Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 104 (2021) 1595–1600

www.elsevier.com/locate/procedia

54th CIRP Conference on Manufacturing Systems

X-PHM: Prognostics and health management knowledge-based framework for SME

Nabil Omri[a,b,*], Zeina Al Masry[a], Nicolas Mairot[b], Sylvian Giampiccolo[b], NoureddineZerhouni[a]

[a] FEMTO-ST institute, Univ. Bourgogne Franche-Comté, CNRS, ENSMM, 24 rue Alain Savary, Besançon cedex, 25000, France
[b] SCODER, 1 rue de la Forêt Z.A. l'Orée du Bois, Pirey 25480, France

* Corresponding author. *E-mail address:* nabil.omri@femto-st.fr

Abstract

Prognostics and Health Management (PHM) is an emerging concept based on industrial data management. The implementation of PHM in small and medium-sized enterprises (SMEs) is currently limited due to data accessibility difficulties. In order to overcome this pitfall, one could formalize the operators' knowledge and integrate it in the SME's information system. Thus, the implementation of the PHM will be based on this information system associating data with knowledge. To this end, we propose a collaborative PHM approach (X-PHM) to ensure the extraction of operators' knowledge and its integration into the PHM process. The decision resulting from this approach is restituted with a concern of explainability. This paper details the proposed approach while focusing on the data management process and its integration in explainable decisions. This new framework is applied in a French SME to understand its production process and facilitate its digital transformation.

Keywords: PHM; Knowledge formalization and integration; Explainable artificial intelligence; SME; Data analysis.

1 Introduction

Small and Medium-sized Enterprises (SMEs) are defined as all companies with less than 250 workers and with less than 40 million euros in terms of turnover [2]. SMEs represent about 90% of all companies, and they have an important contribution to job creation and global economic development [4]. Despite their important role in the global economy, SMEs fail to integrate new technologies that threaten their sustainability. In the era of Industry 4.0, small businesses suffer from many constraints that limit their integration in the world of Industry 4.0. In [3], the authors group the SME's limitations into two classes (resource and organizational constraints). It is true that digitized data, the raw material for technologies based on artificial intelligence (AI), are not readily available in SMEs. However, SMEs have a large amount of data in the form of remarkable know-how that can be exploited for successful digital transfer. The advantage in SMEs is that operators are close to the tool, which allows them to identify their needs and prioritize them in terms of urgency. The versatility of the staff allows them to analyze each problem with its impact on the entire production process. Thus, the phase of modeling the problem and identifying the variables necessary for the resolution seems easier and more efficient in the case of SMEs, which reduces the costs of implementation. The real challenge is of human nature to convince the staff to share their know-how with a set of mysterious black box-type technologies.

In this work, we propose to use Prognostics and Health Management (PHM) as a framework to ensure the extraction and formulation of operators' knowledge and its integration into the information system of SMEs. This knowledge is used in data analysis tasks, and the resulting decisions are restituted with a concern of explicability. PHM is a science that studies the health of a system and predicts its future evolution. This concept makes it possible to control the systems better and to set up adapted maintenance strategies [5]. In [3], the authors define PHM as "a set of tools that can be used in cascade or separately to monitor the health of a system, predict its future evolution and/or optimize decisions ".

The objective of this work is to improve system performance and to enrich the company's information system through explainable analysis results. However, two problems emerge: (i) the treatment of industrial data and (ii) the formalization and

10.1016/j.procir.2021.11.269

integration of human knowledge in the analysis process. We are faced here with two complex problems in the field of data analysis. These problems are machine learning tools that are explainable [1] and informed [8], as known in the literature. The term explanation refers to the ability of the machine learning algorithm to explain the results and models obtained. This makes the algorithm more transparent to the user and creates more trusting relationship with the users. While "informed machine learning" refers to integrating human knowledge into the learning process to improve it.

In summary, this work aims to develop a collaborative PHM approach (X-PHM) to ensure the extraction of operator knowledge and its integration into the PHM process. The decisions resulting from this approach are returned with a concern for explainability. The rest of this document is organized as follows. Section 2 presents a state of the art on the formalization of human knowledge and the explanation of black-box models. Section 3 details the proposed X-PHM framework. This framework is applied to a real case study, the results of which are presented in section 4. The conclusion and perspectives of this work are detailed in section 5.

