

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/128967/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Martins, Isabelle D., Bahiense, Laura, Infante, Carlos E.D and Fernandes De Arruda, Edilson 2020. Dimensionality reduction for multi-criteria problems: an application to the decommissioning of oil and gas installations. *Expert Systems with Applications* 148 , 113236. 10.1016/j.eswa.2020.113236

Publishers page: <https://doi.org/10.1016/j.eswa.2020.113236>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



# Dimensionality reduction for multi-criteria problems: an application to the decommissioning of oil and gas installations

Isabelle D. Martins<sup>a,\*</sup>, Laura Bahiense<sup>b</sup>, Carlos E. D. Infante<sup>a,c</sup>, Edilson F. Arruda<sup>a,d</sup>

<sup>a</sup>*Alberto Luiz Coimbra Institute – Graduate School and Research in Engineering, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil*

<sup>b</sup>*Systems Engineering and Computer Science Program, Transportation Engineering Program, Alberto Luiz Coimbra Institute – Graduate School and Research in Engineering, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil*

<sup>c</sup>*Universidade Federal de São João del Rey, São João del Rey MG, Brazil*

<sup>d</sup>*School of Mathematics, Cardiff University, Senghennydd Rd, Cardiff CF24 4AG, UK*

---

## Abstract

This paper is motivated by decommissioning studies in the field of oil and gas, which comprise a very large number of installations and are of interest to a large number of stakeholders. Generally, the problem gives rise to complicated multi-criteria decision aid tools that rely upon the costly evaluation of multiple criteria for every piece of equipment. We propose the use of machine learning techniques to reduce the number of criteria by feature selection, thereby reducing the number of required evaluations and producing a simplified decision aid tool with no sacrifice in performance. In addition, we also propose the use of machine learning to explore the patterns of the multi-criteria decision aid tool in a training set. Hence, we predict the outcome of the analysis for the remaining pieces of equipment, effectively replacing the multi-criteria analysis by the computational intelligence acquired from running it in the

---

\*corresponding author

*Email addresses:* `isabelle@sage.coppe.ufrj.br` (Isabelle D. Martins), `laura@cos.ufrj.br` (Laura Bahiense), `prof.eduinfante@gmail.com` (Carlos E. D. Infante), `efarruda@po.coppe.ufrj.br` (Edilson F. Arruda)

training set. Computational experiments illustrate the effectiveness of the proposed approach.

*Keywords:* Oil & gas, Decommissioning, Dimensionality reduction, Feature selection, Machine learning, Multi-criteria decision analysis

---

Declaration of interest: none

## 1. Introduction

The end-of-life of oil and gas structures has become a worldwide concern. In fact, many countries that have a significant production of oil and gas have recently been discussing regulations and elaborating guidelines on the issue (Rouse et al., 2018; Oil & Gas UK, 2015; MEI, 2018; DMIRS, 2017). Among the diversity of structures that are part of the offshore system, sub-sea installations demand a special attention due to both their sensitive nature and the logistic challenges associated to the decommissioning process. In addition, decommissioning is becoming more complex as time elapses due to the increased number of pieces of equipment operating in deep waters. Sub-sea structures have different compositions and require different operations and equipment for decommissioning. Initially, possible end-of-life measures include in-situ abandonment and total or partial removal. Furthermore, total or partial removal can be achieved by means of different alternatives, which depend on the available removal techniques. Given the distinct impacts and a possibly large number of criteria to consider, selecting a decommissioning alternative becomes a rather challenging problem. To address such a problem, one must account for the great variety of technologies and materials available, as well as the distinct environmental and socio-economic conditions of each locality. In order to evaluate the alternatives, a multidisciplinary approach is needed which involves economic, envi-

ronmental, technical, social and safety aspects, among others (e.g., Martins et al., 2019b; Oil & Gas UK, 2015). Moreover, the possibly large number of stakeholders and their potentially conflicting interests can result in a very controversial process (e.g. Fowler et al., 2014; Henrion et al., 2015; Martins et al., 2019b). Typically, the decision is aided by a specialised expert system that makes use of the *multi-criteria decision analysis* (MCDA) framework (Martins et al., 2019b). Indeed, such systems have been applied in the literature for a varied set of problems (e.g., Beynon et al., 2001; Del Vasto-Terrientes et al., 2015).

To the best of our knowledge, the majority of published oil and gas decommissioning reports so far have relied on a methodology called comparative assessment (Oil & Gas UK, 2015; MEI, 2018; DMIRS, 2017), often based on subjective judgements by stakeholders with respect to a number of pre-selected criteria and sub-criteria. Generally, the alternatives are ranked by means of a weighted sum of the evaluations with regard to the selected criteria (e.g., Ineos, 2018; Shell, 2017a; Repsol, 2017). The decision making process is commonly conducted individually for each equipment, in a progression that can become rather cumbersome for large offshore systems. Given the possibly large extension of the sub-sea system, the application of multi-criteria methodology on a case-by-case basis often results in a very time-consuming process. Indeed, some reported studies have taken up to ten years to be finalised (Shell, 2017b). In addition, the evaluation of sub-criteria for each alternative is often subjective, and based on the preference of either the decision maker or a group of stakeholders (Duro et al., 2014). Hence, the process tends to become more complex, labour intensive and error prone as the number of criteria/sub-criteria increases (Waegeman et al., 2009).

This paper proposes a novel approach to address such pitfalls that embeds *machine learning* (ML) techniques into the expert systems built to aid decommissioning

decisions by means of MCDA techniques. We propose a MCDA based system with computational intelligence that enables it to learn decision patterns from the successive application of the tool in a training set. The acquired intelligence is then used to (i) rank the criteria in decreasing order of importance, which enables the decision maker to reduce the problem’s dimension by eliminating inconsequential criteria and (ii) to predict the outcome of the MCDA tool based on the criteria evaluations, effectively replacing the MCDA tool by the intelligence acquired from the data collected from its application in a training subset.

More specifically, a novel contribution of this paper is the introduction of a method for dimensionality reduction applied to decommissioning studies in the field of oil and gas. The proposed method is devised in such a way that it can also ensure the selection of the most appropriate decommissioning alternatives. The method is comprised of three steps, the first of which consists in the application of a selected multi-criteria decision analysis (MCDA) tool to identify the most appropriate decommissioning alternative for each piece of equipment within a prescribed training dataset. It is worth emphasising that this analysis takes into account evaluations of the impacts with respect to all criteria/sub-criteria. The second step feeds the training dataset to a selected supervised machine learning technique in order to obtain a model to classify the pieces of equipment, based on similarities in their characteristics and criteria/sub-criteria assessments. Lastly, the method employs feature selection to identify the smallest subset of sub-criteria which are most relevant to reaching a decision, or equivalently the largest subset of criteria that are inconsequential to the analysis, i.e. that will not significantly alter the outcome if removed. The rationale is to reduce not only the computational time required to reach a decision, but also the labour intensive process required to produce an appraisal of the performance of each piece of equipment with respect to each sub-criteria. The second

reduction is attained because only the sub-criteria within the selected subset need to be evaluated in order to estimate the MCDA outcome for a given piece of equipment. Observe also that the decommissioning process can be further abbreviated if the decision maker elects to use the decommissioning alternatives predicted by the machine learning algorithm for all installations outside the training set, instead of conducting an individual MCDA analysis for each of them individually.

### *1.1. Multi-Criteria Decision Analysis and Machine Learning*

Multi-criteria decision analysis (MCDA) provides a comprehensive framework for building computational intelligence capable of emulating the decision making process of experts and stakeholders (Carneiro et al., 2019; Del Vasto-Terrientes et al., 2015; Beynon et al., 2001). This framework has been extensively used to produce expert systems for decision support under multiple criteria (e.g., Lakhani et al., 2019; Glaize et al., 2019; Bystrzanowska and Tobiszewski, 2018; Dehe and Bamford, 2015). However, while the ensuing models are generally focused on the mapping of the decision making process (Carneiro et al., 2019, 2018), occasionally including the comparison of distinct MCDA approaches (Dehe and Bamford, 2015), they seldom exploit the patterns that emerge in multiple successive applications of the decision aid tool. In this work, we exploit these patterns by means of machine learning techniques.

The connection and complementarity of multi-criteria decision analysis (MCDA) and machine learning (ML) have been previously explored in the literature. Indeed, MCDA and ML have been used as competing techniques for flood susceptibility mapping (Tehrany et al., 2019) and fraud detection (Spathis et al., 2002). Nonetheless, combinations of MCDA and ML are comparatively rare in the literature. In an early example, Cheng (2010) employed an MCDA approach for computerised essay assess-

ment that made use of text mining techniques to derive the evaluation parameters. Later on, Hillerman et al. (2017) proposed a ML approach to identify suspicious claims from healthcare providers and employed MCDA to rank the suspicious claims in decreasing order of auditing priority. Other applications include the combination of MCDA and cellular automata to predict land use dynamics (Quesada-Ruiz et al., 2019), the use of ML to estimate the parameters of a multi-objective optimisation model employed to design electrification plans (León et al., 2019), as well as a customer satisfaction prediction model that uses MCDA to aggregate individual preferences and employs data-mining techniques to estimate missing individual evaluations.

Two previous works are particularly related to the approach we propose in this paper, namely (Jedrkiewicz et al., 2018) and (Kartal et al., 2016). The former employs machine learning to reduce the number of alternatives presented to the MCDA tool by exploiting the correlations in the data, whereas the latter uses MCDA to divide inventory products in three classes and then applies ML to predict the classes of incoming products. Our approach differs from that in (Jedrkiewicz et al., 2018) in two main aspects; firstly, we do not seek to reduce the number of available decommissioning alternatives, for the decision maker does not have that discretion. Instead, our objective is to rank the criteria in decreasing order of importance, which may allow the decision maker to reduce the dimension of the evaluation space, bearing in mind that evaluations are costly. Secondly, whilst Jedrkiewicz et al. (2018) seek patterns in the evaluation data, we look for the patterns that emerge from the successive application of the MCDA tool; hence we are also able to identify patterns in the MCDA tool. In addition, in contrast to the model in (Kartal et al., 2016), which is limited to only three classes, our model can account for any number of classes. Moreover, the model in (Kartal et al., 2016) neither ranks the criteria nor allows

dimensionality reduction.

