Journal of Informetrics 15 (2021) 101107

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Use of classification trees and rule-based models to optimize the funding assignment to research projects: A case study of UTPL

Roberto Fernandez Martinez[a,∗], Ruben Lostado Lorza[b], Ana Alexandra Santos Delgado[c], Nelson Piedra[d]

[a] *Department of Electrical Engineering, University of the Basque Country UPV/EHU, Bilbao, Spain*
[b] *Department of Mechanical Engineering, University of La Rioja, Spain*
[c] *Department of Business Studies, Universidad Tecnica Particular de Loja, Loja, Ecuador*
[d] *Department of Computer Science, Universidad Tecnica Particular de Loja, Loja, Ecuador*

ARTICLE INFO

Article history:
Received 6 July 2020
Received in revised form 20 October 2020
Accepted 27 October 2020
Available online 5 December 2020

Keywords:
Automated logical classification
Resampling methods
Funding assignment
Unbalanced dataset
Model validation methodology

ABSTRACT

In the process of funding research projects, two important factors must be studied. First, experts judges the potential value of a project. Secondly, the research ability is judged by the applicants previous research activity. The most appropriate way to assign the appropriate amount of money to project proposals is always a difficult decision. This work focuses on the second factor based on classifying the researchers previous research activity on an automated logical classification (accepted, rejected) resolving conflicts of interests between administration and applicants and helping in the decision-making process. As the class in these kinds of studies is usually unbalanced, because there are fewer accepted projects than rejected projects, how the use of an imbalanced dataset or a balanced dataset affects to the models is investigated by using several resampling methods. Later, several trees and rule-based machine learning techniques are used to create classification models. This is based on information from the faculty members information of the "Technical Particular University of Loja (UTPL)," in cases, with balanced datasets and those with unbalanced datasets. Multivariate analysis, feature selection, algorithm parameter tuning and validation methods are used to achieve robust classification models. The most accurate results are obtained with a rules-based model and use of the C5.0 algorithm. As the latter provides acceptable accuracy, close to 95 % when predicting both classes and to 99 % when predicting the accepted projects class, both the methodology and final model are validated.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

Success in research policies can lead to obtaining competitive advantages and an improvement in the economic situation of institutions or even countries. Thus, the possible increase in the level of investment that is brought about through these kind of projects and the funding distribution has been studied by governments, universities and privates companies in many

∗ Corresponding author at: Department of Electrical Engineering, College of Engineering of Bilbao, University of the Basque Country, UPV/EHU, Engineeer Torres Quevedo Square, 48013, Bilbao, Spain.

E-mail addresses: roberto.fernandezm@ehu.es (R. Fernandez Martinez), rlostado@unirioja.es (R. Lostado Lorza), aasantos@utpl.edu.ec (A.A. Santos Delgado), nopiedra@utpl.edu.ec (N. Piedra).

https://doi.org/10.1016/j.joi.2020.101107
1751-1577/© 2020 Elsevier Ltd. All rights reserved.

countries (Chubb & Reed, 2018; Cruz-Castro & Sanz-Menendez, 2016; Jung & Seo, 2010; Liu et al., 2016; Sandström & Van den Besselaar, 2018). In order to achieve this success, the funding departments of these organizations must decide on the research projects to which resources should be allocated in order to optimize this research. This is especially true for public institutions that need to demonstrate that their funding is directed to excellent scientists who produce excellent results (Cook et al., 2005; European Science Foundation, 2012; Helgesen & Reinhardt, 2012; Hutchison-Krupat & Kavadias, 2015).

Proposals are usually submitted to a funding agency and reviewed by experts in the topic (Braun, 1998; Hirzel et al., 2018; Zhu et al., 2015). Thus, funding departments need to have a transparent and systematic protocol with which to evaluate and prioritize research projects for funding. This protocol usually has two parts. The first is related with criteria that measure the quality of the proposal (Cartier et al., 2018; Henriksen & Traynor, 1999; Tuffaha, Andronis, & Scuffham, 2017, 2018). The second part involves criteria that measure the quality of the curriculum vitae of the research team members (Molleri, Petersen, & Mendes, 2018; Wolf et al., 2013).

The funding decision depends largely on the results of the peer review of the proposal. Although part of the decision depends also given on the work group capacity to perform tasks and solve problems. The process to assess people suitability to carry out the development of a research project is not a simple task. Some studies have evaluated whether the classical evaluation criteria are appropriate for the measurement of research development (Kulczycki, Korzeń, & Korytkowski, 2017; Martensson et al., 2016), although the decision of this part of the proposal is usually based on variables that define the scientific production of the work team in previous projects.

Due to the controversy of this point, different approaches has been studied and discussed. Several studies have discussed whether the publications should be attributes evaluated in the allocation of funds (Butler, 2003; Sandström & Van den Besselaar, 2016) or a panel of experts should decide the worth of the research team (Győrffy, Herman, & Szabó, 2020). At the same time, other studies propose several ways to evaluate publications, whether by quantity features or by quality features (Ahlgren, Colliander, & Persson, 2012; Subochev, Aleskerov, & Pislyakov, 2018). Analyzing metrics through expert score ranking and through data analysis and machine learning techniques (Kulczycki et al., 2017; Saarela & Kärkkäinen, 2020; Tohalino, Quispe, & Amancio, 2020). Also pointing out in many of these works, that a problem with expert-based ranking is the fact that they are not objective in every case due to the great variety of disciplines found within the same evaluation system (Saarela et al., 2016).

Therefore, the purpose of this paper is to introduce an approach that can classify the basis and evolution of the quality of research performance of a faculty member, with the aim of making a more objective decision system based on available research metrics and other relevant metadata characterizing the research performance and diffusion of the faculty members. The following work attempts, by means of automatic learning classification techniques (Bishop, 2006), to generate a model that automatically classifies faculty members or those on the research staff of a university (in this study case, the Technical Particular University of Loja in Ecuador (UTPL)) as suitable or unsuitable to receive research project funding. The decision of the model should not be exclusive, but should be made in combination with the quality evaluation of the project, in view of the information for decision that is given to the university financing department. In this case the classification system is supervised, because it has been six years since the university has gathered information concerning the most significant research indicators. Also, the class of each of the instances is known, as who was granted with funds during each year based on research merits is known.

In recent years, the use of supervised automatic classification techniques has become more pronounced in general problems focused on decision-making, but also in problems similar to the one studied (Treeratpituk & Giles, 2009; Tüselmann, Sinkovics, & Pishchulov, 2015; Zhang et al., 2020). The models created with these techniques can work without any type of human intervention, which allows decisions to be more objective and also reduces the cost associated with the availability of these human resources (Canhoto & Clear, 2020). These techniques also allow working with a higher number of features that explain the research quality of the researcher providing more information to the problem. Therefore, information from various approaches can be applied to the classification to improve the explanation of the variable to be classified (Ebadi & Schiffauerova, 2016; Ebadi et al., 2020; King, 1987).

