Towards the Automation of Industrial Data Science: A Meta-learning based Approach

Moncef Garouani[1,2,3], Adeel Ahmad[1][a], Mourad Bouneffa[1], Arnaud Lewandowski[1], Gregory Bourguin[1] and Mohamed Hamlich[2]

[1]*Univ. Littoral Côte d'Opale, UR 4491, LISIC, Laboratoire d'Informatique Signal et Image de la Côte d'Opale, F-62100 Calais, France*

[2]*ISSIEE Laboratory, ENSAM, University of Hassan II, Casablanca, Morocco*

[3]*Study and Research Center for Engineering and Management(CERIM), HESTIM, Casablanca, Morocco*

Keywords: Automated Machine Learning, Manufacturing Big Data, Industry 4.0, Industrial Data Science, Meta-learning.

Abstract: In context of the fourth industrial revolution (industry 4.0), the industrial big data is subject to grow rapidly to respond the agile industrial computing and manufacturing technologies. This data evolution can be captured using ubiquitous integrated sensors and multiple smart machines. We believe the use of data science methodologies, for the selection of models and configuration of hyper-parameters, may help to better control such data evolution. But, at the same time, the industrial practitioners and researchers often lack machine-learning expertise to directly retrieve the benefit from valuable manufacturing big data. Such a lack poses the major obstacle to yield value from even-though familiar data. In this case, a collaboration with data scientists may become an exigence along with the extensive machine learning knowledge which presumably may result to pursue further delays and effort. Multiple approaches for automating machine learning (AutoML) have been proposed for the past recent years in order to alleviate this deficiency. These approaches are expected to perform better along with accomplishment of computing resources which are mostly not readily accessible. To address this research challenge, in this paper, we propose a meta-learning based approach that may serve an effective decision support system for the AutoML process.

1 INTRODUCTION

Advanced analytics offers new opportunities to improve and innovate manufacturing processes (Wolf et al., 2019). The recent advances, in terms of storage capacity, computing power as well as the rapid development of advanced analytics solutions, have offered manufacturing industries the unprecedented possibilities to extract knowledge and business value from large datasets (Wang et al., 2018). It may provide means to achieve the highest predictive performances instead of traditional predictive modeling approaches. The use of advanced machine learning (ML) methods can play a significant role in production design, quality management, scheduling, etc.

The current competitive environment, with improved availability, sustainability, and quality of manufacturing services in smart factories, has already triggered the requirement of using Artificial Intelligence (AI) solutions to streamline complex operations while improving quality and reducing costs (Thoben et al., 2017; Wuest et al., 2016). The manufacturing sector can benefit greatly from the use of advanced analytics since data is abundantly available (Wolf et al., 2019). Recent manufacturing strategies, such as Industry 4.0 in Germany, Industrial Internet in the United States, and the Made in China 2025 initiative, recognize the crucial importance of utilizing data in order to enhance manufacturing competitiveness (Thoben et al., 2017; Tao et al., 2018). However, the manufacturing industry is not exploiting the full potential of data analytics. We observe the following reasons for such a lack:

- Complexity of unifying the data analytics and micro services,
- Lack of reliable data ingestion chains,
- Lack of collaboration tools between business

[a] https://orcid.org/0000-0003-0132-3808

Garouani, M., Ahmad, A., Bouneffa, M., Lewandowski, A., Bourguin, G. and Hamlich, M.
Towards the Automation of Industrial Data Science: A Meta-learning based Approach.
DOI: 10.5220/0010457107090716
In *Proceedings of the 23rd International Conference on Enterprise Information Systems (ICEIS 2021) - Volume 1*, pages 709-716
ISBN: 978-989-758-509-8

managers, data engineers and data-scientists.

The domain experts (business managers and data engineers) are not necessarily competent to perform data analysis using some data mining techniques. These users often resort to the default implementations or commit to collaborations with machine-learning experts which are mostly complex and time consuming. The process itself, also known as knowledge discovery, consists of several steps such as data preprocessing, features engineering, model selection, hyper-parameters tuning and finally model validation or interpretation. The algorithm selection phase is among the most important steps along with the configuration of its related hyper-parameters, hence, the conduction of related tasks can make them less feasible for the non-experts.