2 Related work

In recent years, many works have been conducted to integrate human knowledge into the data analysis process. However, in reality, the proposed techniques fail due to human resistance in sharing their know-how with black-box models. Thus, we propose to use explainable learning to make the selected models more transparent and to build a form of trust between humans and data analysis technologies. The next paragraphs present a brief review of human knowledge formalization and its integration into the explainable learning process.

2.1 Human knowledge formalization and integration

Human knowledge is an important source of information that should be used to strengthen the information system of SMEs. In this context, several approaches to formalize human knowledge are proposed [8]. This section offers a review of the literature of these approaches while detailing the different types of human knowledge and the means to formalize them for informed data analysis.

Human beings are the most intelligent creatures and possess a vast knowledge. Many studies have been carried out in the literature to characterize this knowledge from different perspectives (e.g., sociology, psychology, etc.). These studies offer several ways of categorizing human knowledge. Following these categorizations, we propose here to classify human knowledge into three main types: (i) General knowledge, (ii) Scientific laws, and (iii) Expertise. These different human knowledge are integrated into machine learning techniques to give birth to the informed learning discipline. Informed learning is defined as a combination of data and human knowledge for a more effective learning process. Overall, a learning process can be divided into three main stages: (i) Data preprocessing, (ii) Data mining, and (iii) Results. Thus, human knowledge can be integrated into the learning process throughout these stages.

- **Informed Data Preprocessing:** The prior knowledge of the studied problem can lead to the phase of selection of the characteristics. In addition, it can assess the accuracy of the data collected and improve it accordingly. Moreover, it can be used as a second data source by feeding a learning algorithm through a collection of rare observations and events.
- **Informed Data Mining:** At the level of data mining, human knowledge can identify the set of parameters of the used algorithm. These parameters can be the architecture of an artificial neural network (ANN) or the depth of decision tree (DT), etc. In addition, prior knowledge of the studied problem can be useful to modify the loss function according to the final objective.
- **Informed Outcomes:** At this level, man can intervene to assess the consistency of the rules extracted by comparing them to existing scientific laws or simply by using his expertise. In addition, known rules that are not learned from the data can be injected into the final results to document them.

Each type of informed learning type requires a different formalization of knowledge. However, it is still challenging to communicate human knowledge to a machine. To this end, many approaches to formalize human knowledge in machine-understandable forms have been developed in the literature. We propose here to classify these approaches into three main categories: (i) Mathematical models, (ii) Logical rules, and (iii) Statistical relationships.

Ideally, knowledge formalization models are the most suitable for communicating between humans and machines because they are understandable from both sides, but not all hu- man knowledge can be modeled according to these forms. To this problem is added the human resistance in sharing their know-how with black-box models. Thus, explanation techniques are used to build trust between human and machine learning algorithms.

2.2 Explanation models

Increasingly, the explanation of black-box decision systems has attracted more attention. This need for explanation is generally due to incomplete problem formalization, creating a fundamental obstacle to optimization and evaluation. Thus, many techniques have recently been proposed to explain black-box decision systems [1]. In this context, the authors of [10] applied decision trees to explain ANN decisions. Indeed, classification rules have been widely adopted to explain the decisions of ANN and support vector machine. These techniques are used to generate a global explanation of the used black-box model and, when the training dataset is available, they can be used as completely transparent classifiers.