Several studies have examined the evaluation of research projects using mathematical modeling or machine learning techniques (Dweiri & Kablan, 2006; Ebadi et al., 2020; Seo, Lee, & Lee, 2017; Wang et al., 2017). Many of these evaluations are not easy to understand and apply to for employees of institutions, since the structure of these models is quite complex. Thus, the idea of this work is to create an easily understandable and applicable model with which to classify each faculty member for his or her suitability to receive funds for research projects. This suitability, or lack thereof, is based on their previous research performance and scientific production. To achieve this goal, machine learning technique-based on decision trees and rules are used, as they can obtain classification models that are easy to understand and provide great accuracy (Breiman, 1996; Pang & Gong, 2009; Quinlan, 1993; Saarela & Kärkkäinen, 2020).

For that purpose, a training-testing methodology was used to calculate the models level of accuracy and to verify the models ability to generalize the prediction. First, a multivariate data analysis was performed to obtain a better understanding of the data that defined the problem, and to adapt the dataset to search for the best classification. Then a study to balance the dataset classes was undertaken, as the most important class to be proper classified was the class with the fewest number of instances in the dataset. Subsequently, a cross-validation training was conducted, in which the most significant algorithm parameters were tuned. The applied algorithms were a mixture of classical and advance decision trees and rule-based techniques. Then the accuracy of the models was calculated as a comparison measure. This criterion facilitated a comparison of the models in order to choose those that were the most accurate. Finally, the chosen models were tested

Table 1
Features related to the level of education and academic career of the faculty members.

Feature name	Feature definition	Factors-Range
Contract	Type of contract with the university	(Full time, Part time, Half time)
Degree	Highest academic degree	(PhD, PhD candidate, Master, Diploma, Degree,. . .)
TheMasDir	Number of master theses supervised	(0−9)
ThePhDDir	Number of PhD theses supervised	(0−12)

Table 2
Features related the projects in which the faculty member took part.

Feature name	Feature definition	Range
PrTotPI	Number of projects as PI role	(0−6)
PrTotTeam	Number of projects in the member teamwork role	(0−6)
PrTotPIHigh	Number of projects as PI with budgets greater than 100,000 USD	(0−3)
PrTotTeamHigh	Number of projects of work team with budgets greater than 100,000 USD	(0−2)
PrTotPIMed	Number of projects as PI with budgets greater than 7000 USD and equal to or less than 100,000 USD	(0−4)
PrTotTeamMed	Number of projects of work team with budgets greater than 7000 USD and equal to or lower than 100,000 USD	(0−7)
PrTotPILow	Number of projects as PI with budgets lower than 7000 USD	(0−2)
PrTotTeamLow	Number of projects of work team with budgets lower than 7000 USD	(0−3)
PrResearch	Number of research projects	(0−8)
PrInnovation	Number of innovation projects	(0−2)
PrLinking	Number of linking projects	(0−3)
PrConsulting	Number of consulting projects	(0−1)
FundInt	Amount of funding (USD) obtained from one's own university	(0–1,020,684)
FundExt	Amount of funding (USD) obtained from external funding sources	(0–608,797)

Table 3
Features related to the scientific production in previous projects.

Feature name	Feature definition	Range
Paper	Total number of papers published in journals indexed in academic research databases during the last year ("PapSci" + "PapReg")	(0−18)
PapSci	Number of papers published in journals indexed in academic research databases like JCR, SCI, SSCI, A&HCI, SJR during the last year	(0−17)
PapReg	Number of papers published in journals indexed in other academic research databases like SCIELO, REDALYC, LILACS, ACM, IEEE or similar during the last year	(0−7)
DisPap	Number of papers published in disclosure journals not indexed in those previously mentioned academic research databases or not indexed in any academic research databases during the last year	(0−15)
PapFirst	Number of papers published as first author of those considered in "Paper"	(0−18)
PapQ1	Number of papers published as Q1 indexed of those considered in "Paper"	(0−6)
PapQ234	Number of indexed papers published as Q2, Q3 or Q4 indexed of those considered in "Paper"	(0−5)
Books	Number of books	(0−5)
ChapBook	Number of book chapters	(0−3)
ArtWork	Number of artistic works	(0−3)
ConfNatOr	Number of attended oral presentations at national conference	(0−3)
ConfIntOr	Number of attended poster expositions at national conference	(0−4)
ConfNatPos	Number of attended oral presentations at international conference	(0−4)
ConfIntPos	Number of attended poster expositions at international conference	(0−5)
ConfPro	Number of conferences proceedings	(0−9)
AwaLoc	Number of local research awards	(0−4)
AwaNat	Number of national research awards	(0−3)
AwaInt	Number of international research awards	(0−2)
PatentNac	Number of national patents in exploitation	(0−1)
PatentInt	Number of international patents in exploitation	(0−2)
UtiModel	Number of utility models in exploitation	(0−8)

to obtain their real classification capacity with data that had not been used previously in the construction of the model (Fernandez Martinez et al., 2014; Lostado-Lorza et al., 2017).

2. Materials

For six years (2012–2017) the Office of the Vice Chancellor for Research of the UTPL gathered information related to research tasks of most of its employees. Tables 1–5show the main indicators that have been considered for this study. This information shows the faculty member level of studies, the quality of his or her scientific production and participation in research projects or evaluation of research works. Also shown are the factors in qualitative features, and the minimum and maximum values of quantitative features each year.

Table 4
Features related to the scientific organization of networks or groups.

Feature name	Feature definition	Range
GroupPI	Number of groups in which a faculty member is PI	(0–1)
GroupPar	Number of groups in which a faculty member is part of the work team	(0–6)
ReNetPar	Number of participations in research networks	(0–2)
ReNetNatCor	Number of times as coordinator in national research networks	(0–2)
ReNetIntCor	Number of times as coordinator in international research networks	(0–3)

Table 5
Features related to the evaluation of research works.

Feature name	Feature definition	Range
CommISI	Number of participations in scientific committees in ISI/Scopus journals	(0–5)
CommOth	Number of participations in scientific committees in other than ISI/Scopus journals	(0–3)
CommArt	Number of participations as art critic	(0–2)
EvaLoc	Number of evaluations as local project evaluator	(0–3)
EvaNac	Number of evaluations as national project evaluator	(0–6)
EvaInt	Number of evaluations as international project evaluator	(0–1)

During this period, the university made a call to grant funds for the research activity of the faculty members, where each of them took part once per year. Then, the university determined whether the merits obtained during the last year were of sufficient research quality or not. Finally, the selected faculty members were granted with research funds for the following year in order to have an extra support in their research activities. Based on this historical data, the university proposed that this information could be added in the process of funding research projects, where at least one of the people who were part of a research team (without limiting whether the member should be the Principal Investigator (PI) or not) would be classified as accepted.