The lack of decision support tools is also a major barrier that may prevent the manufacturing actors, as well as researchers, from harnessing the maximum potential of data analytics. The decision support tools can help to rendre the most suitable data analytic technique from a wide range of possible choices and configurations (Zacarias et al., 2018). While, the available literature has brought forward various techniques capable of solving the manufacturing complex problems, whereas, the necessary decision support tools to implement these techniques on an operative level are yet not sufficient to equip practitioners, decision-makers, and researchers for better data-analytics.

The work, in this paper, hence attempts an approach towards overwhelming this obstacle. The main objective of the proposed approach remains however to assist the industrial practitioners and researchers in data-science with the help of a recommendation system. This recommendation encompass the most convenient machine-learning algorithm and its related hyper-parameters configuration that shall ultimately give the best result of the analysis for the domain-specific problems. In order to achieve that, we make use of the concept of meta-learning (Brazdil et al., 2008), which consists of two main phases; which are learning phase and recommendation phase. For a given dataset and a predictive metric, we suggest the ML algorithms and their related hyper-parameters configuration that once applied yield the best classification performance (e.g., predictive accuracy, Recall, F1 score).

The rest of the paper is organized as follows: Section 2 presents a brief review of the related works in respect of theoretical background about meta-learning as an AutoML solution. Section 3 provides an overview of the proposed Meta-Learning based framework and methodology for supporting the optimization decisions of the AutoML process. We discuss the results and feasibility of the proposed methodology in Section 4. We evaluate the proposed approach with the help of an empirical study in section 5. Finally, we conclude the contents of the paper in Section 6 .

2 RELATED WORKS

A careful literature review reveals that the machine-learning methods have been proposed as a key technology for the industrial data analytics. Many researchers have focused on the instantiation of Data Science methods to retrieve benefits in industry 4.0 automation. However, their main concern has been regarding the ingestion of new data in data lakes to establish compliance with the governance rules for the business managers with or without the assistance of data engineers. Similarly, the data scientists have been concerned with the capabilities to collaborate and deploy, independently, their models in production.

Similarly, a lot of work has been done in recent years to add explanations to Artificial Intelligence (AI) systems (Samek and Müller, 2019; De et al., 2020; Bohanec et al., 2017), in particular those based on deep neural networks, which are mostly very efficient, but also in principle designed and implemented like black boxes. The explanation of reasoning has always been felt since the emergence of decision support systems (De et al., 2020; Shin,) but it is now more desirable to re-assure legitimate confidence on machine learning applications, particularly in real-time systems (such as the industry 4.0 applications). The need for explainability in machine learning requires the development of interrogeable information systems that must allow the transparency of involved concepts according to the level of abstraction of the concerned actors. The innovation of AI has been mostly characterized by the algorithmic advancements with respect to efficient data analysis. The aspects of meta-learning (to find out the factors that play a more critical role to better use limited data) are usually given less focus while the attention is given to the more computational performance and more training data.

We constrain the focus of current work on meta-learning technique due to available space limitation for the contents. Meta learning is an aspect of monitoring the progress of machine learning processes. Meta-learning techniques provide the methodologies to observe the performance of different machine learning models according to the target outcome and its correlation to the meta-data (learning experience). This notion is often referred to the understanding of the reasoning process and involved parameters while the deduction of target outcome. But in its very nature meta-learning differs from meta-reasoning. The fundamental objective of meta-learning in current machine learning literature is to enhance the performance of machine learning models through learning experience. Learning experience in this regard, concerne the exploitation of the neural architectures (machine learning pipelines) to eliminate the less-worthy intermediate decision points. The meta-learning in this aspect helps to evaluate the machine learning model with respect to the accuracy and learning time. It requires transparent model executions which involve the awareness of meta-features (algorithm configurations, network architecture, pipeline compositions, etc.) used to train the model.

Meta-learning or learning to learn, is a commonly used process that supports automation in data mining tools selection and configuration (Brazdil et al., 2008). It is a method that consist of observing relationships between dataset characteristics (Meta-features) and data mining algorithms performances. Later for a given unseen dataset the system should be able to select and rank a pool of learning algorithms that could yield the best predictive performance according to the expected performance metric. The main idea of meta-learning for advanced analytics selection and configuration is based on the simple assumption *Algorithms show similar performance for the same configuration for similar problems*. Particularly, meta-learning paradigm is the process of understanding and adapting learning itself on a higher level instead of starting *from scratch*, we leverage previously gained insights (Lemke et al., 2015).

