IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 12 September 2022, accepted 16 September 2022, date of publication 22 September 2022,
date of current version 30 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3208957

APPLIED RESEARCH

Enhancing Cluster Analysis With Explainable AI and Multidimensional Cluster Prototypes

SZYMON BOBEK[1,2], **(Member, IEEE), MICHAL KUK**[3], **MACIEJ SZELĄŻEK**[3], **AND GRZEGORZ J. NALEPA**[1,2], **(Member, IEEE)**

[1]Jagiellonian Human-Centered Artificial Intelligence Laboratory (JAHCAI), Jagiellonian University, 31-007 Kraków, Poland
[2]Institute of Applied Computer Science, Jagiellonian University, 31-007 Kraków, Poland
[3]Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering, AGH University of Science and Technology, 30-059 Kraków, Poland

Corresponding author: Szymon Bobek (szymon.bobek@uj.edu.pl)

This work was supported in part by the PACMEL Project funded by the National Science Centre under the CHIST-ERA Program, Poland, under Grant NCN 2018/27/Z/ST6/03392; and in part by the Cyfronet, Kraków, for Granting Access to the Computing Infrastructure Built through the Project "Platon—Science Services Platform" and "Deploying High-Availability, Critical Services in Metropolitan Area Networks (MAN-HA)" under Grant POIG.02.03.00-00-028/08 and Grant POIG.02.03.00-00-110/13. The work of Szymon Bobek was supported by a HuLCKA Grant from the Priority Research Area (Digiworld) under the Strategic Program Excellence Initiative at the Jagiellonian University under Grant U1U/P06/NO/02.16.

ABSTRACT Explainable Artificial Intelligence (XAI) aims to introduce transparency and intelligibility into the decision-making process of AI systems. Most often, its application concentrates on supervised machine learning problems such as classification and regression. Nevertheless, in the case of unsupervised algorithms like clustering, XAI can also bring satisfactory results. In most cases, such application is based on the transformation of an unsupervised clustering task into a supervised one and providing generalised global explanations or local explanations based on cluster centroids. However, in many cases, the global explanations are too coarse, while the centroid-based local explanations lose information about cluster shape and distribution. In this paper, we present a novel approach called ClAMP (Cluster Analysis with Multidimensional Prototypes) that aids experts in cluster analysis with human-readable rule-based explanations. The developed state-of-the-art explanation mechanism is based on cluster prototypes represented by multidimensional bounding boxes. This allows representing of arbitrary shaped clusters and combines the strengths of local explanations with the generality of global ones. We demonstrate and evaluate the use of our approach in a real-life industrial case study from the domain of steel manufacturing as well as on the benchmark datasets. The explanations generated with ClAMP were more precise than either centroid-based or global ones.

INDEX TERMS Data mining, clustering, explainable AI, expert's knowledge.

I. INTRODUCTION

In recent years, pattern discovery has been dominated by effective black-box models such as deep neural networks or boosting trees. However, these methods are not easily understandable, which could limit their application in areas where results of machine learning algorithms need to be combined or confronted with domain knowledge and experts' experience. To deal with this, Explainable AI (XAI) methods are being developed to bring transparency to the decision-making process of AI-based systems [1]. This trend is especially visible in the area of Industry 4.0, where a large amount of data is gathered directly from hardware and is used to discover patterns or anomalies in machinery operation as well as to provide decision support based on the results. A human operator is usually involved in the analysis and verification of the decisions of the system because the control of the critical system components cannot be left solely to the AI system. On the other hand, this requires the model to be understandable by a domain expert, as depicted in Figure 1.

The associate editor coordinating the review of this manuscript and approving it for publication was Alberto Cano.

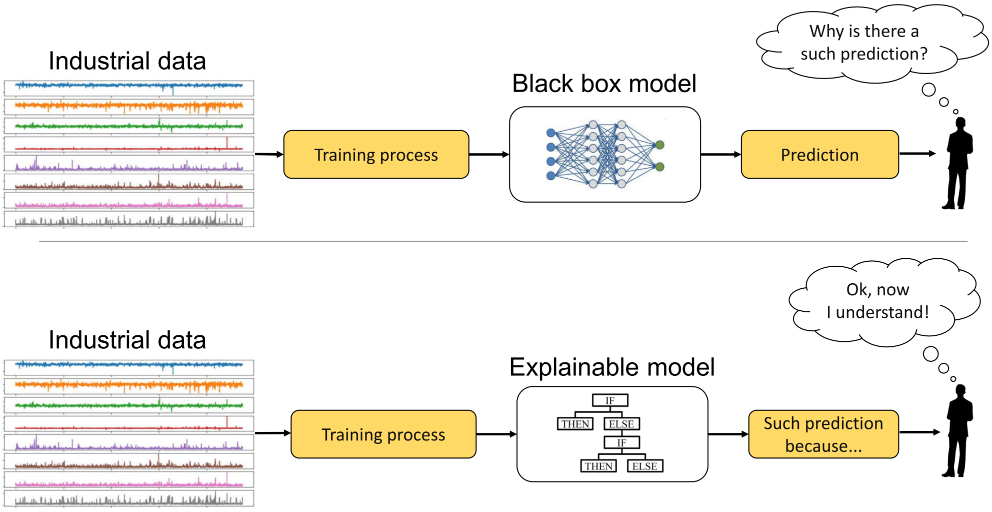

FIGURE 1. Visualization of the role of XAI in Industry 4.0 data analysis. High-stakes decisions have to be understandable to be properly justified.

In practical applications, where large amounts of data have to be analysed, the AI-based (Artificial Intelligence based) decision support is usually implemented utilising machine learning algorithms. Three types of learning can be considered: supervised, semi-supervised, and unsupervised learning. Supervised learning is an approach that makes use of labelled datasets. Semi-supervised learning can be applied in the case of using a training dataset, with both labelled and unlabelled data. Unsupervised learning uses machine learning algorithms to analyse unlabelled datasets. In most cases, XAI methods are considered with respect to a supervised machine learning task such as classification or regression. However, in many industrial applications, data comes with no labelling, making it unfeasible for supervised methods and XAI algorithms. In such cases, data mining techniques such as clustering are often used to reveal patterns hidden in the data. Clustering is defined as unsupervised learning where the objects are grouped on the basis of some similarity between them [2]. In such cases, XAI can be used to explain the differences between unfolded patterns as well as to explain a single instance assignment to a particular cluster. To apply state-of-the-art XAI methods, the considered problem should be reformulated in a manner that fits the supervised task. The main objective of such a reformulation is to obtain the proper representation of the cluster that is delivered to the explanation mechanism. An obvious choice of cluster centroids may not give valid results in the case of clusters that have complicated shapes or do not have Gaussian distribution. On the other hand, using global explanations lacks details which might be crucial for a proper understanding of the differences between clusters.

In this work, we aimed to formulate an XAI methodology that would allow balancing a trade-off between granularity of global explanations and complexity of instance-based explanation for a cluster of arbitrary shape and dimensionality. We adopted the developed methodology in a real-life industrial case of the hot-rolling process from the steel industry. To achieve this, we attempted to represent clusters with multidimensional prototypes and utilise these prototypes in the explanation process. The developed methodology can be divided into the following stages:

- Execute clustering with an arbitrarily selected method;
- Reformulate the problem to the classification task;
- Generate cluster prototypes in the form of multidimensional bounding boxes and obtain rule-based explanations for them.
- Evaluate generated rules with the use of the HeaRTDroid inference engine [3] and experts' knowledge.

This work is carried out in the CHIST-ERA Pacmel project. The project aims to develop novel methods of process mining, knowledge modelling, and intelligent sensor data analysis in Industry 4.0. In the area of rules and inference engines, we build on our previous works including the XTT2 (formalised rule representation) rule-based knowledge representation and the HeaRTDroid inference engine [3], which were developed by us using the Semantic Knowledge Engineering methodology [4].

The reminder of the paper is organised as follows: in Section II, we describe the works concerning the explainable methods. This is the foundation for our motivation and original contribution described in Section III. In Section IV, we concentrate on describing the clustering and classification methods and present a novel approach to building prototypes for clusters. This section also includes the description of a method for obtaining rule-based explanations for discovered prototypes. In Section V, we present a functional evaluation of ClAMP in comparison to centroid-based and global explanations used in state-of-the-art solutions. In Section VI, we perform human-grounded evaluation on synthetic datasets with 24 participants involved in the process. Finally, in Section VII, we move on to the case study and close the evaluation of ClAMP by showing the application of our methodology in an industrial setting. At the end, in Section VIII, we summarise the results of our work.

II. RELATED WORKS

The process which allows explaining clustering generally involves a three-step explanation procedure that changes an unsupervised clustering task into a supervised classification task [5], [6], [7]. First of all, an optimal quality clustering of unlabelled data needs to be obtained. Secondly, a classifier needs to be built that uses discovered cluster labels as values of the target variable. Finally, the classification task should be explained with XAI methods [6]. This gives us information about the differences between clusters, which can help in the final cluster analysis performed by the expert. There have been a variety of XAI methods developed over the last decade which differ in the explanation mechanism used, an explanation granularity, as well as in the form in which they present explanations. In this section, we provide a review of existing approaches for the cluster analysis enhanced with XAI algorithms. Furthermore, we present the original contribution in more detail at the end of the section.

In [8], the authors extend the image prototypes approach presented in [9] by introducing an interpretable image classification model with a pool of prototypes shared by the classes (ProtoPol), which focuses on the crucial image parts. Based on the image, the model discovers the parts of the image (prototypes) which could be useful for further analysis. This allows for a more interpretable model and to discover similarities between classes.

In [10], the authors provide a novel solution that can be used to cluster data. They call it the eUD3.5 algorithm, which relies on inducing a collection of diverse unsupervised decision trees. The main advantage of their solution is that the eUD3.5 algorithm does not require any parameters that control the number of objects in the leaf nodes because the algorithm automatically stops expanding a branch if the evaluation is worse than the best evaluation in that branch. The second important advantage mentioned in [10] is that the algorithm can provide patterns associated with each cluster that can be easily understandable by a human. The patterns describe the whole database with just a few patterns.

The authors in [7] develop the Single Feature Introduction Test (SFIT) method which is run on the model to recognise the statistically significant features which characterise each of the clusters of data. They test their discovered method on a real wealth management compliance case. The method is divided into two steps: the clustering step, and the explaining step. First, data is clustered with the use of a clustering algorithm such as K-means. The second step is to train the classifier to learn how to predict the cluster and run the SFIT procedure on the instances belonging to a considered cluster. This allows obtaining a set of features that are significantly characterising this cluster. The procedure is tested on the 2D and 3D datasets of the Fundamental Clustering Problems Suit (FCPS). In both cases, this method is able to correctly uncover patterns.

In [11], the adopted method concentrates on the centres of the clusters. Discovered Cluster-based sentence utility (CBSU, or utility) refers to the degree of relevance (on a scale from 0 to 10) of a particular sentence to the general topic of the entire cluster. However, such methods are very sensitive to the shape of the clusters and can be executed only in specific cases.

Many explainability approaches consider the use of tree-based clustering models. According to [12], the most popular method is cluster representation with the use of their centroids. However, in the case of the not compact or non-isotropic cluster, such a method cannot be executed successfully. Another common approach is that of visualisation with the use of principal component analysis but, in this case, we lose the relationship between the clusters and the original variable. In [12], the authors propose an unsupervised learning algorithm that solves the task using an optimisation lens while providing the user with more accurate and interpretable results based on the feature vectors. They use Silhouette Metrics and Dunn Index, as the objective function. Tests were executed using datasets from FCPS and real-world examples.

In [5], the authors use methods of supervised machine learning for cluster interpretation by changing the problem into a classification case. Particularly, they analyse which features are necessary to assign instances to the correct cluster. This allows recognising the characteristics relevant to specific cluster structures.