In total, the dataset contains 6822 instances. Each instance defines the research activity of each faculty member of the university per year. These include 50 variables (two qualitative variables and 48 quantitative variables) that define a binary class ("Yes" when funds were granted and "No" when funds were not granted). The main feature of the dataset was that it was unbalanced, with 6618 of instances labeled "No" and 204 instances labeled "Yes."

3. Methods

3.1. Preprocessing sampling methods for unbalanced datasets

The distribution of unbalanced data between classes is a problem to consider when applying machine learning classification algorithms. When the database contains a very high difference of instances of each class, the classifier cannot discriminate effectively between a class with a large number of instances and a class with a low number of instances (Cordon et al., 2018; Hasanin et al., 2019; Khalilpour Darzi, Akhavan Niaki, & Khedmati, 2019). For this reason, the use of preprocessing sampling methods was helpful to improve the final results of the classifier, especially when classifying the minority class (Kim & Kim, 2018).

In this work, several data preprocessing sampling methods (random over-sampling (ROS), random under-sampling (RUS), synthetic minority over-sampling technique (SMOTE), and random over-sampling examples (ROSE)) were used to increase the accuracy of the algorithms, when the minority class was classified in order to select the technique with more accurate results.

With the random over-sampling method, new artificial samples were replicated from the original minority class samples to create a balanced dataset. At the same time, using the random under-sampling method, some original samples were removed from the original majority class samples to create a balanced dataset (Japkowicz & Stephen, 2002). Using the SMOTE method, some new arbitrarily instances of the minority class were created from the closest neighbors of the minority class samples (Chawla et al., 2002). Finally, ROSE method applied smoothed bootstrapping to create new artificial samples from the feature space neighborhood around the minority class (Lunardon, Menardi, & Torelli, 2014; Menardi & Torelli, 2014).

3.2. Supervised classification methods

The proposed approach employed several supervised classification algorithms to obtain the model that allows classifying each faculty member or staff into two categories that determine whether he or she will be able to obtain funding based on the quality of previous research activity. It is presented a short introduction to these methods:

- Basic Classification and Regression Trees (CART): CART is based on a greedy non-backtracking technique where the decision tree o rule is built following a top-down approach using the Gini impurity index as splitting criteria. Gini index value minimization criterion defines the amount of undistributed data in each division (Breiman et al., 1984).

- Bagged CART (BCART): BCART is an improvement of CART algorithm. It joins bagging techniques and CART to enhance the performance of classification models trying to reduce overfitting (Breiman, 1996). Bagging is an ensemble technique where several models are built and performed to predict bootstrapped replicas of the original dataset. Then and based on the majority voting among these predictions, the class that receives the most votes is used as the final classifier.
- C 4.5 (C45): It is the successor of the ID3 algorithm developed by Quinlan (1986). In this case, the splitting criterion uses the normalized information gain measure (GAIN) that is based on Shannon entropy. The feature that gets the highest GAIN at each node of the tree or rule is chosen to make the decision of effectively split the set of instances into subsets enriched in each class (Quinlan, 1993).
- C 5.0 (C50): It is the successor of C45. In the same way, C50 selects the attribute with the highest information gain ratio as split attribute, but now supports boosting technology. The boosting algorithm sets weight for each instance, which presents its importance. The influence of each instance on the decision tree or rule depends of the weight gives by the boosting algorithm (Kuhn & Johnson, 2013; Pang & Gong, 2009).
- Evolutionary learning of globally optimal classification Tree (evTree): CART and ID3 based algorithms are recursive partitioning greedy non-backtracking techniques that build the model in a forward stepwise search. The obtained results of previous recursive methods can lead the tree or rule to a local optimum, as splits are chosen to maximize homogeneity at the next step only. evTree uses an alternative way to search over the parameter space using evolutionary algorithms as optimization method (Grubinger, Zeileis, & Pfeiffer, 2011).
- Random Forest (RF): RF is based on an ensemble of randomized CART in which several decision trees or rules (amount selected during the tuning process) are built using a random subsample of the dataset instances. The final model is one built as a weighted average over all the optimal models selected by its predictive performance, its predictions' small bias and variance, and a low correlation of individual trees or rules (Breiman, 2001).

Then, to implement the process, the dataset that contains the 50 selected features and defines the activity of each faculty member during each year was used as input. The process consisted of two stages. They were training stage and testing stage. During the training stage, feature reduction and parameter tuning were conducted to achieve simpler and more accurate classification models. Later, during the testing stage, the constructed models were tested with instances that had not been used previously in the training to discover their real generalization capacity and possible overtraining issues. Each model was studied under some robustness criteria. Then, the results were compared and the most accurate model was identified.

3.3. Modeling and the validation process

The process that was defined as the methodology to build and validate the model was as follows. First, the dataset was normalized to facilitate a better understanding of the solution and to assign the same influence to each of the features in the output variable of the model. Then, the dataset was divided randomly into two subsets, one containing 80 % of the instances to form the training dataset and one containing 20 % of the instances to form the testing dataset. From this original training dataset, four more training datasets were created according to the four sampling techniques studied for imbalanced datasets (ROS, RUS, SMOTE and ROSE). The classification process was performed for each case, the original training dataset and the new four training datasets.

In each case, the training dataset was used to construct and train models using 50 times repeated ten-fold cross-validation. During this process, since each algorithm relies on a set of parameters that needs to be adjusted to obtain accurate model performance, some significant parameters of each algorithm were tuned to improve the accuracy and optimizing some robustness criteria. Although there are many techniques that can be used to optimize this setting (Rodriguez et al., 2019), in this work, it was decided to perform a tuning of the most influential parameters based on a grid search of their possible values. This point, although it can influence a loss of precision by not finding the optimal values since not all possible values are analyzed, allowed to work under parallel processing across multiple computers or processors, reducing the processing time and the complexity of the work.

Then, the most accurate models of each of the six algorithms for each of the training datasets were selected and tested using instances from the testing dataset. Finally, from the testing stage, the most accurate model using the most suitable balance sampling method was selected as the best classifier to define whether a researcher get into the category to receive funding for a research project according to previous research activity.

The classification models that were developed by applying this methodology were obtained with the use of R statistical software environment v3.6.0 (R Core Team, 2019).

3.4. Robustness criteria

In order to select the most accurate model of those that have been constructed using this methodology, two robustness criteria were chosen. They were the overall classification accuracy (Eq. (1)) and the un-weighted kappa statistic (Eq. (2)).

