Supervised Machine Learning for Knowledge-Based Analysis of Maintenance Impact on Profitability

Kai Schenkelberg* **Ulrich Seidenberg*** **Fazel Ansari****

* *Chair of Production and Logistics Management, University of Siegen, Unteres Schloss 3, 57072 Siegen, Germany (e-mail: kai.schenkelberg@uni-siegen.de, seidenberg@bwl.wiwi.uni-siegen.de).*
** *Research Group of Smart & Knowledge-Based Maintenance, Institute of Management Science, Vienna University of Technology (TU Wien), Theresianumgasse 27, 1040 Vienna, Austria (e-mail: fazel.ansari@tuwien.ac.at)*

Abstract: Recent empirical studies reveal that predictive maintenance is essential for accomplishing business objectives of manufacturing enterprises. Knowledge-based maintenance strategies for optimal operation of industrial machines and physical assets reasonably require explaining and predicting long term economic impacts, based on exploring historical data. This paper examines how supervised machine learning (ML) techniques may enhance anticipating the economic impact of maintenance on profitability (IMP). Planning and monitoring of maintenance activities supported by various statistical learning and supervised ML algorithms have been investigated in the literature of production management. However, data-driven prediction of IMP has not been largely addressed. A novel data-driven framework is proposed comprising cause-and-effect dependencies between maintenance and profitability, which constructs a set of appropriate features as independent variables.

Keywords: Maintenance; Profitability; Supervised learning; Machine learning; Regression; Knowledge-Based Maintenance

1. INTRODUCTION

In the context of the fourth industrial revolution and Industrial Internet of Things (IIoT), Predictive Maintenance (PdM) plays an essential role due to the parallel advances in the field of sensing technologies, intelligent connectivity and data science (IoT Analytics 2019). It is expected, that PdM will save around \$188B of maintenance costs from companies worldwide in 2024, which can in turn lead to an improvement in profitability, measured by Return on Investment (ROI) (IoT Analytics 2019).
Focusing on German manufacturing industry, the majority of 153 companies deal with PdM intensively (81%) (Feldmann et al. 2017). However, gains in performance, as a result of higher availablity, are seen as the main benefit of PdM (79%), while only under one fifth of the respondents view maintenance as an enabler for cost reduction (Feldmann et al. 2017).
In literature, it turned out, that PdM approaches mainly aim at anomaly detection and prediction of upcoming machine failures (Ansari et al. 2019). On the other hand, knowledge deficiencies are the root of failures (Ansari et al. 2019). Therefore, it is nessessary to develop data-driven maintenance strategies based on the concept of Knowledge-Based Maintenance (KBM), where economic aspects of maintenance strategy decisions are taken into account (Pawellek 2016, Ansari et al. 2019, Ansari and Glawar 2019). The economic Impact of Maintenance on Profitability (IMP) has been analyzed in literature (see e.g. Rishel and Canel (2006)) by simulating the impact of variations in maintenance policy on profitability. There is a lack of applying suitable data-driven methods, which enable maintenance management to improve maintenance planning and monitoring, considering economic indicators such as profitability.
The paper aims to examine statistical learning and supervised ML algorithms towards IMP prediction. Considering the prediction of profitability as a regression task, regression models namely Linear, Ridge, Lasso and MARS regression, Regression Tree, Random Forest, k-Nearest-Neighbor and Gradient Boosting Machine are investigated. The rest of the paper paper is structured as follows: Definitions and basic concepts of maintenance and supervised learning are discussed in Section 2. Section 3 gives a literature overview about the application of supervised learning in maintenance. Section 4 presents the results of a case study, where several suitable supervised Machine Learning (ML) algorithms are applied. In Section 5, results are aggregated and advantages as well as drawbacks of the previously mentioned approaches are discussed. Finally, in Section 6, the paper provides recommendations for further research considering technical risks mitigation measures such as data availability and data quality.