The method presented in [6] aims to explain the outcome of unsupervised algorithms. Generally, the framework relies on the expert's knowledge to, i.a., extract the correct features (feature selection). When the data is embedded, EXPLAIN-IT uses unsupervised learning techniques to explore it. In particular, EXPLAIN-IT uses a clustering technique that plays the role of a meta-learning approach, which reduces the complexity of the analysis using the idea of clustering methods – aggregating similar instances.

In [13], the authors outline that there are no effective methods to apply to security tasks. In their paper, they propose a dedicated method that generates a small set of interpretable features to explain how the input sample is classified. The main idea is to approximate the local area of the deep learning decision boundary with the use of a simple interpretable model. The model is specially designed to:

- Handle feature dependency to better work with security applications;
- Handle non-linear local boundaries to boost explanation fidelity.

The method concentrates on identifying a small set of features that are key contributors to the classification of data instances. The method generates a local approximation of the target classifier's decision boundary near a given point. This method does not assume that the local detection boundary is linear and the features are independent. Instead, they introduce a new approach to approximate the non-linear local boundaries based on a mixture regression model enhanced by fused lasso.

In [14], the authors present two novel ''algorithm-agnostic'' explainability methods: Global permutation percent change (G2PC) and Local permutation percent change (L2PC). Their methods use a well-known model-agnostic explainability method that is widely used in the context of supervised machine learning called permutation feature importance. L2PC feature importance extends the permutation to obtain explainability for clustering algorithms. In contrast to G2PC, L2PC permutes each of the features for a single sample-specific time using values that are randomly selected from the same feature of other samples in the dataset. After that, it calculates the percentage of time that the sample changes clusters during the permutations. The permutation percent change values can be used to obtain the statistical significance of each feature. As the percent change increases, the importance of a specific sample increases.

A Boolean decision rules generator [15] is a method that utilises Boolean rules either in their disjunctive normal form (DNF) or conjunctive normal form (CNF) to build predictive models. According to this idea, a low number of rules makes patterns more easily understood and interpreted by humans. The authors outline that in the case of large and complex datasets, the problem with computational time may occur. To avoid this issue, they propose an approximate column generation algorithm that uses randomisation to efficiently search the rule space and learn DNF or CNF classification rules [16].

The authors in [17] introduce a simple and practical framework called Teaching Explanations for Decisions (TED), which provides explanations that match the mental model of the consumer. The idea is based on X (the feature vector), Y (a label), and E (the explanation for each decision, which can take any form) making a classifier where the value of Cartesian product Y and E (YE) is predicted. The next step is to make decoding to partition a YE prediction into its components Y and E.

Most of the aforementioned techniques are based on state-of-the-art model-agnostic XAI algorithms such as LIME, SHAP, Anchor, and others. In the following paragraphs, we introduce them briefly. One of the most popular methods for black-box models is a local interpretable model-agnostic explanation (LIME). This method is able to generate interpretations for a single instance and can be applied to any classifier. LIME generates simulated data points around given instances through random perturbation and provides an explanation by fitting a sparse linear model over the predicted responses from the perturbed points [18].

The authors in [18] propose an extended version of LIME called DLIME (Deterministic Local Interpretable Model-Agnostic Explanations). In comparison to LIME, to find a set of samples and corresponding predictions instead of random perturbation, KNN (k-nearest neighbours) is first used to find the closest neighbours to the instance. Then, the cluster label for the test instance is assigned based on the majority label among the k-nearest neighbours. Finally, the data point belonging to the class is used to train a linear regression model which is used to generate an explanation.

TABLE 1. Summary of related works in the area of explainable clustering.

Paper reference	Explanation type	Classification model	Clustering algorithm	Executable explanations
[7]	global	any	any	no
[10]	local	fixed	fixed	yes
[11]	global	n/a	any	no
[12]	local	fixed	fixed	yes
[5]	local	any	any	no
[6]	local	any	any	yes
[14]	global	fixed	n/a	no
[15]	local	any	any	no
[16]	local	n/a	n/a	yes
[18]	global	n/a	n/a	yes
[19]	local	n/a	any	yes
[20]	local	any	any	yes

In [19], the authors present a novel model-agnostic algorithm called *The Anchor*. Based on the given instance, the Anchor algorithm generates a rule that sufficiently decides the prediction locally. It should be emphasized that changes to other feature values of the instance do not essentially affect the prediction value. For each instance, the Anchor is executed with an empty rule, subsequently, in an iterative fashion, new rules are generated and the previous is replaced if the precision is lower.

III. MOTIVATION AND ORIGINAL CONTRIBUTION

Most of the methods mentioned in the previous section are focused on a specific task and tuned to work with particular clustering algorithms, or with a particular audience. On the other hand, general frameworks such as [5], [6], [7], and [10] focus mostly on global explanations, which limits the details presented to the user and reduces the capabilities of in-depth cluster analysis. In Table 1, we present a summary of related works in the area of explainable clustering. One can observe that there is no solution that will satisfy the hybrid explanations mechanism that will: 1) allow for a balance between the expressiveness and granularity of the generated results, 2) allow the use of an arbitrary selected clustering algorithm, 3) allow the use of an arbitrary selected classification method to discover patterns between clusters, or 4) provide explanations in an executable format that allows for easier, automated integration with other system components.

In our approach, we aim mainly to provide a method that will address all of the above four issues. The starting point of this work was the preliminary results introduced at the IEEE DSAA 2021 Conference [20]. Here, we present a fully developed approach, enclosed within a methodological framework for cluster analysis with multidimensional prototypes (ClAMP) and evaluated on a real-life industrial case and benchmark datasets. The most important aspects of our original contribution include the following:

- We expanded the possibility of cluster representation. We added another method for discovering

cluster prototypes. We were interested in how randomly selected points within the cluster influence the HeaRTDroid results;
- We expanded the number of metrics considered to obtain more reliable results. In the previous approach, only the accuracy metric was calculated. In this work, we added metrics such as precision, recall, and F1-score;
- We added explainability optimization, taking into account several criteria according to which the user is able to choose the best result. In cooperation with the experts, we decided to allow deciding which metric can be treated as a target parameter;
- We created a pipeline (methodology) which as its input takes the dataset without labels and is able to generate explanations and evaluate them with the use of the HeaRTDroid rule-based inference engine;
- We provided the final human-readable rules to the experts for evaluation.

In the following section, more details on the ClAMP methodology will be provided.

IV. CLUSTER ANALYSIS WITH MULTIDIMENSIONAL PROTOTYPES

The main goal of our work on ClAMP was to provide a method for cluster analysis that will be agnostic with respect to the clustering and classification algorithms and will provide explanations in the form of executable and human-readable rules. The ClAMP methodology can be divided into four stages as depicted in Figure 2.

1) **Phase 1:** Clustering of unlabelled data with arbitrary selected clustering algorithm.
2) **Phase 2:** Reformulation of the clustering problem into the classification task and building a classifier that is trained to distinguish labelling discovered in the previous phase.
3) **Phase 3:** Generation of cluster prototypes as multidimensional bounding boxes on top of the clustering performed in the first phase.
4) **Phase 4:** Generation of explainable rules for the cluster prototypes generated in phase 3.

In the following sections, these main phases will be described in detail.

A. PHASE 1: CLUSTERING OF UNLABELLED DATA

Good quality of clusters is crucial in obtaining good quality explanations for them. The choice of the clustering algorithms is highly dependent on the characteristics of the dataset and the shape of the clusters. This is why, in our methodology, we assume that this step should be independent of the explanation mechanism.

There are various different clustering algorithms that can be applied to different kinds of data. One of the advantages of the ClAMP methodology is the possibility of applying different clustering methods, leaving the opportunity to choose the one which gives the best results. In this work, we tested the following clustering methods to assign labels to the analysed datasets: Gaussian Mixture, BIRCH (balanced iterative reducing and clustering using hierarchies), and the Deep temporal clustering algorithm. The first two methods described above are implemented in scikit-learn [21]. The third method is presented in [22]. The algorithm utilises an autoencoder for temporal dimensionality reduction and a novel temporal clustering layer for cluster assignment. Then, the clustering and dimensionality reduction objectives are optimised. To detect the optimal number of clusters, we used silhouette score; however, the choice of the metric used for selecting the number of clusters is not limited.

It is worth noting that this stage is independent of the whole methodology. In fact, one can also apply our approach to the dataset which originally contained labels, or where labels were obtained using expert knowledge instead of a clustering algorithm. This could be particularly useful in cases where the cluster analysis is performed mainly for conformance checking with existing domain knowledge [23].

B. PHASE 2: REFORMULATION OF THE CLUSTERING PROBLEM INTO THE CLASSIFICATION TASK

To reformulate the clustering problem into the classification task it is necessary to find a classifier that reproduces labels obtained during the clustering stage in the best possible way.

In our work, we chose XGBoost (Gradient Boosting framework) classifier [24] as the classification algorithm, an optimised distributed gradient boosting open-source package designed to be highly efficient, flexible, and portable. It implements machine learning algorithms under the Gradient Boosting framework. XGBoost provides parallel tree boosting that solves many data science problems in a fast and accurate way. Results demonstrated in [24] show that the XGBoost classifier can be used for a wide range of problems. Classifiers have great potential and allow the obtaining of good results; however, they have a lot of hyperparameters that directly affect these results. To account for this, hyperparameter tuning should be done during the algorithm performance [25]. There is a possibility to do it manually, but in such a case, the user can not be sure that the best parameter settings have been determined. To do it automatically, a simple GridSearch algorithm can be applied that allows checking each combination of parameter values defined in their domains (ranges) determined by the user. In our case, we applied a RandomizedSearchCV (RandomizedSearch Cross-Validation) available in scikit-learn [21], because this optimiser allows obtaining satisfying results by trying only a fixed number of parameter settings. Random search is actually more practical than grid search [25], as it does not test all parameters but executes the search at random. For the automatic hyperparameter tuning, other optimization methods can also be applied, e.g., the Sequential Model-Based Optimization (SMBO) implemented in the model-based optimisation package (mlrMBO) [25], [26].