Table 6
Number of instances of each class in the original datasets and in the datasets based on the studied resampling methods.

	Class	
	"No"	"Yes"
Original dataset	6618	204
Training original dataset	5309	148
Testing original dataset	1309	56
Training ROS dataset	5309	5309
Training RUS dataset	148	148
Training SMOTE dataset	592	444
Training ROSE dataset	2710	2747

These criteria were used for the set of all instances of the dataset, and also in some cases for each of the classes separately due to the difference in the number of instances per class.

$$Accuracy = \frac{TP + TN}{TI} \tag{1}$$

$$k = \frac{p_a + p_e}{1 + p_e} \tag{2}$$

where for the overall classification accuracy, *TI*, is the number of total instances. *TP* is the number of faculty members that have the condition and test positive for it. *TN* is the number of faculty members who do not have the condition and test negative for it. And for the un-weighted kappa statistic, p_a is the relative observed agreement among raters, and p_e is the hypothetical probability of chance agreement.

4. Results

Prior to the dataset balancing, the variables of the original database were analyzed. A multivariate data analysis was conducted to find possible outlier cases or redundancies between variables. Some of the initial set of features were discarded, since they did not contribute with significant variations to improve the classification model, adding only noise. For example, some discarded features were: the year when the PhD was obtained, the faculty members department or the number of interviews explaining research projects. These features did not provide meaningful information with which to improve the class classification.

In addition, a correlation analysis of the quantitative features was undertaken. Some variables correlation were high, such as that between the total number of papers published in journals indexed in academic research databases during the last year (Paper) and the number of papers published in journals indexed in academic research databases like JCR, SCI, SSCI, A&HCI, SJR during the last year (PapSci). However, although a 94.41 % correlation was found in this case, it was concluded that breaking down the total number of papers into three categories (PapSci, PapReg and DisPap) gives more significant information to the classification models.

Finally, with the 52 selected variables, the original database was randomly divided into two datasets, a training dataset with 80 % of the instances (5309 "Yes" and 148 "No") and a testing dataset with the remaining 20 % of the instances (1309 "Yes" and 56 "No"). The only limitation of this random process is that the ratio between "Yes" instances and "No" instances of the new databases, training dataset and testing dataset, could not differ by more than a 2% from the original dataset ratio to avoid a possible bias related to the highly unbalanced classes.

4.1. Preprocessing sampling methods for unbalanced datasets

In both the original database and the training and testing datasets, there were a problem of unbalanced class (Table 6). As this can cause problems of inaccuracy when classifying the minority class using constructed models with unbalanced datasets, several resampling methods were applied to the original training dataset before beginning construction of models. Nevertheless, the testing dataset was kept as unchanged in order to compare the results obtained from the different models and, in this way, to achieve the real degree of generalization of the models.

Table 6 shows the number of instances of each of the datasets under study, the original datasets and the datasets that were obtained by using resampling methods. All datasets that were labeled as training dataset were used to build and train the models. It was also observed that, after the resampling techniques were applied, the training datasets showed balanced classes. Fig. 1 shows a multidimensional scaling analysis (MDS), a multivariate data analysis approach that was used to visualize the similarity of samples. This was accomplished by plotting points in two dimensional plots, which enabled one to see the number of instances and the relationship between both classes (Borg & Groenen, 2005).

An analysis of the new datasets was conducted at this point, since special cases appeared when using some of the resampling techniques. In some cases, when the technique was applied, the value of all instances of a variable became the same. Therefore, the new variable lost the significance for the prediction of the class. In these cases, these variables were removed from the process and the class was predicted with use of the remaining variables.