Other approaches tackle the problem of explaining the local behavior of a black box [1]. In other words, they explain the

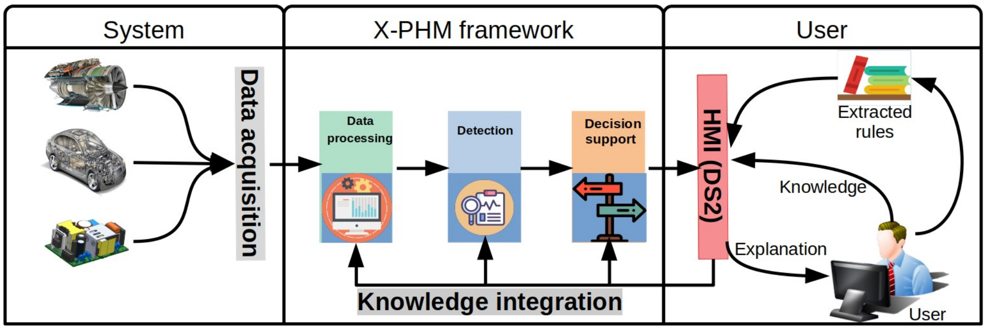

Fig. 1. Illustration of the proposed approach which consists of four steps: (i) improve data quality using human knowledge, (ii) extract knowledge from these data, (iii) explain the extracted knowledge to enrich the human knowledge base, and (iv) use this knowledge to improve data quality in future data analysis tasks.

decision assigned to a specific instance. There are two types of approaches: model-dependent approaches and agnostic approaches. In the first category, most of the previous works aim to explain ANN algorithms. Their explanation is based on salience masks, that is, a subset of the instance that explains what is primarily responsible for the prediction [14]. Examples of protruding masks are parts of an image or words or phrases in the text. On the other hand, agnostic approaches provide explanations for any black-box type. In [6], the authors present the local interpretable model-agnostic explanations (LIME), which starts from instances generated randomly in the vicinity of the instance to be explained. The method deduces linear models from them as well as understandable local predictor models. The importance of a feature in the linear model represents the explanation ultimately given to the user. As a limit of the approach, a random generation of the district does not consider the density of the results of the black box in the district authorities. Therefore, the linear classifiers derived from them may not correctly characterize the values of the results based on the predictive characteristics. We can instead use a genetic algorithm that exploits the generation of the black box, for example. LIME extensions using decision rules (called anchors) and program expression trees are presented in [7] and [9] respectively. The [7] extension uses a bandit algorithm that randomly constructs the anchors with the highest coverage and respecting a precision threshold. [9] takes a simulated annealing approach that randomly increases, decreases, or replaces nodes in an expression tree. The adopted neighborhood generation process is the same as that of LIME. Another crucial weak point of these approaches is the need for user-specified parameters for the desired explanations: the number of features, the level of precision, the maximum depth of the expression tree [7].

On the contrary, our approach is a non-parametric tool that can explain the results of any black-box model. Moreover, it provides a logical rules rather than an impact order of each variable. Thus, the user receives not only an explanation for a specific observation but also a decision rule applicable locally in the neighborhood instances. This informed and explainable learning concept is introduced in the proposed X-PHM framework to ensure the extraction of operators' knowledge and its integration into the PHM process. The decisions resulting from this approach are restituted with a concern of explainability. The following section details the proposed X-PHM framework.

3 Proposed approach

As shown in Figure 1, the proposed approach consists of an interactive framework that allows communication between the user and the black-box machine learning algorithm. This interaction is represented in terms of (i) integrating user expertise into the learning phase and (ii) explaining the proposed decision. Thus, human knowledge can be validated through this process to confirm (or decline) it. The capitalized knowledge is then stored in a library of rules that will be used in the future to support decision-making and explain strange phenomena.

3.1 Knowledge integration

This section presents the phases of knowledge integration and data generation. The first step is to inject human knowledge into the analysis process. Then, this knowledge is used to generate data samples close to the instances whose decisions we want to explain belong to them. Human knowledge can be integrated into the learning process throughout its various stages. However, some tasks are automated by sophisticated algorithms to facilitate the user's mission. Scalable algorithms are used to adapt the architecture of the ANN to the complexity of the problem [13]. Moreover, many variable selection techniques allowing to accomplish this mission automatically are developed and applied in many [11, 12] fields. In addition, other methods are being developed to effectively learn from imbalanced data or to assess the consistency of results automatically. For this, we propose integrating human knowledge in the generation of new data and improving their quality. To do this, we propose in this part to enhance the knowledge of the PHM user in four ways:

- Definition of new variables by merging a set of variables to generate a new features more relevant to the description of the studied problem.
- Definition of the set of variables which can affect the problem (or which has no influence).
- Statistical description of variables (mean, max, etc.).