### *1.2. Dimensionality Reduction*

Dimensionality reduction is often employed in the literature to extract useful information from very large datasets (Sorzano et al., 2014). It is comprised of feature selection and feature extraction (Khalid et al., 2014; Xu et al., 2019). While the former is concerned with selecting a meaningful subset of the original variables (e.g., Guyon and Elisseeff, 2003), the latter is concerned with synthesising a reduced subset of meaningful features from a high-dimensional space (Jolliffe, 2002; Pearson, 1901). When it comes to decommissioning problems, however, the use of dimensionality reduction is still incipient. Some papers proposed selecting criteria based upon expert judgement (ISM, 2011; Ahmed et al., 2016), while Bernstein et al. (2010) advocated a selection centred on the availability of information. Therefore, there seems to be plenty of room for the introduction of formal machine learning techniques to address variable selection in decommissioning problems.

Classification methods, in particular, find applications in a number of fields, such as finances (Ryman-Tubb et al., 2018), internet safety (Huang et al., 2018) and transportation safety (Chai et al., 2019). Specifically, they have been successfully employed in the oil and gas industry. Some examples include the utilisation of Support Vector Machines (SVM), Decision Trees (DT) and Random Forests (RF) to predict corrosion in pipeline inspection (Liu et al., 2019). In a related work, El-Abbasy et al. (2016) make use of regression analysis, artificial neural networks and DT to investigate the causes of pipeline failure. Other works were concerned with predicting oil production (Schuetter et al., 2018; Li and Chan, 2010). The former utilises classification models, such as SVM and Gradient Boosting Machines (GBM), to predict oil potential in unconventional reservoirs. The latter applies a neural-based



decision tree to forecast well production. Further applications include using RF to anticipate the development of oil and gas fields and estimate the impact to animal species (Copeland et al., 2009). Nonetheless, specific applications of machine learning to decommissioning problems have been found lacking. An additional innovation of this work is to fill this gap.

The need for ML techniques in the context of decommissioning studies, which is reinforced in this paper, has already been acknowledged in pioneering decommissioning studies (e.g., MEI, 2018; Oil & Gas UK, 2015), even though it was deferred to later work. Specifically, these references suggest clustering pieces of equipment based on similarities with respect to their characteristics, such as sub-sea status, diameter, installation data, among others. We argue that such a classification may be insufficient and should benefit from the assessment of a reduced number of criteria/sub-criteria, which may act as proxies for the characteristics of the environment surrounding the installation, which certainly plays an important role in the selection of the decommissioning alternative.

To validate the proposed approach, we introduce a dataset based on real-world data from actual sub-sea ducts, which we believe can be used in the future for benchmarking purposes. We also compare the performance of DT, RF, SVM and GBM as classification analysis tools in the proposed method. These procedures are applied in the second and third steps of the framework, and their performance is assessed and contrasted in the light of the selected dataset. In our experimental analysis, the GBM model presented auspicious results, having the best overall performance considering all features, with a mean accuracy of 82% and kappa of 72%. The further feature selection results suggest that it may be possible to use only half of the initial sub-criteria while still maintaining similar levels of performance.

Even though the evaluation and comparison of distinct MCDA approaches is

beyond the scope of this paper, we perform a preliminary experiment designed to illustrate the possible use of distinct MCDA algorithms in a decommissioning study. In the experiment, we make use of three distinct MCDA algorithms and select the alternative based on a vote. The results are preliminary and illustrate the challenges and opportunities in the use of ensemble methods.

The remainder of this paper is organised as follows. Section 2 describes problem setting for this study. Section 3 explains the general method of analysis and features a brief overview of the MCDA and ML techniques employed. The dataset of the numerical experiments is introduced in Section 4, which is followed by the presentation of the experimental results in Section 5. In that section, the ML methods are compared and a variable relevance analysis is reported. Finally, Section 6 concludes the paper.

## 2. Problem definition

Decision making for decommissioning in oil and gas platforms is generally carried out from the analysis of each piece of equipment individually (e.g. Shell, 2017a; BG Group, 2016; Xodus, 2017). Assume there is a set of decommissioning alternatives  $A = \{a_1, a_2, \dots, a_n\}$  for a given piece of equipment, and suppose the decision is to be reached based on a set of sub-criteria  $G = \{1, 2, \dots, m\}$ . In a decommissioning study, generally the sub-criteria are related to environmental, social, economic, safety and technical issues (Oil & Gas UK, 2015; MEI, 2018). The evaluation gives rise to an  $N \times M$  matrix comprised of the evaluations of each alternative with respect to each sub-criterion. As previously mentioned, the majority of published decommissioning studies so far have relied on a methodology called comparative assessment (e.g., Oil & Gas UK, 2015; MEI, 2018), which produces a performance index  $I_i$  for each

alternative  $a_i$  as follows:

$$I_i = \sum_{j=1}^m l_j g_j(a_i), \quad (1)$$

where  $l_j$  is the weight attributed to sub-criterion  $j$ ,  $1 \leq j \leq m$ , and  $g_j(a_i)$  is the evaluation of criterion  $j$  for alternative  $a_i$ ,  $1 \leq i \leq n$ . The alternatives are then ranked in decreasing order of performance index. Other MCDA methods are also applied in different decommissioning studies, such as AHP (e.g. Xodus, 2017; Repsol, 2017; Na et al., 2017), ELECTRE (e.g. Dimitrijevic et al., 2014; Soltanmohammadi et al., 2008) and PROMETHEE (e.g. Kerkvliet and Polatidis, 2016; Mergias et al., 2007). The latter methods are somewhat more complex, but can also be employed to generate a ranking of the alternatives based on performance indices. For more details on MCDA methods and analyses, refer to (Greco et al., 2005).

Regardless of the MCDA approach, a common issue in oil and gas decommissioning studies is that collecting information and producing each evaluation  $g_j(a_i)$  for each piece of equipment often takes time. Furthermore, some criteria evaluations may require multiple assessments by different, possibly conflicting, stakeholders. On top of that, the abundance of pieces of equipment in the seabed leads to an increase in complexity, especially in deep waters. Some authors also point out that a large number of criteria can render the decision making process more difficult and may generate confusion among the stakeholders (Amirshenava and Osanloo, 2018). In order to address these issues, this paper builds upon a conjecture found in (MEI, 2018; Oil & Gas UK, 2015) that similar features in distinct pieces of equipment may lead to similar choices of decommissioning alternatives. We argue, however, that similar features may not be enough, since the selection also depends upon other factors, such as the environment surrounding the piece of equipment. Fortunately, the evaluations of the alternatives for each installation can serve as proxies for those

factors. In addition, it is possible that the evaluation of a reduced set of sub-criteria may be enough to produce an accurate estimation of the course of action that would be selected if all criteria were accounted for.

Let  $E = \{e_1, e_2, \dots, e_r\}$ ,  $1 \leq r \leq \infty$ , be a small albeit representative subset of the pieces of equipment found in a given oil field pending decommission. Let  $g_j^k(a_i)$ ,  $1 \leq j \leq n$ ,  $1 \leq i \leq m$ , be the evaluation of the sub-criterion  $j$  for alternative  $i$  relative to piece of equipment  $e_k$ ,  $1 \leq k \leq r$ . Finally, denote by  $t_p$  the total number of parameters of each piece of equipment. The proposed procedure includes the three steps detailed in Algorithm 1 below:

---

**Algorithm 1 (Classification Analysis of the Training Set).**

1. *For each piece of equipment  $e_k$ , create a row vector  $e_k = (x_k, y_k)$ , where  $x_k$  is the feature vector and contains all  $t_p$  parameters and  $(m \times n)$  sub-criteria evaluations  $g_j^k(a_i)$ ,  $1 \leq j \leq n$ ,  $1 \leq i \leq m$ . Hence  $x_k$  is of dimension  $(m \times n + t_p)$ , and  $y_k$  is the alternative  $a_i^k$  assigned by the MCDA algorithm for this piece of equipment, i.e.  $y_k$  is the class label. Observe that  $e_k$  holds the entries matrix  $P$  defined in Step 2;*
  2. *Use supervised classification to divide the population  $P = e_k$ ,  $k = 1, 2, \dots, r$ , into  $m$  groups, one for each decommissioning alternative; and*
  3. *Find the characteristics and sub-criteria that most impact in the action selection by the MCDA algorithm.*
- 

It is worth of emphasizing that matrix  $P$  in Algorithm 1 is the training set of a supervised learning algorithm (e.g., Hastie et al., 2001). Note also that the training labels  $y_k$ ,  $k = 1, \dots, r$ , are assigned by a selected MCDA algorithm.

Observe that the decision maker can employ the output of Step 3 to simplify the assessment of the decommissioning alternatives for pieces of equipment outside the training set. The assessment can now be performed as a function of the subset of most relevant characteristics and sub-criteria. It is our conjecture that, in most cases, such a subset will have a rather reduced dimension when compared to the original set.

### 3. General method of analysis

The method in Algorithm 1, further detailed in Figure 1, comprises two distinct tasks: dataset pre-processing and classification. Note that Step 1 of Algorithm 1 generates a dataset where each piece of equipment is associated with both its characteristics and the scores for each alternative/sub-criterion pair, as well as the decommissioning alternative recommended by the selected MCDA tool.

