Improving Understandability of Explanations with a Usage of Expert Knowledge

Maciej Szelążek[1]([✉]) [ID], Szymon Bobek[2] [ID], and Grzegorz J. Nalepa[2] [ID]

[1] AGH University of Science and Technology, Kraków, Poland
szelazek.m@gmail.com

[2] Jagiellonian Human-Centered Artificial Intelligence Laboratory (JAHCAI) and Institute of Applied Computer Science, Jagiellonian University, Kraków, Poland
{szymon.bobek,grzegorz.j.nalepa}@uj.edu.pl

Abstract. Data analysis is one of the most important parts of data mining and machine learning tasks. In recent years, explainable artificial intelligence methods have been used very often to support this phase. However, the explanations themselves are very often difficult to understand by domain experts, who play one of the most important roles in the phase of data analysis. In this work, we proposed a procedure to combine domain knowledge with ML and XAI methods to improve the understandability of explanations. We demonstrated the feasibility of our approach on a publicly available medical dataset. We describe a procedure for obtaining intuitively interpretable information about distinguishable groups of patients and defining differences between them with the usage of clustering, rule–based encoded domain knowledge, and SHAP values.

1 Introduction

Explainability (XAI) methods are becoming increasingly important with the spread of practical applications of Machine Learning (ML). XAI could be used to recognize internal dependencies of a model by determining the influence of features on the prediction result, and finally help get better insight into data and the model in pre–modeling phase of data mining (DM) and ML workflows. In this phase, the main focus is on domain understanding, data cleansing, feature selection, etc. The success of this stage is determined by a proper understanding of the data and the domain. Therefore, in most cases, the pre–modeling phase is a combined effort of different stakeholders including domain experts and data scientists. In recent years, XAI methods have been used to improve the communication between data scientists and nontechnical stakeholders by explaining the model decisions. However, the target users of most current XAI algorithms such as SHAP, Lime, Anchor, Lore, etc. are usually data scientists [15, 19, 23]. This limits their usage to model–debugging, or feature–selection tasks, while their potential of generating explanation for broader audience including domain experts is not fully utilized. In our research, we focus on augmenting existing XAI methods with an additional layer of domain knowledge that will increase their understandability. We base our procedure on the SHAP algorithms, which are among the most mature and widely used for both research and practical applications.

SHAP importance score allows us to determine the contribution of individual input parameters to ML model decisions. Expert knowledge is usually not limited to recognizing the importance of features but also encodes complex relationships between them. Thus, the transition from a simple feature attribution scheme, which most modern XAI methods follow, to more expressive explanations is not straightforward and limits the usefulness of such explanations from the expert perspective. One of the goals of the research presented in this article was to fill this gap, developing a method that will improve the understandability of XAI explanations for non–technical users, but also provide a formalized way to encode expert knowledge into the explanation process to let data scientists get better insight into the domain [22]. To achieve this goal, we combined rule-based knowledge representation, clustering, and explainability methods according to the Semantic Data Mining paradigm [13].

The rest of the paper is organized as follows. In the 2 section, we describe our motivations. The articles related to our research are presented in Sect. 3. The methodology of is shown in 4. The results obtained and their evaluation are presented in Sect. 5.

2   Motivation

One of the goals of our research was to develop a method to integrate model explanations with additional domain knowledge to enhance the interpretability of XAI results. One way to do this is to recognize the relationship between instance values and their impact on the model's decision [11]. We applied it by designing a set of rules for the features to capture expert knowledge of certain relationships that are hidden without experience in a particular area. Such an approach has at least two advantages, the one–time implementation of additional information allows its later reuse. The second is the ability to automate the entire procedure whose results will be understandable to a non–specialist user.

In order to find patterns in the data and relate them to the explainability results, we had to consider several issues like:

1. Identify features that are not intuitive or even impossible to understand without having domain knowledge.
2. Implementation of domain knowledge in the form of rules
3. Transformation of the rules into new, binary features set
4. Defining patterns in data to recognize similarities between instances (Clustering)
5. Defining relations between the clusters using: a) ML model and SHAP results to determine distinctive patterns of features significance for each group b) Rule-based encoded domain knowledge c) Data distribution analysis

In Fig. 1 are the workflows of two approaches for explaining unsupervised learning models. In our study, we used the schema shown in 1A. Schema 1B represents the basic workflow in which the relations in the dataset are defined by the importance of SHAP features on the basis of unsupervised model decisions. Approach 1B, despite its simplicity, may be less generic. Depending on the model used for the unsupervised learning task, the appropriate method to create explanations should be selected.

