On the Application of Machine Learning Techniques in Condition Monitoring Systems of Complex Machines

Marcin Hinz[a], Dominik Brueggemann[a], and Stefan Bracke[a]
[a] University of Wuppertal, Wuppertal, Germany

Abstract: The aim of this paper is to provide an overview about machine learning algorithms that can be applied for condition monitoring systems. Therefore, machine learning theory is discussed first with regard to the given goals. In accordance with the selected algorithms, the input data as well as the output data and the possibility of the prognosis of certain machine states are outlined with regard to the applicability of the mentioned algorithm types.
The presented theory is finally applied to a real usage case for a complex machine out of the process industry. Here, first the types of recorded data as well as the data interpretation are discussed in detail. Subsequently, the application of chosen machine learning algorithms and the results of the analysis are presented in order to show the possibilities of the condition monitoring system with regard to the increase of the machine availability and adaption of the usability.
All performed analysis steps are executed with a software package written for the purpose of this study. The software package is written in python including all analysis steps and graphical user interface and can be used as a framework for further condition monitoring purposes. Though, the data used for the publication studies is artificial but has a real case data character.

Keywords: Predictive Maintenance, Machine Learning, Inductive Learning, Condition Monitoring System, Process Engineering.

1. INTRODUCTION

Knowledge about the technical state of a complex machine is fundamental regarding the reliability and availability within the entire usage phase of a certain product. By knowing the exact condition state, it is possible to improve the maintenance, increase the operability of the system and at the end the satisfaction of the customer. Obviously, in case of safety critical systems, a precise condition state leads to increase of the system safety.

Efficient monitoring of components, products and systems has become a very challenging task for reliability engineers within all technical disciplines. In the very most cases, machines are equipped with plenty of sensors producing a high amount of data. The challenge is the proper collection and storage of the data but also its access and analysis. With regard to the big data problems, the classical statistical methods are not applicable anymore doe to too high effort put in the analysis process itself. Therefore, data mining techniques are needed for an efficient data interpretation and subsequently decision making.

Machine learning addresses the issue of recognizing patterns and other regularities of data with the aim to find general rules and to make predictions. Machine learning algorithms are available in a wide range and can be used for statistical analysis and data mining, especially in case of big data problems.

2. MACHINE LEARNING – FUNDAMENTALS

Machine learning is counted among the branches of artificial intelligence since it gives the machine the ability to learn from experience without being explicitly programmed [1]. Since it is a process of solving problems by analyzing data that is already present in a database while discovering patterns in that data, Machine learning is a field of data mining, according to [2].
As stated by [DB2] Machine learning can be divided into three major sub branches:

- Supervised learning: The underlying connection between input data and the upfront known target variable is established using training data.
- Unsupervised learning: Used on unlabeled datasets with no knowledge about the required target variables. Hidden structures within the data are revealed.
- Reinforcement learning: A sequence of actions gets explored by an intelligent agent that either rewards or punishes certain behavior. The aim is to maximize the cumulative gained reward.

For further information regarding classification of Machine learning algorithms refer to [1] or [2].

One of the most commonly used Machine learning algorithms within inductive learning is the C4.5. The C4.5 algorithm is a decision tree classifier that belongs to the class of supervised learning algorithms [1].
Classifiers in form of a decision tree are modeled within the algorithm using a divide-and-conquer method which works as follows:
If a set of instances is uniform with regard to the class it belongs to, or if the set is small enough, there is no need for further dividing. A labeled leaf node is created.
If the previous case does not occur, the algorithm selects a test based on the given attributes to split the set into multiple sub-branches. Here, the algorithm applies the information gain criterion, i.e., the gain criterion selects the test that maximizes the information gain according to the entropy out of the information theory. Note, that the information carried by a message is dependent on its probability and is defined as

$$I(S) = -\log_2\left(\frac{freq(C_j,S)}{|S|}\right) \tag{1}$$

where $freq(C_j, S)$ is the number of messages in S that belong to the class C_j.
This procedure continues with all non-labeled branches until no further splitting is possible or necessary. For further information on the C4.5 refer to [1, 3]

Another well-known algorithm is the NNge (Non Nested Generalized Exemplars). The aim of the procedure is to create a set of hyperrectangles by using training data. Those hyperrectangles represent generalized examples and aggregate multiple or single examples from the training data into a class. The rectangles are created following three steps:

- Classification - Find the hyperrectangle that is the closest to the example from the training data. The distance between the example and the hyperrectangle is defined by a modified Euclidean distance equation:
  The distance between a rectangle H and an example from the training data is defined by:

$$D_{EH} = W_H\sqrt{\sum_{i=1}^{n}\left(W_i\,\frac{E_i - H_i}{max_i - min_i}\right)^2} \tag{2}$$

  Here, W_H is a weight that's defined by the observed accuracy of the predictions of the rectangle H.
  W_i is a factor on the different features of the examples, weighing their importance.
  min_i and max_i define the range of values with regard to the *i*-th feature.
  Since H_i denotes an interval (width of the hyperrectangle with regard to the *i*-th feature) the difference $E_i - H_i$ is defined as the minimal difference between E_i and a value from the interval H_i.

- Model adjustment - Split the hyperrectangle if it covers a conflicting example, i.e., an example that does not belong in the class that is represented by the hyperrectangle.
- Generalization - Expand the hyperrectangle in order to cover the example that was associated to the hyperrectangle but only if it does not overlap or cover other hyperrectangles afterwards.

For further explanations as well as a deeper look into the mathematical background consult [4]

3. CONDITION MONITORING – STATE OF THE ART

„Modern maintenance is in a phase of change from pure maintenance prevention-prediction to a pro-active and operators-driven maintenance concept. It is simply not enough to prevent and predict failures, whether potential or functional. The aim is to build systems where maintenance and operators will join their know-how and actions to provide a proactive condition monitoring system where even the smallest correction will be undertaken on time to eliminate factors negatively affecting equipment condition even to the least extend." as stated in [5] gives a good summary of where the idea of maintenance in conjunction with condition monitoring was taken in the past years.
According to [5] some activities that were considered the job of researchers and experts are now considered a regular maintenance job. It has become an „ordinary maintenance activity" to check on the condition of a machines due to advanced testing tools and enhanced statistical analysis methods.
On the other hand, advanced testing tools and especially a majority of these lead to massive amounts of data, which is why big data is nowadays a well-known idea. There is no doubt, that big data are now rapidly expanding in all science and engineering domains. While the potential information that can be gathered from those data increased massively, the need for new ways of thinking and novel learning techniques emerges. [6] Thus, the interdisciplinary field of machine learning building upon ideas from many different kinds of fields such as artificial intelligence, optimization theory, information theory, statistics, cognitive science, optimal control, and many other disciplines of science, engineering, and mathematics grew rapidly in the past years.
While the classical fields of machine learning, i.e., supervised, unsupervised and reinforcement learning still form the basic framework, new advanced learning methods were needed to cope with the still growing amounts of data. Here is a selection of recent approaches of machine learning:

Representation Learning:
Representation Learning is used on high-dimensional datasets, which has become increasingly common nowadays. By finding useful representations of the data, it becomes easier to extract useful information when building classifiers or other predictors. Representation Learning has achieved impressive performance on many dimensionality reduction tasks which is why the scientific interest in this learning technique has increased rapidly. Furthermore, a series of empirical successes occurred simultaneously. Examples include remarkable successes in speech recognition, natural language processing and intelligent vehicle systems. [7-9]

Deep learning:
In contrast to most traditional learning techniques, which are considered using shallow-structured learning architectures, deep learning mainly uses supervised and/or unsupervised strategies in deep architectures. By that, more complicated, hierarchically launched statistical patterns may be captured. Due to that, it is highly adaptive to new areas.
Examples for deep learning approaches are Deep belief networks (DBNs) [10, 11] and convolutional neural networks (CNNs) [12, 13] which have both been proposed over the past decade. They have been well established and have shown great promise for future work.

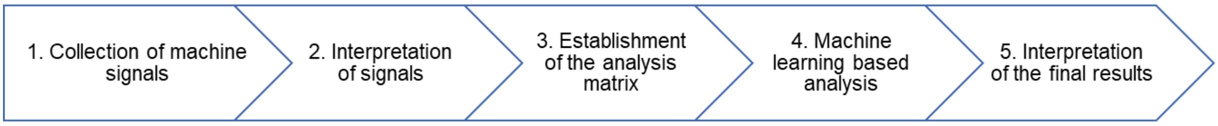
Transfer learning:
A major assumption in many machine learning algorithms is, that training and test data are drawn from the same feature space and have the same distribution. Since that requirement isn't always met in practice, transfer learning has been proposed to allow the domains, tasks, and distributions to be different. Hence, one can extract knowledge from one or more source tasks and apply the knowledge to a target task.
One of the main advantages of transfer learning is the ability to intelligently apply knowledge learned previously to solve new problems faster.