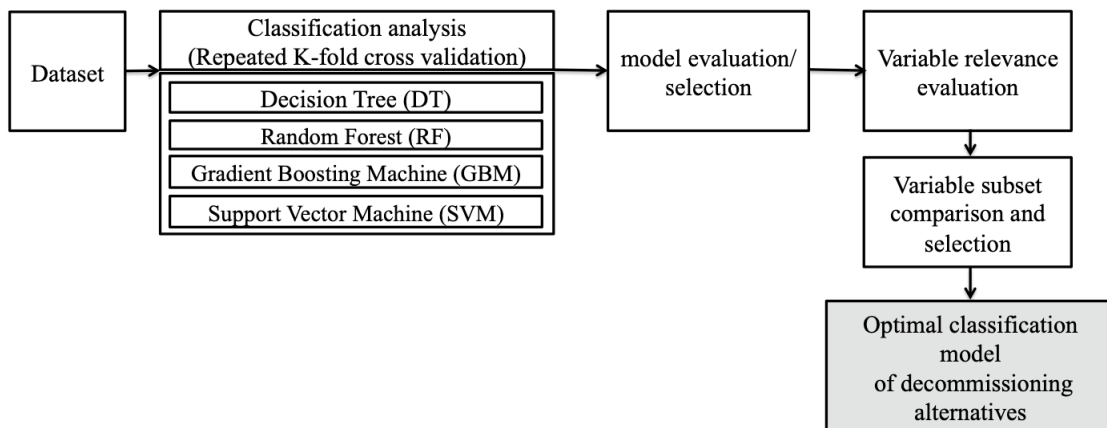


Figure 1: Methodology framework regarding steps 2-3 of Algorithm 1.

Observe in Figure 1 that some ML algorithms are applied to the dataset under a  $k$ -fold cross validation scheme (Steps 2-3 of Algorithm 1). Our case study, in particular, makes use of decision trees (DT), random forests (RF), gradient boosting

machines (GBM) and support vector machines (SVM). In the model selection step, we compare these algorithms according to selected evaluation metrics and statistical analyses, further explained in Section 3.5, and select a single model to be used in the remaining steps. A variable relevance analysis then follows for the selected model. After that, the decision maker chooses the smallest possible subset of the most relevant variables that maintains the accuracy of the model. Finally, this subset is then elected to comprise the reduced classification model in the last step.

The application of the framework in the context of a decommissioning study of an oil and gas field is depicted in Figure 2. Let  $E_s$  be the set of pieces of equipment left out of the training set, with  $s \gg r$ . It is this set that is fed to the first step in Figure 2. For each piece of equipment  $e_k \in E_s$ , a vector  $p_k$  is formed with the evaluation of all relevant sub-criteria selected in Step 3 of Algorithm 1. Then, the reduced model is applied to forecast the selected decommissioning alternative (classification step), which is the output of the last step in Figure 2.

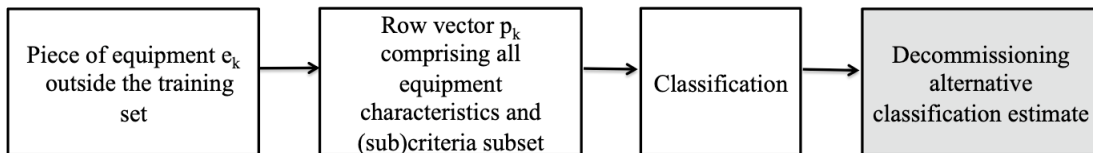


Figure 2: Classification model deployment for decommissioning.

The following sections present a brief description of the supervised methods previously mentioned.

### 3.1. Decision trees (DT)

Decision trees (Rokach and Maimon, 2008) utilise a recursive partitioning algorithm to construct the tree and have the advantage of accepting both numerical and categorical variables. It is a top-down approach that models decisions and possi-

ble consequences, including growing and pruning stages. Basically, a decision tree comprises:

- Nodes - a split performed on an attribute;
- Root - the topmost node;
- Branch - an outcome of the split; and
- Leaves - a class label.

The process consists in continuously splitting the training data into two or more descendent subsets, until a stopping criteria is reached or all classes are split. The splitting criterion is selected to identify the partition which is closest to having all elements belonging to their correct class.

In order to avoid over-fitting, one can limit the growth of the tree or prune it. Pruning is often attained by recursively snipping off the least important splits, based on the complexity parameter ( $cp$ ). On the one hand, because they depend on the observation values, decision trees are robust to outliers. They are also fairly easy to understand and interpret. On the other hand, they can be computationally expensive due to the need of identifying splits from multiple variables (Hodeghatta and Nayak, 2016).

### *3.2. Random forests (RF)*

Random forests (Breiman, 2001) belong to the class of ensemble methods. Briefly, the method consists in first bootstrapping the dataset, and selecting random samples to construct the training sample of each tree. After that, the method randomly selects the features of each tree and makes use of the Gini index, a measure of heterogeneity, to produce splits. Finally, the model utilises the samples that were not part of the

training set, the so called *out-of-bag (OOB) data*, for testing purposes. The final response is based on a vote on the decisions generated by each of the constructed trees.

It presents some benefits compared to other machine learning methods, such as requiring only two parameters, namely the number of trees (*ntree*) to be generated and the number of variables selected for each tree (*mtry*). Breiman (2001) recommends that *mtry* be equal to the square root of the total number of variables. Other benefits include reducing over-fitting when compared to decision trees, since the response is the average of several trees. Additionally, the random selection of variables has the advantage of reducing the correlation between trees (Feng et al., 2015). It may be argued that the technique is more robust by virtue of creating multiple decision trees and optimising the output to obtain a better-performing classifier (Hodeghatta and Nayak, 2016).

### 3.3. Gradient Boosting Machines (GBM)

Similarly to RF, Gradient Boosting Machines (GBM) (Friedman, 2001) are also a type of an ensemble approach. The method relies on combining a large number of weak trees to obtain a stronger ensemble prediction. In contrast to random forests (RF), whose output is produced by a voting of the constituent trees, GBM can be characterized as a boosting method (Natekin and Knoll, 2013). In order to boost GBM models, the approach sequentially introduces new models to the ensemble with the aim of mitigating the error of the current ensemble. Each new model is dubbed *weak learner*.

There are different approaches that can be introduced to GBM in order to avoid over-fitting (Natekin and Knoll, 2013). One is sub-sampling and involves the selection of a random subset of the training set at each learning interaction, whose length is



defined by a parameter called *bag fraction*. This means implementing a stochastic gradient descent that steers the model away from local minima. Another approach is to properly adjust the learning rate that controls how fast the gradient descent is. In spite of potentially providing better generalisation, reduced learning rates increase the computational cost. Hence, a compromise must be pursued. Furthermore, one can also optimise the number of trees in the ensemble. Finally, the last aspect that can be controlled is the interaction depth, which can be defined as the number of nodes in each tree.

### 3.4. Support Vector Machines ( $k$ -SVM)

SVM (Boser et al., 1992) are conceived to identify the hyperplane that maximises the distance between the two classes in a binary classification problem. The configuration of the hyperplane depends on the distance to the training samples at the edge of the class, dubbed *support vector points*. The method was originally applicable only to linearly separable data, but can now be generalised by means of a transformation into a higher dimensional space (Witten et al., 2016). Multi-class problems can benefit from a “one-to-one” approach, which gives rise to multiple binary classifiers, each separating the training samples of a pair of different classes (Kim et al., 2003). The appropriate class is decided by a vote.

There are two parameters to be optimised in  $k$ -SVM, namely the cost of constraint violation ( $C$ ) and sigma ( $\sigma$ ), a parameter associated with the kernel function (Gacquer et al., 2011). Compared with other supervised methods,  $k$ -SVM is effective in high dimensional spaces and memory efficient (Braga et al., 2019).

### 3.5. Model evaluation and selection

Several evaluation metrics can be used for model assessment. In this paper, we consider both accuracy and a measure of agreement to be defined below. Firstly, we

need to define the *confusion matrix* (Razi et al., 2019)  $Q = [q_{ij}]$ ,  $1 \leq i \leq m$ ,  $1 \leq j \leq m$ , where  $m$  is the total number of classes. Each element  $q_{ij}$  denotes the number of times that an element of class  $i$  was assigned to class  $j$  according to the classification method. Also, recall from Section 2 that the total number of pieces of equipment in the dataset is represented by  $r$ .

Accuracy is the percentage of correct predictions in a classifier when the classifier is applied to unseen data (Hossin and Sulaiman, 2015; Razi et al., 2019). It is given by:

$$P(0) = \frac{1}{r} \sum_{i=1}^m q_{i,i} \times 100\%. \quad (2)$$

A measure that compares accuracy with the probability of agreement (Razi et al., 2019; Cohen, 1960) is given by  $\kappa$ , and is defined as:

$$\kappa = \frac{P(0) - P(E)}{1 - P(E)}, \quad (3)$$

where  $P(0)$  is the accuracy of the classification model and  $P(E)$  is the chance of agreement, which is obtained as follows:

$$P(E) = \frac{\sum_{i=1}^m (q_{:,i} q_{i,:})}{r^2}. \quad (4)$$

In the expression above,  $q_{:,i}$  and  $q_{i,:}$  are, respectively, sum of  $i$ -th row and  $i$ -th column of the confusion matrix.

In order to obtain reliable estimates of classifier effectiveness, the model should be tested in a different data sample than the one used at the training stage. One method commonly utilised is called holdout, and consists in using a fraction of the dataset (for example  $\frac{2}{3}$  of it) for training and the remaining fraction for testing. The drawback is that the training set may be considerably reduced (Han et al., 2011).

Another option is  $k$ -fold cross-validation, which consists in  $k$  training stages, each with  $k$  samples used for testing and the remaining ones for training (Witten et al., 2016). By doing that, one makes sure that all samples are considered in the training phase.

Finally, statistical tests can be used for comparing machine learning methods with respect to certain performance measures (Hothorn et al., 2005; Han et al., 2011). We compare the models based on the values of  $P(0)$  and  $\kappa$  resulting from the  $k$ -fold cross validation. To determine whether a model is superior to another, we apply a  $t$ -test with Bonferroni correction (Bland and Altman, 1995) and use a 95% confidence level.

#### 4. Numerical experiments

For the sake of validation, we applied the framework proposed in Section 2, and further specified in Section 3, to pipeline data from the Brent field (Shell, 2017a). The decommissioning alternatives are presented in Table 1. A single alternative is to be selected considering the set of twelve sub-criteria described in Table 2.