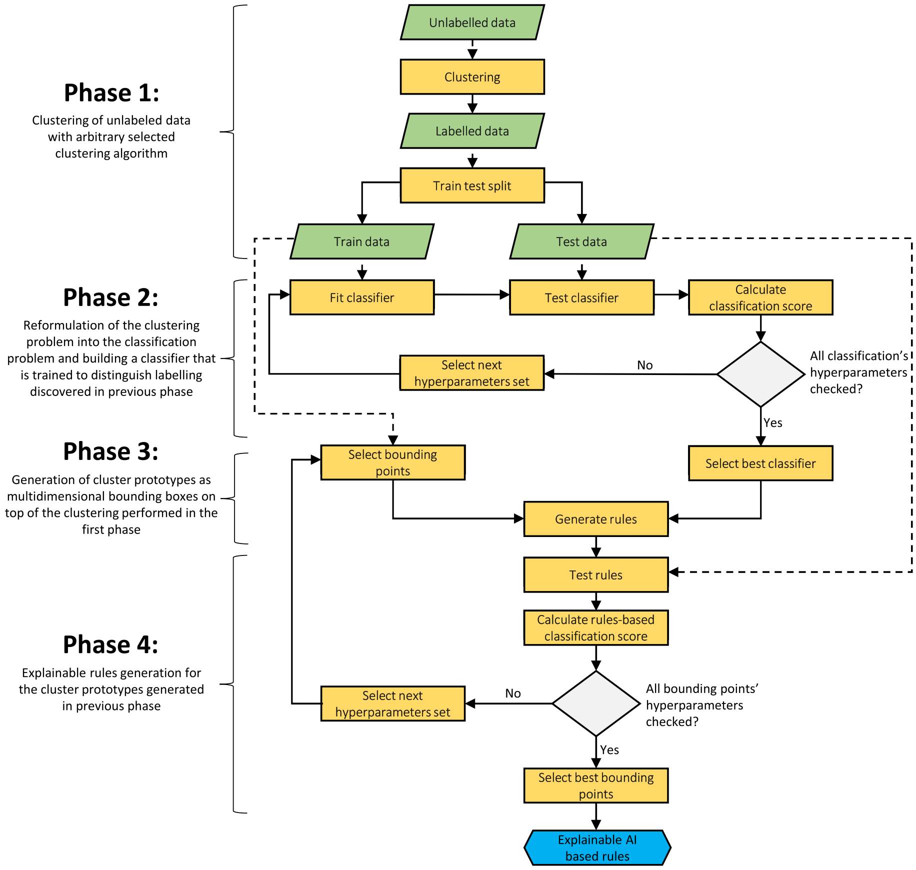
Phase 1:
Clustering of unlabeled data with arbitrary selected clustering algorithm
Phase 2:
Reformulation of the clustering problem into the classification problem and building a classifier that is trained to distinguish labelling discovered in previous phase
Phase 3:
Generation of cluster prototypes as multidimensional bounding boxes on top of the clustering performed in the first phase
Phase 4:
Explainable rules generation for the cluster prototypes generated in previous phase
Unlabelled data
Clustering
Labelled data
Train test split
Train data
Test data
Fit classifier
Test classifier
Calculate classification score
All classification's hyperparameters checked?
No
Select next hyperparameters set
Yes
Select best classifier
Select bounding points
Generate rules
Test rules
Calculate rules-based classification score
All bounding points' hyperparameters checked?
No
Select next hyperparameters set
Yes
Select best bounding points
Explainable AI based rules

FIGURE 2. ClAMP **methodology diagram.**

To validate the effectiveness of the classification methods built on top of cluster labels, several metrics can be used. In our case, we used recall, precision, F1-score, and accuracy.

C. PHASE 3: MULTIDIMENSIONAL BOUNDING BOXES AS CLUSTER PROTOTYPES

The classifier that allows the correct assignment of instances to the previously discovered clusters is the main requirement for phase 3 of the ClAMP methodology. This classifier will be used later to generate explanations for a particular cluster, based on the instances that form a (potentially multidimensional) bounding box around it. The selection of the bounding box points that form the cluster representation (prototype) is the main objective of this phase of the ClAMP methodology. The idea of such an approach is presented on the two-dimensional dataset in Figure 3.

In the case of a real dataset with many features, the shape of the cluster may be unimaginable and the selection of a method determining the proper description points can be difficult. That is why, in the proposed methodology, we treat the method for discovering cluster prototypes as a tuning parameter that can be adjusted to the specific case. Three different methods for discovering *cluster prototypes* are considered in this paper: Random selection, K-D tree (k-dimensional tree), and Isolation forest; all are described next.

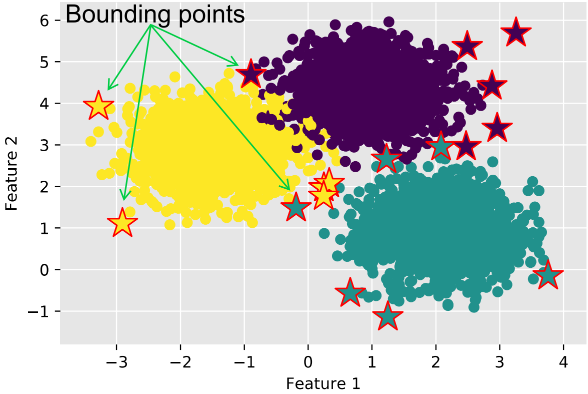

FIGURE 3. The idea behind determining a bounding box. Relatively outer (bounding) points can give more information about the boundary of each cluster than choosing the centroids of each cluster.

It is worth noting that for clusters of different shapes, different bounding boxes may be suitable for different clusters. Therefore, in ClAMP, we optimise the selection of the method for each of the clusters separately. The selection of CLAMP hyper-parameters can be done automatically with any optimisation algorithm and with respect to the target metric we want to optimise (for instance, the accuracy of the explanations obtained). Selection of a metric and optimisation algorithm depends on the task we want to solve and the data we use (e.g., balanced, imbalanced, etc.) and, therefore,

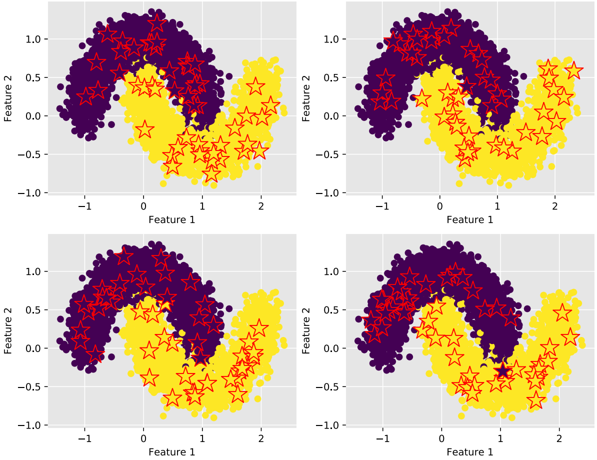

FIGURE 4. Exemplary points determined by the random selection.

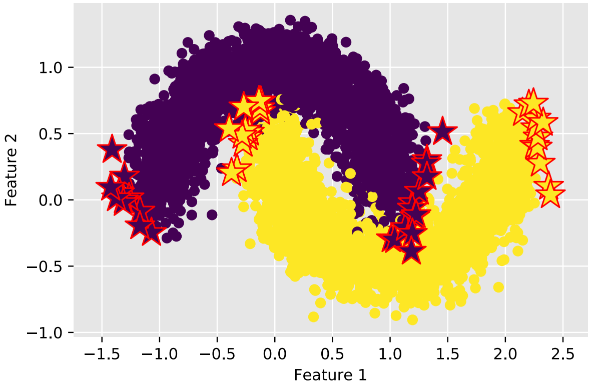

FIGURE 5. Exemplary points determined by the K-D tree.

is considered outside the scope of this work, which focuses on the ClAMP methodology itself, not on particular domain-dependent applications.

1) RANDOM SELECTION APPROACH

The random selection method considered in this paper generates a randomly selected set of points belonging to each cluster. The number of points to be selected from each cluster is treated as a hyperparameter which should be optimised. To choose points, the "sample" function, built in the Pandas library in Python, was used. To obtain randomly selected points, only the number of items was passed and other parameters were used with their default values.

Exemplary points determined by the random selection approach are presented in Figure 4, further divided into 4 separate charts denoting different runs of the random selection procedure. As can be seen for each of the charts, the determined points which are used for rule generation are different. Therefore, for evaluation purposes, several runs are used and averaged to obtain reliable results.

2) K-D TREE APPROACH

The K-D Tree algorithm addresses the computational inefficiencies of the brute-force approach. This algorithm allows a general reduction of the required number of distance calculations with the use of encoding aggregate distance information for the sample. In particular, if point "A" is very far from point "B", and point "B" is much closer to point "C" than to point "A" then the algorithm knows that points "A" and "C" are very distant. The main advantage of such a conception is obtaining information about the distance between points "A and "C" without calculating the distance between them." The K-D tree is a binary tree structure that recursively divides the parameters space along the data axes, dividing it into nested orthotropic regions into which data points are filled. Dividing is executed only along the data axes and no D-dimensional distances need to be computed, that is why the K-D tree is very fast. It should be outlined that this method is fast in relatively low-dimensional cases $D < 20$ and becomes inefficient when D grows above the mentioned value [21], [27].

Implementation of the K-D tree algorithm requires tuning of some of the hyperparameters like leaf size and metric. According to the documentation, the "leaf size" parameter does not affect the results of the algorithm, so the default value was used. For the "metric" parameter, two possible values were considered in this paper: "minkowski" and "manhattan". Because the bounding box we are looking for consists of the outremost points, we added one more hyperparameter which is the percentage of the farthest points from the centre of each cluster. Exemplary points determined by the K-D tree approach are presented in Figure 5. As can be seen, for each of the clusters, the KD-tree algorithm found the outermost points (boundaries of each cluster), which was one of the goals of our developed methodology.

3) ISOLATION FOREST APPROACH

The isolation forest method is one of the ways to execute outlier detection in high-dimensional datasets. The principle of operation is to "isolate" observations by randomly selecting a feature and then randomly selecting a split value between the maximum and minimum values of the selected feature [21], [28]. In the algorithm, the recursive partitioning can be represented by a tree structure, while the number of splittings required to isolate each sample is equivalent to the path length from the root node to the terminating node. The length of the path mentioned above is the measure of normality and our decision function, and this length is averaged over a forest of random trees. Thanks to random portioning, shorter paths for anomalies are produced. Hence, when the random trees collectively produce shorter paths, it is more probable to assign a sample as an anomaly [29], or bounding box point.

Implementation of an Isolation Forest requires tuning of some of the hyperparameters such as:

- the number of base estimators in the ensemble;
- the number of samples to draw from X to train each base estimator;

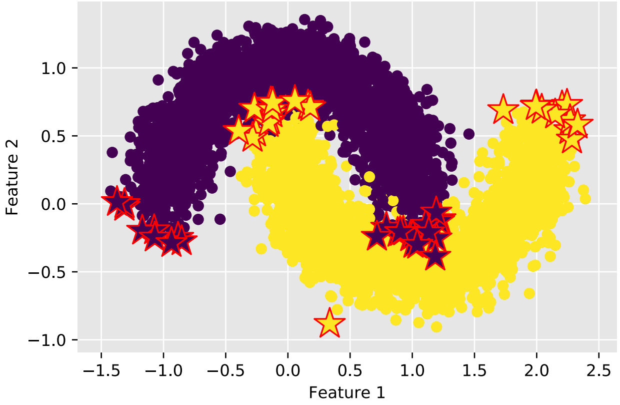

FIGURE 6. Exemplary points determined by Isolation forest.

- the number of contaminations of the dataset, i.e., the proportion of outliers in the dataset.

and several others. That is why, in this case, hyperparameter optimisation is necessary.

In the methodology developed in this paper, the Isolation Forest algorithm is applied to detect the outer points belonging to the specified cluster, which can be used to execute rules. The set of hyperparameters allows adjusting the algorithm to detect describing points. In our case, we decided to adjust only the contamination which is the proportion of outliers in the dataset to all points in the dataset and it directly affects the number of describing points obtained. Exemplary points determined by the Isolation Forest approach are presented in Figure 6.

D. PHASE 4: GENERATION OF EXPLAINABLE RULES

In this paper, to generate explainable rules, we use the Anchor explainer. The quality of the rules is evaluated not only by a human expert but also automatically with the HeaRTDroid rule-based inference engine [3]. This allows comparing our method to other approaches with well-known metrics such as F1, accuracy, precision, and recall.

1) THE ANCHOR EXPLAINER – RULES GENERATOR

Three methods for generating bounding box representations of clusters are used to provide the input to the explanation algorithm. In this work, we used Anchor, which is a novel model-agnostic algorithm that is able to explain the behaviour of complex models with high-precision rules representing local conditions for prediction. The Anchor explainer introduces explanations based on "if-then" rules, called "anchors". The algorithm generates human-readable rules which do not depend on the rest of the feature values of the instance. Furthermore, Anchor's rules are executed only if all conditions presented in the rule are satisfied. As the Anchor algorithm is model-agnostic, it can be applied to any class model [30]. Contrary to the LIME algorithm [18], which creates a linear decision boundary that best approximates the model given a perturbation space, the Anchor explainer is able to construct an explanation whose coverage is adapted to the model's behaviour, and clearly determine their boundary [30].