For further information on machine learning approaches as well as descriptions on additional approaches such as distributed and parallel learning, active learning and kernel-based learning, refer to [6].

4. CONDITION MONITORING – FRAMEWORK

In this chapter, a machine learning based framework for condition monitoring system is presented based on Figure 1.

Fig. 1. – Framework of the condition monitoring system



Obviously, a condition monitoring system of a technical product is based on the collected and in a proper way interpreted data. Therefore, it is very important to collect as much information about the condition of a particular product as possible. The very most of the technical products developed or produced nowadays are equipped with a number of sensors which provide valuable information that can be used for the description of the current health state. This information can be stored e.g. with a data logger (lots of easy to handle and low budget loggers are available and suitable from the technical point of view), if no internal logging possibility is not given. Tough, the key attribute of such a logger is the storage which has to be chosen according to the possible amount of the recorded data and recording time.

The data itself can be extracted from the following data sources:

- Product information in form of a constant data
- Cumulative data as a single variable
- Signals representing the change of a variable over time

The overall information about the product, such as product name, product type, manufacturing date, equipment variant, country of distribution, or usage place provide a solid starting point of gathering the product data and establish a sufficient data base. All these information are constant for a given product excluding minor exceptions like possible change of the usage place after reselling) and do not change over the time.

Cumulative data such as number of actuations, number of cycles (e.g. door openings), overall runtime, time spent in a certain mode of a product (e.g. driving a car in the first gear) represent variables that change during the operation period. These variables need to be gathered either directly from the product (only if equipped with a control unit, a memory storage and any data connection interface) or can be derived from other variables (e.g. driven mileage is the integration of the velocity signal). Cumulative variables change over time during the usage phase. Therefore, every single storage of the data of the same product shall be treated separately

Signals can be grabbed out of all available sensors of a product. It should be noticed that the temporal resolution differs for various data loggers. The temporal accuracy of the recorded data has no influence on the presented concept, but might have a significant impact on the overall results. Obviously, the higher the temporal resolution of a data logger, the more information can be provided for the analysis and the higher is the possibility of a proper health state description.

Once all the signals are collected, they need to be interpreted in a proper way. Basically, there is a wide range of possibilities to interpret the signals and provide attributes which can be used for the further analysis. The main aim of such an interpretation is the quantification of the health state if form of (easy to interpret) attributes. For example, one of an attributes of an interpretation of the car speed signal can be the maximum speed reached during a certain car ride. The interpretation tools can be divided into the following groups:

- Standardized statistical and mathematical methods for the analysis of signals like fast Fourier transform (FFT), time series analysis, descriptive statistics (mean, median, dispersion, variance), trend analysis and many more
- Specified analysis of certain signals in combination with the specific knowledge of a given product like e.g. maximum speed or maximum acceleration of a car derived from the velocity signal