An important part of the proposed framework is the MCDA analysis for each piece of equipment in the sample, refer to Step 1 of Algorithm 1 for details. However, the approach is designed to work with any MCDA technique available to the decision maker. At this point, it is worth emphasising that a discussion about the choice of the MCDA approach to be employed is besides the scope of this paper. In our experiments, we utilised the classical ELECTRE III (Roy, 1985; Figueira et al., 2013; Del Vasto-Terrientes et al., 2015) method, which makes use of outranking to select an available course of action (Rowley et al., 2012).

A second experiment is also presented to illustrate the possible use of ensemble methods, i.e. methods that make use of multiple MCDA algorithms. In our experi-

Table 1: Pipeline decommissioning alternatives.

| Alternatives | Description  |
|--------------|--|
| $a_1$        | Leave in situ with no further remediation required |
| $a_2$        | Leave tied-in at platform; remote and trenched     |
| $a_3$        | Leave tied-in at platform; remote and rock-dumped  |
| $a_4$        | Trench and backfill whole length                   |
| $a_5$        | Rock-dump whole length                             |
| $a_6$        | Recover whole length by cut and lift               |
| $a_7$        | Recover whole length by reverse S-lay              |

Source: Adapted from Shell (2017a).

ment, we applied ELECTRE III and PROMETHEE II (Brans and Vincke, 1985) to the data and maintained their choice when the alternative coincided. On the other hand, when both algorithms differed, we made use of comparative assessment (Greco et al., 2005) to break the tie.

All machine learning experimental results were generated with 10-fold cross validation. The computational experiments were performed in *R* and made use of some public machine learning libraries, namely *rpart*, *caret*, *kernlab* and *gbm*. The following section describes the data pre-processing and Section 5 features an overview of the results.

#### 4.1. Dataset

To validate our approach, the original intent was to use real-world data from decommissioning reports. In that sense, we found a dataset of sub-sea ducts in the context of a classical report (Shell, 2017a,c). Unfortunately, the dataset is comprised of only 14 samples, which are insufficient for our purposes (Chen et al., 2017; Chang

Table 2: Sub-criteria for the decommissioning of the Brent field.

| Criteria      | Label | Weight          | Sub-criterion                             |
|---------------|-------|-----------------|---|
| Safety        | 1     | $\frac{0.2}{3}$ | Safety risk to offshore project personnel |
|               | 2     | $\frac{0.2}{3}$ | Safety risk to other users of the sea     |
|               | 3     | $\frac{0.2}{3}$ | Safety risk to onshore project personnel  |
| Environmental | 4     | $\frac{0.2}{4}$ | Operational environmental impacts         |
|               | 5     | $\frac{0.2}{4}$ | Legacy environmental impacts              |
|               | 6     | $\frac{0.2}{4}$ | Energy use                                |
|               | 7     | $\frac{0.2}{4}$ | Emissions                                 |
| Technical     | 8     | 0.2             | Technical feasibility                     |
| Social        | 9     | $\frac{0.2}{3}$ | Effects on commercial fisheries           |
|               | 10    | $\frac{0.2}{3}$ | Employment                                |
|               | 11    | $\frac{0.2}{3}$ | Communities                               |
| Economic      | 12    | 0.2             | Cost                                      |

Source: Adapted from Shell (2017a).

et al., 2014). Fortunately, the literature contains many reports of insufficient datasets in diverse fields, such as medicine and manufacturing (Chen et al., 2017; Lateh et al., 2017), and methodological suggestions on how to address this issue. To circumvent small datasets, researchers often resource to synthetic data, which can be generated by a number of techniques, such as fuzzy theory (Huang and Moraga, 2004), synthetic minority over-sampling technique (SMOTE) (Chawla et al., 2002) and bootstrapping (Ivănescu et al., 2006; Tsai and Li, 2008; Chao et al., 2011; Jiménez et al., 2014). In this study, we apply the latter alternative to generate a synthetic dataset for validation. This dataset with 1313 synthetic pipelines was made public (Martins

et al., 2019a) for benchmarking purposes.

The decommissioning guidelines in (Oil & Gas UK, 2015; MEI, 2018) gave rise to the conjecture that the following characteristics could be used as variables for pipeline classification: type (e.g. rigid, flexible); fluid (e.g oil, gas, water); size; length; coated/uncoated; installation date; on bottom status (e.g. fully exposed, rock dumped), proximity to other infrastructure; residues (likely/ability to clean); condition (e.g. good and recoverable, damaged). Bearing that in mind, and in possession of the dataset in (Shell, 2017c), we selected the following parameters of sub-sea ducts: diameter; length; concrete, steel and coat composition; weight; fluid; proximity to other infrastructure (number of crossings); installation date; cleaning type and on bottom status.

The initial dataset in (Shell, 2017a,c) contained all of the eleven characteristics (parameters) mentioned above for each piece of equipment. In addition, it also included the evaluation of each of the seven alternatives with respect to each of the twelve sub-criteria described in Table 2. Initially, for each piece of equipment, we applied the ELECTRE III tool, with the same set of weights used in the original decommissioning report - which appears in Table 2, to produce the recommended decommissioning alternative. The values of indifference, preference, and veto thresholds were set to zero (0). In the second experiment, we applied both ELECTRE III and PROMETHEE II with the same threshold values and weights. When these methods recommended different alternatives, comparative assessment (CA) was conducted, with the same set of weights. In that case, between the two alternatives recommended by ELECTRE III and PROMETHEE II, we select that with the highest CA score.

The recommended decommissioning alternative acts as the independent variable in the supervised training routine in Step 2 of Algorithm 1. In that routine, each

equipment  $e_k = (x_k, y_k)$  is an entry in the dataset, where  $x_k$  is a vector containing all the parameters and sub-criteria assessments and  $y_k$  is the independent variable, i.e. the decommissioning action the MCDA tool(s) recommended for the respective installation.

## 5. Experimental results

This section is divided into three subsections. The following two subsections report results of the first experiment. In the following subsection, we report the results of each selected machine learning technique in the supervised classification problem of Step 2 in Algorithm 1. Then, Section 5.2 illustrates the results of Step 3 of Algorithm 1, that was performed only for the GBM model, which outperformed the competing methods in Step 2. Because of its superior performance, GBM was the selected method for the second experiment, which is reported in Section Section 5.3.

### 5.1. Model comparison

The aim of this section is to compare the performance of selected machine learning techniques, namely DT, RF,  $k$ -SVM and GBM, for predicting decommissioning decisions based on the input variables described in Section 4.

We implemented Grid search (Bergstra and Bengio, 2012), a method for hyper-parameter optimisation, to optimise the performance of each evaluated technique. The parameters are briefly discussed below:

- DT: The only specification to be optimised for decision trees is the complexity parameter and it was set as 0.047;
- RF: The number of trees was set as 1000. We tried each value in the set  $\{8, 9, \dots, 13\}$  for the parameter  $mtry$ , i.e. the optimal number of variables selected for each tree. The optimal value obtained was 13;

- $k$ -SVM: We used the radial basis kernel function and the parameter search was for  $\sigma \in \{0.01, 0.02, 0.025, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.25, 0.5\}$  and  $C \in \{1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5\}$ . The selected values were  $\sigma = 0.04$  and  $C = 5$ ; and
- GBM: We evaluated all combinations involving interaction depth in the set  $\{6, 7, 8\}$ , number of trees in the set  $\{130, 131, \dots, 140\}$  and shrinkage in the set  $\{0.1, 0.15, 0.2, 0.3\}$ . The best results were obtained with 140 trees, interaction depth equal to 7 and shrinkage equal to 0.1. The bag fraction was set to 0.8.

Table 3 and Figure 3 summarise the evaluations of accuracy ( $P(0)$ ) and  $\kappa$  for the optimised models. Also, Table 4 shows the  $p$ -values of pairwise  $t$ -test results. The significance threshold was  $\alpha = 0.05$ . Each element in Table 4 is the  $p$ -value of the null hypothesis, according to which the algorithms in the corresponding line and column, respectively, are indifferent with respect to the performance measures. One can see from the referred table that this hypothesis is rejected in all pairwise comparisons.

An inspection in the preceding results yield that the GBM model presents the best overall performance considering all features. It boasts a mean accuracy of 82% and  $\kappa = 72\%$ . At the other end, the worst performance is attained by DT, with significantly lower accuracy and  $\kappa$ . Bearing that in mind, the GBM algorithm was singled out for feature selection in Step 3 of Algorithm 1.

## 5.2. Feature selection

GBM’s feature selection tool is hybrid and combines learning and feature selection (Liso, 2016). The measure of relative importance is due to Friedman (2001) and is



Table 3: Model comparison through accuracy and kappa evaluation metrics.

***Accuracy***

|                       | Min  | 1st Qu. | Median | Mean | 3rd Qu. | Max  |
|-----------------------|------|---------|--------|------|---------|------|
| <b>GBM</b>            | 0.78 | 0.80    | 0.82   | 0.82 | 0.83    | 0.88 |
| <b>RF</b>             | 0.74 | 0.77    | 0.79   | 0.79 | 0.81    | 0.84 |
| <i>k</i> – <b>SVM</b> | 0.67 | 0.70    | 0.70   | 0.71 | 0.74    | 0.75 |
| <b>DT</b>             | 0.59 | 0.61    | 0.62   | 0.62 | 0.63    | 0.64 |

***Kappa***

|                       | Min  | 1st Qu. | Median | Mean | 3rd Qu. | Max  |
|-----------------------|------|---------|--------|------|---------|------|
| <b>GBM</b>            | 0.71 | 0.73    | 0.76   | 0.76 | 0.78    | 0.84 |
| <b>RF</b>             | 0.64 | 0.69    | 0.71   | 0.72 | 0.74    | 0.78 |
| <i>k</i> – <b>SVM</b> | 0.55 | 0.60    | 0.60   | 0.61 | 0.65    | 0.66 |
| <b>DT</b>             | 0.43 | 0.45    | 0.47   | 0.47 | 0.48    | 0.49 |

a function of the number of times that a variable is selected for splitting nodes, modulated by the model improvement resulting from each split. For the sake of comparison, the measures of importance are standardised.