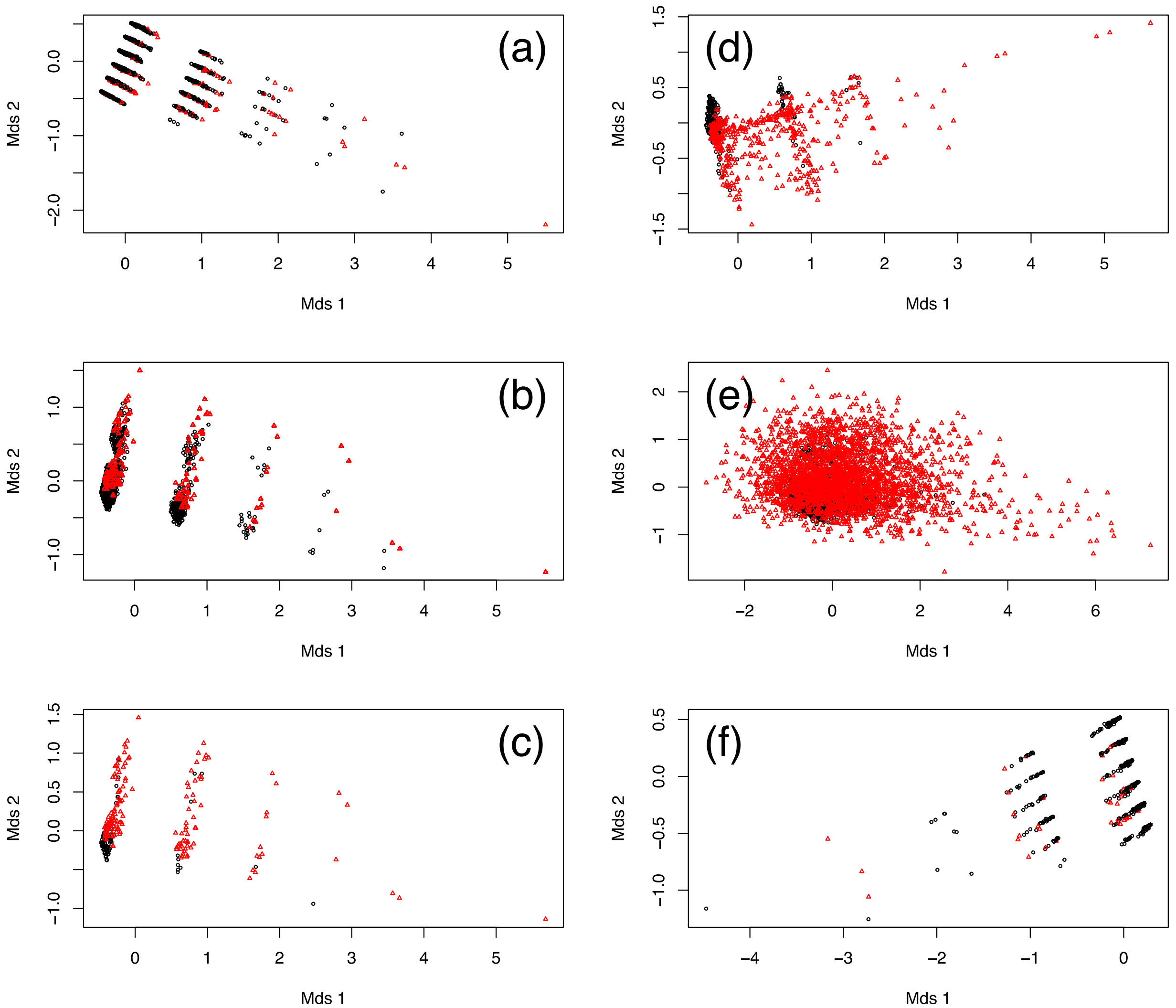

Fig. 1. MDS that shows the relationship between all dataset classes in a 2-D plot. (a) Training dataset. (b) Balanced training dataset using ROS. (c) Balanced training dataset using RUS. (d) Balanced training dataset using SMOTE. (e) Balanced training dataset using ROSE. (f) Testing dataset. (Black circles denote "No" labeled instances and red triangles denote "Yes" labeled instances). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.2. Supervised classification methods

At this point, supervised classification was undertaken with all of the training datasets (Training original dataset, Training ROS dataset, Training RUS dataset, Training SMOTE dataset and Training ROSE dataset). Several models were built and train with the proposed algorithms. The most significant parameters for each algorithm were tuned to optimize the model and improve the accuracy. These parameters are shown in Table 7, along with the range of values for each parameter. This tuning process was performed at the same time that a 50 times repeated ten-fold cross-validation process. In this way, the models accuracy was calculated and the value of the algorithm parameters was optimized at the same time. One example of this process appears in Fig. 2. It shows results of algorithm C50 that have been optimized. The process in this case uses the RUS training dataset and tunes the number of boosting Iterations, the option of making feature selection or not (winnowing or no winnowing), and the type of model (based on trees or based on rules). The overall accuracy was calculated as main decision criterion during the tuning. Also, in order to differentiate in similar cases, the 95 % confidence interval (CI) and the Kappa statistic were calculated as well. In the case of Fig. 2, it is shown that the most accurate model is based on decision rules, not applying winnowing and with a number of boosting iterations equal to 80.

During the training stage, the most accurate models built with each training dataset were selected for testing with the testing dataset that was formed by instances that were not previously used in the model training, in order to determine its real generalization capacity. Table 8 shows the results obtained by using the original dataset, Table 9 shows the results obtained by using the ROS subsampling dataset, Table 10 shows the results obtained by using the RUS subsampling dataset,

Table 7
Classification techniques applied in the analysis, with a brief definition and the studied range of parameters tuned during the training stage.

Classification technique	Parameters	Range
CART	**complexity parameter**: controls the size of the decision tree or the decision rules	0–1
BCART	No tuning parameters for this model	–
C45	**C:** confidence threshold for pruning the model	0–1
	M: minimum number of instances permissible at any leaf	10–50
C50	**boosting Iterations**: adaptive boosting number of trials	0–100
	model type: classification tree models or rule-based models	trees - rules
	winnow: feature selection	Yes - No
evTree	**complexity parameter**: controls the size of the decision tree or the decision rules	0–1
RF	**number of trees**: number of trees to grow	0–500
	randomly selected predictors: number of variables randomly sampled at each stage	0–52

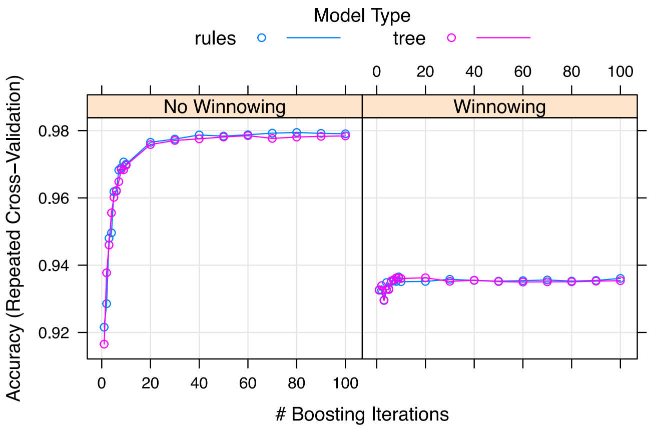

Fig. 2. Accuracy obtained during the training of the C50 algorithm, using RUS training dataset. Boosting iterations, winnowing and type of model are tuned.

Table 8
Results of the training and testing stages by using the original training dataset to build the models.

	Training			Testing		
	Accuracy	95 % CI	Kappa	Accuracy	95 % CI	Kappa
CART	99.18	(98.90, 99.40)	0.8401	98.10	(97.22, 98.75)	0.7402
BCART	99.98	(99.90, 100.00)	0.9965	98.17	(97.31, 98.81)	0.7524
C45	98.75	(98.42, 99.03)	0.7361	98.32	(97.48, 98.93)	0.7636
C50	100.00	(99.93, 100.00)	1	98.39	(97.57, 98.99)	0.7801
evTree	98.75	(98.42, 99.03)	0.7361	98.32	(97.48, 98.93)	0.7636
RF	100.00	(99.93, 100.00)	1	98.61	(97.83, 99.16)	0.8047

Table 9
Results of the training and testing stages by using the ROS training dataset to build the models.

	Training			Testing		
	Accuracy	95 % CI	Kappa	Accuracy	95 % CI	Kappa
CART	99.59	(99.44, 99.70)	0.9917	97.29	(96.28, 98.08)	0.6750
BCART	99.99	(99.95, 100.00)	0.9998	97.58	(96.62, 98.33)	0.7053
C45	99.14	(98.95, 99.31)	0.9829	97.44	(96.45, 98.21)	0.7155
C50	100.00	(99.97 100.00)	1	98.32	(97.48, 98.93)	0.7840
evTree	99.14	(98.95, 99.31)	0.9829	97.44	(96.45, 98.21)	0.7155
RF	100.00	(99.97 100.00)	1	98.17	(97.31, 98.81)	0.7611

Table 10
Results of the training and testing stages by using the RUS training dataset to build the models.

	Training			Testing		
	Accuracy	95 % CI	Kappa	Accuracy	95 % CI	Kappa
CART	93.92	(90.56, 96.36)	0.8784	89.96	(88.25, 91.51)	0.4030
BCART	100.00	(98.76 100.00)	1	92.75	(91.24, 94.07)	0.4960
C45	93.92	(90.56, 96.36)	0.8784	89.96	(88.25, 91.51)	0.4030
C50	100.00	(98.76 100.00)	1	94.65	(93.32, 95.78)	0.5770
evTree	93.92	(90.56, 96.36)	0.8784	89.96	(88.25, 91.51)	0.4030
RF	100.00	(98.76 100.00)	1	93.11	(91.64, 94.40)	0.5101

Table 11
Results of the training and testing stages by using the SMOTE training dataset to build the models.