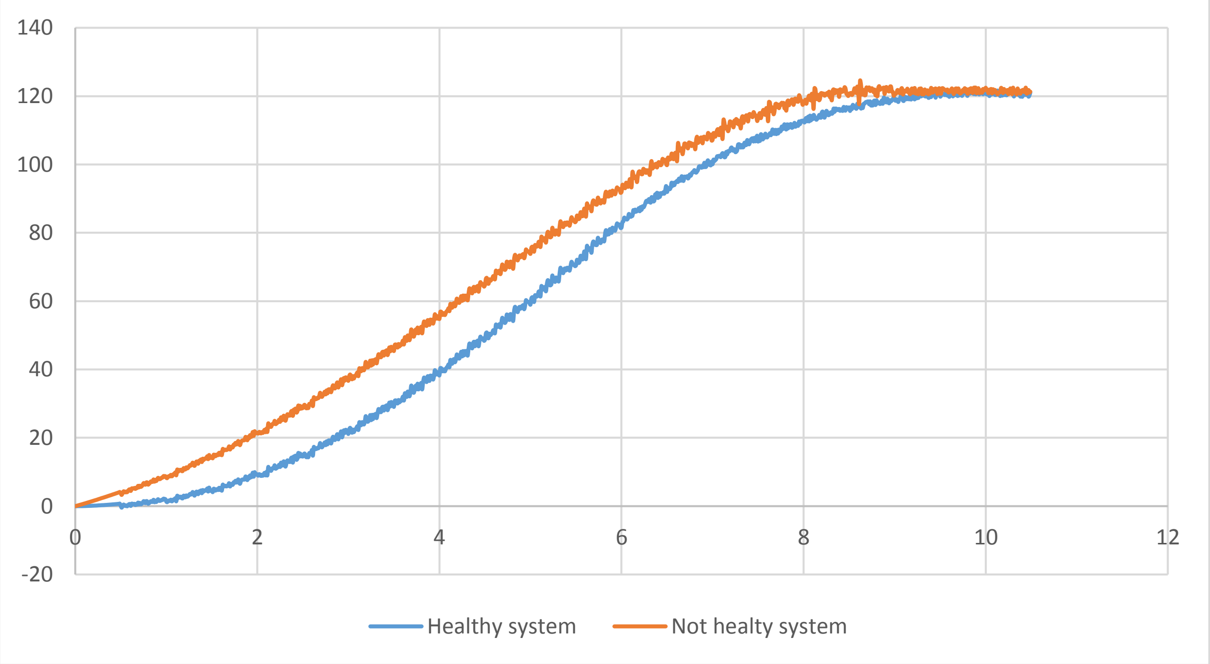
Subsequently to the entire signal interpretation process, a corresponding data matrix as a result of the analysis needs to be developed in a structured way. In simple terms, the temporal development or a propagation of a possible failure has to be trackable and easy accessible. If the failure cause, based for example on previous experiments, shall be detected and provided to the user, a corresponding database needs to be available for all known failure causes. For the proper detection of a failure, one of the algorithms introduced in chapter 2 can be applied easily. It is of note that the algorithms provide results in different form. C4.5 develops a rule which consist of as many attributes as needed for the entire description of a distinction. In other words, if the rule can be described by only one attribute, all remaining ones will be omitted. In contrary to that, NNge provides a rule which always consist of all analyzed attributes.

Finally, once the rules are determined and development of a certain attribute drifts in the direction of an experimental forced failure, the health state of the product needs to be changed from healthy to not healthy. Based on this data driven decision, a proper maintenance action can be planed or already performed.

4. REAL CASE – APPLICATION

The framework of the condition monitoring system described in previous chapter is now applied exemplary on an artificial, simulated data which represents a real case scenario and shows a very similar behavior to a real problem. For the sake of simplicity, the example is shown on one particular signal gathered from both, healthy and not healthy system. Since the data is artificial, there is no need of description of the collection of the signals (cf. first step of the presented approach).