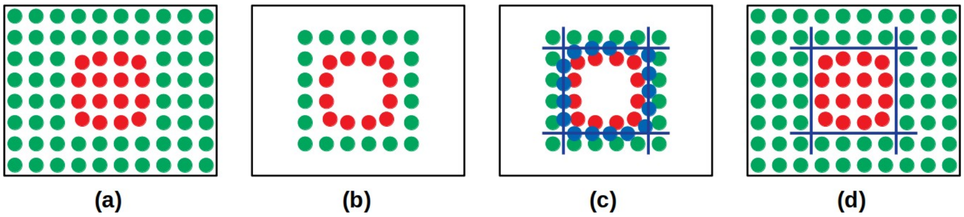

Fig. 2. Details of the explanation process.

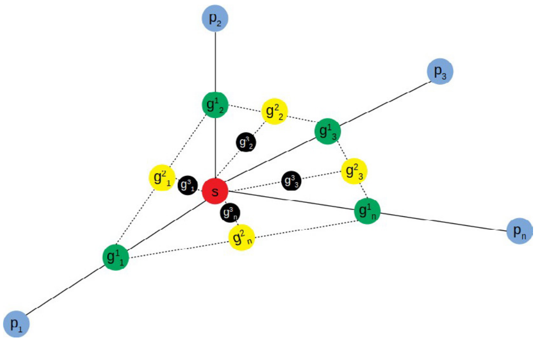

Fig. 3. The process of generating data.

- Definition of statistical relationships between variables (correlation, statistical distribution, etc.).

The information provided by the user will be used to generate new data, which is an important step in explaining the results. Thus, we consider a data instance s that we would like to explain the decision that belongs to it. To do this, the n nearest neighbors (p_i, $i = 1..n$) of the data point s are selected for use in the generation process. These data points are used to generate the first set of n new data samples g_i^1 as follows:

$$g_i^1 = \frac{p_i + s}{2}, \forall i = 1, .., n. \quad (1)$$

The generated data is then used to generate a second set of n new data samples g_i^2 as follows:

$$g_i^2 = \frac{g_i^1 + g_o^1}{2}, \forall o, i = 1, .., n. \quad (2)$$

where i and o are consecutive and $i \neq o$.

Using the initial data point s and the points generated at the second level g_i^2, a third set of n new data samples g_i^3 are generated:

$$g_i^3 = \frac{g_i^2 + s}{2} \forall i = 1, .., n. \quad (3)$$

The same process is repeated until the required N data samples are generated. The data generated in this step will be used later in the explanation of the results. The following paragraph details the explanation process.

3.2 Semi-global explanation of results

The last step of the proposed approach is the explanation phase. To do this, a semi-global explanation algorithm is used to obtain an explanation for each decision. For an explanatory task, the most important thing is to determine a clear and understandable rule allowing to separate the different classes. Thus, the work surface is the one that separates these classes. As shown in Figure 2, the proposed approach is based on the generation of new instances at the boundary between classes. These new instances are then used to separate the classes linearly.

Let ζ(., .) be a function which takes as inputs a black box model b and an observation of the system $\boldsymbol{V_k} = \{v_{k1}, v_{k2}, .., v_{km}\}$, where v_{kj}, $k = 1, . . . , l$, $j = 1, . . . , m$ is an instance of the variable v_j. The function ζ(., .) is called an explanatory model when it relates the decision of $b(.)$ to a physical reality of the system. The proposed approach can be summed up in three points:

- **Identification of the decision plan:** For a data point s whose decision we want to explain, the first step in identifying its decision plan consists of determining the n instances closest to it and which belong to the inverse class of that of the system (see Figure 2 (a)).
- **Determination of the separation plan:** Using the n neighboring instances identified, the decision plan is determined. To do this, for each instance identified, we determine the instance closest to it and belonging to the same class as that of s (see Figure 2 (b)).
- **Explanation of the decision:** The space between these two groups of identified instances is called the "decision plane". In the rest of this work, the explanation will be given in this plan. Thus, artificial data is generated on this surface using the generation algorithm defined above (see Figure 2 (c)).