TABLE 2. Exemplary rules generated by the Anchor explainer for the artificial dataset.

No	Cluster	Coverage	Precision	Rule
1	0	0.25	0.90	$F2 > 2.93$ and $0.96 < F1 \leq 1.83$
2	0	0.89	0.89	$F2 > 2.93$ and $F1 > 0.96$
3	0	0.25	1	$F2 > 4.11$ and $F1 > 1.83$
4	1	0.50	1	$F2 \leq 1.11$ and $F1 > 0.96$
5	1	0.50	1	$F2 \leq 2.93$ and $F1 > 0.96$
6	2	0.25	1	$F1 \leq -1.36$
7	2	0.50	0.91	$F1 \leq 0.96$ and $1.11 < F2 \leq 2.93$

Exemplary rules generated by the Anchor explainer are presented in Table 2. The *Cluster* column determines the number of the cluster which is determined by the rule. The *Coverage* and *Precision* columns describe respectively: the ratio of the number of instances for which the rule holds in the whole dataset and its precision on this subset of instances.

The rules obtained with the Anchor algorithm can be directly analysed by the expert but can also be formalised and executed. This allows for automatic evaluation of the rules obtained within the ClAMP methodology as well as easier integration with other system components. For the purpose of representation and execution of the rules, we use the HMR+ rule language and HeaRTDroid inference engine described in the following paragraphs.

2) HeaRTDroid RULE-BASED INFERENCE ENGINE

The HeaRTDroid is a rule-based engine that uses the rule-based language HMR+, which allows reasoning and handling of uncertain and incomplete knowledge. The HMR+ language used by the HeaRTDroid also allows for modelling uncertainty with certainty factor algebra [3].

In our methodology, HeaRTDroid is used for executing a rule-based model consisting of rules, precision and coverage parameters, and cluster numbers determined by the rule, as shown in Table 2. The key idea of using HeaRTDroid is to evaluate the effectiveness of the rule-based model which is provided by the Anchor algorithm.

More specifically, the rule-based model with the above-mentioned parameters and data points without any labels is treated as an input to the HeaRTDroid interference engine. Then, the HeaRTDroid is executed and the main task of this stage is to predict the cluster number based on the given rule, precision, and coverage parameters, and the point under test. This action is executed for each point in the tested dataset. As a result, a cluster number is predicted for each tested point.

3) EVALUATION METRICS

The HeaRTDroid allows obtaining the labels created on the basis of the rule, precision and coverage parameters, and a given instance. As a result, we are able to compare the original labels obtained from clustering with the labels predicted by the HeaRTDroid. In the developed methodology, we use the following evaluation metrics: Precision, F1-score, Accuracy, and Recall with micro average.[1]

4) RULES EVALUATION BY EXPERTS

Along with the functional evaluation of the explanations' quality presented in previous paragraphs, the human-grounded or task-grounded evaluation in cooperation with experts is also possible in the ClAMP methodology. We provide rules which consist of the features names, values, and inequality signs – human-readable form, to the experts. The task is to check the rules generated by the Anchor algorithm and evaluate them. An important issue for our methodology is to obtain rules which would be understandable and useful for the experts, which means that after looking at them, the expert should be able to clearly assign which rules concern which cluster and determine how well these rules describe the cluster. Additionally, the expert should be able to determine whether these rules bring information that allows separating the clusters and how complicated this separation is. To do this, the expert should also take into consideration similarities between the rules. To fully evaluate our methodology, we want to gain information about the structure of the rules; if they are short or too long, or whether the number of rules is not too large. Due to the number of iterations in the rule generation optimisation process, and the number of hyperparameters of the stages described above, only rules with the best scores are delivered to the experts. By default, in the developed methodology, the optimisation of the F1-score metric is applied. However, the choice of the optimisation metric can always be modified depending on the needs as well as experts' suggestions. This should allow obtaining the best scores and rules for a specific example.

To evaluate the developed methodology, we tested it on three cases. The first case concerns multiple publicly available benchmark detests, where functional evaluation was performed to check the quality of explanations in comparison to state-of-the-art methods based on centroids or global explanation. The second case uses human grounded evaluation on synthetic, reproducible datasets. The third case uses real industrial data from the hot-rolling process in the steel manufacturing industry. All of the cases are described in the following sections.

V. EVALUATION ON BENCHMARK DATASETS

In this section, we present results obtained from the evaluation of the ClAMP methodology on the artificial and publicly available datasets. The goal of this section was to confront the novel ClAMP methods of generating explanations with state-of-the-art approaches that are based on cluster centroids or global explanations. This forms a reproducible set of tests, focused on the functional evaluation (no human factor involved) that can be used to achieve an unbiased comparison of our method with other approaches.[2] The factor that we took into consideration in this type of evaluation was the quality of the explanations in terms of accuracy. We wanted to prove that ClAMP provides more accurate explanations at a similar level of complexity (e.g., length of the rule, number of rules) compared to centroid-based and global explanations.

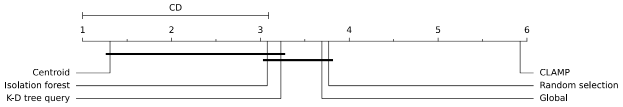

FIGURE 7. Critical difference for Nemenyi test with $\alpha = 0.05$.

All of the phases of ClAMP (see Figure 2) were fully automated and optimised with the GridSearch algorithm. The generated rules were tested against selected quality metrics (i.e., accuracy, F1, precision and recall) in a 10-fold cross-validation approach. As a result, we obtained 10 measurements for each of the combinations of dataset and bounding box selection methods. The summarised results for the F1 metric are presented in Table 3.

Our goal was to show that ClAMP selection methods are better than centroid-based and global ones. Therefore, we performed a Friedman test followed by a Nemenyi pairwise post-hot test for multiple comparisons of mean rank sums.

From the Friedman test, we obtained statistics equal to 28.0, with a p-value equal to 0.000008. With 6 algorithms and 14 datasets, we have 5 and 65 degrees of freedom respectively, which allows us to determine that the critical value for $F(5, 65)$ for $\alpha = 0.05$ is 2.35. This allows us to reject the null hypothesis.

After this, we performed a Nemenyi test to observe how the algorithms differ, and between which algorithms the difference is statistically significant. The results from the post-hoc Nemenyi test are presented in Tabel 4 and also visualised in Figure 7.

It can be observed that the critical distance is 2.015, and we can prove that ClAMP is significantly better than other methods in achieving good quality explanations. It is worth noting that each of the bounding box methods taken separately (i.e., Isolation forest, Random selection, K-D tree query) might not be significantly better than the others; it depends on the cluster shapes and, thus, the dataset used for clustering. It also depends on the clustering algorithms used (e.g., K-means produce similarly shaped clusters, while DBSCAN might produce arbitrarily shaped groups). Therefore, using ClAMP in order to optimise the selection of the bounding box is a reasonable approach.

[1] This is the Recall metric for multi-class classification that aggregates contributions of true positives, and for all classes and averages them over the global sum of true positives and false negatives, hence, taking into account possible class imbalance.

[2] The datasets along with the source code of the benchmark were made publicly available at https://github.com/sbobek/clamp

TABLE 3. Comparison of F1 performance. Column denoted as ClAMP represents the combined approach that integrates all of the bounding box methods, including Isolation forest, K-D tree query and Random selection optimised against selected quality measures. The values after ± denote standard deviation in 10-fold cross-validation.

Dataset	CLAMP	Centroid	Global	Isolation forest	K-D tree query	Random selection
balance	0.88 ± 0.06	0.70 ± 0.16	0.85 ± 0.12	0.78 ± 0.12	0.83 ± 0.11	0.79 ± 0.12
breast_tissue	0.93 ± 0.11	0.47 ± 0.21	0.52 ± 0.07	0.64 ± 0.17	0.78 ± 0.14	0.88 ± 0.14
bupa	0.79 ± 0.19	0.46 ± 0.13	0.72 ± 0.18	0.47 ± 0.09	0.42 ± 0.06	0.63 ± 0.09
ecoli	0.81 ± 0.03	0.57 ± 0.07	0.70 ± 0.04	0.71 ± 0.06	0.72 ± 0.06	0.74 ± 0.04
glass	0.91 ± 0.04	0.66 ± 0.14	0.90 ± 0.03	0.85 ± 0.06	0.79 ± 0.07	0.88 ± 0.04
iris	0.97 ± 0.01	0.81 ± 0.04	0.92 ± 0.02	0.95 ± 0.03	0.90 ± 0.04	0.87 ± 0.07
lung_cancer	0.64 ± 0.01	0.00 ± 0.00	0.64 ± 0.01	0.58 ± 0.12	0.55 ± 0.19	0.08 ± 0.20
lymphography	0.91 ± 0.10	0.76 ± 0.15	0.78 ± 0.09	0.76 ± 0.18	0.81 ± 0.18	0.81 ± 0.19
parkinsons	0.86 ± 0.04	0.62 ± 0.10	0.85 ± 0.05	0.77 ± 0.06	0.80 ± 0.05	0.78 ± 0.06
primary_tumor	0.90 ± 0.10	0.66 ± 0.21	0.71 ± 0.04	0.74 ± 0.14	0.78 ± 0.12	0.84 ± 0.12
seeds	0.95 ± 0.03	0.80 ± 0.10	0.92 ± 0.03	0.93 ± 0.03	0.90 ± 0.04	0.87 ± 0.04
vote	0.99 ± 0.03	0.97 ± 0.04	0.98 ± 0.01	0.94 ± 0.09	0.97 ± 0.05	0.98 ± 0.03
wdbc	0.95 ± 0.01	0.78 ± 0.06	0.95 ± 0.01	0.92 ± 0.02	0.93 ± 0.03	0.94 ± 0.01
wine	0.95 ± 0.03	0.68 ± 0.11	0.93 ± 0.03	0.94 ± 0.03	0.90 ± 0.02	0.91 ± 0.03

TABLE 4. Nemeny post-hoc test results.

	CLAMP	Centroid	Global	Isolation forest	K-D tree query	Random selection
CLAMP	1.00	0.00	0.04	0.00	0.00	0.01
Centroid	0.00	1.00	0.00	0.07	0.04	0.01
Global	0.04	0.00	1.00	0.81	0.90	0.90
Isolation forest	0.00	0.07	0.81	1.00	0.90	0.90
K-D tree query	0.00	0.04	0.90	0.90	1.00	0.90
Random selection	0.01	0.01	0.90	0.90	0.90	1.00

In the following sections, we evaluate the explanations with human-grounded evaluations to observe if the quality of explanations in human perception are also at a satisfying level.

VI. EVALUATION ON SYNTHETIC DATASETS

A. THE ANALYSED DATASETS

We established four artificially generated data samples: Gaussian blobs in two-dimensional space, Gaussian blobs in three-dimensional space, values randomly generated in two-dimensional space, and an Iris dataset. We decided to use such simple and obvious datasets because they are well known, and most users should possess skills that allow them to interpret and evaluate the generated rules based on the knowledge of the data or visualised charts.

To make this evaluation more reliable, we added noise to each of the datasets. For both Gaussian blobs datasets values, we tuned the noise by increasing the standard deviation. In the case of a randomly generated dataset, we changed the range of each cluster by the use of rules which change cluster assignments to another cluster. We did only one exception concerning the Iris dataset. In this case, we didn't change any cluster assignment. The datasets used are presented in the following figures:

- Gaussian blobs dataset in two-dimensional space is presented in Figure 8
- Gaussian blobs dataset in three-dimensional space is presented in Figure 9
- Randomly generated values dataset in two-dimensional space is presented in Figure 10

The last step was to provide the dataset to the participants who were asked to use ClAMP methodology to generate explanations for discovered clusters by tuning hyperparameters of ClAMP and finally evaluate their quality.