Fig. 2. – Gathered signals from healthy and not healthy systems

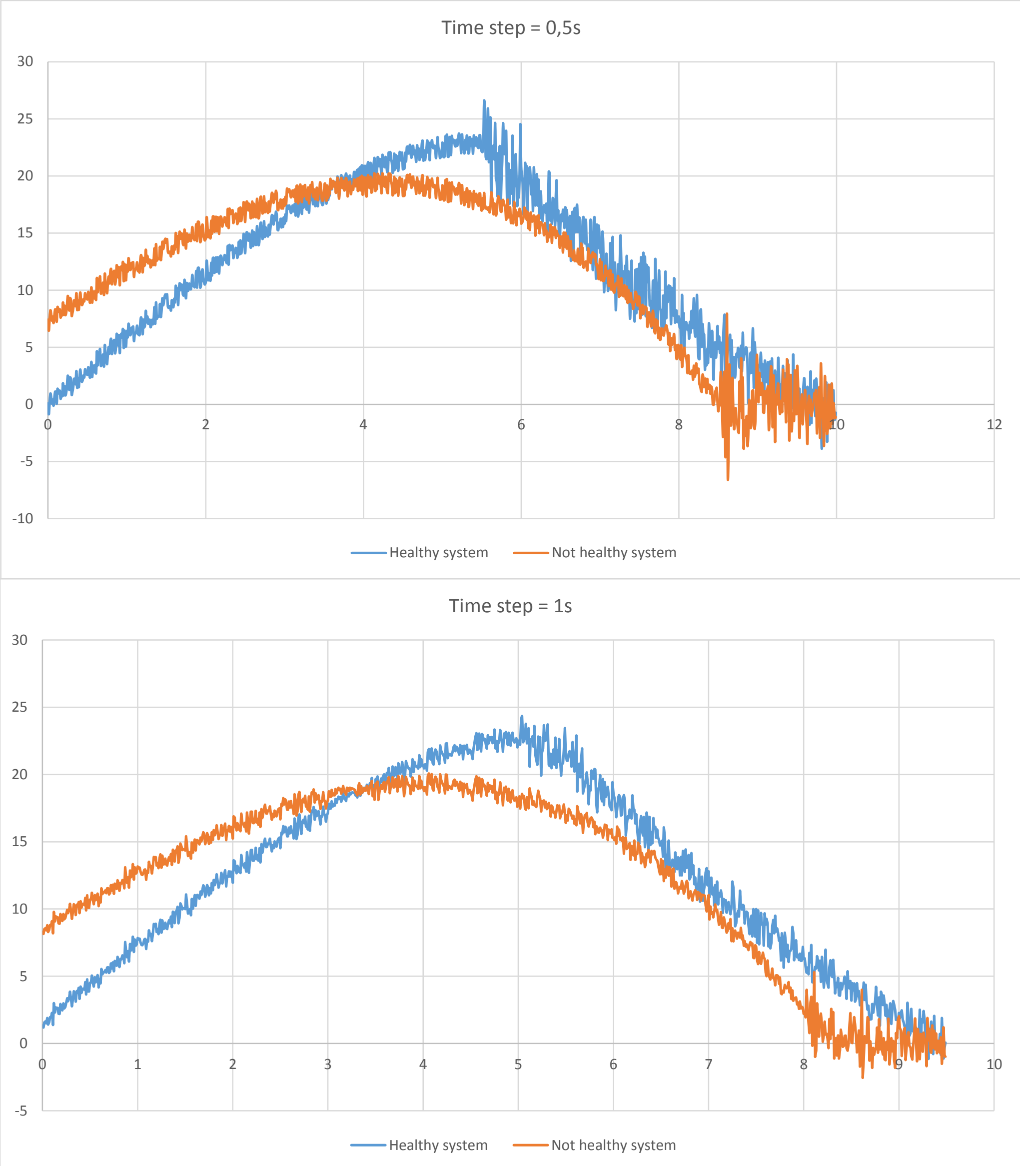


The example is shown based on a change of signals from the state 0 to 120. Note that there is no need of a precise description of the particular signal including its unit since the proposed method shall be applicable for all signal types. It can be easily observed that the healthy system changes the state later than the not healthy one. Therefore, there is a need of the quantification of this certain signal development. A temporal shift of a signal is a very common type of a recognition of the distinction between the mentioned types of systems. Though, easy to observe shift is slightly more complicated in terms of an establishment of a proper training data set. Here, the gradient of the curve can be calculated easily in terms of the first derivative. Since it is not a continuous signal, the derivative can be calculated by means of the forward difference quotient scheme:

$$N = \frac{f(x+h)-f(x)}{h} \tag{3}$$

Where h is a time step which can be selected with regard to the particular signal. It is of note that the time step must not be too short in order to avoid random scattering. This shall be explained based on the following example which is the differentiation of the signals presented in figure 2.