2. FUNDAMENTALS

2.1 Knowledge-Based Maintenance

Maintenance covers all technical, administrative and management actions during a life cycle of an object, i.e. retaining and restoring, so that it can fulfil its required function (DIN 13306 2019). Following DIN 13306 (2019), a maintenance strategy is defined as a management method, which is applied to accomplish maintenance objectives. According to Ansari and Glawar (2019), maintenance strategies and approaches can be separated into three groups, namely 1. Management strategies in the field of maintenance like Total Productive Maintenance or Reliability Centered Maintenance, which provide recommendations and standard procedures for goal setting as well as appropriate definition and implementation of maintenance activities, 2. Maintenance strategies without sensing and computing technologies, which are subdivided in a) Reactive (or Run-to-failure or corrective) Maintenance, b) Preventive and c) Proactive Maintenance, 3. Maintenance strategies with sensing and computing technologies, which can be broken down further as follows: a) Condition Based Maintenance, b) Predictive Maintenance and c) Prescriptive Maintenance.

Knowledge-based Maintenance (KBM) (Sturm 2001, Reiner et al. 2005, Pawellek 2016, Ansari et al. 2019) is a system-oriented and holistic maintenance management concept, that identifies critical elements and examines measures with regard to their potential effect on results (Pawellek 2016). Furthermore, because maintenance strategy decisions have a major impact on overall maintenance costs, long-term economic effects are taken into consideration (Pawellek 2016). KBM comprises three interconnected areas, from which data is collected: 1. Maintenance Management, 2. Plant condition and 3. Economic consequences. After transmission to an appropriate management concept in each case, namely 1. risk-based maintenance, 2. condition as well as time-based maintenance and 3. Total Productive and Lean maintenance, they can be summarised to an overall knowledge-based strategy (Pawellek 2016).

KBM can be defined "as a functional unit responsible to i) continuously support value generation and ii) facilitate developing and protecting maintenance collective knowledge across smart factories, which is enhanced by need- or opportunity-driven knowledge detection, discovery, modelling and representation approaches" (Ansari et al. 2019). KBM can be classified into four approaches based on complexity and maturity level: 1. Descriptive maintenance (What happened?), 2. Diagnostic maintenance (Why did it happen?), 3. Predictive Maintenance (What will happen, when?), 4. Prescriptive Maintenance (How should it happen?). By connecting 3. and 4. with a feedback loop, synergies of predicting future events and giving recommendations for improving upcoming maintenance processes can be created.

2.2 Supervised Learning

Statistical learning theory was first introduced in the late 1960's (Vapnik 1999). A *(supervised) learning model* consists of 1.) a generator of random vectors $X \in \mathbb{R}^p$, also called inputs, predictors or features, 2.) a supervisor, which returns an output vector (or response) $Y \in \mathbb{R}$ for an input vector X, 3.) a learning machine which implements a set of functions $\{f(X)\} \in \Lambda$. The *learning problem* is to choose a function from Λ which predicts Y in the best possible way based on a training dataset with l independent identically distributed (i.i.d.) observations $(X_1, Y_1), \ldots, (X_l, Y_l)$, drawn from a joint probability distribution $P(X, Y)$ (Vapnik 1999). This can be described by a statistical model including an error term ϵ with mean zero (James et al. 2013, Russell et al. 2016, Hastie et al. 2017):

$$Y = f(X) + \epsilon \tag{1}$$

If Y is quantitative, the learning problem is called *regression* and *classification* if Y is qualitative (Hastie et al. 2017). The best possible way to choose a function can be mathematically expressed by minimization of the expected value of the loss (Risk)

$$R = \int L(Y, f(X))dP(X, Y) \tag{2}$$

with a loss function $L(Y, f(X))$, measuring the loss or discrepancy between Y and $f(X)$. In order to minimize the risk with an unknown joint probability distribution, the empirical risk

$$R_{emp} = \frac{1}{l}\sum_{i=1}^{l} L(Y_i, f(X_i)) \tag{3}$$

will be minimized instead (Vapnik 1999).

In literature, several supervised learning methods with respect to regression learning problems exist, so that only a few of them will be described briefly in the following section. In order to cover a wide range, the following algorithms have been selected: 1. Linear, Lasso and Ridge Regression as linear learner, 2. MARS as a non-linear learner, 3. Regression Trees (RT) as a tree-based learner, 4. Random Forest (RF), Gradient Boosting Machine (GBM) and Stacking as ensemble learner and 5. k Nearest Neighbor (kNN)

A *Linear Regression (LR)* model (Hastie et al. 2017) assumes that a linear relationship between the variables is a reasonable approximation for $f(X)$. It has the form

$$f(X) = \beta_0 + \sum_{i=1}^{p} X_i * \beta_i \tag{4}$$

with coefficients $\beta_j \in \mathbb{R}, j = 0, \ldots, p$. For estimating the unknown parameter $\beta = (\beta_0, \ldots, \beta_p)^T$, the least squares method, which minimizes the Residual Sum of Squares (RSS), is the most common approach.