The dataset was randomly chosen for each participant. After the programming task was completed, the participants were obliged to fill in a survey containing evaluation questions.[3]

The next section presents the obtained results from the evaluation on synthetic datasets.

B. RULES ANALYSIS BY PARTICIPANTS

In the following section, we present results obtained by the 25 participants who took part in the study. Each participant was asked to evaluate the clustering results and explanations according to the 4 criteria listed below. Additionally, we asked the participants several questions concerning each of the criteria used to obtain the evaluation.

1) Adequacy of granularity level of explanations:
   a) Are the rules adequate to explain a given cluster or more individual instances in the cluster?
   b) How many rules (maximum) can each cluster be described with so that the rules are still understandable?
2) Evaluation time in comparison to cluster analysis without explanations:
   a) What would be more time-consuming to distinguish and describe the clusters: using rules or using available cluster labels?

[3] The script and evaluation survey is available at: https://github.com/sbobek/clamp

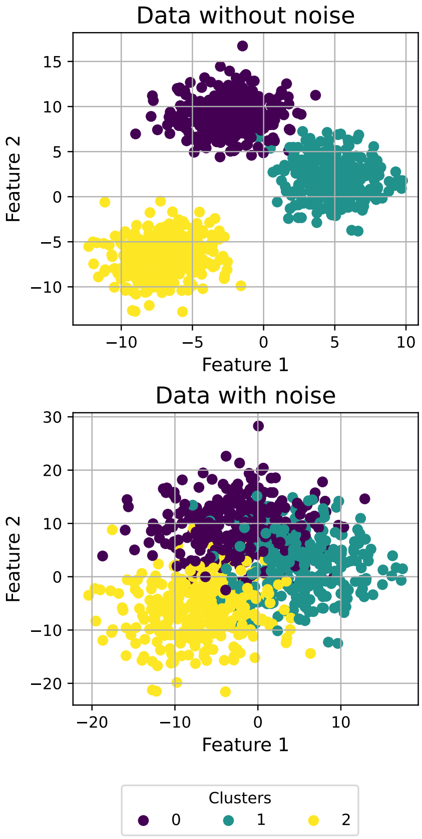

FIGURE 8. Gaussian blobs in two-dimensional space.

3) Data science or domain knowledge experience required to properly interpret explanation results:
   a) How understandable to you are the rules, i.e., do they provide information on the basis of which you are able to draw dependencies between them?
4) Overall usefulness of the rules:
   a) How do the rules help distinguish clusters and understand how they differ?
   b) How does overlap between rules make it difficult to interpret them?
   c) Have you noticed dependencies in the rules?
   d) How do these dependencies help you to understand the rules?

We prepared an online form which was provided to the participants to evaluate the rules. The participants' answers were collected with a 5-point bipolar scaling method, analogous to the Likert scale. We additionally asked the participants to put optional comments related to each of the criteria. These comments were analysed in Section VII-E. All obtained answers are presented in the following bar and box plots, with a triangle marked green as a mean value.

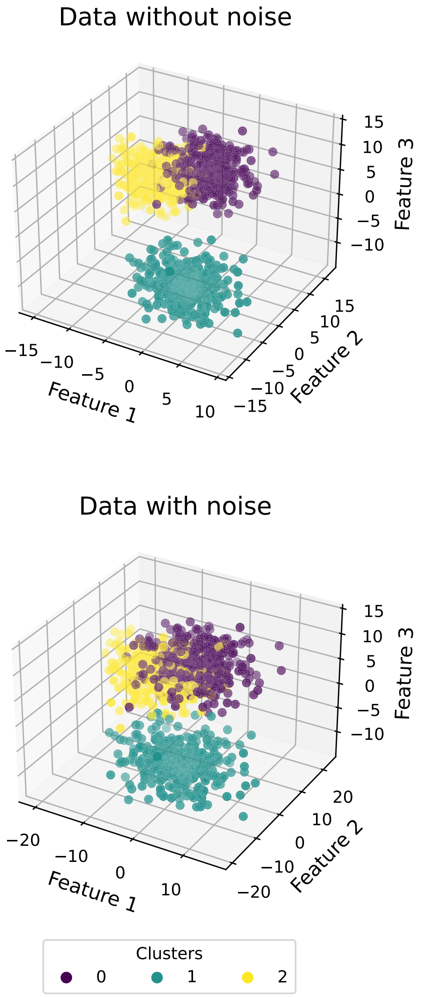

FIGURE 9. Gaussian blobs dataset in three-dimensional space.

1) ADEQUACY OF GRANULARITY LEVEL OF EXPLANATIONS

This criterion was selected to investigate if the method allows for a good trade-off between the generality of the explanation and the amount of detail required to properly analyse the cluster results. To evaluate this, we asked two questions to the participants.

Question asked to the participants: Are the rules adequate to explain a given cluster or more individual instances in the cluster?

Answer: The answer is presented in Figure 11, where **1** corresponds to the **cluster** and **5** to the **instance**. The majority of the participants decided that the explanations better explain the whole cluster rather than a single instance. The median value of the response is equal to 2. Therefore, they assure a good level of generality.

Question asked to the participants: How many rules (maximum) can each cluster be described with so that the rules are still understandable?

Answer: The answer is presented in Figure 12. Most participants who provided rather a low number of rules are better at explaining the cluster. This is the premise for the conclusion that the rules generated by our method are expressive enough to give a sufficient amount of information (details) to participants. Two of the participants significantly stand out from the rest of the results.

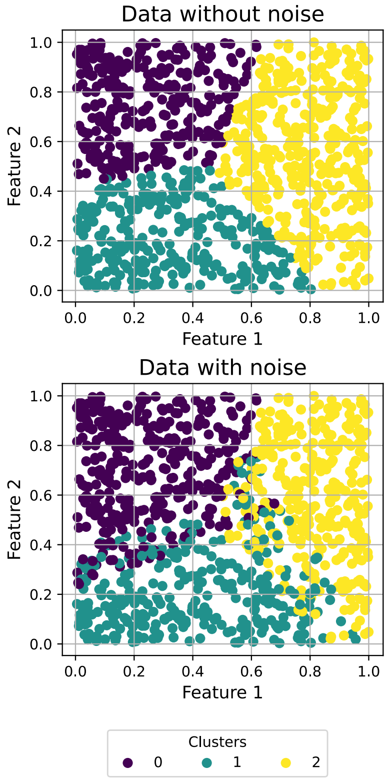

FIGURE 10. Randomly generated values dataset in two-dimensional space.

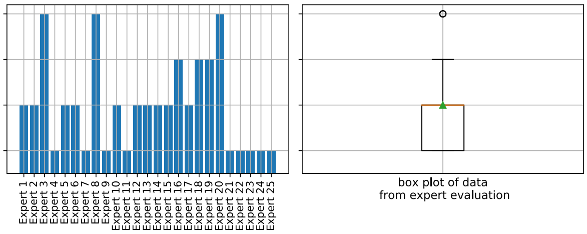

FIGURE 11. Are the rules adequate to explain a given cluster or more individual instances in the cluster?

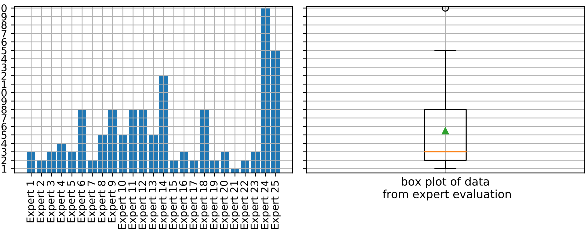

FIGURE 12. How many rules (maximum) can each cluster be described with so that the rules are still understandable?

2) EVALUATION TIME IN COMPARISON TO CLUSTER ANALYSIS WITHOUT EXPLANATIONS

The goal of this criterion was to determine if the method decreases the analysis time.

Question asked to the participants: What would be more time-consuming to distinguish and describe the clusters: using rules or using available cluster labels?

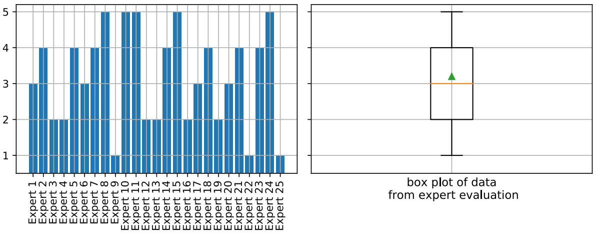

FIGURE 13. What would be more time-consuming to distinguish and describe the clusters: using rules or using available cluster labels?

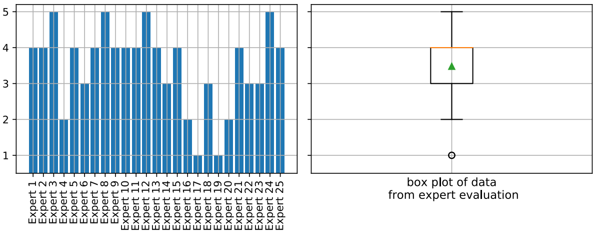

FIGURE 14. How are the rules understandable to you, i.e., do they provide information on the basis of which you are able to draw dependencies between them?

Answer: The answer is presented in Figure 13. Where **1** corresponds to the **rules** and **5** to the **labels**. The median value of the results is equal to 3. This means that the participants state that in both cases time could be comparable. However, it could be caused by the fact that we allow considering this question to relatively easy-to-understand datasets.

3) DATA SCIENCE OR DOMAIN KNOWLEDGE EXPERIENCE REQUIRED TO PROPERLY INTERPRET EXPLANATION RESULTS

The goal of this criterion was to determine if the method can be evaluated by participants who possessed only domain knowledge and not having any experience connected with data science.

Question asked to the participants: How are the rules understandable to you, i.e., do they provide information on the basis of which you are able to draw dependencies between them?

Answer: The answer is presented in Figure 14. Where **1** corresponds to **non-understandable** and **5** to **understandable**. Most of the participants agreed that the generated rules were understandable for them.

4) OVERALL USEFULNESS OF THE RULES

The goal of this criterion was to evaluate the overall usefulness of the rules. In the case of very similar rules, the challenge is to separate the rules among clusters, and thus their analysis is complicated. If the rules differ significantly, it is much easier to assign the rules to a specific cluster and determine the differences in the clusters.

Question asked to the participants: How do the rules help distinguish clusters and understand how they differ?

Answer: The answer is presented in Figure 15. Where **1** corresponds to **not helpful at all** and **5** to **very helpful**. All

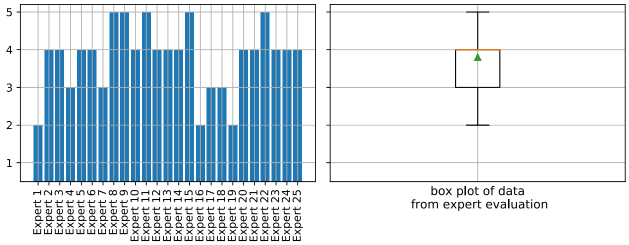

FIGURE 15. How do the rules help distinguish clusters and understand how they differ?

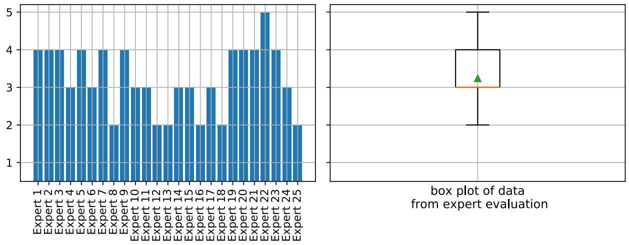

FIGURE 16. How does overlap between rules make it difficult to interpret them?