The first step in the explanation process is to use the black box model $b(.)$ to label the newly generated data. This task will identify the behavior of $b(.)$ in the decision surface. In this context, a decision d is a function that serves to differentiate the instances in the decision plane. In 2D, the function used to classify between instances is a row, while the function used to classify instances in 3D is called a plane, just as the function which classifies the point in a higher dimension is called hyperplane (see Figure 2 (d)). In general, the equation of the hyperplane in

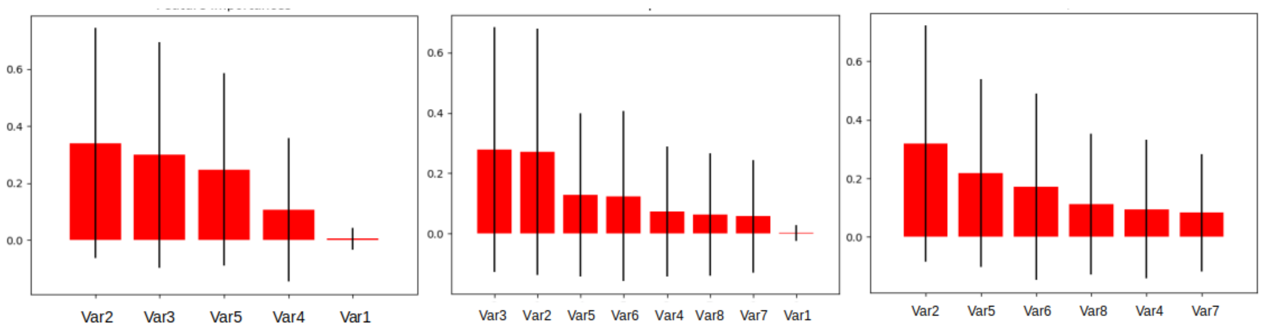

Fig. 4. Features importance at each step of the data generation. Based on the strength of materials theory, new features are defined. These variables are: $Var_6 = f(Var_2, Var_3)$, $Var_7 = f(Var_1, Var_2)$ and $Var_8 = f(Var_1, Var_2)$. From the SCODER's staff point of view, these new features are very important but they create a redundancy in the data. For that, Var_1 and Var_3 are eliminated since they are represented by Var_2, Var_6 and Var_7.

n dimensions can be given by:

$$\alpha^T \times \mathbf{V} + c. \tag{4}$$

where c is a constant, $\mathbf{V} = (V_1, \ldots, V_l)$ is the vector of variables and α^T is the leading vector of the decision plane.

The numerical formula for α^T is then determined using the "Soft margin" technique. To do this, we consider the case where there are two classes $Class1$ and $Class2$ that refer to the healthy and faulty classes which corresponds to a detection problem. A decision d_k for an observation V_k can take two values: $d_1 = -1$ when $\{V_k \in Class1\}$ and $d_2 = 1$ if $\{V_k \in Class2\}$. Thus, a decision that belong to an observation V_k is well explained if the following condition is satisfied:

$$d_k \times (\alpha^T \times V_k + c) \geq 1, \quad d_k \in \{-1, 1\}. \tag{5}$$

This condition requires that the decision plan properly explains all decisions. This requirement is therefore difficult to meet in reality. For this reason, we suggest allowing some bad explanations in the dataset. To do this, we will grant a constant $s_k \geq 0$ which for each observation V_k, we have:

$$d_k \times (\alpha^T \times V_k + c) \geq 1 - s_k, \quad s_k \geq 0. \tag{6}$$

This new constraint makes it possible to accept imperfect explanations. However, the objective is to minimize these imperfect explanations. Thus, the leading vector of the decision plane α^T is determined as follows:

$$\begin{cases} \min \sum_{k=1}^{l} s_k \\ Constraints: \\ \quad d_k \times (\alpha^T \times V_k + c) > 1 - s_k. \\ \quad s_k \geq 0. \end{cases} \tag{7}$$

The proposed explanation approach allows explaining decisions in a semi-global way for binary classification problems (only two classes). However, multi-class problems can be treated in the same way as binary classification problems. Thus, a multi-class problem can be transformed into a two-class problem by considering only the class of the instance s to be explained and the rest of the classes as a single class. In the following section, a validation of the approach is conducted in a real case study.