A number of studies explore the application of meta-learning in various levels. These approaches range from automatic data pre-processing (Nargesian et al., 2017), automatic features extraction (Bilalli et al., 2016; Bilalli et al., 2018) to automatic model selection and hyper-parameters tuning (Laadan et al., 2019; Dyrmishi et al., 2019).

In (Bilalli et al., 2016), the authors propose a meta-learning based approach for automated data preprocessing. The authors used 28 features which are extracted from the datasets to train a meta-model. The meta-model is able to predict the impact of a list of 7 data transformations strategies on the final performance of 5 classification algorithms (Logistic Regression, Naive Bayes, IBk, PART, J48). For each dataset-algorithm pair, possible transformations are classified as either good, bad, or neutral by the meta-model, which corresponds to whether the transformation increases the prediction accuracy, decreases it or doesn't have a significant contribution.

Recently, in (Laadan et al., 2019), the authors propose RankML a meta-learning based approach for ML pipeline performance prediction. The RankML produces a ranked list of all pipelines based on their predicted performance for the given dataset, evaluation metric and the set of candidate pipelines. However, this approach may not be practical in all situations, since the system asks to provide the list of pipelines to rank, that a non-ML expert cannot produce.

Moreover, some studies (Cohen-Shapira et al., 2019; Feurer et al., 2019) propose the use of meta-features and learning to improve the AutoML process. However, these frameworks do not achieve the goal of identifying the promising analytics tools as well as configurations as a prompt and powerful support for the manufacturing application areas in the first place. Therefore, they are not suitable for decision-making at the managerial level (Zacarias et al., 2018).

3 FRAMEWORK OF THE PROTOTYPE OF VALIDATION

The framework in its actual form includes two independent main phases; which are the *Learning phase* and the *Recommendation phase*. The high-level architectural description of prototype of the proposed solution is illustrated in Fig. 1. We discuss in detail the different components of the framework's life cycle, in the following sub-sections.

3.1 Learning Phase

The *learning phase* is performed offline and consists of two main steps. In the first step a meta-dataset is established. We extracted 42 dataset characteristics (meta-features)–detailed in section 4.3 from each dataset. Furthermore, on each dataset, we executed 8 classification algorithms (meta-learners), by generating different predictive performance metrics (predictive accuracy, Recall, precision, F1-score) values. We primarily evaluate that with a 5-fold stratified cross validation strategy. For each data mining algorithm, we obtained a meta-dataset that is fed to the meta-knowledge base.

Dataset characteristics and performance measures altogether are referred to as metadata. In the second step, meta-learning is performed on top of the meta-knowledge base. As a result, a predictive meta-model is generated that can be used to predict a ranked list

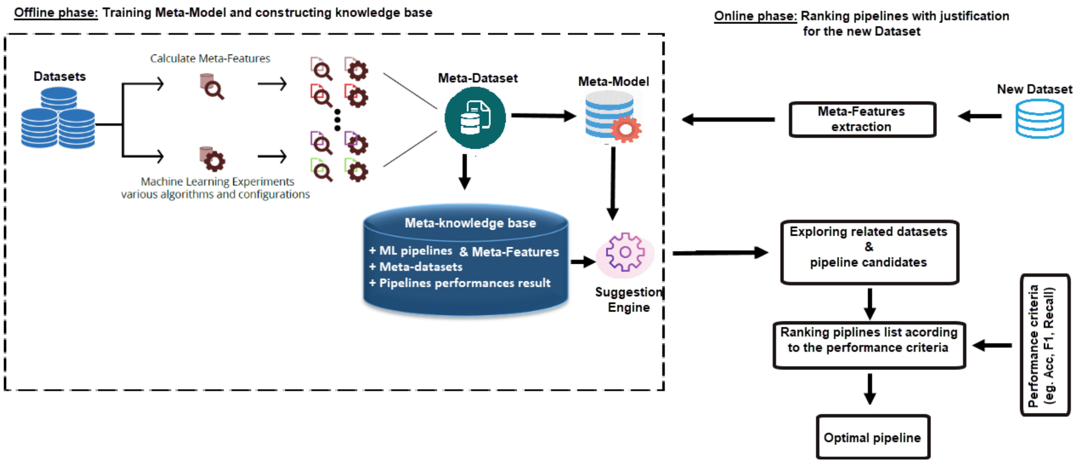

Figure 1: The workflow of the proposed framework.

of a classification algorithms and their related hyper-parameters configuration on any new dataset.