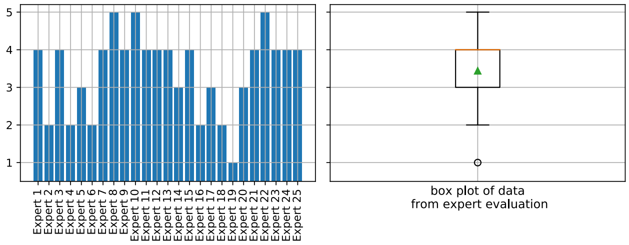

FIGURE 17. Have you noticed dependencies in the rules?

of the participants agreed that the rules are helpful in spotting the differences between clusters.

Question asked to the participants: How does the overlap between rules makes it difficult to interpret them?

Answer: The answer is presented in Figure 16. Where **1** corresponds to **does not limit** and **5** to **limit**. Most of the participants decided that the overlap of rules' conditions limits interpretability. We discuss these results in Section VI-C.

Question asked to the participants: Have you noticed dependencies in the rules? The goal of this question is to determine if there are some patterns in the rules. For example, are there rules that contain the same features in their conditional part?

Answer: The answer is presented in Figure 17. Where **1** corresponds to **hard to see** and **5** to **can be seen**. Most of the participants noticed such dependencies which in most cases were helpful to determine clusters - according to the following question.

Question asked to the participants: How do these dependencies help you to understand the rules?

Answer: The answer is presented in Figure 18. Where **1** corresponds to **not helpful** and **5** to **very helpful**.

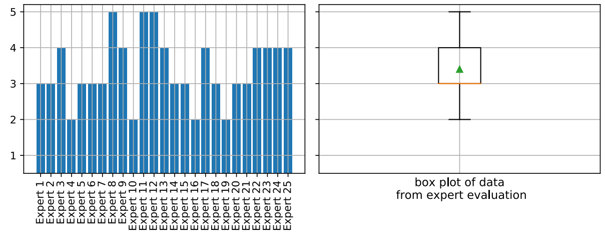

FIGURE 18. How do these dependencies help you to understand the rules?

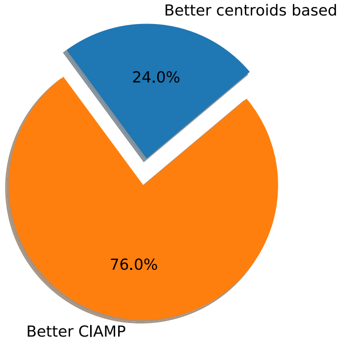

FIGURE 19. In comparison to a benchmark (centroids based) are the ClAMP **results better?**

5) OVERALL EVALUATION

To finally evaluate our methodology based on the artificially generated datasets we decided to ask another question. In the script, there was a possibility to generate rules not based on the bounding box prototypes but based on the centroid point in each of the clusters. As a result, the participants were able to compare results obtained for each of the methods and answer the following question.

Question asked to the participants: In comparison to a benchmark (centroids based) are the ClAMP results better?

Answer: The answer is presented in Figure 19. Almost 80% of the responders answer that the ClAMP methodology allows obtaining rules that better describe clusters and help to understand them.

C. DISCUSSION ON THE RESULTS OBTAINED ON THE ARTIFICIALLY GENERATED DATASETS

The overall evaluation results suggest that our methodology is useful for participants in cluster analysis. Taking the obtained results into consideration, we are able to state that the developed methodology delivers satisfactory results in the case of application to artificially generated datasets.

Participants agreed that the ClAMP methodology allows describing better clusters than each of the instances, which was one of our goals (see Figure 11). Additionally, Figure 12 depicts the answers to the question of how many rules the participants considered satisfactory to maintain clarity of the rules. It shows the maximum number of rules that are satisfactory to understand clusters well. Only three of the participants answered that this number could be greater than 10. This answer is aligned with our observation that interpretable models are not always explainable due to their complexity.

What is surprising is there is no clear answer to what is a more time-consuming cluster description, using the rules or using instances, see Figure 13. We assume that the use of rules should be much faster than the use of each instance. However, we provided relatively uncomplex datasets to the evaluation and simple analysis of the values concerning each of the clusters allows understanding of the relations between clusters. On the other hand, in the case of datasets with more features, this case could be not so easy to determine, and in such examples, the ClAMP methodology works significantly better.

One of the issues observed by both authors and participants is overlapping rules, which occurs when the generated rules possess some common parts. This issue limits interpretability and causes that participants have to spend more time to understand rules, see Figure 16. However, once spotted, they improve the overall understanding. Based on the results presented in Figures 17 and 18, we assume that participants noticed some dependencies and in generated rules, which helped them better understand the whole explanation. This observation could be an argument for creating explanations that take advantage of different forms of visual analytics to help users spot important patterns within the explanations [31].

To sum up, in general, almost 80% of the participants agreed that the ClAMP methodology is better than that based on the cluster centroids. This enables us to state that the application of our methodology is useful in the case of artificially generated datasets. In the following section, we consider a real industrial case and present the results obtained by the experts.

VII. EVALUATION USING AN INDUSTRIAL CASE STUDY

A. THE HOT-ROLLING PROCESS

The considered real industrial case refers to the Hot Rolling Mill which is located in Krakow (Poland) as part of the company ArcelorMittal Poland. The hot-rolling process used in this case relies on the production of steel coils from flat slabs. At the beginning of the process, the slabs have a thickness that has to be reduced in the transverse section. As a result, the final thickness of the product is typically 10 to 100 times smaller than the original. The quality of the final product depends on many factors and is directly connected with the manufactured material results from material science, control engineering, mechanical engineering, and knowledge of production engineering.

The hot rolling process is based on the metal's ductility at high temperatures and consists of several steps. Initially, the raw slab's thickness is reduced from about 220 mm to about 30 mm, while the temperature of the slab reaches almost 1200° C. Later, the strips are moved to the finishing mill, where more precise process control is applied. The finishing mill is composed of six stands. Each of the stands has a lower distance between the rollers whereby the thickness of the strips is reduced. At the final stage, the temperature reaches almost 900° C. Finally, each prepared product is coiled and transferred to storage [32]. Figure 20 shows a schematic diagram of the hot rolling process.

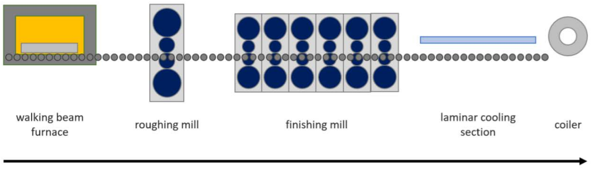

FIGURE 20. Simplified hot rolling mill process flow [32].

For the analysis, we took into consideration 10 000 different slabs with four parameters for each of them: width, profile, tempexit, and tempcoil with calculated average and standard deviation for each of these parameters. These parameters were chosen as key parameters in the case of final product quality. The choice was made by the experts. Our assumption was to treat the case as an unsupervised machine learning problem because such an approach gives opportunities to discover data patterns that would be imperceptible to the experts. The industrial problem considered in this paper is directly connected with the hot-rolling process described in this section. Based on the obtained parameters, we performed clustering. As a result, all considered slabs were divided into three groups to allow us to suppose that in the production phase, occurring processes affect the final quality of the product. In cooperation with the experts, we decided to use the ClAMP methodology to uncover differences between these three groups. Such classification and fully understanding the dependencies between groups may result in better process management.

As the analysed problem is treated as an unsupervised problem, we decided to present the results separating them into two stages, resolved as clustering and classification; rules creation and evaluation.

B. EVALUATION OF CLUSTERING AND CLASSIFICATION

The initial step in the methodology includes data clustering for obtaining good quality clusters. We tested three different methods as mentioned in Section IV-A. We used three different types of algorithms: temporal (DTC), density-based (Gaussian mixture), and one from the K-means family (BIRCH). In the case of the DTC (deep temporal clustering) clustering method and the BIRCH algorithm, we used the silhouette score in the Gaussian mixture method and determined the number of clusters based on the BIC (Bayesian Information Criteria) and AIC (Akaike's Information Criteria) metrics [21], [33]. The comparison of obtained results is presented in Table 5. It is visible that the BIRCH algorithm performed the best over all the others, hence it was chosen as the method for further analysis.

After the cluster labels were obtained, we used the XGBoost classifier with hyperparameter optimisation to build a model that will be able to distinguish clusters as accurately as possible. For BIRCH clustering, we obtained

TABLE 5. Clustering metrics comparison between three different classes of algorithms.

Clustering method	Silhouette score	Number of clusters
DTC	-0.029	3
Gaussian Mixture	0.079	4
BIRCH	0.560	3

TABLE 6. Average scores for explanation evaluation n 5-fold cross-validation dataset. Values after ± represent standard deviation in the obtained results.

Dataset	F1	Accuracy	Precision
CLAMP	0.98 ± 0.004	0.98 ± 0.004	0.98 ± 0.002
Centroids	0.94 ± 0.079	0.93 ± 0.105	0.97 ± 0.025
Global	0.95 ± 0.003	0.97 ± 0.002	0.95 ± 0.004
Isolation forest	0.98 ± 0.003	0.98 ± 0.003	0.98 ± 0.003
K-D tree query	0.98 ± 0.004	0.98 ± 0.004	0.98 ± 0.002
Random selection	0.98 ± 0.004	0.98 ± 0.004	0.98 ± 0.003

the following classification results: accuracy of 0.99 and an F1 score equal to 0.96.

In the final step of this stage, we obtained a dataset with labels split into 3 classes with distribution: 0.43, 0.43, 0.14 respectively to classes 0, 1, and 2.

C. CREATION OF RULES AND EVALUATION

After the first stage of the analysis, we moved on to building bounding boxes, rule generation, and the rule evaluation stage. This stage is fully automated. We applied the Weights&Biases platform[4] to optimize the hyperparameters and save the obtained results for each run. In this phase, all hyperparameters connected with the bounding box generation mentioned in section IV-C have been optimised.

At the beginning of this stage, the whole dataset was split into train and test subsets. The training subset was used for generating bounding boxes. The points which were treated as a bounding box for each cluster were passed to the LUX [34] explainer algorithm to obtain rules in human-readable form. this is an algorithm similar to Anchor, however, it includes information about uncertainty of explanations, which allows better selection of final rules that describe the cluster. This algorithm, based on the feature values for bounding data points and cluster labels assigned to these points, produced rules which were able to describe each cluster. Additionally, in parallel with the rules, the confidence of each rule was calculated, which allows determining the initial quality of the rule. In the next step, a preliminary rule analysis was performed. The idea of this stage is to reduce the number of rules by dropping duplicates (the same rules generated for different data points). Among the duplicated rules, only the one with the highest value of precision and coverage parameters was left allowing the number of rules and the computation time during the HeaRTDroid application to be reduced. The resulting rule set was then translated to XTT2 format which is executable by HeaRTDroid. Additionally, coverage and precision parameters were also taken into consideration by HeaRTDroid but in the form of a product of these two.

Therefore, the generated rules, a product of precision and coverage parameters, and a testing subset were treated as the input to HeaRTDroid, based on which, HeaRTDroid predicted the cluster label for each testing point. As a result, we obtained an array of predicted labels based on the rules generated for the considered bounding boxes. These labels were then compared with the labels generated by the clustering algorithm for the testing subset.

To check the effectiveness of the generated rules, we proposed four metrics particularly described in Section IV-D3. The final results of the evaluation are presented in Table 6, which contains the final results of the developed methodology obtained for each of the considered methods used to define the bounding box.

As we decided to treat the F1-score as the target metric, the best score, equal to 0.98, was obtained for the ClAMP method, which combines all of the bounding box description methods, as presented in Table 6.

The quality of rules in terms of classification metrics can be adjusted by changing the number of points used to form a bounding box. In general, the more points we create, the more precise the classification we obtain. As a trade-off, we lose the interpretability due to the increased number of rules analysed by the expert. Figure 21 presents charts that show how classification scores change, with the number of considered bounding points defined in percentages. Additionally, we are able to present such dependencies for two different metrics applied for the KD-tree description method. The number of points for which we obtained the best F1-score for each of the considered methods is marked by the red dashed line.