	Training			Testing		
	Accuracy	95 % CI	Kappa	Accuracy	95 % CI	Kappa
CART	98.26	(97.27, 98.97)	0.9646	95.38	(94.13, 96.44)	0.6056
BCART	100.00	(99.64 100.00)	1	95.75	(94.54, 96.76)	0.6217
C45	97.20	(96.00, 98.12)	0.9427	95.90	(94.71, 96.89)	0.6257
C50	100.00	(99.64 100.00)	1	97.36	(96.37, 98.15)	0.7295
evTree	97.20	(96.00, 98.12)	0.9427	95.90	(94.71, 96.89)	0.6257
RF	100.00	(99.64 100.00)	1	95.97	(94.79, 96.95)	0.6302

Table 12
Results of the training and testing stages by using the ROSE training dataset to build the models.

	Training			Testing		
	Accuracy	95 % CI	Kappa	Accuracy	95 % CI	Kappa
CART	99.71	(99.52, 99.83)	0.9941	89.74	(88.01, 91.30)	0.4019
BCART	100.00	(99.93 100.00)	1	90.04	(88.32, 91.57)	0.4097
C45	99.41	(99.17, 99.60)	0.9883	95.90	(94.71, 96.89)	0.6303
C50	100.00	(99.93 100.00)	1	44.70	(34.40, 57.00)	0.1403
evTree	99.41	(99.17, 99.60)	0.9883	95.90	(94.71, 96.89)	0
RF	100.00	(99.93 100.00)	1	92.01	(90.45, 93.40)	0.4746

Table 13
Number of correctly predicted instances labeled as 'Yes' during the testing stage (In brackets the dataset with which the model was built).

	Total number of instances	TP (Original dataset)	TP (ROS dataset)	TP (RUS dataset)	TP (SMOTE dataset)	TP (ROSE dataset)
CART	56	39	41	54	53	55
BCART	56	40	42	55	52	55
C45	56	39	47	54	51	52
C50	56	41	44	55	52	55
evTree	56	39	47	54	51	0
RF	56	41	42	55	51	56

Table 11 shows the results obtained by using the SMOTE subsampling dataset and Table 12 shows the results obtained by using the ROSE subsampling dataset. The final model selection was based on the results of the training and testing stages.

These results show that, in general, all of the methods produced accurate results during the training stage and achieved a prediction success rate of close to 100 % in most of the cases. Also, when the new instances of the testing dataset were applied to the models, the accuracy was really high as well.

It was seen that during the training and testing stages that the global results showed high success rate, although some special cases suggested that the classes be analyzed separately during the testing stage. For example, the model that was built with the evTree algorithm (using the ROSE training dataset) predicted the "No" class very well (i.e., predicted that class with values close to an accuracy rate of 100 %), but did not predict the instances that were labeled as "Yes" in the same way. In another case, the model that was built with the C50 algorithm (using the ROSE training dataset) produced great training results, but very poor test results. Despite predicting the "Yes" class very well, the "No" class prediction was very inaccurate. In addition, the testing dataset is quite unbalanced. Approximately 95 % of the instances are labeled as N̈o.T̈hus, the overall results do not show clearly the model accuracy for each of its classes. Consequently, the number of instances that were correctly predicted during the testing stage from those labeled as "Yes" were studied independently. This was done in an attempt to prevent that the prediction rate of instances labeled as "No" influence the prediction rate of instances labeled as "Yes" in the global results (Table 13).

In order to select the most accurate model of those analyzed, special attention was given to the total accuracy and the kappa statistic, and to the number of True Positives (TP) obtained in the class labeled as "Yes" during the testing stage. In order to achieve this, the discussion first focused on several cases. First was the RF model that was trained with the ROSE resampled dataset, which managed to predict 100 % of the samples labeled as "Yes", although it lost reliability in class "No" prediction (91.70 % rate of TP, with 109 wrong predicted instances) during the testing stage. These 109 False Positives (FP) instances in which it was considered that the funding should be allocated to faculty members based on their previous research performance and scientific production, although they did not adequately merit it for this purpose. In addition, the other models that were trained with the ROSE resampled dataset were discarded, as despite having 55 T P instances, a high percentage of FP instances also appeared. The second, the C50 model trained with the RUS resampled dataset, was also analyzed. Despite having one False Negative (FN), with this model, a lower percentage of FP was obtained (94.50 % rate of TP, with 72 wrong predicted instances). Finally, based on these results, it was decided to select the latter, since the number of TP was 55 and the number of FP was reduced by 37 compared to the former. Table 14 shows the testing stage confusion matrix of the model that was chosen and which was based on the C50 algorithm using the training RUS dataset. This model also

Table 14
Confusion matrix obtained using the model based on the C50 algorithm (built using the RUS training dataset) during the testing stage.

	No	Yes
No	1237	1
Yes	72	55

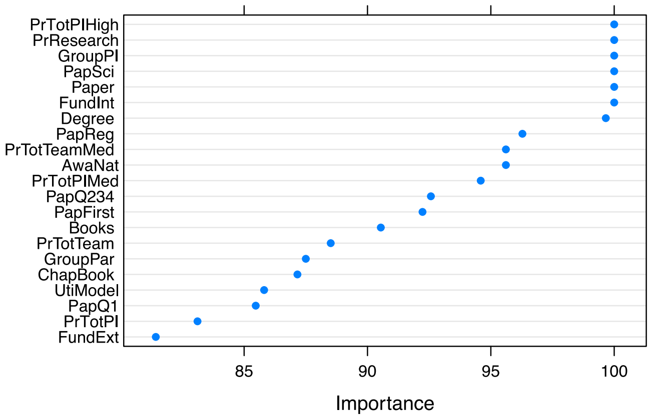

Fig. 3. Contribution of individual features to the selected classifier.