3.2 Recommendation Phase

In the recommendation phase, when a user wants to analyze a new dataset, she selects a predictive metric to be used for the analysis and then the system automatically recommends a machine-learning algorithm and its related hyper-parameters configuration to be applied such that the predictive performance is the first-rate. In order to do that, the system first extracts the dataset meta-features through the meta-features executor module. Then, the extracted meta-features are fed to the meta-model to provide the candidate pipelines. Finally, the suggestion engine, according to the meta-knowledge base, rank the pipelines in respect to the provided metric. The recommendation phase algorithm is listed as follows:

Algorithm 1: The recommendation phase algorithm.

Input:
D = The Dataset
M = Predictive metric
Output:
P<$P_1, P_2, P_3, \ldots, P_n$> // The suggested pipelines
Method:
Begin
// characterize the new dataset D
$Y_D \xleftarrow{\text{(Metafeatures extraction)}}$ Metafeatures(D)
//select nearest neighbor from the
// meta-dataset
Neighbors $\leftarrow$ $min(distance(Y_D, Y_I)_{i=1}^{m}$
// Y_i (actually 42) is the vector of the
// m as the size (actually 200)
// of training dataset metafeatures
//characterize the candidate neighbors
FT $\leftarrow$ Neighbors Metadata
//Suggested pipelines ranking of
//the new dataset D
//based on the provided performance criteria
Γ = Suggestion_Engine ($P_1, P_2, P_3, \ldots, P_n$)
End

4 IMPLEMENTATION ARCHITECTURE

In this section, we discuss the implementation of the proposed approach into a prototype solution. As we already know, the principal phases of the proposed frame work are *Learning phase* and *Recommendation phase*; these are implemented independently of each other. In the following, we give the detailed description for each of these phases.

4.1 Datasets

We collected 200 real-world manufacturing classification datasets. These have been collected from the popular UCI[1], OpenML[2], Kaggle[3] repositories along with some other real-world scenarios which were used in the learning phase. These datasets represent a mix of binary (54%) and multi-class (46%) classification tasks.

Although not limited to the problems in machines level, the data set includes many manufacturing classification problems, including tasks such as predictive maintenance, anomaly detection, shop floor applications, among others.

[1]https://archive.ics.uci.edu
[2]https://www.openml.org
[3]https://www.kaggle.com

4.2 Meta-learners

In the context of this work, we used 8 popular ML algorithms from *scikit-learn*, a widely used ML library implemented in Python. Each algorithm and its related hyper-parameters are described in Table 1.

4.3 Meta-features

The used meta-features should well describe the datasets to create an effective meta-model able to recommend the most suitable pipeline with a high precision. The features should be good predictors of the relative performance of algorithms. Several categories of meta-features have been developed. These range from simple features such as the number of instances in a dataset to more complex ones. Most meta-features belong to one of the following categories (Vilalta et al., 2004):

4.3.1 Simple, Statistical and Information-theoretic

Simple features can be rapidly extracted such as the number of instances, attributes and classes in the dataset. To some extent, they are designed to measure the complexity of the underlying problem. *Statistical and Information-theoretic features* are designed to describe the numerical properties of data distribution in a dataset sample and informations about the numeric features such as the class entropy, mean skewness of attributes.

4.3.2 Landmarking

It characterizes the extent of datasets when basic machine learning algorithms are performed on them. Some of the examples include the performance of a Decision Trees (DT), Gaussian Naive Bayes (GNB) or Linear Discriminant Analysis landmarker (LDA).

4.4 Meta-model

We used the k-Nearest Neighbor (k-NN) algorithm to induce meta-model able to predict top performer pipeline. It is often used in recommendation systems based on meta-learning (Laadan et al., 2019; Dyrmishi et al., 2019). After identifying the closest neighbors of the dataset using a distance metric such as "Euclidean Distance", a weighted average of each individual neighbor's actual ranking is used for computing the candidate dataset's predicted ranking of modeling algorithms based on the relevant metric. Thus, when the meta-learning system is applied to a new dataset, the suggestion engine returns a list of the most suitable pipelines, based on the meta-feature values extracted from the dataset and the evaluation metric.