In Figure 21, the highest changes in values occur for the random selection method and increase monotonically, which is understandable as it covers more and more points from the dataset. This increases accuracy but makes the explanation model more complex and thus less useful in practical applications. For centroid-based methods, the change is not visible as one point is always used as the ''bounding box''. In comparison with the rest of the charts, using random selection and isolation forest methods, we can see that the KD-tree method needs fewer points to obtain comparable results than the others. Such a presentation, together with the knowledge of how the presented methods work, may be very informative for the experts.

D. RULES ANALYSIS BY EXPERTS

With the 26 obtained rules, we performed an evaluation with three domain experts recruited from ArcelorMittal who were asked to use CLAMP to obtain explanations for clusters and later answer the same set of questions as presented in Sections VI-B2. In Table 7, we present the synthesised results we obtained both from domain experts in the industrial case, and from participants involved in the evaluation on artificial datasets.

For the analysis of the results, one can see that answers are slightly different depending on the dataset. This is especially

[4] Weights & biases platform: https://wandb.ai/site.

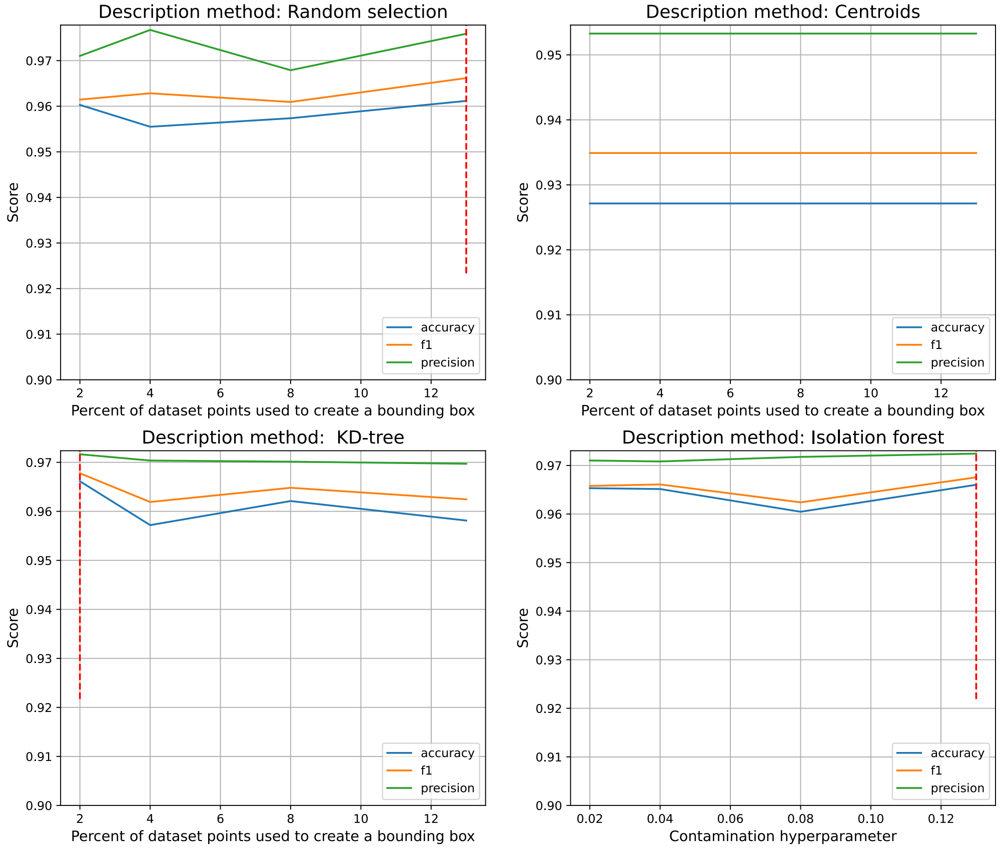

FIGURE 21. How classification score changes with the number of considered bounding points defined in percentages.

TABLE 7. Average answers for survey questions with respect to the dataset.

Question Dataset	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
Industrial	2.0	18.33	3.33	3.33	3.66	4.00	3.66	3.33
Gaussian Blobs 2D	2.00	8.00	3.00	4.40	4.40	2.80	4.40	4.40
Gaussian Blobs 3D	1.75	5.75	3.50	2.50	3.50	3.0	3.25	2.75
Iris	2.40	3.60	3.80	4.00	3.80	3.40	3.40	3.20
Random values	1.91	5.09	2.90	3.18	3.63	3.45	3.09	3.27

visible in the case of the number of rules required to understand clusters (Q2), while the other answers are similar.

E. DISCUSSION OF THE RESULTS

The overall evaluation results suggest that our methodology is useful for experts in cluster analysis. Most of the crucial features of our method gained high scores from the experts. In Section VI-B1, two experts agreed with the statement that the rules presented in our approach were better for explaining clusters, not single instances, which was our intention. One of the experts outlined that such rules could explain an instance as well. In the case of the number of generated rules, one of the experts decided to divide the answers into two cases. The number of rules to comprehend by the human user is heavily dependent on the way the results are presented.

- After the rules are initially grouped according to common parts, conclusions can be drawn from > 50 rules.
- When all rules are presented in tabular form, conclusions can be drawn from < 10 rules.

In addition, the expert pointed out that the set of rules that describe limits for a single parameter at once is more intuitive than the interpretation of one rule describing the relation between 3 or more features. Rules built on the basis of 2 parameters seem to have the right balance between the amount of information and the availability of its evaluation.

In comparison to the analysis without explanations (Section VI-B2), two experts agreed that our method can strongly decrease the time needed to distinguish and describe the clusters. One of the experts pointed out that the method allows adjusting the complexity of the conditions set depending on the time constraints of the user for which they are prepared. For fast evaluation, the generated conditions could be visualised and put in the context of product measurements. For the purposes of advanced study, the amount of information is more important than the time of analysis. In that case, the computed rules could be an intermediate dataset for further analysis.

The domain knowledge is as important as the data science background in analysing the results of the explanations (Section VI-B3). The data, which were treated as an input to the ClAMP methodology in the presented use case, were delivered to the data scientist and based on statistical

TABLE 8. Rules overlapping example. Rule 1 subsumes rule 2.

No	Rule
1	$F1 > 1$ and $F2 < 2.5$
2	$F1 > 1.2$ and $F2 < 2.5$

properties such as standard deviations, variance, etc. We also consulted experts without data science experience about the results. They pointed out that they prefer rules which would be generated based on real production parameters, not on the statistics. One of the experts additionally pointed out that the complexity of the rule should be correlated with the expected cognitive abilities of the end-user. The user evaluating the rules should be experienced and understand the process to which they relate. In their current form, they can be addressed to technologists and process specialists. Interpretation of the conditions by an inexperienced user such as management staff requires giving context to the rules, e.g., what is the probability of a defect if the condition will be exceeded. In general, the rule-based form is friendly to interpretation by a specialist, however, it is strongly dependent on the selection of features in the model input dataset. In this case, the explanations are understandable as the parameters features like standard deviation or mean are easy to comprehend by specialists. The use of 'quadrilles' or 'percentiles' would not be so clear to the human user.

However, this comment relates to the whole family of XAI methods that do not consider the type or characteristics of the user as an important factor in preparing explanations. We believe that this should be a trigger for extensive research in the area of XAI and HCI (Human Computer Interaction). However, it is outside the scope of this work.

In the last Section VI-B4, the experts noted that there are overlapping rules that make the task challenging because it is difficult to determine the cluster's boundaries. There are several rules which consist of the same feature, with the same inequality sign but different values. It is hard to define which of these rules is the best, that is, which rules allow obtaining the best balance in recognising points that concern a specific cluster and does not allow too many points to be recognised that concern a cluster which this rule do not define. An example of such is presented in Table 8. In the example, Rule 1 is more general than Rule 2 (Rule 1 subsumes Rule 2). This, in some cases, might cause the analysis of the explanations to be difficult, as the decision regarding which rule is more appropriate to describe the cluster might not be entirely clear.

VIII. SUMMARY

In this paper, we presented a novel approach for cluster analysis with multidimensional prototypes called ClAMP that aids experts in cluster analysis with human-readable rule-based explanations. ClAMP methodology is divided into two stages: clustering and classification and rules creation and evaluation.

In the first stage, clustering and classification are executed. The goal of these two steps is to convert unsupervised learning problems into supervised one. In our case, we tested three different clustering methods. As a classification algorithm, we used XGBoost with hyperparameter optimisation. In the second stage, we implemented three methods that can be used to determine the prototypes (bounding boxes) that are then treated as an input to the explainer algorithm. Thanks to the application of these methods, the clusters are described only by the most representative points that allow avoiding the generation of unnecessary rules which can introduce noise into the explainability mechanism. Such an approach limits the computational time needed to generate explanations and increase explainability transparency. Hence, the proposed approach increases the effectiveness and efficiency of the rules generation. The generated bounding boxes are treated as an input to the Anchor explainers to generate rules for each cluster. The Anchor explainer is able to generate precision and coverage parameters as well. Thanks to these, there is a possibility to determine which rules describe the cluster better. These parameters are also useful for checking the effectiveness of the created rules. To do that, we used the HeaRTDroid rule-based inference engine which allows predicting labels based on the generated Anchor explainer rules and parameters returned. It also allows for the integration of the knowledge discovered using the XAI method with other system components. As a result, we implemented an approach that allows delivering human-readable rules to the experts taking into consideration different clustering methods, hyperparameter optimisation, and a novel approach to generating bounding boxes for evaluation.

In comparison to the methods presented in Section II, we noticed two main differences. Firstly, the developed methodology allows obtaining a cluster representation in the form of human-readable rules. It is worth emphasising that we do not concentrate on explaining particular instances but rather on the whole group. This gives the opportunity to deliver to the experts information about the considered groups and understand the data division into clusters. Secondly, our methodology can verify the obtained rules with the use of HeaRTDroid, which allows predicting labels based on the instances and rules obtained based on the train set.

We demonstrated our approach using two cases. The first concerns publicly available, artificially generated datasets that can be considered benchmark cases. The second case concentrates on the real-life use case scenario with confidential data shared for the purpose of the PACMEL project from the company ArcelorMittal. Taking into account the assessment of the rules by experts, the idea of the proposed methodology proved that it is useful. According to the completed questionnaire, the experts pointed out that the rules help them describe the clusters, and such understanding is less time-consuming than in the case of the labels themselves. Although participants noticed that descriptions of clusters usually contain overlapping rules, they were able to identify the redundant parts and correctly interpret the explanations. The bottom line is that the generated rules are able to provide useful information about clustering.

Furthermore, one of the limitations of the developed methodology pointed out by the experts is the fact that some of the generated rules overlap. In future work, we are planning to find a solution concerning that limitation, based on our prior works in this area [35]. We also plan to adjust the bounding box generation by selecting the most suitable method for discovering cluster prototypes, not for the whole data but for each cluster separately. Taking into account the approach for Knowledge Augmented Clustering (KnAC) presented in [23], which is based on the clusters' centroids, ClAMP will be considered as an extension of KnAC.

ABBREVIATIONS

TABLE 9. Abbreviations explanation.