As previously stated, GBM was the chosen method for variable selection because it performed best in the classification step. The goal here is to eliminate irrelevant features, i.e. installation parameters or sub-criteria assessments with very limited impact on the output of the MCDA method. By doing so, one can build a lower dimensional model with comparable performance and decreased computational cost (Guyon and Elisseeff, 2003). From a practical standpoint, this means that we can find an alternative model that requires a reduced number of sub-criteria assessments and data collection. This, in turn, implies an accelerated decision making process,

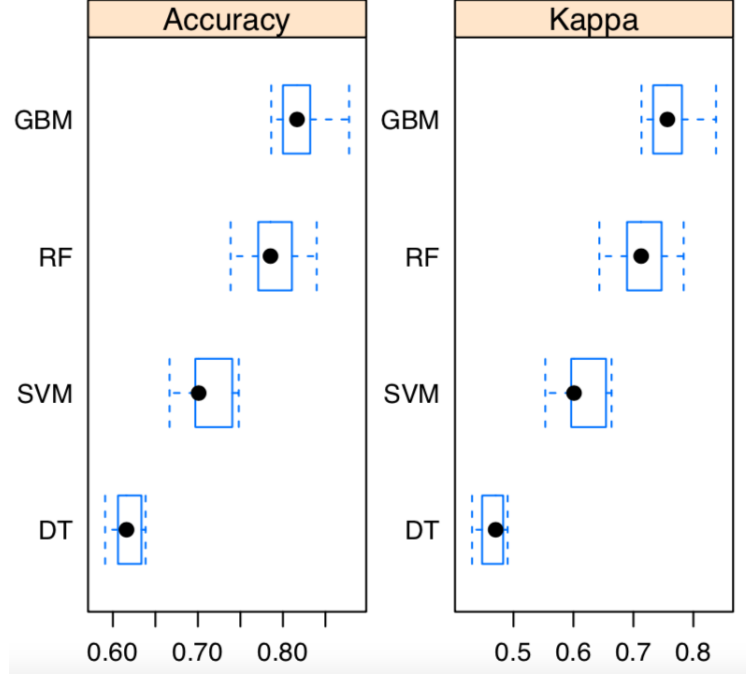


Figure 3: Model comparison.

with potentially considerable reductions in both costs and times for data acquisition and sub-criteria evaluation.

In our case study, we opted to maintain all eleven features associated with the installation parameters, which were enumerated in Section 4. This is because this information is very easy to obtain, and hence the omission would not bring any relevant benefit. In contrast, we ranked the sub-criteria in decreasing order of importance, and tested dimensionality reduction scenarios whereby some of them were eliminated, as detailed below.

Figure 4 conveys the relative importance of each sub-criterion that appears in Table 2. It stands out that the sub-criteria 12 (Cost) was the most important. In addition, we found that 12 (Cost), 6 (Energy Use), 11 (Communities), 1 (Safety risk to offshore project personnel), 7 (Emissions), 4 (Operational environmental impacts),

Table 4: The p-values corresponding to pairwise comparison of different classification models.

***Accuracy***

|                     | <b>RF</b> | <b><i>k</i>–SVM</b> | <b>DT</b> |
|---------------------|-----------|---------------------|-----------|
| <b>GBM</b>          | 0.026     | 2.71e-05            | 1.758e-08 |
| <b>RF</b>           |           | 2.5e-04             | 1.26e-07  |
| <b><i>k</i>–SVM</b> |           |                     | 3.58e-06  |

***Kappa***

|                     | <b>RF</b> | <b><i>k</i>–SVM</b> | <b>DT</b> |
|---------------------|-----------|---------------------|-----------|
| <b>GBM</b>          | 0.019     | 3.18e-05            | 5.56e-09  |
| <b>RF</b>           |           | 3.6e-04             | 3.56e-08  |
| <b><i>k</i>–SVM</b> |           |                     | 1.06e-06  |

2 (Safety risk to other users of the sea) and 9 (Effects on commercial fisheries), are responsible for about 91.85% of the total importance. Furthermore, it is also striking that the sub-criteria 8 (Technical feasibility) presents a very limited relevance, representing only 0.24% of relative influence.

To assess the effect of removing the less significant variables from the model, we tested the GBM model with the subset of most relevant sub-criteria that account for 92%, 83%, 68% and 56% of relative importance, respectively. Obviously, the objective is to come up with the smallest possible subset of sub-criteria that produces no significant decrease in performance.

Table 5 summarises accuracy and kappa evaluation metrics for each subset considered. The  $p$ -values produced by each  $t$ -test with respect to a pair of models are unveiled in Table 6. One can easily see from the latter table that the models with 92% and 83% of the total importance are indistinguishable from the original model.

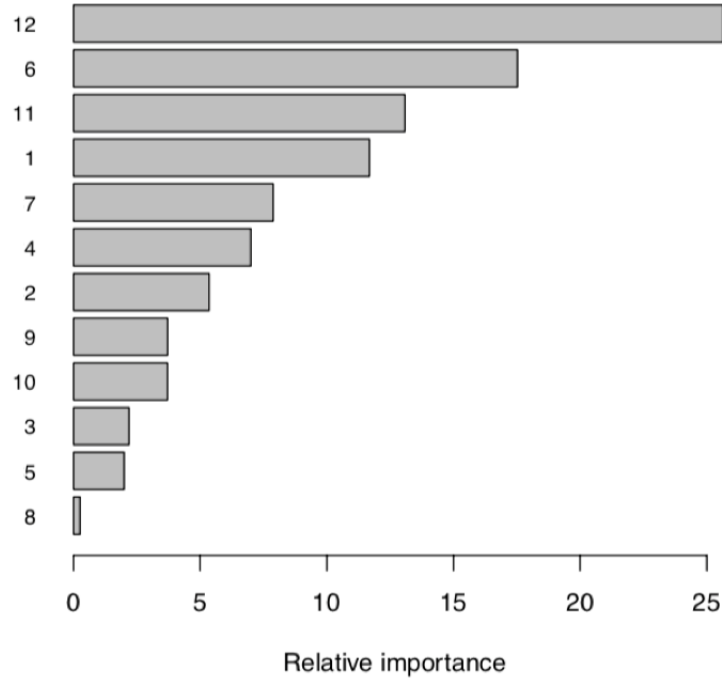


Figure 4: Relative importance of criteria in the first experiment.

This means that we can keep only half of the sub-criteria assessments, namely sub-criteria 12, 6, 11, 1, 7 and 4, with virtually no impact on the performance of the classification method.

The results seem very promising, and suggest that the decommissioning study of an oil field can be considerably simplified with the application of the proposed method. Indeed, eliminating half of the sub-criteria assessments with no significant loss in performance would be a very welcome development in a long, complex process. Arguably, one can expect considerable reduction in the number of sub-criteria since, in a complex process such as that which sets up the sub-criteria, it is possible that many criteria and sub-criteria assessments be highly correlated. Such a correlation can be captured by machine learning techniques in the process of generating a simplified analysis tool with comparable results.

Table 5: GBM models performance comparison considering different variables subsets through accuracy and kappa evaluation metrics.

***Accuracy***

|                   | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|-------------------|------|---------|--------|------|---------|------|
| <b>GBM (100%)</b> | 0.79 | 0.80    | 0.82   | 0.82 | 0.83    | 0.88 |
| <b>GBM (92%)</b>  | 0.78 | 0.79    | 0.80   | 0.81 | 0.83    | 0.88 |
| <b>GBM (83%)</b>  | 0.76 | 0.80    | 0.81   | 0.81 | 0.83    | 0.86 |
| <b>GBM (68%)</b>  | 0.76 | 0.78    | 0.80   | 0.80 | 0.81    | 0.83 |
| <b>GBM (56%)</b>  | 0.75 | 0.76    | 0.78   | 0.78 | 0.80    | 0.82 |

***Kappa***

|                   | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|-------------------|------|---------|--------|------|---------|------|
| <b>GBM (100%)</b> | 0.71 | 0.73    | 0.76   | 0.76 | 0.78    | 0.84 |
| <b>GBM (92%)</b>  | 0.71 | 0.73    | 0.74   | 0.75 | 0.77    | 0.84 |
| <b>GBM (83%)</b>  | 0.68 | 0.74    | 0.75   | 0.75 | 0.77    | 0.81 |
| <b>GBM (68%)</b>  | 0.67 | 0.71    | 0.74   | 0.73 | 0.75    | 0.78 |
| <b>GBM (56%)</b>  | 0.67 | 0.68    | 0.71   | 0.71 | 0.74    | 0.76 |

Another potential gain of the proposed approach is that one does not need to deploy the MCDA tool for all installations. Instead, it is only utilised in the training set. For all installations outside this set, the decision maker can simply make use of the predictions provided by the machine learning classification tool.

Finally, the ranking of the sub-criteria is instrumental to the analysis, since it enables the decision maker to verify the consistency of the proposed MCDA tool. For instance, it will confirm or not that variables which are deemed important to the analysis have a real impact on the outcome of the MCDA tool.

Table 6:  $p$ -values corresponding to pairwise comparison of the GBM model performance considering different variables subsets according to the percentage of relative importance of the criteria given by Friedman (2001) method.

***Accuracy***

|            | GBM (92%) | GBM (83%) | GBM (68%) | GBM (56%) |
|------------|-----------|-----------|-----------|-----------|
| GBM (100%) | 1         | 1         | 0.18      | 0.054     |
| GBM (92%)  |           | 1         | 1         | 0.259     |
| GBM (83%)  |           |           | 0.563     | 0.029     |
| GBM (68%)  |           |           |           | 0.28      |

***Kappa***

|            | GBM (92%) | GBM (83%) | GBM (68%) | GBM (56%) |
|------------|-----------|-----------|-----------|-----------|
| GBM (100%) | 1         | 1         | 0.20      | 0.064     |
| GBM (92%)  |           | 1         | 1         | 0.279     |
| GBM (83%)  |           |           | 0.560     | 0.030     |
| GBM (68%)  |           |           |           | 0.31      |

*5.3. Second experiment - ensemble method*

In the second experiment, we present a preliminary evaluation of the use of an ensemble MCDA method in the generation of the training set for the ML approach. We made use of two outranking methods, namely ELECTRE III and PROMETHEE II and a *simple average weighting* (SAW) algorithm referred to as *comparative assessment (CA)* in the decommissioning literature (Martins et al., 2019b; Greco et al., 2005). We selected the alternative to be inserted in the training set based on a vote. Hence, whenever the two outranking methods selected the same alternative, this alternative was inserted in the training set. On the other hand, whenever the outrank-

ing methods diverged, the alternatives selected by ECTREE III and PROMETHEE II were fed to the CA framework, which selected the best of the two.

