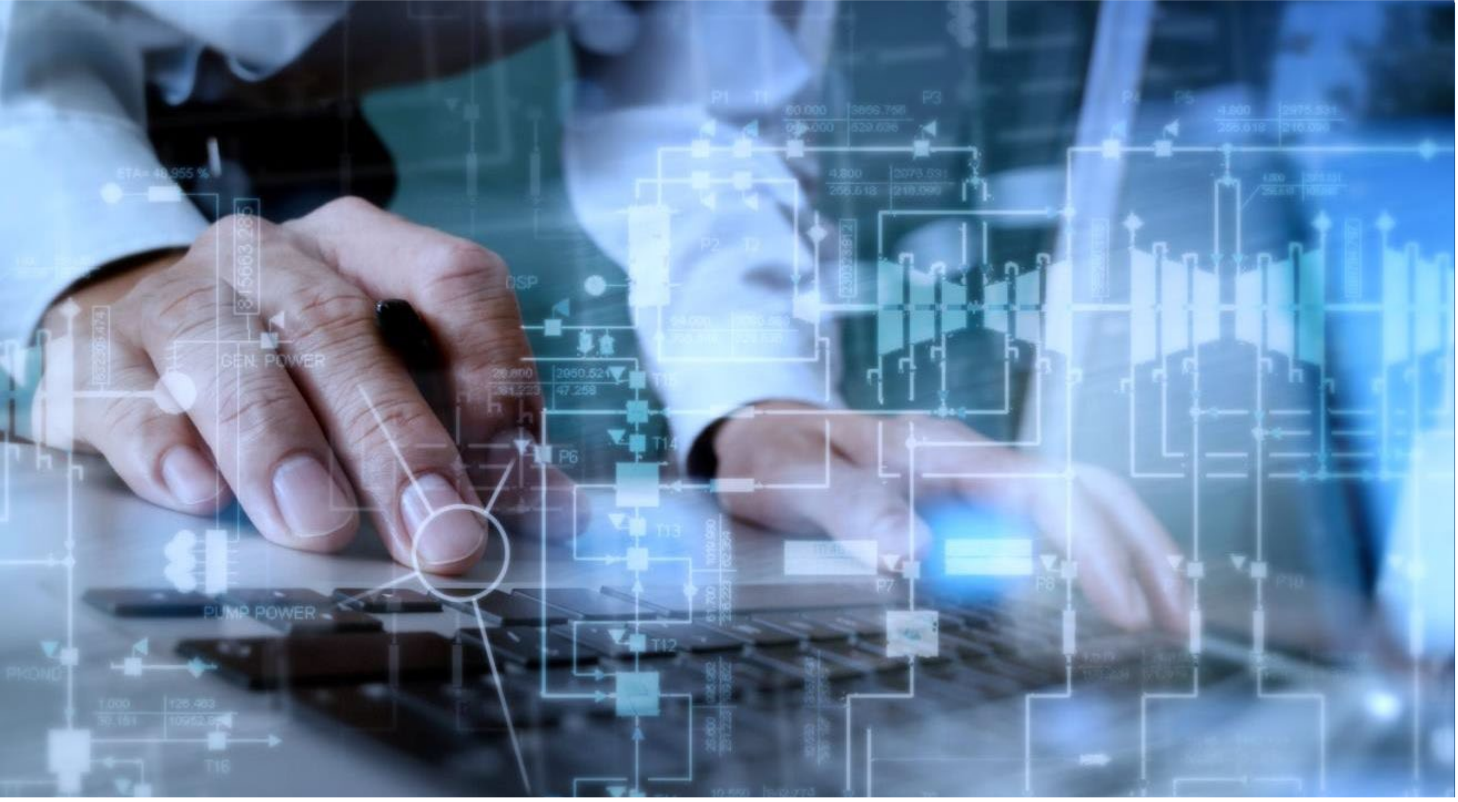
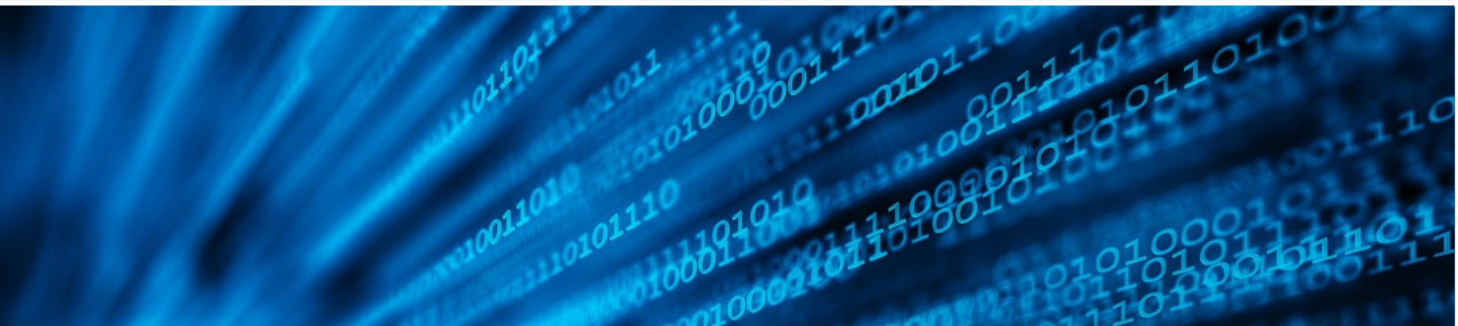