No	Abbreviation	Explanation
1	AIC	Akaike's Information Criteria
2	BIC	Bayesian Information Criteria
3	BIRCH	Balanced iterative reducing and clustering using hierarchies
4	CBSU	Cluster-based sentence utility
5	CLAMP	Cluster Analysis with Multidimensional Prototypes
6	CNF	Conjunctive normal form
7	DLIME	Deterministic Local Interpretable Model-Agnostic Explanations
8	DNF	Disjunctive normal form
9	DTC	Deep temporal clustering
10	FCPS	Fundamental Clustering Problems
11	G2PC	Global permutation percent change
12	HCI	Human Computer Interaction
13	HEARTDROID	Rule-based inference engine
14	HMR+	Rule-based language
15	K-D tree	K-dimensional tree
16	KNAC	Knowledge Augmented Clustering
17	KNN	K-nearest neighbours
18	L2PC	Local permutation percent change
19	LIME	Local interpretable model-agnostic explanation
20	mlrMBO	Model-based optimization package
21	Randomized SearchCV	RandomizedSearch Cross-Validation
22	SFIT	Single Feature Introduction
23	SMBO	Sequential Model-Based Optimization
24	TED	Teaching Explanations for Decisions
25	XGBoost	Gradient Boosting framework
26	XTT2	Formalized rule representation

ACKNOWLEDGMENT

The authors would like to thank all of the persons involved in the evaluation process of the paper.

REFERENCES

[1] C. Molnar, *Interpretable Machine Learning*. Abu Dhabi, United Arab Emirates: Lulu, 2020.

[2] A. Saxena, M. Prasad, A. Gupta, N. Bharill, O. P. Patel, A. Tiwari, M. J. Er, W. Ding, and C. T. Lin, "A review of clustering techniques and developments," *Neurocomputing*, vol. 267, pp. 664–681, Dec. 2017, doi: 10.1016/j.neucom.2017.06.053.

[3] S. Bobek, G. J. Nalepa, and M. Slazynski, "HEARTDROID—Rule engine for mobile and context-aware expert systems," *Expert Syst.*, vol. 36, no. 1, 2019, Art. no. e12328, doi: 10.1111/exsy.12328.

[4] G. J. Nalepa, *Modeling With Rules Using Semantic Knowledge Engineering* (Intelligent Systems Reference Library), vol. 130. Cham, Switzerland: Springer, 2018.

[5] J. Lötsch and S. Malkusch, "Interpretation of cluster structures in pain-related phenotype data using explainable artificial intelligence (XAI)," *Eur. J. Pain*, vol. 25, no. 2, pp. 442–465, 2021, doi: 10.1002/ejp.1683.

[6] A. Morichetta, P. Casas, and M. Mellia, "EXPLAIN-IT: Towards explainable AI for unsupervised network traffic analysis," in *Proc. 3rd ACM CoNEXT Workshop Big DAta, Mach. Learn. Artif. Intell. Data Commun. Netw.*, New York, NY, USA, Dec. 2019, pp. 22–28, doi: 10.1145/3359992.3366639.

[7] E. Horel, K. Giesecke, V. Storchan, and N. Chittar, "Explainable clustering and application to wealth management compliance," in *Proc. 1st ACM Int. Conf. AI Finance*, 2020, pp. 1–6.

[8] D. Rymarczyk, Ł. Struski, M. Górszczak, K. Lewandowska, J. Tabor, and B. Zielinski, "Interpretable image classification with differentiable prototypes assignment," 2021, *arXiv:2112.02902*.

[9] C. Chen, O. Li, C. Tao, A. J. Barnett, J. Su, and C. Rudin, *This Looks Like That: Deep Learning for Interpretable Image Recognition*. Red Hook, NY, USA: Curran Associates, 2019.

[10] O. Loyola-Gonzalez, A. E. Gutierrez-Rodriguez, M. A. Medina-Perez, R. Monroy, J. F. Martinez-Trinidad, J. A. Carrasco-Ochoa, and M. Garcia-Borroto, "An explainable artificial intelligence model for clustering numerical databases," *IEEE Access*, vol. 8, pp. 52370–52384, 2020.

[11] D. R. Radev, H. Jing, M. Styś, and D. Tam, "Centroid-based summarization of multiple documents," *Inf. Process. Manage.*, vol. 40, no. 6, pp. 919–938, 2004.

[12] D. Bertsimas, A. Orfanoudaki, and H. Wiberg, "Interpretable clustering via optimal trees," 2018, *arXiv:1812.00539*.

[13] W. Guo, D. Mu, J. Xu, P. Su, G. Wang, and X. Xing, "Lemna: Explaining deep learning based security applications," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2018, pp. 364–379, doi: 10.1145/3243734.3243792.

[14] C. A. Ellis, M. S. E. Sendi, E. P. T. Geenjaar, S. M. Plis, R. L. Miller, and V. D. Calhoun, "Algorithm-agnostic explainability for unsupervised clustering," 2021, *arXiv:2105.08053*.

[15] S. Dash, O. Günlük, and D. Wei, "Boolean decision rules via column generation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 4655–4665.

[16] P. Linardatos, V. S. Papastefanopoulos, and S. Kotsiantis, "Explainable AI: A review of machine learning interpretability methods," *Entropy*, vol. 23, no. 1, pp. 1–45, 2021.

[17] M. Hind, D. Wei, M. Campbell, N. C. Codella, A. Dhurandhar, A. Mojsilović, K. Natesan Ramamurthy, and K. R. Varshney, "TED: Teaching AI to explain its decisions," in *Proc. AAAI/ACM Conf. AI, Ethics, Soc. (AIES)*, 2019, pp. 123–129.

[18] M. R. Zafar and N. M. Khan, "DLIME: A deterministic local interpretable model-agnostic explanations approach for computer-aided diagnosis systems," 2019, *arXiv:1906.10263*.

[19] M. T. Ribeiro, S. Singh, and C. Guestrin, "Anchors: High-precision model-agnostic explanations," in *Proc. 32nd AAAI Conf. Artif. Intell. (AAAI)*, 2018, pp. 1527–1535.

[20] M. Kuk, S. Bobek, and G. J. Nalepa, "Explainable clustering with multidimensional bounding boxes," in *Proc. IEEE 8th Int. Conf. Data Sci. Adv. Anal. (DSAA)*, Oct. 2021, pp. 1–10.

[21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, and M. Blondel, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, no. 10, pp. 2825–2830, Jul. 2017.

[22] N. S. Madiraju, S. M. Sadat, D. Fisher, and H. Karimabadi, "Deep temporal clustering: Fully unsupervised learning of time-domain features," 2018, *arXiv:1802.01059*.

[23] S. Bobek, M. Kuk, J. Brzegowski, E. Brzychczy, and G. J. Nalepa, "KnAC: An approach for enhancing cluster analysis with background knowledge and explanations," 2021, *arXiv:2112.08759*.

[24] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, New York, NY, USA, Aug. 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[25] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, pp. 281–305, Feb. 2012. [Online]. Available: http://jmlr.org/papers/v13/bergstra12a.html

[26] B. Bischl, J. Richter, J. Bossek, D. Horn, J. Thomas, and M. Lang, "MlrMBO: A modular framework for model-based optimization of expensive black-box functions," 2017, *arXiv:1703.03373*.

[27] P. Ram and K. Sinha, "Revisiting kd-tree for nearest neighbor search," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2019, pp. 1378–1388.

[28] G. Staerman, P. Mozharovskyi, S. Clémençon, and F. d'Alché Buc, "Functional isolation forest," in *Proc. Asian Conf. Mach. Learn.*, 2019, pp. 332–347.

[29] F. Tony Liu, K. Ming Ting, and Z.-H. Zhou, "Isolation Forest," in *Proc. ICDM*, 2008, pp. 332–347.

[30] C. Molnar, *Interpretable Machine Learning*. Lulu.com, 2019. [Online]. Available: https://christophm.github.io/interpretable-ml-book/cite.html

[31] S. Bobek, S. K. Tadeja, Ł. Struski, P. Stachura, T. Kipouros, J. Tabor, G. J. Nalepa, and P. O. Kristensson, "Virtual reality-based parallel coordinates plots enhanced with explainable AI and data-science analytics for decision-making processes," *Appl. Sci.*, vol. 12, no. 1, p. 331, Dec. 2021. [Online]. Available: https://www.mdpi.com/2076-3417/12/1/331

[32] J. Jakubowski, P. Stanisz, S. Bobek, and G. Nalepa, "Explainable anomaly detection for hot-rolling industrial process," in *Proc. IEEE 8th Int. Conf. Data Sci. Adv. Anal. (DSAA)*, Oct. 2021, pp. 1–10.

[33] S. Akogul and M. Erisoglu, "A comparison of information criteria in clustering based on mixture of multivariate normal distributions," *Math. Comput. Appl.*, vol. 21, no. 3, p. 34, Aug. 2016.

[34] S. Bobek and G. J. Nalepa, "Introducing uncertainty into explainable ai methods," in *Computational Science–(ICCS)*, M. Paszynski, K. Dieter, V. V. Krzhizhanovskaya, J. J. Dongarra, and P. M. A. Sloot, Eds. Berlin, Germany: Springer, 2021, pp. 444–457.

[35] G. Nalepa, S. Bobek, A. Ligęza, and K. Kaczor, "HalVA—Rule analysis framework for XTT2 rules," in *Rule-Based Reasoning, Programming, and Applications* (Lecture Notes in Computer Science), vol. 6826, N. Bassiliades, G. Governatori, and A. Paschke, Eds. Berlin, Germany: Springer, 2011, pp. 337–344. [Online]. Available: http://www.springerlink.com/content/c276374nh9682jm6/

SZYMON BOBEK (Member, IEEE) received the Ph.D. degree. He is currently an Assistant Professor at Jagiellonian University. His work includes, most recently, hybrid models for explainable AI. He is the author of over 50 research papers in international journals, books, and conferences, and has participated in eight research projects. He has reviewed research papers for a number of international journals. He has cooperated with several companies in applied projects, especially with massive big data processing using machine learning methods. He is in active cooperation at Cambridge University in the area of utilizing XAI in healthcare applications. He collaborates with the AFFCAI (affcai.geist.re) Group in the area of application of XAI to processing biomedical signals for affective computing. Recently, he has been leading a team that developed a series of original tools in the area of explainable AI for applications in industrial AI. He is a Co-Organizer of the Practical Applications of Explainable Artificial Methods (PRAXAI) special session (praxai.geist.re), an Editor of XAILA proceedings, and a PC member of multiple workshops and conferences.

MICHAL KUK received the M.Sc. degree from the Faculty of Drilling, Oil and Gas, AGH University of Science and Technology, where he is currently pursuing the Ph.D. degree with the Faculty of Drilling, Oil and Gas, with a specialization in mining and geology. In 2017, he realized his master's thesis in which he developed an algorithm optimizing the location of new wells. His scientific research interests include optimization of oil and gas production. He uses machine learning algorithms to improve production from the reservoirs. One of his methods has been presented at SPE student paper contest, where he achieved second place in the Ph.D. division. Since November 2020, he has been a member of the GEIST Team. He was involved in Process-Aware Analytics Support based on Conceptual Models for Event Logs—PACMEL Project and is currently participating in the XPM Project (explainable predictive maintenance).

MACIEJ SZELĄŻEK received the M.Sc. degree in automation and metrology from AGH UST, Kraków, Poland, in 2010, where he is currently pursuing the Ph.D. degree with the Department of Applied Computer Science. He has worked as a Data Analyst in the Office of Statistical Process Control (SPC) at ArcelorMittal Poland. He currently participates in the creation and development of an analytical system based on central database integrating distributed data sources, reporting system, and statistical data mining software. He has conducted big-data multidimensional analyses related to searching for bottlenecks, logistics, cost optimization, and limiting the variability of processes.

GRZEGORZ J. NALEPA (Member, IEEE) is currently a Full Professor at Jagiellonian University, Kraków, Poland. He has coauthored over 200 research papers in international conferences and journals. He has been involved in tens of projects, including research and development projects with a number of companies. His recent research interests include applications of AI in industry 4.0 and business, explainable AI, affective computing, context awareness, and the intersection of AI with law.