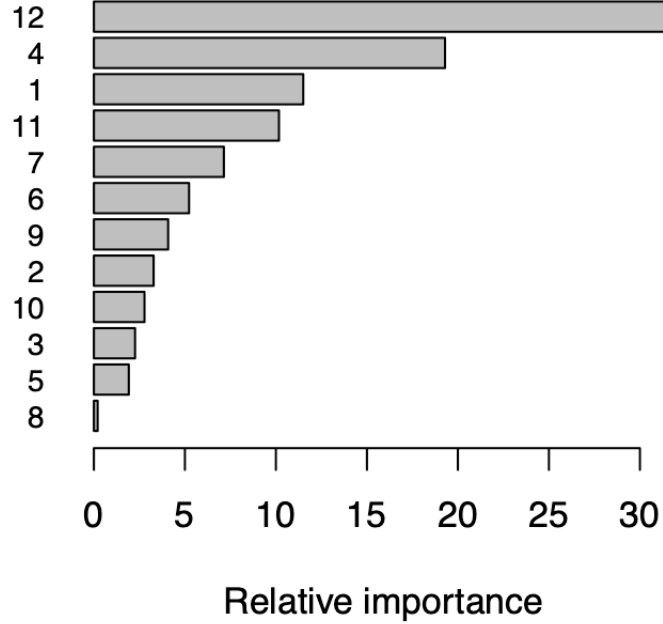


Figure 5: Relative importance of criteria in the second experiment.

In the numerical experiment, the outranking methods elected the same alternative for 80% of the pieces of equipment. Hence, we needed CA to break the tie 20% of the time. As previously mentioned, we opted to use GBM for feature selection and classification, since it was the superior method in the first experiment. Once again we conducted grid search for the parameter optimisation, and this time the best results were obtained with 130 trees, interaction depth equal to 7 and shrinkage equal to 0.1. For this set of parameters, the mean accuracy reached 76% and  $\kappa = 68\%$ . Observe that these values are slightly reduced with respect to the first experiment, which registered a mean accuracy of 82% and  $\kappa = 76\%$ .

Regarding feature selection, the results in Figure 5 show that sub-criteria 12

(Cost) and 8 (Technical Feasibility) were, once again, respectively the most and least important. One can also notice that the results were similar to those in Figure 4, however the importance seems to decrease slightly more steeply.

Table 7: GBM models performance comparison considering different variables subsets through accuracy and kappa evaluation metrics considering the labelling determined by an ensemble MCDA method.

***Accuracy***

|                   | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|-------------------|------|---------|--------|------|---------|------|
| <b>GBM (100%)</b> | 0.73 | 0.75    | 0.77   | 0.76 | 0.78    | 0.80 |
| <b>GBM (90%)</b>  | 0.73 | 0.74    | 0.75   | 0.76 | 0.77    | 0.82 |
| <b>GBM (80%)</b>  | 0.72 | 0.73    | 0.74   | 0.75 | 0.76    | 0.79 |
| <b>GBM (73%)</b>  | 0.69 | 0.72    | 0.73   | 0.73 | 0.75    | 0.76 |
| <b>GBM (63%)</b>  | 0.68 | 0.72    | 0.73   | 0.74 | 0.76    | 0.80 |

***Kappa***

|                   | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|-------------------|------|---------|--------|------|---------|------|
| <b>GBM (100%)</b> | 0.63 | 0.67    | 0.69   | 0.68 | 0.70    | 0.73 |
| <b>GBM (90%)</b>  | 0.64 | 0.65    | 0.66   | 0.67 | 0.69    | 0.76 |
| <b>GBM (80%)</b>  | 0.61 | 0.64    | 0.66   | 0.66 | 0.68    | 0.71 |
| <b>GBM (73%)</b>  | 0.58 | 0.62    | 0.64   | 0.64 | 0.66    | 0.68 |
| <b>GBM (63%)</b>  | 0.57 | 0.62    | 0.64   | 0.65 | 0.68    | 0.73 |

To evaluate the precision of reduced models, we tested GBM with subsets that totalled 90%, 80%, 73% and 63% of relative importance, respectively. The evaluation metrics for each of the models generated with the subsets as well as the  $p$ -values obtained from the pairwise comparisons between the models are shown in Tables 7 and



Table 8:  $p$ -values corresponding to pairwise comparison of the GBM model performance considering different variables subsets according to the percentage of relative importance of the criteria given by Friedman (2001) method considering the labelling determined by different MCDA outranking methods.

***Accuracy***

|            | GBM (90%) | GBM (80%) | GBM (73%) | GBM (63%) |
|------------|-----------|-----------|-----------|-----------|
| GBM (100%) | 1         | 1         | 1         | 1         |
| GBM (90%)  |           | 1         | 0.2122    | 1         |
| GBM (80%)  |           |           | 0.6323    | 1         |
| GBM (73%)  |           |           |           | 1         |

***Kappa***

|            | GBM (90%) | GBM (80%) | GBM (73%) | GBM (63%) |
|------------|-----------|-----------|-----------|-----------|
| GBM (100%) | 1         | 1         | 1         | 1         |
| GBM (90%)  |           | 1         | 0.2554    | 1         |
| GBM (80%)  |           |           | 0.6132    | 1         |
| GBM (73%)  |           |           |           | 1         |

8. From the results, we can see that a model with only the 4 (four) most important sub-criteria, namely 12 (Cost), 4 (Operational environmental impacts), 1 (Safety risk to offshore project personnel) and 11 (Communities), has an indistinguishable performance when compared to the complete model. It is worth pointing out that these four criteria account for around 63% of the relative importance. Hence, the ensemble method enables the elimination of two thirds of the sub-criteria (from 12 to 4), whereas a 50% reduction (12 to 6) was attained with the use of ELECTREE III on its own.

To sum up, while the proposed ensemble method performed slightly worse in terms of  $\kappa$  and accuracy, it boasts an improved performance in terms of dimensionality reduction. Future research could be devoted to find out if these preliminary results would be maintained in different ensemble configurations. While the investigation of the effects of the ensemble in the performance of the proposed approach falls beyond the scope of this problem, our results suggest that future research may aim at investigating some questions that are raised by these preliminary results. For example, one may be interested in investigating what the best configuration of an ensemble method would be. Such an investigation would encompass a very large number of possible combinations of MCDA approaches and could delve into aspects such as the compatibility of the approaches and their consistencies and similarities in terms of outputs. In addition, distinct ways of producing an outcome of the ensemble could be investigated. For instance, instead of a vote, the method could make use of regret analysis (Bubeck and Cesa-Bianchi, 2012) for labelling the training set.

## 6. Summary and conclusions

This study developed a framework based on supervised algorithms and dimensionality reduction techniques aiming to reduce the time and effort of sub-criteria evaluations in a decommissioning study. The framework makes use of a reduced dataset comprised of installations characteristics and sub-criteria evaluations. A number of machine learning algorithms are then applied to the dataset and that with the best overall performance is singled out for variable selection. This latter step then produces a reduced subset of relevant sub-criteria that should be measured for the pieces of equipment left out of the training dataset. The reduced model can be used to predict the decommissioning alternative for these pieces of equipment, thus circumventing the need for a case by case MCDA analysis.

The framework was validated through numerical analyses for a synthetic dataset based on real data for pipelines in the Brent field (Shell, 2017c). The dataset variables included eleven characteristics, such as diameter and fluid type, and the evaluation of twelve sub-criteria for each of the seven decommissioning alternatives. The numerical experiments, which were performed using ELECTREE III, as well as an ensemble of three MCDA approaches to generate the labelling of the training set, suggest that a significant reduction in the number of assessed criteria can be obtained with virtually no impact on the performance, thus reducing the overall effort and cost of the decommissioning study. One significant contribution is the suggestion that the decision maker can accurately predict the recommended decommissioning alternatives with a reduced number of sub-criteria evaluations. In addition, the use of machine learning precludes the need for a case by case MCDA analysis, for the recommended alternatives for the installations outside the training set can be accurately forecast by the classification method.

One limitation of the work is the use of a synthetic dataset, which was necessary given the absence of real-world data. Such a limitation can be addressed in the future if companies agree to make their data available, even if discharacterised. In addition, the generality of the results is limited by the fact that the decommissioning criteria may vary depending on the locality and on the regulatory bodies (Martins et al., 2019b), hence the conclusions on the significance of a particular subset of criteria may not be applicable to distinct decommissioning projects.

The current work can be extended and complemented in a number of directions. One possible direction involves the investigation of the effects of changing weights and parameters in the MCDA tool, whereby machine learning could be applied to predict the response of the MCDA tool given the input of a weight vector and a set of parameters. Such an investigation could also aim to find which MCDA approaches

are less sensitive to changes in parameters and weights, with a view to providing the decision maker with a dependable and reproducible decision aid tool. Another possible topic of future research is the design of ensemble methods of similar or different families of MCDA approaches to generate the training set. One could investigate if different families tend to produce different decisions, as well as if some specific methods are complementary or consistent with other approaches. In addition, the possibility of applying regret analysis (Bubeck and Cesa-Bianchi, 2012) for labelling could be evaluated.

## **Acknowledgements**

We are indebted and would like to express our gratitude to the referees for their many suggestions and helpful comments, which have certainly improved the quality of the paper.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. This study was partially supported by the Nacional Council for Scientific and Technological Development - CNPq, under grants 305180/2016-9 and 311075/2018-5.