In our study, we do not analyze the explainability scores of the model that creates the labels. This approach (Fig. 1A.) is universal due to the independence of the internal structure of the model, as we only use the input data and the output results for the Semantic Data Mining analysis.

3   Related Works

In our research, we want to expand the available set of information in order to create explanations that are easier to interpret, especially without detailed knowledge of the data. This approach is inspired by the concept of Semantic Data Mining [20]. It relies on the Semantic Data Mining tasks that systematically incorporate domain knowledge, especially formal semantics, into the process. As noted by the authors of [17], the main aspect of Semantic Data Mining is the explicit integration of this knowledge into algorithms for modeling or post-processing. In our study, we use domain knowledge that describes medical data for diabetes in the form of additional rules to code the original data and recognize its impact on the outcome (in our case, health status) [14]. One of the methods to look for relationships in the data is to compare the distributions between groups [4,6]. We used information obtained on the original and additional feature distributions as a basis for interpreting the XAI results. [7]
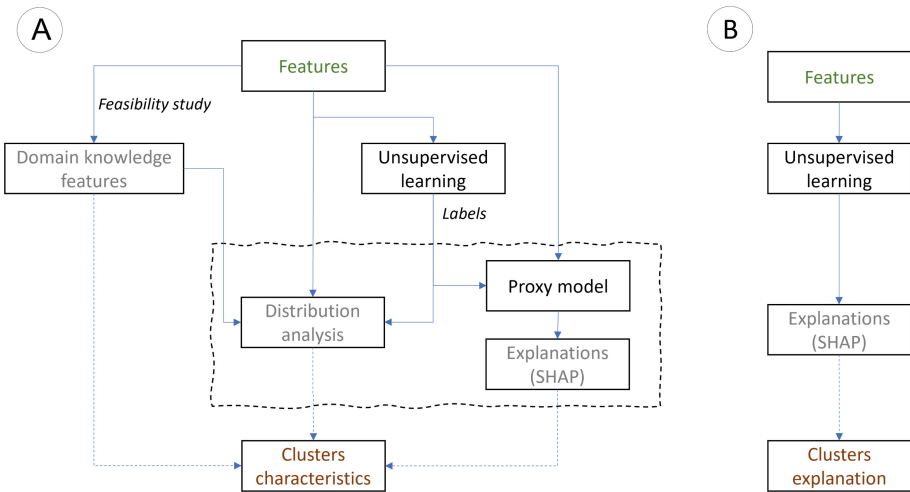

Fig. 1. Comparison of workflow used in our study (A) vs basic workflow for unsupervised learning explanations (B)

Despite many existing explainability algorithms, there is still no scientific consensus on the explicit recognition of multifaceted dependencies of model input parameters [9]. This problem was noticed by the authors of [10], especially in the case of multi–modal relationships, where different aspects of a phenomenon are represented by individual modalities/channels.

In our study, we used the proxy model to determine differences between clusters (Fig. 1A.). This term is defined by the authors of [3] under the name "surrogate model" as "usually directly interpretable model that approximates a more complex model. We evaluated the results of explainability from various perspectives in a similar way to the idea described by the authors in [5]. In our research, we explore the possibilities of recognizing relations between the instances using additional features coded from domain knowledge–based rules, original dataset distributions, and SHAP importance score.

4   Transforming Domain Knowledge into a Set of Rules

4.1   Dataset

The 'Pima Indians Diabetes' (PID) dataset [21] is a collection of data about the incidence of diabetes, originally from the National Institute of Diabetes and Digestive and Kidney Diseases. PID dataset was designed to detect relations between diabetes and parameters that describe the health status of the subjects. There are 768 cases in the database. Dataset includes information if a patient has or does not have diabetes, and 8 features:

– Pregnancies – Number of pregnancies
– Glucose – Plasma glucose concentration a 2 h in an oral glucose tolerance test
– BloodPressure – Diastolic blood pressure [mm Hg]
– SkinThickness – Triceps skin fold thickness [mm]
– Insulin – 2–Hour serum insulin [mu U/ml]
– BMI – Body mass index
– DiabetesPedigreeFunction – Indicator based on the diabetes mellitus history in relatives and the genetic relationship of those relatives to the patient.
– Age