4 Case study

The proposed approach is applied to the Scoder company. Scoder is a French SME specialized in ultra-precise stamping for automotive applications. The application consists of a stamping line where sheet metal properties are controlled to identify their impact on production performance. The class is a binary variable that indicates if the metal coil is suitable for production or not. The *X_PHM* framework is used to understand the impact of each metal propriety in the production performance and to identify the optimal sheet-metal characteristics for stable production. In this context, the Scoder's staff have excellent know-how that can be used to guide the knowledge extraction phase. The previously detailed approach is used in order to make the extracted rules transparent. Below, the steps of this application are detailed:

1. Data Preparation. The metal coils proprieties are collected and crossed with the machine breakdowns to train a black-box machine learning model.

2. Knowledge integration. In this step, operators know-how is used in different levels:

- Features selection: from the collected data, only five variables are chosen as important. For a confidentiality reason, these variables will be noted $Var_j, (j = 1, \ldots, m)$.
- New variable definition: based on the theory of strength of materials, new variables are defined. These variables are: $Var_6 = f(Var_2, Var_3)$, $Var_7 = f(Var_1, Var_2)$ and $Var_8 = f(Var_1, Var_2)$. From the Scoder's staff point of view, these new variables are significant, but they create a redundancy in the data. For that, Var_1 and Var_3 are eliminated since they are represented by Var_2, Var_6 and Var_7. Figure 4 shows the importance of the features using the Gini criteria at each step.

- Variables bounds: at this level, human knowledge is used to set the variables' bounds to avoid outliers in the data generation step for the explanation phase.

3. Train *b*(.). The collected data is used to train a black-box algorithm to solve the problem. In this application, the ANN algorithm is used to predict each entry class.

4. Explanation of results. Using the process described in Section 3.2 the learned model by the ANN algorithm is explained for each data instance. Figure 5 shows an explanation example of the decision belonging to the instance *s*. This explanation consists of the separation plan, which separates the different classes (see Equation (8)). This explanation is valid to each data point in the neighborhood of the instance *s*.

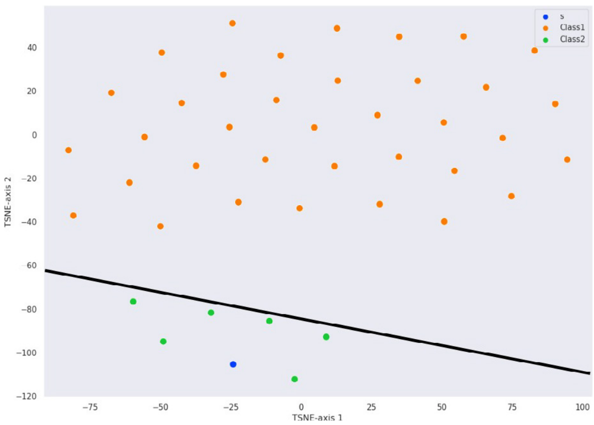

Fig. 5. Explanation results. For an initial instance *s* (in green), the proposed approach is applied to generate new data and identify the decision plane.

$$\zeta(b,s) = \begin{cases} Class1 \; if \; 0.66 \times Var_2 - 0.26 \times Var_4 + 0.26 \times Var_5 - 0.50 \\ \quad \times Var_6 - 0.19 \times Var_7 - 0.19 \times Var_8 < 233 \\ Class2 \; if \; 0.66 \times Var_2 - 0.26 \times Var4 + 0.26 \times Var_5 - 0.50 \\ \quad \times Var_6 - 0.19 \times Var_7 - 0.19 \times Var_8 \geq 233 \end{cases} \quad (8)$$

where $Class1$ is the class of suitable metal coils and $Class2$ is the class of bad metal coils.