Predictive Maintenance on the Basis of “Big Data“ and ”Machine Learning“

Monitoring with Extra Intelligence, Efficiency, and Profitability

Capturing and storing sensor-based data from power plants in real time is now state of the art. However, volatile external factors like e.g. weather, fuel quality, load, etc. make it difficult to assess the condition of the plant components based on these data alone. For this, the data have to be transformed into valuable information by means of intelligent evaluation. With a monitoring system based on Big Data and machine learning, STEAG Energy Services has made a crucial step towards achieving this goal.

The ever-inceasing performance capability of the IT hardware is an essential driver for developments like Big Data, digitalization, and the Internet of Things. A key technology of digitalization is machine learning, i.e. algorithms that automatically analyze correlations existing in data and thereby, for instance, make projections ("Predictive Analytics") or detect recurring patterns ("Advanced Pattern Recognition").

Digital twins by means of machine learning

In the context of power generation, machine learning can be used to generate digital copies ('digital twins', Fig. 1) of plants and their components. For quite some time, machine learning has been a component of software solutions supporting predictive maintenance by means of data-based, digital twins of the process and the plant components. Additionally, current developments in the environment of Big Data have the potential to reduce the engineering effort for implementing a solution for this maintenance strategy.

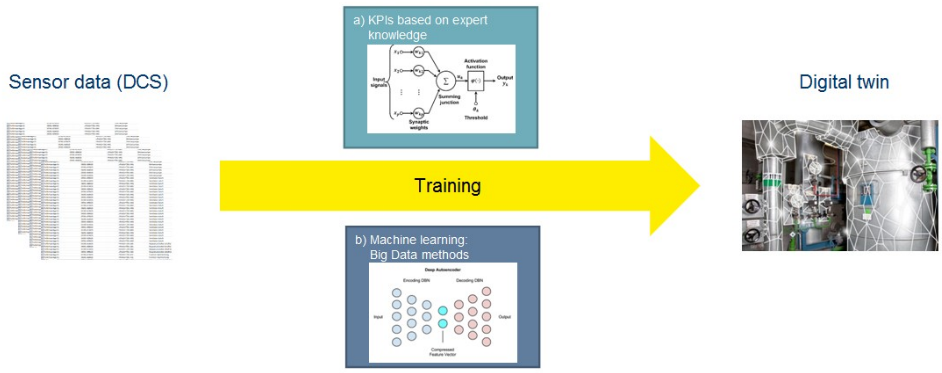


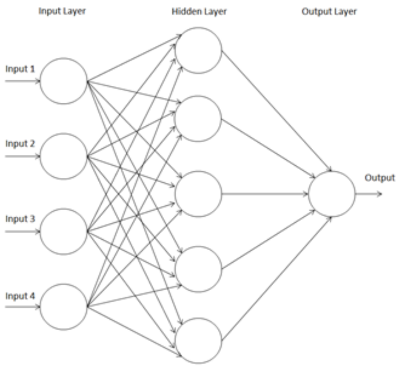
Fig.1: Digital twins for predictive maintenance