After conducting exploratory analysis, we have noticed a significant amount of zeros values in the original dataset. The most important feature in this regard was Insulin with 374 zero values out of 768. In Fig. 2 we can see silhouette scores for all features (A), data excluding Insulin measurements (B) and after removing a small number of zeros (C) of the BMI (11 measurements), Blood Pressure (35 measurements) and Glucose (5 measurements) features. After cleaning the data and excluding Insulin, 724 cases remained as input to the unsupervised learning model. Based on the 2C plot, model was set to find similarity for 4 clusters.
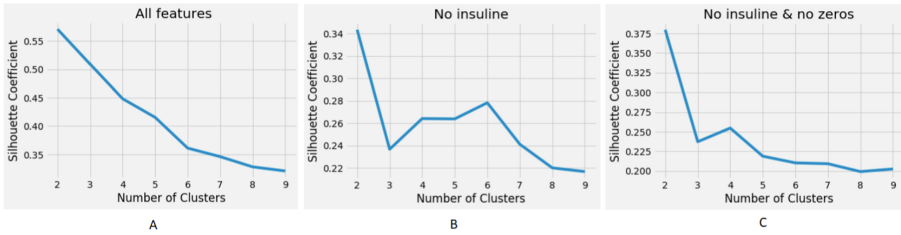
Fig. 2. Silhouette scores for different subsets of data

4.2   Feasibility Study of Rules for Features Assessment

The most time–consuming part of our approach is the preparation of rules coding the original dataset. Incorporation of domain knowledge requires a thorough exploration of the possible sources of this knowledge. In practice, decision support systems are often developed in collaboration with end users or domain experts. In such cases, users are an important source of expertise.

In our case, the two main goals of domain data mining were: 1) Determine the thresholds that would allow assessing the value of features, shown in Fig. 3, 2) Obtain information on the real impact of the feature on the outcome, in this case susceptibility to diabetes.

Let's consider influence assessment issue using the BMI feature as an example. BMI is one of the most important features related to diabetes susceptibility. It is a universal factor that is often used as a health index based on the weight–to–height ratio. The main advantage is its simplicity of calculation and interpretation. Some interpretations of the BMI factor separate the correct limits for men and women due to differences in body build [2], but this approach does not follow the guidelines of the World Health Organization (WHO), the Centers for Disease Control and Prevention (CDC) [1]. As described by the authors of [18], BMI is a useful index associated with susceptibility to various diseases such as asthma. The controversy related to BMI is oversimplification, which does not distinguish between muscle mass and fat mass. This leads to a situation where similar, high values can be obtained by a person with excessive, unhealthy weight and an athlete with a lot of muscles and a negligible share of body fat. One of the postulates of physicians analyzing this topic is the proposal to take into account the circumference of the waist as a factor that distinguishes these cases.

In a detailed analysis of this feature, we can observe that the limits determining the correct values do not depend on age, although with age there is a tendency for a statistical increase in BMI [8]. In the dataset we use in our research, all the respondents are older than 21 years, therefore the following thresholds describing obesity are universal for the entire dataset (Fig. 3), however, considerations concerning children and adolescents are not so clear. In adolescents and children, the different growth rates and generally large changes over several years mean that BMI should be assessed in relation to the BMI distribution of the local peer group and with the use of additional measurements [12]. One of the goals of our research was to assess patient health based on medical measurements and to determine the factors of influence that are the most risky

in terms of health loss. The ambiguity in the interpretation of the BMI value means that as a single information is not crucial for pro–health activities recommendations. However, BMI is important information in relation to additional indicators that describe the patient's health condition.

The rules have been prepared to give semantic meaning to the values in the dataset. Without additional knowledge, in basic approach, XAI algorithms let us obtain information such as "BMI was the most influential feature considered by the model for the cluster assignment decision". The same data, after being processed by domain–based rules, allow us to designate the most influential groups of features in the context of dependencies understandable to a non–specialist user. In practice, the number of classes and the complexity of the rules depend on the individual goals of the analysis. The next section describes the results we obtained after applying our procedure.