Ridge regression (Hoerl and Kennard 2000) is based on RSS minimization of the linear model and appends a term $\lambda * \sum_{i=1}^{p} \beta_i^2$ to the target function, containing a complexity parameter $\lambda \geq 0$, which influences the amount of shrinkage of the regression coefficients by penalizing the sum-of-squares of β.

LASSO (Least Absolute Shrinkage and Selection Operator) regression (Tibshirani 1996) is also a shrinkage method, which is similar to ridge regression, where the L2-norm of the penalty term is replaced by the L1-norm.

The *MARS (Multivariate Adaptive Regression Splines)*-model (Friedman 1991) can be desscribed by a weighted sum of M basis functions $B_m(x)$: $\hat{f}(x) = \sum_{m=1}^{M} \alpha_m *$

$B_m(x)$. Each basis function can either be a constant, a hinge function or a product of multiple hinge functions.
kNN (Fix and Hodges 1989) is a non-parametric fundamental method for classification, but can also be applied to regression problems. In the latter case, it averages over the responses of the k closest neighbors.
Ensemble methods (or ensemble learning) (Dietterich 2000) are learning algorithms, which help at predicting new examples by creating an ensemble of predictors, which are combined in some way. The most common approaches are 1. Bagging (bootstrap aggregation) (Breiman 1996a) combining predictors, each of them is generated from a data subset, sampled with replacement, 2. Boosting (Schapire 1990) as an iterative approach, where several weak learners are generated iteratively and combined to a final prediction, 3. Stacking (or stacked regression) (Wolpert 1992, Breiman 1996b) combines predictions of several learners and uses them as an input for an ensemble learner in a higher space.
In case of a classification task, a *Random Forest (RF)* (Breiman 2001), as a bagging method, consists of a collection of tree-based classifiers, where each tree generats a unit vote for the most popular class. Random Forests for regression have tree predictors with numerical output and the final predictor averages over all predictions of the trees.
A *GBM* (Friedman 2001) is an iterative boosting method, where in each step a tree (base learner) is trained based on pseudo-residuals as responses and added to the final prediction.

3. REGRESSION-BASED PREDICTIVE MAINTENANCE: A BRIEF LITERATURE REVIEW

The following literatur review has been conducted on two different scientific databases, namely IEEE Digital Library and ScienceDirect. Only papers with a regression task and at least one of the aforementioned approaches (see chapter 2.2) have been taken into account. The following keyword string was formulated and used:
("predictive maintenance" OR "maintenance") AND ("machine learning" OR "supervised learning") AND "regression".
Due to a large body of existing approaches, only publications from 2013 to 2019 have been considered and the most remarkable examples will be described.
Mathew et al. (2017) predict the Remaining Useful Lifetime (RUL) of turbo fan engines on four datasets. Each engine has different sensor values. They compare the Root Mean Squared Error (RMSE) of several supervised learning methods like a Decision Tree, Support Vector Machine, Random Forest, kNN, K Means, Gradient Boosting Machine, Adaboost, Deep Learning and Anova. As a result, Random Forest generates the smallest error.
In a case study based on vibration monitoring data, Amihai et al. (2018) use the Random Forest algorithm to predict KCIs (Key Condition Indices), which indicate the severity level of the observed failure mode. Compared to a standard persistence technique, the error of random forests, measured by RMSE, is significantly lower.
In order to predict the RUL of aero-propulsion engines, Hsu and Jiang (2018) compared recurrent neural networks with multi-layer perceptron, support vector regression, relevance vector machine and convolutional neural network on the NASA C-MAPSS data set. It turned out that the first mentioned approach has the lowest RMSE.
Susto and Beghi (2016) investigate time-series maintenance data for predicting the time before a failure occurs. Several feature extraction techniques, namely Supervised Aggregative Feature Extraction (SAFE), Statistical Moments (SM) and Median Values (VM), have been applied before performing Ridge Regression on an industrial dataset. As a result, SAFE outperformed the other approaches.
For estimating fuel cell duration time, features have been extracted from both real and imaginary parts of the impedance spectrum in Onanena et al. (2009). Finally, a linear regression model, which uses different subsets of extracted features, has been trained and for considering features from both real and imaginary part, the mean error was the lowest.
Orozco et al. (2018) present a model for diagnostics of wind turbine gearbox failures. Linear Regression, Multivariate Polynomial Regression, Random Forest and Neural Network have been evaluated by three different metrics (RMSE, Pearson and Shapiro-Wilk normality), whereby the first two approaches performed best.
Schlechtingen et al. (2013) compare Cluster Center Fuzzy logic, Neural Network, k Nearest Neighbor and Adaptive Neuro-Fuzzy-Inference System (ANFIS) model for wind turbine power curve monitoring. When enhancing the model, which is commonly used in literature, with two additional features (ambient temperature and wind direction), an earlier detection of abnormal turbine performance is possible. Considering the metrics, namely RMSE, Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Standard Deviation (SD), the differences between the models are small.
A Random forest was applied in Wu et al. (2016) to predict tool wear in dry milling operations. Experimental results showed that random forest is able to predict very accurate. Additionally, a parallel random forest algorithm was developed, which allows accleration in computation.