5 EVALUATION

We performed an experimental study to evaluate the performance that can be achieved by using the proposed approach on various manufacturing related problems. After specifying the experimental environment, we evaluate the systems ability to predict the ML algorithms with its hyper-parameters configuration that shall provide the best result of the analysis.

We benchmark on a highly varied selection of 20 more curated datasets to ensure meaningful evaluation. It covers binary and multi-class classification problems from different industry 4.0 levels. These data are gathered from state of the art papers dealing with industry 4.0 related problems using machine-learning solutions as described in the Table 2. It is important to note that, the selected data sets were never exploited by any learning method on the offline phase.

The proposed system exploits the meta-model to predict all the pipelines in the meta-knowledge base with respect to the analyzed dataset and then returns its top-ranked pipelines according to the provided performance criteria. These pipelines are then fitted on the datasets train set and evaluated on the test set using the 70% / 30% splitting ratio.

The performance of the proposed system is comparable to the results of the datasets treated by the related papers. Lets us explain this with the help of data in Table 3. The first column of the table cites the original papers from which we borrow the datasets for testing and comparison purpose. The second column gives the recommended ML configurations (ML algorithm and its related hyper-parameters configurations) results generated by the proposed model on same datasets. The third column indicates the achieved accuracy of each dataset on the related original paper. The fourth column shows the evaluation results of applying the default configuration on the recommended ML algorithm. Evidently, as shown in Table 3, the obtained results are more accurate than the results from the related papers. It can be observed that some machine learning oriented manufacturing works could be improved simply through the use of a better ML algorithm configuration.

The illustrated results reveal the effectiveness of the AutoML solution in manufacturing data mining processes. The suggested model configurations exhibit better performance than the classic supervised learning techniques with default hyper-parameters

Table 1: ML algorithms and related hyperparameters tuned in the experiments.

ML algorithm	Hyper-parameters
Support Vector Classifier (SVC)	C: range (1e-10, 500) gamma: range (0.001, 1.01) kernel: ['poly', 'rbf'] degree: [2, 3] coef0: range (0., 10.)
AdaBoost (AB)	max_depth: range(1, 11) algorithm: ['SAMME', 'SAMME.R'] n_estimators: range(50, 501) learning_rate: [0.01, 2]
Gradient Boosting (GB)	learning_rate: [0.01, 1] criterion': ['friedman_mse', 'mse'] n_estimators: range (50, 501) max_depth: range (1, 11) min_samples_split: range (2, 21) min_samples_leaf: range (1, 21) max_features: [0.1, 0.9]
Extra Trees (ET) & Random Forest (RF)	n_estimators:[100] bootstrap: [True, False] max_features: range (0.1, 0.9) min_samples_leaf: range (1, 21) min_samples_split: range (2, 21) criterion: ['entropy', 'gini']
Decision Tree (DT)	max_features: range (0.1, 0.9) min_samples_leaf: range (1, 21) min_samples_split: range (2, 21) criterion: ['entropy', 'gini']
Logistic Regression (LR)	C: range (1e-10, 10.) penalty: ['l2', 'l1'] fit_intercept: [True, False]
Stochastic Gradient Descent (SGD)	loss: ['hinge', 'log', 'modified_huber', 'squared_hinge', 'perceptron'] penalty:['l2', 'l1', 'elasticnet'] learning_rate:['constant', 'optimal', 'invscaling'] fit_intercept: [True, False] l1_ratio: range (0., 1.) eta0: range (0., 5.) power_t: range (0., 5.)

settings and the configurations by non-ML experts, as shown in Figure 2.

6 CONCLUSIONS

In this paper, we studied the effectiveness of automated machine-learning techniques for the selection and parametrization of ML for the problems more often related to manufacturing industry. The main objective of the current work has been focused towards the design of a decision support system in order to enable the non-expert practitioners and data engineers, prospectively in the domain of industry 4.0 to take maximum benefit of ML models. The proposed approach validates the automated selection of ML models and suggest the optimized hyper-parameters for their configurations. The contents of the paper briefly describe the potential use of automatic machine learning methods in the 4th industrial revolution field. The proposed approach eventually aims to improve the confidence level of industrial practitioners to specify the appropriate configuration in the ML tools as well as to improve the reliability generalizing the high-risk and dynamic manufacturing environment.