4.3   Creating Rules

In our research, we consider the way of implementing the domain knowledge information that describes the underlying dataset to obtain meaningful results. For this purpose, we conducted a feasibility study to acquire thresholds for features real–life influence assessment. Next, we turned that information into a set of rules to classify the original data in a form of binary–coded features. Each rule r_f takes form of:

$$r_f = \prod_{j \in T_f} I(x_j \in s_{jf})$$

where T_f is a set of thresholds for feature f, I is the indicator function that is equal to '1' when threshold x_j is in a specified subset of values s for the j–th threshold and '0' otherwise. We are using numerical features, so s_{jf} is an interval in the value range of the threshold:

$$(x_{jf,lower} < x_j) \ AND \ (x_j < x_{jf,upper})$$

For example, if we need to incorporate domain–based assessment of the BMI feature that states 'BMI less than 35 and more than 30', we add a new feature – 'obesity type 1' in the instance that will be equal to '1' if it satisfies that condition and '0' otherwise. Visualization of additional features related to the level of healthy value exceedances is shown in Fig. 3.

After choosing the appropriate way to encode additional information, we proceeded to analyze the dataset. We use the clustering technique to identify patterns in the data. This is a form of unsupervised learning that allows us to group instances according to their similarity. In the next step, we create a proxy model using the original data as input. The model was aimed at predicting the clusters formed by unsupervised learning. Then we used XAI scores and distribution statistics to relate all available data in the context of the clusters. The main stages of our research include: 1) Transformation of domain knowledge into a form of rules, and then features, used for ML methods. 2) Unsupervised learning to determine patterns in the data. 3) Training ML model for identifying the relations between model input and designated clusters. 4) Explainability analysis using SHAP values. 5) Recognition of patterns and relations between clusters using explainability results, distribution statistics, and rule–based features analysis for individual clusters.
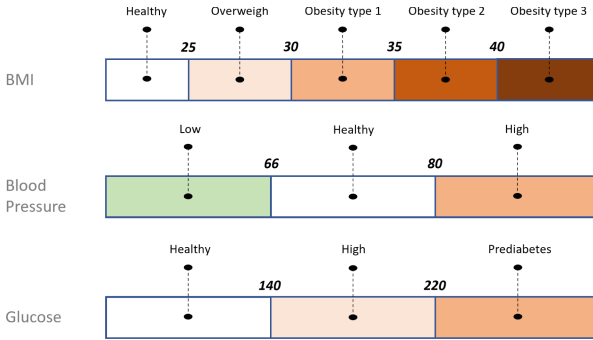
Fig. 3. Visualization of domain–based rules used to create an additional set of features

Clustering is used as a source of patterns that we want to identify and determine the internal relationships that distinguish clusters from each other. The objective of the case study presented in Sect. 5 was to divide patients into groups according to their susceptibility to diabetes and recognize the characteristics of each group.

5 Evaluation

In our procedure, we draw conclusions about relationships between patients by merging information from several sources (shown in Fig. 1): 1) Distribution characteristics of the original data 2) Distribution characteristics of the features derived from domain knowledge 3) XAI results of the proxy model

During the study, we tested 4 types of models: Logistic Regression, K–Nearest Neighbors (KNN), Random Forest Classifier, and Support Vector Machine. We chose the Random Forest Classifier for further analysis because it achieved the highest accuracy.

The Table 1 contains clustering results. Cluster 0 has the highest number of people with diagnosed diabetes (45% of all) in terms of both the number of people with diabetes and their proportion. Cluster 3 represents the most "healthy" group. Clusters 1 and 2 have similar proportions when it comes to diabetics, and to identify preliminary differences between these groups, we analyzed the distributions of feature measurements.

Table 2 shows the mean feature values according to the clusters. Bold font indicates the highest values for column, italic – the lowest. The results show that in clusters 0 and 3 people with the worst and the best scores were collected, respectively, which is consistent with the shares of diabetics in these clusters. Compared to the information in Table 1, we can observe differences between clusters 1 and 2. In cluster 2 patients have a higher average for the number of pregnancies and age, while noticeably lower values for features BMI and Diabetes Pedegree Function (represent genetic susceptibility to diabetes). The largest difference (almost 15 times) in favor of patients in cluster 2 compared to cluster 1 is noticeable for the Skin Thickness feature, which represents information about adipose tissue.

Table 1. Division of patients into clusters of the unsupervised learning model.

Cluster	Sum	Outcome	Quantity	Diabetes share
0	153	0	40	
		1	113	74%
1	194	0	127	
		1	67	35%
2	147	0	100	
		1	47	32%
3	230	0	208	
		1	22	10%

Table 2. Mean values of features for clusters.