It can be concluded, that in the area of predictive maintenance, RUL estimation is a central problem discussed in the literature, inter alia, by focusing on various use-cases not limited to the manufacturing sector. With restriction to regression learning problems, several supervised learning methods have been applied to predict the time span before a failure occurs.

4. METHODOLOGY

The methodology of the present study is based on the CRISP-DM (Cross Industry Standard Process for Data Mining)-model (Chapman et al. 2000) consisting of six steps: 1. Business understanding, 2. Data Understanding, 3. Data Preparation, 4. Modeling, 5. Evaluation, 6. Deployment. The following three sections implement step 1 to 5.

4.1 Business and data understanding

In an industrial company, historical data of a production machine were collected over time. The dataset contains about 1500 records from a three-shift operation over a period of two years about machine and failure states as well as number of failures and production volume. In order to translate the general content-related problem of analyzing IMP into a data mining problem, it is essential to identify dependent and independent variables. Because the dataset does not comprise any information about economic indicators, it is necessary to make some reasonable assumptions. In the present case, unplanned maintenance costs are opportunity cost, which can be calculated by the product of production volume, failure duration and a unit contribution margin. Planned maintenance costs are assumed to be positvely linear dependent on availability. Profit is defined as the difference between revenue and total costs, whereby the former is computed by the product of production volume and a unit price. The latter is limited to planned, unplanned maintenance and production costs.

4.2 Data preparation

In order to create a final dataset from raw data for applying supervised learning methods, R (R Core Team 2019) and R Studio (R Studio Inc. 2019) have been used. As a first step, influencing factors on profitability should be identified as independent variables. With failure time, time without failure and number of failures, Mean Time Between Failures (MTBF) and Mean Time To Repair (MTTR) are calculated in order to compute (inherent) Availability for each period. Afterwards, with the above-mentioned assumptions, revenue, depending on Production Volume (PV), as well as planned and unplanned maintenance costs are determined with the ultimate goal of estimating profit. Table 1 gives an overview about variable selection for further data analysis, whereby profit represents the dependent variable and the others are to be understood as independent variables.

Table 1. Variables for supervised ML in order to analyze IMP

Name	Unit
Availability	Percent (%)
PV	Quantity Units (QU)
Profit	Monetary Units (MU)

4.3 Modeling and Evaluation

For model selection, nine different ML methods have been applied. 70 percent of the dataset has been selected randomly and used for training while the rest served for testing/validation. As Performance measures, MAE, RMSE and Normalized Root Mean Squared Error (NRMSE), as well as R squared, were computed with 10-fold cross validation. All models with hyperparameters have been tuned with 100 iterations. The results are aggregated in Table 2.