## **References**

Ahmed, S., Ahmed, S., Shumon, M., Falatoonitoosi, E., Quader, M., 2016. A comparative decision-making model for sustainable end-of-life vehicle management alternative selection using AHP and extent analysis method on fuzzy AHP. *International Journal of Sustainable Development & World Ecology* 23, 83–97. doi:10.1080/13504509.2015.1062814.

- Amirshenava, S., Osanloo, M., 2018. Mine closure risk management: An integration of 3D risk model and MCDM techniques. *Journal of Cleaner Production* 184, 389–401.
- Bergstra, J., Bengio, Y., 2012. Random search for hyper-parameter optimization. *Journal of Machine Learning Research* 13, 281–305. doi:10.1016/j.jclepro.2018.01.186.
- Bernstein, B., Bressler, A., Cattle, P., Henrion, M., John, D., Kruse, S., Pondella, D., Scholz, A., Setnicka, T., Swamy, S., Fink, L., McCann, B., 2010. Evaluating alternatives for decommissioning California’s offshore oil and gas platforms: a technical analysis to inform state policy. Technical Report. California Ocean Science Trust. California.
- Beynon, M., Cosker, D., Marshall, D., 2001. An expert system for multi-criteria decision making using dempster shafer theory. *Expert Systems with Applications* 20, 357 – 367. doi:10.1016/S0957-4174(01)00020-3.
- BG Group, 2016. Atlantic & Cromarty fields - Decommissioning programmes and comparative assessment report. Technical Report. BG Group. Aberdeen, UK.
- Bland, J., Altman, D., 1995. Multiple significance tests: the Bonferroni method. *BMJ* 310, 170. doi:10.1136/bmj.310.6973.170.
- Boser, B., Guyon, I., Vapnik, V., 1992. A training algorithm for optimal margin classifiers, in: *Proceedings of the fifth annual workshop on Computational Learning Theory*, ACM. pp. 144–152. doi:10.1145/130385.130401.
- Braga, D., Madureira, A., Coelho, L., Ajith, R., 2019. Automatic detection of

- Parkinson's disease based on acoustic analysis of speech. *Engineering Applications of Artificial Intelligence* 77, 148–158. doi:10.1016/j.engappai.2018.09.018.
- Brans, J.P., Vincke, P., 1985. A preference ranking organisation method: (the promethee method for multiple criteria decision-making). *Management Science* 31, 647–656.
- Breiman, L., 2001. Random forests. *Machine learning* 45, 5–32. doi:10.1023/A:1010933404324.
- Bubeck, S., Cesa-Bianchi, N., 2012. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning* 5, 1–122. doi:10.1561/22000000024.
- Bystrzanowska, M., Tobiszewski, M., 2018. How can analysts use multicriteria decision analysis? *Trends in Analytical Chemistry* 105, 98–105. doi:10.1016/j.trac.2018.05.003.
- Carneiro, J., Conceição, L., Martinho, D., Marreiros, G., Novais, P., 2018. Including cognitive aspects in multiple criteria decision analysis. *Annals of Operations Research* 265, 269–291. doi:10.1007/s10479-016-2391-1.
- Carneiro, J., Martinho, D., Marreiros, G., Novais, P., 2019. Arguing with behavior influence: A model for web-based group decision support systems. *International Journal of Information Technology & Decision Making* 18, 517–553. doi:10.1142/S0219622018500542.
- Chai, M., Li, S., Sun, W., Guo, M., Huang, M., 2019. Drowsiness monitoring based on steering wheel status. *Transportation Research Part D: Transport and Environment* 66, 95–103. doi:10.1016/j.trd.2018.07.007.

- Chang, C.J., Li, D.C., Dai, W.L., Chen, C.C., 2014. A latent information function to extend domain attributes to improve the accuracy of small-data-set forecasting. *Neurocomputing* 129, 343–349. doi:10.1016/j.neucom.2013.09.024.
- Chao, G., Tsai, T., Lu, T., Hsu, H., Bao, B., Wu, W., Lin, M., Lu, T., 2011. A new approach to prediction of radiotherapy of bladder cancer cells in small dataset analysis. *Expert Systems with Applications* 38, 7963–7969. doi:10.1016/j.eswa.2010.12.035.
- Chawla, N., Bowyer, K., Hall, L., Kegelmeyer, W., 2002. SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research* 16, 321–357. doi:10.1613/jair.953.
- Chen, Z., Zhuand, B., He, Y., Yu, L., 2017. A PSO based virtual sample generation method for small sample sets: Applications to regression datasets. *Engineering Applications of Artificial Intelligence* 59, 236–243. doi:10.1016/j.engappai.2016.12.024.
- Cheng, S., 2010. Intelligent Cognition-Based Systems Approach to Multiple-Criteria Computerized Essay Assessment. *Systems Research and Behavioral Science* 27, 680–696. doi:10.1002/sres.
- Cohen, J., 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20, 37–46. doi:10.1177/001316446002000104.
- Copeland, H., Doherty, K., Naugle, D., Pocewicz, A., Kiesecker, J., 2009. Mapping oil and gas development potential in the US Intermountain West and estimating impacts to species. *PloS One* 4, e7400. doi:10.1371/journal.pone.0007400.

- Dehe, B., Bamford, D., 2015. Development, test and comparison of two multiple criteria decision analysis (MCDA) models: A case of healthcare infrastructure location. *Expert Systems with Applications* 42, 6717–6727. doi:10.1016/j.eswa.2015.04.059.
- Del Vasto-Terrientes, L., Valls, A., Slowinski, R., Zielniewicz, P., 2015. ELECTRE-III-H. *Expert Systems with Applications* 42, 4910–4926. doi:10.1016/j.eswa.2015.02.016.
- Dimitrijevic, B., Vujic, S., Matic, I., Majianac, S., Praštalo, J., Radosavljevic, M., Čolakovic, V., 2014. Multi-criterion analysis of land reclamation methods at Klenovnik open pit mine, Kostolac coal basin. *Journal of Mining Science* 50, 319–325. doi:10.1134/S106273911402015X.
- DMIRS, 2017. Petroleum Decommissioning Guideline. Technical Report October. Department of Mines, Industry Regulation and Safety - Government of Western Australia. Western Australia.
- Duro, J., Saxena, D., Deb, K., Zhang, Q., 2014. Machine learning based decision support for many-objective optimization problems. *Neurocomputing* 146, 30–47. doi:10.1016/j.neucom.2014.06.076.
- El-Abbasy, M., Senouci, A., Zayed, T., Parvizsedghy, L., Mirahadi, F., 2016. Unpiggable oil and gas pipeline condition forecasting models. *Journal of Performance of Constructed Facilities* 30, 04014202. doi:10.1061/(ASCE)CF.1943-5509.0000716.
- Feng, Q., Liu, J., Gong, J., 2015. UAV remote sensing for urban vegetation mapping using random forest and texture analysis. *Remote Sensing* 7, 1074–1094. doi:10.3390/rs70101074.



- Figueira, J.R., Greco, S., Roy, B., Slowinski, R., 2013. An overview of ELECTRE methods and their recent extensions. *Journal of Multi-Criteria Decision Analysis* 20, 61–85. doi:10.1002/mcda.1482.
- Fowler, A., Macreadie, P., Jones, D., Booth, D., 2014. A multi-criteria decision approach to decommissioning of offshore oil and gas infrastructure. *Ocean & Coastal Management* 87, 20–29. doi:10.1016/j.ocecoaman.2013.10.019.
- Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. *Annals of Statistics* , 1189–1232doi:10.1214/aos/1013203451.
- Gacquer, D., Delcroix, V., Delmotte, F., Piechowiak, S., 2011. Comparative study of supervised classification algorithms for the detection of atmospheric pollution. *Engineering Applications of Artificial Intelligence* 24, 1070–1083. doi:10.1016/j.engappai.2011.04.002.
- Glaize, A., Duenas, A., Di Martinelly, C., Fagnot, I., 2019. Healthcare decision-making applications using multicriteria decision analysis: A scoping review. *Journal of Multi-Criteria Decision Analysis* 26, 62–83. doi:10.1002/mcda.1659.
- Greco, S., Figueira, J., Ehrgott, M., 2005. Multiple criteria decision analysis - State of the art - Surveys. Springer, New York, NY. doi:10.1007/b100605.
- Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. *Journal of Machine Learning Research* 3, 1157–1182.
- Han, J., Pei, J., Kamber, M., 2011. Data mining: concepts and techniques. Elsevier. doi:10.1016/C2009-0-61819-5.

- Hastie, T., Tibshirani, R., Friedman, J., 2001. The elements of statistical learning: Data mining, inference, and prediction. Springer Series in Statistics, Springer, New York. doi:10.1007/978-0-387-84858-7.
- Henrion, M., Bernstein, B., Swamy, S., 2015. A multi-attribute decision analysis for decommissioning offshore oil and gas platforms. Integrated Environmental Assessment and Management 11, 594–609. doi:10.1002/ieam.1693.
- Hillerman, T., Souza, J., Reis, A., Carvalho, R., 2017. Applying clustering and AHP methods for evaluating suspect healthcare claims. Journal of Computational Science 19, 97–111. doi:10.1016/j.jocs.2017.02.007.
- Hodeghatta, U., Nayak, U., 2016. Business analytics using R-A practical approach. Springer. doi:10.1007/978-1-4842-2514-1.
- Hossin, M., Sulaiman, M., 2015. A review on evaluation metrics for data classification evaluations. International Journal of Data Mining & Knowledge Management Process 5, 1–11. doi:10.5121/ijdkp.2015.5201.
- Hothorn, T., Leisch, F., Zeileis, A., Hornik, K., 2005. The design and analysis of benchmark experiments. Journal of Computational and Graphical Statistics 14, 675–699. doi:10.1198/106186005X59630.
- Huang, C., Moraga, C., 2004. A diffusion-neural-network for learning from small samples. International Journal of Approximate Reasoning 35, 137–161. doi:10.1016/j.ijar.2003.06.001.
- Huang, J.W., Chiang, C.W., Chang, J.W., 2018. Email security level classification of imbalanced data using artificial neural network: The real case in a world-leading