Cluster	Pregnancies	Glucose	BloodPressure	SkinThickness	BMI	DiabetesPedigreeFunction	Age
0	**5,0**	**168,7**	**76,6**	25,1	**35,1**	**0,55**	**38,7**
1	3,5	121,7	73,7	**33,6**	34,9	0,50	31,9
2	4,6	121,9	75,9	*2,3*	30,7	*0,42*	37,7
3	*2,9*	*90,8*	*66,2*	21	29,8	0,44	*28,2*

Table 3 shows the summarized values of binary features according to the clusters. We can use this to provide context to the original dataset. The SHAP explainability results for the proxy model are shown in Fig. 4 The Glucose_preDiabetes feature in all clusters was 0.

The major differences considering the domain knowledge features include the following:

– Glucose – The most significant differences between the Cluster 0 and others groups are observed on the Glucose parameter due to the highest mean value. This is due to the presence of 83% of all patients with excesses in this group. Clusters 1 and 2 contain only 16 cases each (out of 185 observations), and 0 cases in cluster 3.
– BMI – people in clusters 0 and 1 have a significantly higher average BMI than clusters 2 and 3. The distinction between these groups can be made by analyzing additional data. We can see that despite the highest average for cluster 0, the largest number of people with exceeded the Obesity type 1 threshold and above are in cluster 1. The lower level of BMI is cluster numbers 2 and 3. The smallest number of overruns is in cluster 2, but it is cluster 3 that has the lowest average BMI. This is due to the fact that cluster 3 has the highest share of people without excesses of this parameter (27% compared to 22% in cluster 2).
– Blood Pressure – the most distinctive group is cluster 3 which has the highest proportion of people with problems with too low blood pressure (54% of all exceedances of this type). Cluster 0 has the largest share of people with too high blood pressure (35%). In cluster 1, the proportions of excesses related to too high and too low blood pressure are similar.

Table 3. Sum of domain–based binary features for clusters.

Cluster	BMI_obese3	BMI_obese2	BMI_obese1	BMI_overweight	Glucose_hi	BloodPressure_hi	BloodPressure_lo
0	32	33	57	20	**153**	**54**	24
1	**34**	**53**	**58**	40	16	51	46
2	12	18	42	42	16	40	25
3	12	37	55	**64**	0	18	**111**

The major interpretation differences of SHAP explainability results include:

– Glucose – SHAP results show inverse proportion to the highest measurement values. More informative is the assessment based on absolute SHAP values. For cluster 0 the results are more than 2x higher than for clusters 1 and 2, and significantly lower in cluster 3. This is consistent with the interpretation of Glucose as one of the most significant characteristics that differentiates the clusters, both in terms of mean and number of exceedances.
– BMI – For this feature, we can observe a relation between the SHAP sign and the high and low values of BMI. Clusters 0 and 1 have a similarly high BMI average, but cluster 1 has significantly higher SHAP results. This may be related to the highest overall number of cases with Obesity type 2 in cluster 1. The situation is similar for the two clusters with the lowest average BMI. In both cases, SHAP values are positive, but for cluster 3 they are significantly higher than for cluster 2. Cluster 3 also has a higher number of cases of type 2 Obesity and Overweight compared to cluster 2.
– Age – Interpretation of the results for Age is not consistent. In cluster 3 it was the most significant parameter, proportional to low age. Analysing the results for cluster 1 (low mean Age), SHAP values are negative, but for the high mean Age in cluster 2 they are also negative. This means that the significance of Age is related to other parameters that characterise the cluster.

Summarizing the results obtained, the patient groups determined by the unsupervised model are characterized by:

– Cluster 0 – Glucose problems, higher Age, genetic susceptibility to diabetes, highest average Pregnancies, high percentage of diabetics. The most significant feature, according to SHAP, defining this cluster was the high values of the Glucose parameter.
– Cluster 1 – Despite the small proportion of participants with increased Glucose, this group is distinguished by the highest proportion of patients with excess BMI and significantly higher average Skin Thickness feature than in the other groups. The most significant characteristics according to SHAP that define this cluster were low Glucose values and high BMI values.
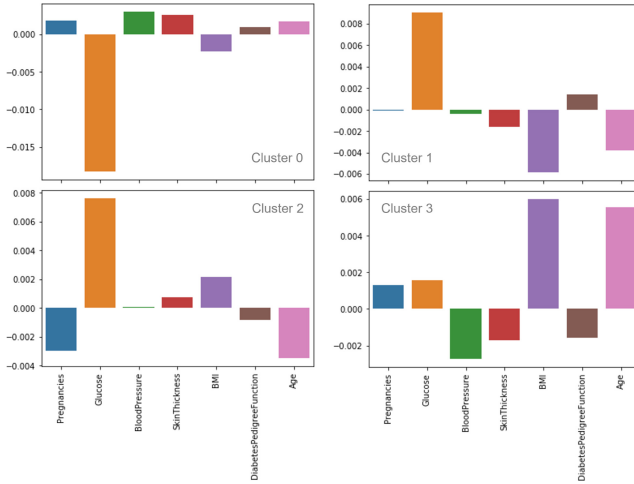
Fig. 4. SHAP explanations of proxy model

– Cluster 2 – The group with the correct glucose results, the lowest number of excess BMI, and the lowest average Skin Thickness. These features and Diabetes Pedigree Function are the most significant differentiation from cluster 0 because the mean number of pregnancies and the age are similar. The most significant features according to SHAP that defined this group were low Glucose values and high Pregnancy values.

– Cluster 3 – The most 'healthy' group, with the lowest averages for almost all parameters. Despite the relatively high number of excess BMI, there are a small number of patients with the highest obesity (type 3 Obese). A distinguishing feature of this group is also the significantly higher number of cases with too low Blood Pressure, as well as the lowest average Age and Pregnancies. The most significant characteristics, according to SHAP, defining this cluster were low BMI, Blood Pressure and Age.

In the course of the research, the method in which domain knowledge was transformed into an additional set of features proved to be an important factor in facilitating the interpretation of the original data. Parameters on a binary scale can be represented in 2 ways. Taking into account 4 levels of exceedances for the BMI feature (equivalent to creating 4 additional parameters), so that each exceeded threshold means 1 and a measurement below the threshold – 0, when describing a person with maximum exceed of BMI (Obesity type 3), all 4 parameters will have a value of 1. This approach to the creation of the dataset turned out to be better in terms of training the prediction model. The second way to represent binary characteristics is to assign 1 value only to the highest exceedance threshold for an individual. In this case, the value of the Obesity type 3 parameter will be equal to 1, while the other features describing BMI exceedances will be equal to 0. This approach is more intuitive for analyzing the distributions of exceedances within clusters and was used to create a Table 3.

6   Summary

We described a procedure for obtaining intuitively interpretable information about distinguishable groups of patients and defining differences between them with a usage of clustering, rule–based encoded domain knowledge, and SHAP values. On the basis of the data patterns found and the analysis of information with different origins, we determined the characteristic attributes of each group. This potentially allows us to automate the procedure and provide final conclusions in the form of easy–to–understand information even by non–specialist user.

This type of analysis is an important part of, for example, recommendation systems. Users of patient prediagnostic decision support are healthcare facilities that have an established set of treatment actions, the equivalent of the clusters in our study. The evaluation of the health of the clients is performed based on the knowledge of specialists and the diagnosis scripts prepared for them. By determining the associations between patient health indicators and the medical services they use, we can shorten the time between identifying the needed treatment and actually starting it. Applying our procedure in practice would require an in–depth analysis of the available domain knowledge as well as the company's experience.

One of the possibilities for further enhancement of our research is to use counterfactual explanations to estimate how much the value of a feature should change to assign a subject to a different category/cluster, other than the original. One of the implementations of that method is Diverse Counterfactual Explanations (DiCE) [16]. With a properly designed procedure, we can use the DiCE algorithm to estimate the goals of a treatment. This information could be useful in planning the intensity of therapy. Development in this direction may allow for faster and automated assessment of the patient's health. This type of information can be used to shorten the diagnostic process and limit the number of intermediate stages before the patient can reach a suitable specialist for his needs.