The optimal value of lambda for Ridge and LASSO regression was set to 774.26 and 2151.43. MARS regression was performed with a maximum degree of 9. A RT was constructed with a cp value of 0.001, minsplit of 5 and maxdepth of 9. For the RF, 1885 trees were selected. The results of GBM were achieved with the following parameters: 3181 trees, interaction depth of 1, minimum of 5 observations in a node, shrinkage of 0.459. The optimum number of neighbors for kNN was set to 5. A stacked learner has been generated with a kNN, a RF and a GBM model as inputs. For the ensemble learning model, MARS has been chosen. As a result, kNN as well as the stacked learner outperform the other approaches in all performance measures, followed by (boosted or bagged) tree learner and MARS. Linear (regularized) regression models have the largest error or rather the worst model fit in the present case. The coefficients of the linear models and MARS are displayed in Table 3 and 4. Taking Variance Inflation Factors (VIF) close to 1 and high *t-values* into account, no independent variable should be affected by multicollinearity.

Table 2. Performance comparison of applied ML algorithms

	MAE	RMSE	NRMSE	R^2
LR	97270.391	127099.39	42.5	0.8263759
Lasso	95942.773	126626.74	43.1	0.8261304
Ridge	90385.452	125221.69	45.6	0.8299676
Mars	18587.606	33647.89	11.0	0.9877231
RT	21132.963	29707.16	9.8	0.9904304
RF	9536.007	29787.36	10.3	0.9903786
GBM	7827.685	20128.47	6.6	0.9956067
KNN	5752.381	13750.83	4.6	0.9979496
Stacked	5614.890	12286.71	4.1	0.9983630

Table 3. Coefficients of linear (regularized) regression models

	LR	LASSO	Ridge
(Intercept)	-1539267.6099	-1076279	-1356604
Availability	2032988.0068	1802161	1792776
PV	16.1289	0	15.43235

Table 4. Coefficients of MARS regression model

	coefficients
(Intercept)	-23177.6
h(0.505777-Availability)	-5217328.1
h(Availability-0.505777)	1514617.2
h(Availability-0.769546)	-1126480.1
h(18687-PV)	-16.1
h(PV-18687)	28.2

The results can be interpreted as follows: If an independent variable is changed by one unit and the other variables remain fixed, profit will change by its weight. MARS regression selects 2 of 2 features. The coefficients of linear and MARS regression models enable some kind of what-if analysis. As an example for top-down (or predictive) inference, Chief Maintenance Officer (CMO) can set availability from a given to a desired value. This effects the profit directly by the corresponding weight. Performing bottom-up (or diagnostic) inference is only possible to a limited extent. In order to determine the needed value of availability for a fixed value of profit prescribed by CMO, the other independent variables have to be known.

Figure 1 shows the resulting RT, where top-down inference can be performed. CMO is also able to carry out bottom-up inference, provided that desired profit is represented by a leaf in the RT.

Model-agnostic methods (Ribeiro et al. 2016a) such as Partial Dependence Plots (PDP) (Friedman 2001), Local Interpretable Model-agnostic Explanations (LIME)

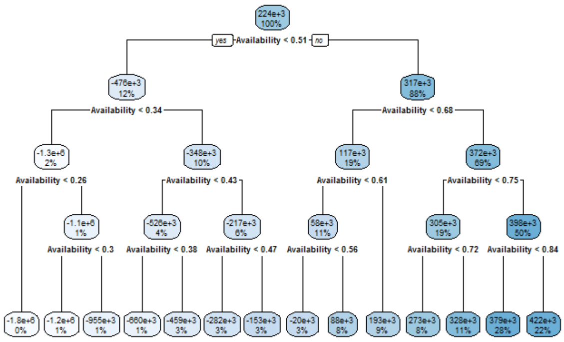

Fig. 1. Decision Tree for analyzing IMP

(Ribeiro et al. 2016b), and Feature (or Variable) Importance (Breiman 2001) provide an alternative to interpretable models as described before. PDP give an overview about the relationship (e.g. linear) between the output and one or two input variables. LIME aims at explaining a prediction by approximation with an interpretable model. Feature Importance, as a special case of Model Reliance (Fisher et al. 2019) in RF, measures the increase in prediction error, when an independent variable is permuted.
As an example of PDP, Figure 2 reveals the relationship between availability or production volume and profit, which is obviously not linear in both cases. Another important finding for CMO is, that increasing availability from approximately 80% will not increase profit significantly. Figure 3 gives CMO the opportunity to draw some inferences, for instance, maximization of availability and PV will not lead to a maximum of profit. Furthermore, considering Feature Importance as shown in Figure 4, availability is clearly the most important feature.