Entry 1:
Rules: Default 'No'

Rule 1: (133, lift 1.955)
PrTotTeam < 0.166
PrTotPIMed ≤ 0
Paper ≤ 0 → **class 'No'** [0.984]

Rule 2: (133, lift 1.956)
PrTotPIMed ≤ 0
Paper ≤ 0
DisPap ≤ 0 → **class 'No'** [0.984]

Rule 3: (121, lift 1.918)
PrTotTeam ≤ 0
PrTotPIHigh ≤ 0
PrTotTeamMed ≤ 0
Paper ≤ 0.111
DisPap ≤ 0 → **class 'No'** [0.966]

Rule 4: (96, lift 1.979)
PrTotTeamMed > 0
Paper > 0 → **class 'Yes'** [1.000]

Rule 5: (91, lift 1.978)
Paper > 0.111 → **class 'Yes'** [1.000]

Rule 6: (33, lift 1.942)
Paper > 0
DisPap > 0 → **class 'Yes'** [1.000]

Rule 7: (20, lift 1.909)
PrTotPIHigh > 0 → **class 'Yes'** [1.000]

Rule 8: (14, lift 1.875)
PrTotTeam > 0.166
DisPap > 0 → **class 'Yes'** [1.000]

Rule 9: (67, lift 1.855)
PrTotPIMed > 0 → **class 'Yes'** [0.940]

Fig. 4. First of the entries that define the model.

had the advantage that it enabled one to discover the influence of the each variable in the model based on the contribution of each individual feature to the classifier (Fig. 3).

One of the main properties for which the C50 algorithm obtained such accurate results was the adaptive boosting technique (Schapire & Freund, 2012) that was applied during the tuning process. Based on this technique, several classifiers were generated. In this case 80 entries were the selected attribute value to improve the accuracy of results. Then, when it was necessary to classify a new instance, each classifier entry voted for its predicted class. Then, the votes were counted to predict the final class.

For example, the first of the entries in the selected model was formed by nine rules (Fig. 4). With these nine rules, seven instances were incorrectly classified during the training (2.4 %.) Subsequent entries gave more attention to these seven failed instances and attempted to classify them correctly. Finally, when all entries were combined by voting, the final predictions provided a lower rate of error.

Fig. 4 shows the nine rules of the first entry of the model. Each of the rules provides the following information. First to be shown is the rule number. The number of training instances covered by the rule appears in brackets. The lift is obtained by dividing the rule estimated accuracy by the relative frequency of the predicted class in the training dataset. Also shown are the conditions that must be satisfied for the rule to be applicable. Next is the class that is predicted by the rule. Finally, prediction confidence of each rule appears within the rectangular brackets.

5. Conclusions

This research proposes an intelligent decision support approach for helping in the decision-making process in funding assignments. In many funding organizations when evaluating previous research activity, the current decision support system is defined by a ranking with different values and levels, which sometimes are scored subjectively. The decision evaluation here is defined by machine learning classification methods according to several features that define the research curriculum during a last period for each applicant. The approach uses a decision model based on the C5.0 classification technique. It determines the best solution when evaluating previous research activity for helping in decision-making about funding a project. This model maximizes the efficiency in the evaluation of previous research activity related with the funding assigned to each proposal applied in the Technical Particular University of Loja with great accuracy. The results are close to 99 % in accuracy when validating with former, accepted projects. The use of various algorithms shows that each problem is different and that some algorithms adapt better than others in a given situation. The proposed approach also allows one to choose the technique that best suits the problem. It has also been confirmed that unbalanced classes are difficult to classify and that the application of resampling techniques to balance the classes improves the results. The accuracy improves from around 74 % to approximately 99 % in the significant class (i.e., the class that is labeled "Yes"). In final conclusion, the work shows how research ability can be judged automatically by a classification model that is based on an applicant former research activity with few errors and making funding decisions more transparent. Therefore, this approach holds the promise of resolving conflicts of interests between the administration and applicants.

Author contributions

Roberto Fernandez Martinez: Conceived and designed the analysis; Collected the data; Performed the analysis; Wrote the paper.

Ruben Lostado Lorza: Conceived and designed the analysis.

Ana Alexandra Santos Delgado: Collected the data; Contributed data or analysis tools.

Nelson Oswaldo Piedra Pullaguari: Conceived and designed the analysis; Collected the data; Contributed data or analysis tools.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgments

The authors wish to thanks to the Basque Government for its support through the KK-2019-00033 METALCR2.

References

Ahlgren, P., Colliander, C., & Persson, O. (2012). Field normalized citation rates, field normalized journal impact and Norwegian weights for allocation of university research funds. *Scientometrics*, *92*(3), 767–780.
Bishop, C. (2006). *Pattern recognition and machine learning*. Springer-Verlag.
Borg, I., & Groenen, P. (2005). *Modern multidimensional scaling: Theory and applications* (2nd ed.). New York): Springer-Verlag.
Braun, D. (1998). The role of funding agencies in the cognitive development of science. *Research Policy*, *27*, 807–821.
Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140.
Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.
Breiman, L., et al. (1984). *Classification and regression trees*. CRC Press.
Butler, L. (2003). Explaining Australia's increased share of ISI publications—The effects of a funding formula based on publication counts. *Research Policy*, *32*, 143–155.
Canhoto, A. I., & Clear, F. (2020). Artificial intelligence and machine learning as business tools: A framework for diagnosing value destruction potential. *Business Horizons*, *63*(2), 183–193.
Cartier, Y., et al. (2018). Priority-setting in public health research funding organisations: An exploratory qualitative study among five high-profile funders. *Health Research Policy and Systems*, *16*(1), 53.
Chawla, N. V., et al. (2002). SMOTE: Synthetic minority over-sampling technique. *The Journal of Artificial Intelligence Research*, *16*, 321–357.
Chubb, J., & Reed, M. S. (2018). The politics of research impact: Academic perceptions of the implications for research funding, motivation and quality. *British Politics*, *13*(3), 295–311.
Cook, W. D., et al. (2005). Optimal allocation of proposals to reviewers to facilitate effective ranking. *Management Science*, *51*(4), 655–661.
Cordon, I., et al. (2018). Imbalance: Oversampling algorithms for imbalanced classification in R. *Knowledge-Based Systems*, *161*, 329–341.
Cruz-Castro, L., & Sanz-Menendez, L. (2016). The effects of the economic crisis on public research: Spanish budgetary policies and research organizations. *Technological Forecasting and Social Change*, *113*(Part B), 157–167.
Dweiri, F. T., & Kablan, M. M. (2006). Using fuzzy decision making for the evaluation of the project management internal efficiency. *Decision Support Systems*, *42*(2), 712–726.
Ebadi, A., & Schiffauerova, A. (2016). How to boost scientific production? A statistical analysis of research funding and other influencing factors. *Scientometrics*, *106*(3), 1093–1116.
Ebadi, A., et al. (2020). Application of machine learning techniques to assess the trends and alignment of the funded research output. *Journal of Informetrics*, *14*(2), Article 101018.
European Science Foundation. (2012). *Evaluation in research and research funding organisations: European practices. A report by the ESF Member Organisation Forum on Evaluation of Publicly Funded Research*.