The obtained explanation gives greater importance to Var_2. In addition, the greater this variable, the more suitable the metal of the coil is for production.

5 Conclusion

This paper presents a new framework, X-PHM, which allows ensuring the extraction and formulation of operators' knowledge and its integration into the information system of SMEs. This knowledge is used in data analysis tasks, and the resulting decisions are restituted with a concern of explicability to enrich the human knowledge. The proposed approach is applied in a real SME to understand its production process and facilitate its digital transformation.

The proposed framework combines the integration of human knowledge, and the explanation of analysis results in a collaborative approach of PHM. This combination allows to improve the data quality and to enrich the human knowledge. To improve the proposed approach, we offer a non-exhaustive list of points that should be considered as future works:

- Further formalization of the data quality as a function of the knowledge even using an empirical approach.
- Knowledge uncertainties quantification and assessment of their impact on data quality improvement results.
- Uncertainties formalization and proposition of correction approaches.

References

[1] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., Pedreschi, D., 2018. A survey of methods for explaining black box models. ACM computing surveys (CSUR) 51, 1–42.
[2] Mittal, S., Khan, M.A., Romero, D., Wuest, T., 2018. A critical review of smart manufacturing & industry 4.0 maturity models: Implications for small and medium-sized enterprises (smes). Journal of manufacturing systems 49, 194–214.
[3] Omri, N., Al Masry, Z., Giampiccolo, S., Mairot, N., Zerhouni, N., 2019. Data management requirements for phm implementation in smes, in: 2019 Prognostics and System Health Management Conference (PHM-Paris), IEEE. pp. 232–238.
[4] Omri, N., Al Masry, Z., Mairot, N., Giampiccolo, S., Zerhouni, N., 2020. Industrial data management strategy towards an sme-oriented phm. Journal of Manufacturing Systems 56, 23–36.
[5] Omri, N., Al Masry, Z., Mairot, N., Giampiccolo, S., Zerhouni, N., 2021. Towards an adapted phm approach: Data quality requirements methodology for fault detection applications. Computers in Industry 127, 103414.
[6] Ribeiro, M.T., Singh, S., Guestrin, C., 2016. " why should i trust you?" explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp. 1135–1144.
[7] Ribeiro, M.T., Singh, S., Guestrin, C., 2018. Anchors: High-precision model-agnostic explanations., in: AAAI, pp. 1527–1535.
[8] von Rueden, L., Mayer, S., Beckh, K., Georgiev, B., Giesselbach, S., Heese, R., Kirsch, B., Pfrommer, J., Pick, A., Ramamurthy, R., et al., 2019. Informed machine learning–a taxonomy and survey of integrating knowledge into learning systems. arXiv preprint arXiv:1903.12394 .
[9] Singh, S., Ribeiro, M.T., Guestrin, C., 2016. Programs as black-box explanations. arXiv preprint arXiv:1611.07579 .
[10] Tan, S., Soloviev, M., Hooker, G., Wells, M.T., 2020. Tree space prototypes: Another look at making tree ensembles interpretable, in: Proceedings of the 2020 ACM-IMS on Foundations of Data Science Conference, pp. 23–34.
[11] Trabelsi, I., Zeddini, B., Zolghadri, M., Barkallah, M., Haddar, M., 2021. Obsolescence prediction based on joint feature selection and machine learning techniques., in: ICAART (2), pp. 787–794.
[12] Zemouri, R., Omri, N., Devalland, C., Arnould, L., Morello, B., Zerhouni, N., Fnaiech, F., 2018. Breast cancer diagnosis based on joint variable selection and constructive deep neural network, in: 2018 IEEE 4th Middle East Conference on Biomedical Engineering (MECBME), IEEE. pp. 159–164.
[13] Zemouri, R., Omri, N., Fnaiech, F., Zerhouni, N., Fnaiech, N., 2019. A new growing pruning deep learning neural network algorithm (gp-dlnn). Neural Computing and Applications , 1–17.
[14] Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A., 2016. Learning deep features for discriminative localization, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2921–2929.