References

1. Body mass index : considerations for practitioners, August 2. https://stacks.cdc.gov/view/cdc/25368, pamphlet (or booklet)
2. Akindele, M., Phillips, J., Igumbor, E.: The relationship between body fat percentage and body mass index in overweight and obese individuals in an urban African setting. J. Public Health Africa **7**, 15–19 (2016)
3. Arya, V., et al.: One explanation does not fit all: a toolkit and taxonomy of AI explainability techniques. CoRR abs/1909.03012 (2019). https://arxiv.org/abs/1909.03012
4. Breiman, L.: Statistical modeling: the two cultures (with comments and a rejoinder by the author). Stat. Sci. **16**(3), 199–231 (2001)
5. Collaris, D., van Wijk, J.J.: Explainexplore: visual exploration of machine learning explanations. In: 2020 IEEE Pacific Visualization Symposium (PacificVis), pp. 26–35 (2020)

6. Ćwiek-Kupczyńska, H., et al.: Semantic concept schema of the linear mixed model of experimental observations. Sci. Data **7**(1), 70 (2020). https://doi.org/10.1038/s41597-020-0409-7

7. Hastie, T., Tibshirani, R., Friedman, J., Franklin, J.: The elements of statistical learning: data mining, inference, and prediction. Math. Intell. **27**, 83–85 (2004)

8. of Health, U.S.D., for Disease Control, H.S.C., for Health Statistics, P.N.C.: National health and nutrition examination survey (nhanes), 1999–2000 (2012). https://doi.org/10.3886/ICPSR25501.v4

9. Holzinger, A.: Explainable AI and multi-modal causability in medicine. i-com **19**(3), 171–179 (2020). https://doi.org/10.1515/icom-2020-0024

10. Jin, W., Li, X., Hamarneh, G.: Evaluating explainable AI on a multi-modal medical imaging task: Can existing algorithms fulfill clinical requirements? In: AAAI Conference on Artificial Intelligence, March 2022

11. Klein, L., El-Assady, M., Jäger, P.F.: From correlation to causation: formalizing interpretable machine learning as a statistical process (2022). https://arxiv.org/abs/2207.04969

12. Kleiser, C., Schaffrath Rosario, A., Mensink, G.B., Prinz-Langenohl, R., Kurth, B.M.: Potential determinants of obesity among children and adolescents in germany: results from the cross-sectional kiggs study. BMC Public Health **9**(1), 46 (2009). https://doi.org/10.1186/1471-2458-9-46

13. Lawrynowicz, A.: Semantic data mining: an ontology-based approach, April 2017

14. Li, Z., Zhang, C., Zhang, Y., Zhang, J.: Semanticaxis: exploring multi-attribute data by semantic construction and ranking analysis. J. Vis. **24**(5), 1065–1081 (2021). https://doi.org/10.1007/s12650-020-00733-z

15. Lundberg, S.M., et al.: Explainable AI for trees: From local explanations to global understanding. CoRR abs/1905.04610 (2019). https://arxiv.org/abs/1905.04610

16. Mothilal, R.K., Sharma, A., Tan, C.: Explaining machine learning classifiers through diverse counterfactual explanations. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, ACM, January 2020. https://doi.org/10.1145/3351095.3372850

17. Nalepa, G.J., Bobek, S., Kutt, K., Atzmueller, M.: Semantic data mining in ubiquitous sensing: a survey. Sensors **21**(13) (2021). https://www.mdpi.com/1424-8220/21/13/4322

18. Novosad, S., Khan, S., Wolfe, B.N., Khan, A.: Role of obesity in asthma control, the obesity-asthma phenotype. J. Allergy **2013**, 538642 (2013)

19. Ribeiro, M.T., Singh, S., Guestrin, C.: Anchors: high-precision model-agnostic explanations. In: AAAI Conference on Artificial Intelligence (2018)

20. Sirichanya, C., Kraisak, K.: Semantic data mining in the information age: a systematic review. Int. J. Intell. Syst. **36**(8), 3880–3916 (2021). https://onlinelibrary.wiley.com/doi/abs/10.1002/int.22443

21. Smith, J.W., Everhart, J.E., Dickson, W., Knowler, W.C., Johannes, R.S.: Using the adap learning algorithm to forecast the onset of diabetes mellitus. In: Proceedings of Annual Symposium on Computer Application in Medical Care, p. 261. American Medical Informatics Ass. (1988)

22. Suchanek, F., Weikum, G.: Knowledge bases in the age of big data analytics. Proc. VLDB Endowment **7**, 1713–1714 (2014)

23. Visani, G., Bagli, E., Chesani, F.: Optilime: optimized LIME explanations for diagnostic computer algorithms. CoRR abs/2006.05714 (2020). https://arxiv.org/abs/2006.05714