In the current work, we mainly focus on the clas-

Table 2: The sample list of datasets used in the evaluation.

Dataset	Number of Classes	Number of Instances	Task
(Mazumder et al.,)	4	959	Failure risk analysis of pipeline networks
(Benkedjouh et al., 2015)	2	61000	RUL prediction
(Saravanamurugan et al., 2017)	3	2000	Chatter prediction
(Costa and Nascimento, 2016)	2	60000	APS system failure prediction
(Baldi et al., 2014)	2	98050	high-energy physics data analyses
(Tian et al., 2015)	7	1941	Faults detection

Table 3: The comparative performance analysis of the proposed framework.

Dataset	Recommended configuration result	Original paper result	ML pipeline with default configuration
(Mazumder et al.,)	93.74	85	80.24
(Benkedjouh et al., 2015)	99.41	98.95	93.88
(Saravanamurugan et al., 2017)	97.06	95	86.12
(Costa and Nascimento, 2016)	99.10	92.56	92.34
(Baldi et al., 2014)	85.59	88	69.45
(Tian et al., 2015)	99.54	80.74	76.23

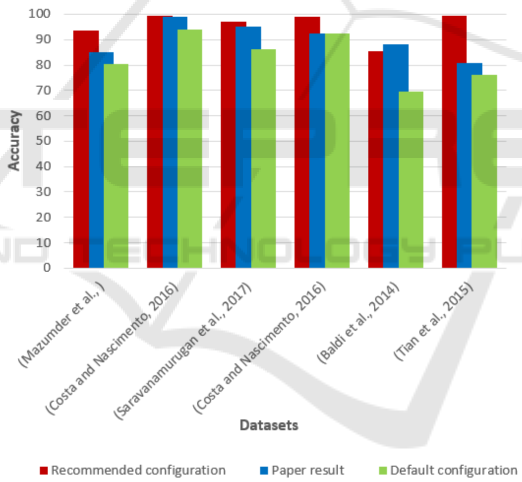

Figure 2: Comparative results of the effectiveness of AutoML over default classic ML configurations and domain expert (industrial researchers) configurations.

sification problems for the sake of clarity of the experimental evaluation. The results obtained from the proposed framework evidently exhibits improved performance of the automated selection of ML algorithms and optimization of hyper-parameters instead of the default values in manufacturing applications. The comparative analysis reveals in the majority of the cases that the recommended configurations yield better performances to enhance the utility of ML methods for the domain experts, notably the non-ML experts. It, thence, can be observed that AutoML paradigms and tools may help manufacturing practitioners—both neophyte and experts in the field of data analysis.

Moreover, we believe that such tools should not lack transparency while using the AI models; the performance of which, most of the times comes from the black box algorithms. It is essential to incorporate features that enable the interpretability and explainability of the produced results to a certain extent. We, hence, endeavor in the future works, the better understanding of the recommended configuration to gain more confidence on the rational of the obtained results. It shall assure the adoption of the proposed solution for the real-time systems even in critical situations.

ACKNOWLEDGEMENTS

This work has been supported, in part, by Hestim, CNRST Morocco, and the Université du Littoral Côte d'Opale, Calais France.

REFERENCES

Baldi, P., Sadowski, P., and Whiteson, D. (2014). Searching for exotic particles in high-energy physics with deep learning. *Nature communications*, 5(1):1–9.

Benkedjouh, T., Medjaher, K., Zerhouni, N., and Rechak, S. (2015). Health assessment and life prediction of cutting tools based on support vector regression. *Journal of Intelligent Manufacturing*, 26(2):213–223.

Bilalli, B., Abelló, A., Aluja-Banet, T., Munir, R. F., and Wrembel, R. (2018). Presistant: data pre-processing assistant. In *International Conference on Advanced Information Systems Engineering*, pages 57–65. Springer.

Bilalli, B., Abelló, A., Aluja-Banet, T., and Wrembel, R. (2016). Automated data pre-processing via meta-learning. In *International Conference on Model and Data Engineering*, pages 194–208. Springer.