- enterprise. *Engineering Applications of Artificial Intelligence* 75, 11–21. doi:10.1016/j.engappai.2018.07.010.
- Ineos, 2018. Windermere Decommissioning Project Comparative Assessment. Technical Report April. INEOS UK SNS Limited. UK.
- ISM, 2011. Submarine Dismantling Project - Operational Effectiveness (OE) Report - interim version to support the Submarine Dismantling Consultation. Technical Report. ISM. Bristol, UK. URL: <https://www.gov.uk/government/collections/submarine-dismantling-project>.
- Ivănescu, V., Bertrand, J., Fransoo, J., Kleijnen, J., 2006. Bootstrapping to solve the limited data problem in production control: an application in batch process industries. *Journal of the Operational Research Society* 57, 2–9. doi:10.1057/palgrave.jors.2601966.
- Jedrkiewicz, R., Tsakovski, S., Lavenu, A., Namiesnik, J., Tobiszewski, M., 2018. Simultaneous grouping and ranking with combination of som and topsis for selection of preferable analytical procedure for furan determination in food. *Talanta* 178, 928–933. doi:10.1016/j.talanta.2017.10.044.
- Jiménez, F., Serradilla, F., Román, A., Naranjo, J., 2014. Bus line classification using neural networks. *Transportation Research Part D: Transport and Environment* 30, 32–37. doi:10.1016/j.trd.2014.05.008.
- Jolliffe, I.T., 2002. *Principal component analysis*. Springer. doi:10.1007/b98835.
- Kartal, H., Oztekin, A., Gunasekaran, A., Cebi, F., 2016. An integrated decision analytic framework of machine learning with multi-criteria decision making for

- multi-attribute inventory classification. *Computers & Industrial Engineering* 101, 599–613. doi:10.1016/j.cie.2016.06.004.
- Kerkvliet, H., Polatidis, H., 2016. Offshore wind farms’ decommissioning: a semi quantitative multi-criteria decision aid framework. *Sustainable Energy Technologies and Assessments* 18, 69–79. doi:10.1016/j.seta.2016.09.008.
- Khalid, S., Khalil, T., Nasreen, S., 2014. A survey of feature selection and feature extraction techniques in machine learning, in: 2014 Science and Information Conference, IEEE. pp. 372–378. doi:10.1109/SAI.2014.6918213.
- Kim, H., Pang, S., Je, H., Kim, D., Bang, S., 2003. Constructing support vector machine ensemble. *Pattern Recognition* 36, 2757–2767. doi:10.1016/S0031-3203(03)00175-4.
- Lakhani, A., Zeeman, H., Wright, C.J., Watling, D.P., Smith, D., Islam, R., 2019. Stakeholder priorities for inclusive accessible housing: A systematic review and multicriteria decision analysis. *Journal of Multi-Criteria Decision Analysis* doi:10.1002/mcda.1689.
- Lateh, M., Muda, A., Yusof, Z., Muda, N., Azmi, M., 2017. Handling a small dataset problem in prediction model by employ artificial data generation approach: A review. *Journal of Physics: Conference Series* 892, 012016. doi:10.1088/1742-6596/892/1/012016.
- León, J., Martín-Campo, F., Ortuño, M., Vitoriano, B., Carrasco, L., Navarte, L., 2019. A methodology for designing electrification programs for remote areas. *Central European Journal of Operations Research* doi:10.1007/s10100-019-00649-6.

- Li, X., Chan, C.W., 2010. Application of an enhanced decision tree learning approach for prediction of petroleum production. *Engineering Applications of Artificial Intelligence* 23, 102–109. doi:10.1016/j.engappai.2009.06.003.
- Liso, A., 2016. Feature selection with Random Forest and Gradient Boosting. Master's degree of Investigation and Innovation in Information and Communications Technology. Universidad Autónoma de Madrid.
- Liu, H., Liu, Z., Taylor, B., Dong, H., 2019. Matching pipeline in-line inspection data for corrosion characterization. *NDT & E International* 101, 44–52. doi:10.1016/j.ndteint.2018.10.004.
- Martins, I., Bahiense, L., Arruda, E., 2019a. Supplementary data - Subsea pipelines synthetic data. Mendeley Data, v1. URL: <http://dx.doi.org/10.17632/2ns23b2ybr.1>.
- Martins, I., Moraes, F., Távora, G., Soares, H., Infante, C., Arruda, E., Bahiense, L., Caprace, J., Lourenço, M., 2019b. A review of the multicriteria decision analysis applied to oil and gas decommissioning problems. *Ocean & Coastal Management* , 105000doi:10.1016/j.ocecoaman.2019.105000.
- MEI, 2018. Brunei Darussalam Decommissioning and Restoration Guidelines for Onshore and Offshore Facilities - Issue for Industry Use. Volume 9: Decommissioning and restoration guidelines. Technical Report. Ministry of Energy and Industry. Negara Brunei Darussalam.
- Mergias, I., Moustakas, K., Papadopoulos, A., Loizidou, M., 2007. Multi-criteria decision aid approach for the selection of the best compromise management scheme

- for ELVs: The case of Cyprus. *Journal of Hazardous Materials* 147, 706–717. doi:10.1016/j.jhazmat.2007.01.071.
- Na, K., Lee, H., Liew, M., Zawawi, N., 2017. An expert knowledge based decommissioning alternative selection system for fixed oil and gas assets in the South China Sea. *Ocean Engineering* 130, 645–658. doi:10.1016/j.oceaneng.2016.11.053.
- Natekin, A., Knoll, A., 2013. Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics* 7, 21. doi:10.3389/fnbot.2013.00021.
- Oil & Gas UK, 2015. Guidelines for Comparative Assessment in Decommissioning Programmes.
- Pearson, K., 1901. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2, 559–572. doi:10.1080/14786440109462720.
- Quesada-Ruiz, L., Perez, L., Rodriguez-Galiano, V., 2019. Spatiotemporal analysis of the housing bubble’s contribution to the proliferation of illegal landfills – The case of Gran Canaria. *Science of the Total Environment* 687, 104–117. doi:10.1016/j.scitotenv.2019.06.079.
- Razi, S., Mollaei, M., Ghasemi, J., 2019. A novel method for classification of BCI multi-class motor imagery task based on Dempster-Shafer theory. *Information Sciences* 484, 14–26. doi:10.1016/j.ins.2019.01.053.
- Repsol, 2017. Rev UKCS Decommissioning Project. Technical Report. Repsol Norge AS.

- Rokach, L., Maimon, O., 2008. Data mining with decision trees: theory and applications. Series in Machine Perception and Artificial Intelligence: Volume 69, World scientific. doi:10.1142/6604.
- Rouse, S., Hayes, P., Davies, I., Wilding, T., 2018. Offshore pipeline decommissioning: Scale and context. Marine Pollution Bulletin 129, 241–244. doi:10.1016/j.marpolbul.2018.02.041.
- Rowley, H., Peters, G., Lundie, S., Moore, S., 2012. Aggregating sustainability indicators: beyond the weighted sum. Journal of Environmental Management 111, 24–33. doi:10.1016/j.jenvman.2012.05.004.
- Roy, B., 1985. Méthodologie multicritère d’aide à la décision. Economica, Paris.
- Ryman-Tubb, N.F., Krause, P., Garn, W., 2018. How artificial intelligence and machine learning research impacts payment card fraud detection: A survey and industry benchmark. Engineering Applications of Artificial Intelligence 76, 130–157. doi:10.1016/j.engappai.2018.07.008.
- Schuetter, J., Mishra, S., Zhong, M., LaFollette, R., 2018. A data-analytics tutorial: Building predictive models for oil production in an unconventional shale reservoir. SPE Journal 23, 1–075. doi:10.2118/189969-PA.
- Shell, 2017a. Brent Field Decommissioning: Comparative Assessment Procedure. Technical Report. Shell U. K. Limited.
- Shell, 2017b. Brent field decommissioning programmes. Technical Report February. Shell U. K. Limited.
- Shell, 2017c. Brent field pipelines decommissioning technical document. Technical Report February. Shell UK Limited. UK.

- Soltanmohammadi, H., Osanloo, M., Rezaei, B., Bazzazi, A., 2008. Achieving to some outranking relationships between post mining land uses through mined land suitability analysis. *International Journal of Environmental Science & Technology* 5, 535–546. doi:10.1007/BF03326051.
- Sorzano, C., Vargas, J., Montano, A., 2014. A survey of dimensionality reduction techniques. *arXiv preprint arXiv:1403.2877* .
- Spathis, C., Doumpos, M., Zopounidis, C., 2002. Detecting falsified financial statements: a comparative study using multicriteria analysis and multivariate statistical techniques. *European Accounting Review* 11, 509–535. doi:10.1080/0963818022000000966.
- Tehrany, M., Kumar, L., Jebur, M., 2019. Evaluating the application of the statistical index method in flood susceptibility mapping and its comparison with frequency ratio and logistic regression methods. *Geomatics, Natural Hazards and Risk* 10, 79–101. doi:10.1080/19475705.2018.1506509.
- Tsai, T., Li, D., 2008. Utilize bootstrap in small data set learning for pilot run modeling of manufacturing systems. *Expert Systems with Applications* 35, 1293–1300. doi:10.1016/j.eswa.2007.08.043.
- Waegeman, W., De Baets, B., Boullart, L., 2009. Kernel-based learning methods for preference aggregation. *4OR* 7, 169–189. doi:10.1007/s10288-008-0085-5.
- Witten, I., Frank, E., Hall, M., Pal, C., 2016. *Data Mining: Practical machine learning tools and techniques*. The Morgan Kaufmann Series in Data Management Systems, Morgan Kaufmann. doi:10.1016/C2009-0-19715-5.



Xodus, 2017. Osprey Field Subsea Infrastructure Comparative Assessment. Technical Report. Xodus Group Limited.

Xu, X., Liang, T., Zhu, J., Zheng, D., Sun, T., 2019. Review of classical dimensionality reduction and sample selection methods for large-scale data processing. *Neurocomputing* 328, 5–15. doi:10.1016/j.neucom.2018.02.100.