Fig. 3. – Calculation of the first derivative based on different time steps



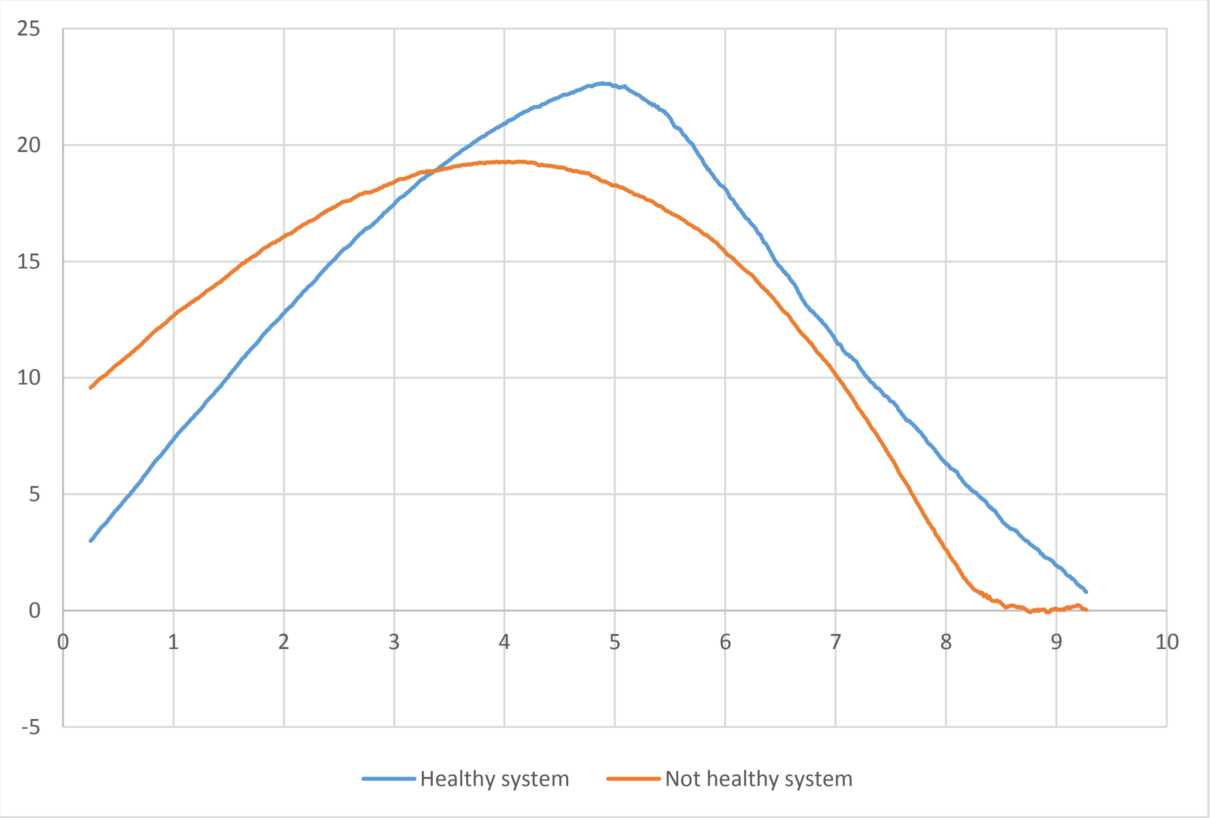
It can be observed easily that the differentiation based on time step of t=1s scatters much less than t=0.5s. If the time step is chosen to small, nothing but scattering can be observed (the curve becomes a straight line).

Now, to quantify the shift of the signal, the maximum of the differentiation as well as the according time can be saved in the training matrix. For this purpose, the differentiation curve has to be smoothened first in order to reach explicit attributes for the description of the shift. For the smoothing of the curve, the moving average can be applied [14]:

$$\bar{p}_{SM} = \frac{1}{n}\sum_{i=0}^{n-1} p_{M-1} \tag{4}$$

A smoothened curve based on the signal with 1s differentiation and calculated with the formula 4 is shown in the figure 4.

Fig. 4. – Moving average of differentiated signals



Based on the gathered curves, the maximum as well as the time of reaching the maximum can be defined easily. These final attributes are now stored in form of a matrix which is the training data set for the machine learning process (cf. step three of figure 1).

The training data based on this simple interpretation is shown in table 1.

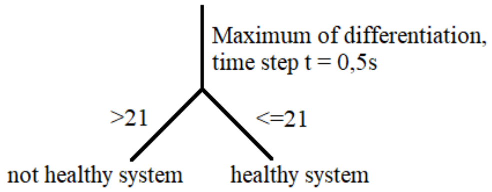
Table 1: Training data set - example

Max differentiate t=0.5s	Max differentiate t=1s	t_max of max diff t=0.5s	t_max of max diff t=1s	Tagret variable
19,37002604	19,28925506	4,13	4,03	Healthy system
23,01009687	22,65621314	5,36	4,89	Not healthy system

Obviously, the training data set presented in table 1 is trivial and consists of only ten entries. Typical data sets gathered from an overall process with both, healthy and not healthy units may consist of thousands attributes (or even much more, depending on the complexity of the process and number of signals) and millions of split process parts. Though, the presented approach will work in the exact same manner.