Bohanec, M., Borštnar, M. K., and Robnik-Šikonja, M. (2017). Explaining machine learning models in sales predictions. *Expert Systems with Applications*, 71:416–428.

Brazdil, P., Carrier, C. G., Soares, C., and Vilalta, R. (2008). *Metalearning: Applications to data mining*. Springer Science & Business Media.

Cohen-Shapira, N., Rokach, L., Shapira, B., Katz, G., and Vainshtein, R. (2019). Autogrd: Model recommendation through graphical dataset representation. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pages 821–830.

Costa, C. F. and Nascimento, M. A. (2016). Ida 2016 industrial challenge: Using machine learning for predicting failures. In *International Symposium on Intelligent Data Analysis*, pages 381–386. Springer.

De, T., Giri, P., Mevawala, A., Nemani, R., and Deo, A. (2020). Explainable ai: A hybrid approach to generate human-interpretable explanation for deep learning prediction. *Procedia Computer Science*, 168:40–48.

Dyrmishi, S., Elshawi, R., and Sakr, S. (2019). A decision support framework for automl systems: A meta-learning approach. In *2019 International Conference on Data Mining Workshops (ICDMW)*, pages 97–106. IEEE.

Feurer, M., Klein, A., Eggensperger, K., Springenberg, J. T., Blum, M., and Hutter, F. (2019). Auto-sklearn: efficient and robust automated machine learning. In *Automated Machine Learning*, pages 113–134. Springer, Cham.

Laadan, D., Vainshtein, R., Curiel, Y., Katz, G., and Rokach, L. (2019). Rankml: a meta learning-based approach for pre-ranking machine learning pipelines. *arXiv preprint arXiv:1911.00108*.

Lemke, C., Budka, M., and Gabrys, B. (2015). Metalearning: a survey of trends and technologies. *Artificial intelligence review*, 44(1):117–130.

Mazumder, R. K., Salman, A. M., and Li, Y. Failure risk analysis of pipelines using data-driven machine learning algorithms. *Structural Safety*, 89:102047.

Nargesian, F., Samulowitz, H., Khurana, U., Khalil, E. B., and Turaga, D. S. (2017). Learning feature engineering for classification. In *IJCAI*, pages 2529–2535.

Samek, W. and Müller, K.-R. (2019). Towards explainable artificial intelligence. In *Explainable AI: interpreting, explaining and visualizing deep learning*, pages 5–22. Springer.

Saravanamurugan, S., Thiyagu, S., Sakthivel, N., and Nair, B. B. (2017). Chatter prediction in boring process using machine learning technique. *International Journal of Manufacturing Research*, 12(4):405–422.

Shin, D. The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable ai. *International Journal of Human-Computer Studies*, 146:102551.

Tao, F., Qi, Q., Liu, A., and Kusiak, A. (2018). Data-driven smart manufacturing. *Journal of Manufacturing Systems*, 48:157–169.

Thoben, K.-D., Wiesner, S., and Wuest, T. (2017). "industrie 4.0" and smart manufacturing-a review of research issues and application examples. *International journal of automation technology*, 11(1):4–16.

Tian, Y., Fu, M., and Wu, F. (2015). Steel plates fault diagnosis on the basis of support vector machines. *Neurocomputing*, 151:296–303.

Vilalta, R., Giraud-Carrier, C. G., Brazdil, P., and Soares, C. (2004). Using meta-learning to support data mining. *IJCSA*, 1(1):31–45.

Wang, J., Ma, Y., Zhang, L., Gao, R. X., and Wu, D. (2018). Deep learning for smart manufacturing: Methods and applications. *Journal of Manufacturing Systems*, 48:144–156.

Wolf, H., Lorenz, R., Kraus, M., Feuerriegel, S., and Netland, T. H. (2019). Bringing advanced analytics to manufacturing: A systematic mapping. In *IFIP International Conference on Advances in Production Management Systems*, pages 333–340. Springer.

Wuest, T., Weimer, D., Irgens, C., and Thoben, K.-D. (2016). Machine learning in manufacturing: advantages, challenges, and applications. *Production and Manufacturing Research*, 4(1):23–45.

Zacarias, A. G. V., Reimann, P., and Mitschang, B. (2018). A framework to guide the selection and configuration of machine-learning-based data analytics solutions in manufacturing. *Procedia CIRP*, 72:153–158.