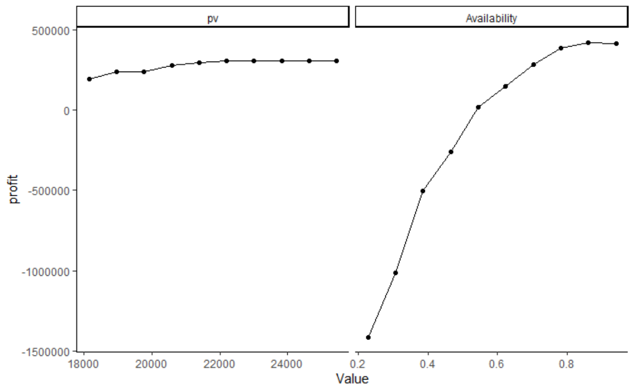

Fig. 2. PDP between profit and availability or PV

5. CONCLUSION AND DISCUSSION

A maintenance strategy should not only consider failure or remaining life prediction and anomaly detection or availability maximization on the operational level, but also take long-term economic effects like profitability into account. Based on a historical dataset, a learning model with profit as dependent variable has been constructed. Several learning algorithms were applied and compared. Supervised learning allows IMP prediction with relatively satisfactory results and can facilitate maintenance planning activities. With regard to performance, kNN and the

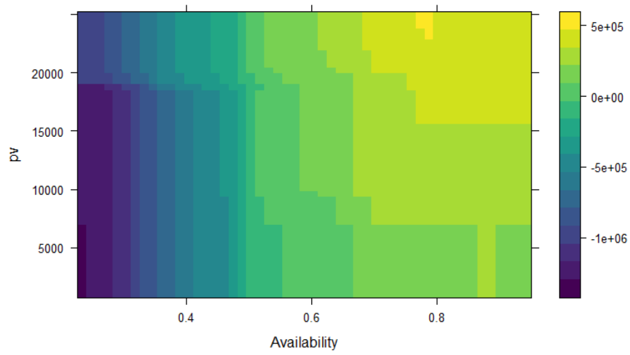

Fig. 3. Multi-Predictor PDP between profit, availability and PV

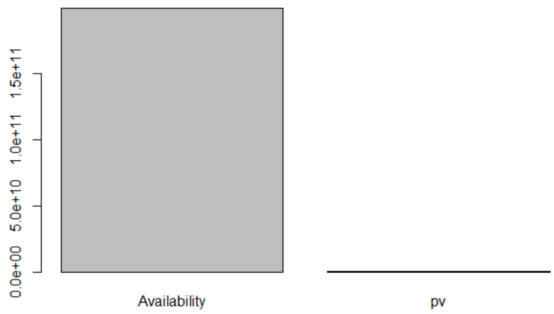

Fig. 4. Permutation Feature Importance of RF

stacked model had the lowest error. In consideration of the coefficients from the linear models as well as Feature Importance, availability has the strongest impact on profitability. While linear (regularized) models, MARS and RT allowed interpretation of the coefficients directly and enable what-if analysis, ensemble-based algorithms were not suitable for that purpose due to their nature. However, PDP as an example of model-independent approaches enable CMO to carry out inference in both directions as long as there are no more than two independent variables. In Schenkelberg et al. (2020), a Dynamic Bayesian Network (DBN) model for predicting IMP is presented, which enhances probability based prediction of profitability, in contrast to deterministic and regression based approaches and reinforces semantic learning.
Table 5 gives an overview about applied supervised learning algorithms as well as probabilistic graphical models for analyzing IMP with respect to parameter reaction (deterministic or stochastic) and capability of what-if analysis. A "✓" symbolizes, that inference can be performed without constraints (any number of evidence variables at the same time), "(✓)" stands for inference with constraints and "-" means, that inference is not possible. For the former, the capability of top-down inference is depending on the model or the number of independent variables. Bottom-up inference is only possible, when the values of all other variables are known. The latter is able to perform predictive (top-down) and diagnostic (bottom-up) inferences with single or multiple evidence setting. In addition, a DBN can be generated based on a short period of time with a small

dataset, wheras supervised learning methods need a lot of data points.