A typical result of the machine learning analysis in form of a decision tree and based on the discussed example performed with the C4.5 algorithm is shown in figure 5.

Fig. 5. – Decision tree – result of the C4.5 analysis



The very simple example consist of only two leaves due to the simplicity of the training data. In fact, the split criteria based on the data entropy and information gain ratio is exactly the same for all attributes. Therefore, the first attribute is chosen for the establishment of the decision tree.

Obviously, the results gathered form a complex analysis consisting of many attributes and many processes are much more complicated in terms of the proper interpretation. Though, they are established in the exact same manner as the described example. The only difference is the size of the tree.

4. CONCLUSION AND OUTLOOK

This paper presents a machine learning concept for a condition monitoring system of technical products. The applied machine learning techniques are focused on the inductive learning. For the purpose of an establishment of a corresponding training data set, some methods for the interpretation of signals gathered directly from the monitored machines are discussed and applied on a simple example.

The concept is applicable for complex machines and high data amounts. Realization of it was performed within a developed software package written in python and applied on big industrial machines out of the process industry. Though, there is still a need of a further development, especially regarding big data problematics. Here, corresponding amount of the data needs to be gathered first. Finally, the proposed algorithms, including machine learning, have to be tested with regard to their efficiency and performance and possibly optimized for the particular problem.

References

[1] M. Awad, R. Khanna, "*Efficient learning machines: Theories, concepts, and applications for engineers and system Designers*", Apress Media, 2015, New York
[2] I.H. Witten and E. Frank, "*Data mining: Practical machine learning tools and techniques" in: "The Morgan Kaufmann series in data management systems*", Kaufmann, 2001, San Francisco.
[3] J.R. Quinlan, "*C4.5: Programs for machine learning*" in: "*The Morgan Kaufmann series in machine learning*", Kaufmann, 1993, San Mateo.
[4] B. Martin, "*INSTANCE-BASED LEARNING: Nearest Neighbour with Generalisation*", Master thesis, University of Waikato, 1995, Hamilton.
[5] Z. Stamboliska, E Rusiński, P. Moczko, "*Proactive Condition Monitoring of Low-Speed Machines*", Springer Verlag, 2015, New York.
[6] J. Qiu, Q. Wu, G. Ding, Y. Xu, S. Feng, "*A survey of machine learning for big data processising*", EURASIP Jornal on Advances in Signal Processing,
[7] A. Bordes, X. Glorot, J. Weston, Y. Bengio, "*Joint learning of words and meaning representations for open-text semantic parsing*", in Proceedings of 15th International Conference on Artificial Intelligence and Statistics , pp. 127–135 (2012)
[8] N. Boulanger-Lewandowski, Y. Bengio, P. Vincent, "*Modeling temporal dependencies in high-dimensional sequences: application to polyphonic music generation and transcription*", arXiv preprint (2012). arXiv:1206.6392
[9] K. Dwivedi, K. Biswaranjan, A Sethi, "*Drowsy driver detection using representation learning*", in Proceedings of the IEEE International Advance Computing Conference, pp. 995–999, (2014)

[10] D. Yu, L. Deng, “*Deep learning and its applications to signal and information Processing*”. IEEE Signal Proc Mag 28(1), pp. 145–154 (2011)
[11] Y. Bengio, “*Learning deep architectures for AI.*”, in Foundations Trends Mach Learn 2(1), pp. 1–127 (2009)
[12] R Collobert, J Weston, L Bottou, M Karlen, K Kavukcuoglu, P Kuksa, “*Natural language processing (almost) from scratch*”.in Journal of Machine Learning Research 12 (2011)
[13] P Le Callet, C Viard-Gaudin, D Barba, “*A convolutional neural network approach for objective video quality assessment*”. IEEE Trans Neural Network 17(5), pp. 1316–1327 (2006).
[14 L. Sachs and J. Hedderich. *“Angewandte Statistik Methodensammlung mit R.“*, 13. Edition, Springer, 2009, Berlin.