catastrophe and impoverishment in paying for health care: with applications to Vietnam 1993-1998. *Health Economics*, 12(11), 921-934.

38. Wang, Z.; Li, X. and Chen, M. (2015). Catastrophic health expenditures and its inequality in elderly households with chronic disease patients in China. *International Journal for Equity in Health*, 14:8.

39. World Health Organization. (2016). *World Health Statistics 2016: Monitoring Health for the SDGs Sustainable Development Goals*, World Health Organization, Geneva, Switzerland.

40. Wright, M. N. and Ziegler, A. (2017). Ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software*, 77(1), 14-30. https://doi.org/10.18637/jss.v077.i01.

41. Wyszewianski, L. (1986). Families with catastrophic health care expenditures. *Health Services Research*, 21(5), 617.

42. Xu, K.; Evans, D. B.; Kawabata, R.; Zeramdini, J.; Klavus, H. and Murray, C. J. (2003). Household catastrophic health expenditure: a multicountry analysis. *Lancet*, 362(9378), 111-117.

43. Xu, K.; Evans, D. B.; Carrin, G.; Aguilar Rivera, A. M.; Musgrove, P. and Evans, T. (2007). Protecting households from catastrophic health spending. *Health Affairs*, 26, 972-983

44. Yabroff, K. R.; Zhao, J.; Han, X. and Zheng, Z. (2019). Prevalence and correlates of medical financial hardship in the USA. *Journal of General Internal Medicine*, 34(8), 1494-1502.

45. Zou, H. and Hastie, T. (2005). Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67 (2), 301–20. https://doi.org/10.1111/j.1467-9868.2005.00503.x.