Fernandez Martinez, R., et al. (2014). Use of decision tree models based on evolutionary algorithms for the morphological classification of reinforcing nano-particle aggregates. *Computational Materials Science*, *92*, 102–113.
Grubinger, T., Zeileis, A., & Pfeiffer, K. P. (2011). *evtree: Evolutionary learning of globally optimal classification and regression trees in R. Research Platform Empirical and Experimental Economics*. Universitt Innsbruck.
Győrffy, B., Herman, P., & Szabó, I. (2020). Research funding: Past performance is a stronger predictor of future scientific output than reviewer scores. *Journal of Informetrics*, *14*(3), Article 101050.
Hasanin, T., et al. (2019). Examining characteristics of predictive models with imbalanced big data. *Journal of Big Data*, *6*, 69.
Helgesen, G., & Reinhardt, A. (2012). *A guide to evaluation activities in funding agencies. Project report*. Strasbourg: ESF Member Organisation Forum on Evaluation of Publicly Funded Research.
Henriksen, A. D., & Traynor, A. J. (1999). A practical R&D project-selection scoring tool. *IEEE Transactions on Engineering Management*, *46*(2), 158–170.
Hirzel, S., et al. (2018). A decision support system for public funding of experimental development in energy research. *Energies*, *11*, 1357.
Hutchison-Krupat, J., & Kavadias, S. (2015). Strategic resource allocation. *Management Science*, *61*(2), 391–412.
Japkowicz, N., & Stephen, S. (2002). The class imbalance problem: A systematic study. *Intelligent Data Analysis*, *6*(5), 429–449.
Jung, U., & Seo, D. W. (2010). An ANP approach for R&D project evaluation based on interdependencies between research objectives and evaluation criteria. *Decision Support Systems*, *49*(3), 335–342.
Khalilpour Darzi, M. R., Akhavan Niaki, S. T., & Khedmati, M. (2019). Binary classification of imbalanced datasets: The case of CoIL challenge 2000. *Expert Systems with Applications*, *128*, 169–186.
Kim, J., & Kim, J. (2018). The impact of imbalanced training data on machine learning for author name disambiguation. *Scientometrics*, *117*, 511–526.
King, J. (1987). A review of bibliometric and other science indicators and their role in research evaluation. *Journal of Information Science*, *13*(5), 261–276.
Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. New York: Springer.
Kulczycki, E., Korzeń, M., & Korytkowski, P. (2017). Toward an excellence-based research funding system: Evidence from Poland. *Journal of Informetrics*, *11*(1), 282–298.
Liu, O., et al. (2016). An intelligent decision support approach for reviewer assignment in R&D project selection. *Computers in Industry*, *76*, 1–10.
Lostado-Lorza, R., et al. (2017). Using the finite element method and data mining techniques as an alternative method to determine the maximum load capacity in tapered roller bearings. *Journal of Applied Logic*, *24*(Part A), 4–14.
Lunardon, N., Menardi, G., & Torelli, N. (2014). ROSE: A package for binary imbalanced learning. *The R Journal*, *6*(1), 8292.
Martensson, P., et al. (2016). Evaluating research: A multidisciplinary approach to assessing research practice and quality. *Research Policy*, *45*(3), 593–603.
Menardi, G., & Torelli, N. (2014). Training and assessing classification rules with imbalanced data. *Data Mining and Knowledge Discovery*, *28*, 92122.
Molleri, J. S., Petersen, K., & Mendes, E. (2018). Towards understanding the relation between citations and research quality in software engineering studies. *Scientometrics*, *117*(3), 1453–1478.
Pang, S., & Gong, J. (2009). C5.0 classification algorithm and application on individual credit evaluation of banks. *Systems Engineering – Theory & Practice*, *29*(12), 94–104.
Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, *1*, 81–106.
Quinlan, R. (1993). *C4.5: Programs for machine learning*. San Francisco, USA: Morgan Kaufmann Publishers.
R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
Rodriguez, M. Z., et al. (2019). Clustering algorithms: A comparative approach. *PLoS One*, *14*(1), Article e0210236.
Saarela, M., & Kärkkäinen, T. (2020). Can we automate expert-based journal rankings? Analysis of the Finnish publication indicator. *Journal of Informetrics*, *14*(2), Article 101008.
Saarela, M., et al. (2016). Expert-based versus citation-based ranking of scholarly and scientific publication channels. *Journal of Informetrics*, *10*, 693–718.
Sandström, U., & Van den Besselaar, P. (2016). Quantity and/or quality? The importance of publishing many papers. *PLoS One*, *11*(11), Article e0166149.
Sandström, U., & Van den Besselaar, P. (2018). Funding, evaluation, and the performance of national research systems. *Journal of Informetrics*, *12*(1), 365–384.
Schapire, R., & Freund, Y. (2012). *Boosting: Foundations and algorithms (MIT)*.
Seo, Y., Lee, K., & Lee, S. (2017). Decision quality of the research project evaluation mechanism by using particle swarm optimization. *Management Decision*, *55*(4), 745–765.
Subochev, A., Aleskerov, F., & Pislyakov, V. (2018). Ranking journals using social choice theory methods: A novel approach in bibliometrics. *Journal of Informetrics*, *12*, 416–429.
Tohalino, J. A. V., Quispe, L. V. C., & Amancio, D. R. (2020). *Analyzing the relationship between text features and research proposal success*. https://arxiv.org/abs/2005.08254
Treeratpituk, P., & Giles, C. L. (2009). Disambiguating authors in academic publications using random forests. *Proceedings of the 9th ACM/IEEE-CS joint conference on digital libraries*, 39–48.
Tuffaha, H. W., Andronis, L., & Scuffham, P. A. (2017). Setting medical research future fund priorities: Assessing the value of research. *The Medical Journal of Australia*, *206*(2), 64–65.
Tuffaha, H. W., et al. (2018). Directing research funds to the right research projects: A review of criteria used by research organisations in Australia in prioritising health research projects for funding. *BMJ Open*, *8*(12), Article e026207.
Tüselmann, H., Sinkovics, R. R., & Pishchulov, G. (2015). Towards a consolidation of worldwide journal rankings – A classification using random forests and aggregate rating via data envelopment analysis. *Omega*, *51*, 11–23.
Wang, Q., et al. (2017). A context-aware researcher recommendation system for university-industry collaboration on R&D projects. *Decision Support Systems*, *103*, 46–57.
Wolf, B., et al. (2013). How to include criteria for productive interactions and impact on practice and society. *GAIA*, *22*(2), 104–114.
Zhang, C., et al. (2020). Applying feature-weighted gradient decent k-nearest neighbor to select promising projects for scientific funding. *Computers, Materials and Continua*, *64*(3), 1741–1753.
Zhu, W. D., et al. (2015). Research project evaluation and selection: An evidential reasoning rule-based method for aggregating peer review information with reliabilities. *Scientometrics*, *105*, 1469–1490.