Table 5. Comparison of Supervised Learning (SL) and Probabilistic Graphical Models (PGM)

	Reaction of parameters	top-down inference	bottom-up inference
SL	deterministic	(✓)	(✓)
PGM	stochastic	✓	✓

6. OUTLOOK

Future research should focus on the following aspects:

(1) The performance of supervised ML is strongly dependent on data availability and quality. Therefore, for further validation, the methodology should be applied on other use case scenarios, where some additional information about maintenance policies as well as economic indicators like profitability is available.
(2) In addition, because only structured data has been taken into account, the supervised ML model should be extended in order to integrate multimodal data. As an example, it should be investigated, whether and how unstructured data like maintenance records can be analyzed, transformed into structured data and finally added to the learning model as independent variables (c.f. Ansari et al. (2014)). This could lead to more accurate prediction results, i.e. decrease the prediction error.
(3) Furthermore, in this paper, the application of Artificial Neural Networks (ANN) for analyzing IMP has been intentionally neglected, due to lack of big data. Especially, a hybrid approach of ANN and DBN could be investigated on large datasets.
(4) Another approach, which could be taken into account, is a simulation-based model for analyzing IMP. For each independent variable, random data should be generated in order to observe IMP. Simulation could overcome the limitations of supervised ML, as discussed in the previous section and in the aforementioned issues, to some extent. Unlike supervised ML, it is independent of data availability and quality problems as well as choosing the right learning algorithm. Additionally, it enables construction of other variables like maintenance policy and top-down inference could be performed.
(5) Finally, for comparing several statistical, ML and simulation-based models with respect to application focus on KBM, the development of suitable criteria is required.

REFERENCES

Amihai, I., Gitzel, R., Kotriwala, A.M., Pareschi, D., Subbiah, S., and Sosale, G. (2018). An industrial case study using vibration data and machine learning to predict asset health. In E. Proper, S. Strecker, and C. Huemer (eds.), *2018 20th IEEE International Conference on Business Informatics*, 178–185. IEEE, Piscataway, NJ.

Ansari, F. and Glawar, R. (2019). Knowledge-based maintenance. In K. Matyas (ed.), *Instandhaltungslogistik*, Praxisreihe Qualitätswissen, 318–342. Hanser, München.

Ansari, F., Glawar, R., and Nemeth, T. (2019). Prima: a prescriptive maintenance model for cyber-physical production systems. *International Journal of Computer Integrated Manufacturing*, 32(4-5), 482–503.

Ansari, F., Uhr, P., and Fathi, M. (2014). Textual meta-analysis of maintenance management's knowledge assets. *International Journal of Services, Economics and Management*, 6(1), 14.

Breiman, L. (1996a). Bagging predictors. *Machine Learning*, 24(2), 123–140.

Breiman, L. (1996b). Stacked regressions. *Machine Learning*, 24(1), 49–64.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinart, T., Sheare, C. Russell H.r, and Wirth, R. (2000). Crisp-dm 1.0: Step-by-step data mining guide. In CRISP-DM Consortium (ed.), *Crisp-Dm 1.0*.

Dietterich, T.G. (2000). Ensemble methods in machine learning. In *Proceedings of the First International Workshop on Multiple Classifier Systems*, MCS '00, 1–15. Springer-Verlag, London, UK, UK.

DIN 13306 (2019). Maintenance - maintenance terminology.

Feldmann, S., Lässig, R., Herweg, O., Rauen, H., and Synek, P.M. (2017). Predictive maintenance - serving tomorrow - and where we are really today.

Fisher, A., Rudin, C., and Dominici, F. (2019). All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *Journal of Machine Learning Research 20 (177)*.

Fix, E. and Hodges, J.L. (1989). Discriminatory analysis. nonparametric discrimination: Consistency properties. *International Statistical Review / Revue Internationale de Statistique*, 57(3), 238.

Friedman, J.H. (1991). Multivariate adaptive regression splines. *Ann. Statist.*, 19(1), 1–67.

Friedman, J.H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189–1232.

Hastie, T., Tibshirani, R., and Friedman, J.H. (2017). *The elements of statistical learning: Data mining, inference, and prediction.* Springer series in statistics. Springer, New York, NY, second edition, corrected at 12th printing 2017 edition.

Hoerl, A.E. and Kennard, R.W. (2000). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 42(1), 80.

Hsu, C.S. and Jiang, J.R. (2018). Remaining useful life estimation using long short-term memory deep learning. In A.D.K.T. Lam, T.H. Meen, and S.D. Prior (eds.), *Applied system innovation for modern technology*, 58–61. IEEE, Piscataway, NJ.

IoT Analytics (2019). Predictive maintenance report 2019-2024.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning: With Applications in R*, volume 103 of *Springer Texts in*

Statistics. Springer, New York, NY.

Mathew, V., Toby, T., Singh, V., Rao, B.M., and Kumar, M.G. (2017). Prediction of remaining useful lifetime (rul) of turbofan engine using machine learning. In *IEEE International Conference on Circuits and Systems Conference, ICCS 2017*, 306–311. IEEE, Piscataway, NJ.

Onanena, R., Faicel Chamroukhi, Latifa Oukhellou, Denis Candusso, Patrice Aknin, and Daniel Hissel (2009). Supervised learning of a regression model based on latent process. application to the estimation of fuel cell life time. In M.A. Wani (ed.), *International Conference on Machine Learning and Applications, 2009*, 632–637. IEEE, Piscataway, NJ.

Orozco, R., Sheng, S., and Phillips, C. (2018). Diagnostic models for wind turbine gearbox components using scada time series data. In *2018 IEEE International Conference on Prognostics and Health Management (ICPHM)*, 1–9. IEEE, Piscataway, NJ.

Pawellek, G. (2016). *Integrierte Instandhaltung und Ersatzteillogistik: Vorgehensweisen, Methoden, Tools*. VDI-Buch. Springer Vieweg, Berlin and Heidelberg, 2. auflage edition.

R Core Team (2019). R. URL `https://cran.r-project.org/`.

R Studio Inc. (2019). R studio. URL `https://www.rstudio.com`.

Reiner, J., Koch, J., Krebs, I., Schnabel, S., and Siech, T. (2005). Knowledge management issues for maintenance of automated production systems. In G. Zülch, H.S. Jagdev, and P. Stock (eds.), *Integrating Human Aspects in Production Management*, volume 160 of *IFIP International Conference for Information Processing*, 229–237. International Federation for Information Processing, Boston, MA.

Ribeiro, M.T., Singh, S., and Guestrin, C. (2016a). Model-agnostic interpretability of machine learning.

Ribeiro, M.T., Singh, S., and Guestrin, C. (2016b). "why should i trust you?": Explaining the predictions of any classifier.

Rishel, T.D. and Canel, C. (2006). Using a maintenance contribution model to predict the impact of maintenance on profitability. *Journal of Information and Optimization Sciences*, 27(1), 21–34.

Russell, S.J., Norvig, P., Davis, E., and Edwards, D. (2016). *Artificial intelligence: A modern approach*. Always learning. Pearson, Upper Saddle River, New Jersey, third edition, global edition edition.

Schapire, R.E. (1990). The strength of weak learnability. *Machine Learning*, 5(2), 197–227.

Schenkelberg, K., Seidenberg, U., and Ansari, F. (2020). Analyzing the impact of maintenance on profitability using dynamic bayesian networks: 13th cirp conference on intelligent computation in manufacturing engineering. In R. Teti and D.M. D'addona (eds.), *13th CIRP Conference on Intelligent Computation in Manufacturing Engineering*. Gulf of Naples, Italy.

Schlechtingen, M., Santos, I.F., and Achiche, S. (2013). Using data-mining approaches for wind turbine power curve monitoring: A comparative study. *IEEE Transactions on Sustainable Energy*, 4(3), 671–679.

Sturm, A. (2001). *Wissen basierte Betriebsführung und Instandhaltung*. VGB-PowerTech Service GmbH, Essen.

Susto, G.A. and Beghi, A. (2016). Dealing with time-series data in predictive maintenance problems. In *2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA)*, 1–4. IEEE, Piscataway, NJ.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), 267–288.

Vapnik, V.N. (1999). An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, 10(5), 988–999.

Wolpert, D.H. (1992). Stacked generalization. *Neural Networks*, 5(2), 241–259.

Wu, D., Jennings, C., Terpenny, J., and Kumara, S. (2016). Cloud-based machine learning for predictive analytics: Tool wear prediction in milling. In J. Joshi (ed.), *2016 IEEE International Conference on Big Data*, 2062–2069. IEEE, Piscataway, NJ.