Digital Twins show relevant dependencies

After all, a digital twin of the plant is a mathematical model representing how relevant process parameters (outputs, temperatures, mass flows, etc.) depend on the ambient conditions and actuating variables of the plant. The model is either based on the physical basic equations describing the plant or it can be derived from historical data by means of machine learning.
The current values of operational measurements can then be compared with the projections of the digital twins to detect also creeping changes in the plant condition early and reliably. On the basis of this information, it is possible to act predictively in terms of predictive maintenance in order to reduce efficiency losses by using the resources efficiently and increase the plant availability. The challenge in this context, however, is to detect anomalies in the plant behavior truly reliably while at the same time avoiding phantom anomalies.

Using expert knowledge to obtain “High Quality KPIs“

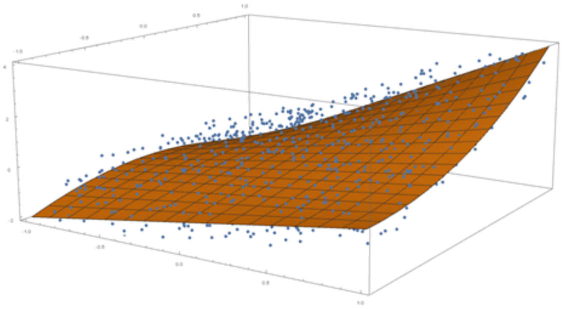
Different approaches which, however, complement each other are possible when creating data-based models: modeling by combining expert knowledge with machine learning or modeling by means of autonomous, unsupervised learning (see Fig. 2), largely based on mathematical algorithms with minimal engineering groundwork.
An expert knows the significant key variables for the condition of the process or of a plant section. He also knows which influencing variables are required for describing the expected behavior of this key variable and which periods of time in the historical data are suitable as reference, i.e. which causalities exist in the data. The expert selects the input and output variables of his model accordingly. Supervised learning methods will then form the model. This way, digital twins with a high accuracy emerge (”High Quality KPIs“). Owing to the selection of particularly informative measured values by engineers, significant changes detected by the system can be easily attributed to possible faults in order to subsequently decide about further examinations and specific maintenance measures respectively.



Supervised learning

Expert knowledge on cause and effect is displayed.

Engineering in the modelling but simplified evaluation.



Unsupervised learning

Correlations between the measured values are detected and independent key variables are identified.

Automatic modelling but engineers' analysis in the evaluation.

Fig. 2: Supervised and unsupervised learning

Autoencoder learns autonomously

In autonomous learning, in contrast, there is no specification of causalities in the data by the expert. It is the task of "machine learning" to identify existing correlations and to form corresponding models. In doing so, the algorithms detect that the data from a plant are not independent of each other but are determined by a few key values. Thus algorithms like e.g. the "deep autoencoder" identify such variables autonomously and learn the correlations between those and the measured values of a plant in an "unsupervised" way.

Identification of essential influencing variables

An autoencoder (see Fig. 3) is a neural network trained to copy the input values onto the output values. The network has several levels of hidden layers of neurons and consists of two parts with a "bottleneck" in the middle.

It prevents the neural network from just learning input = output and enforces that essential influencing variables ("features") describing the behavior of the input data are identified.

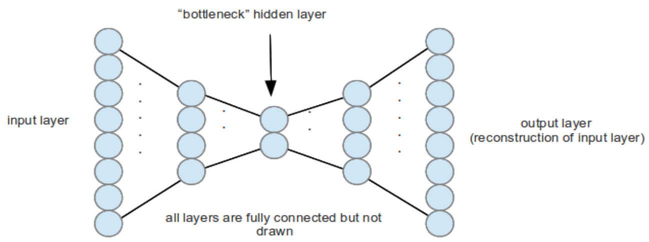


Fig. 3: Structure of an autoencoder

Automatic check of significant deviations

In an online application for supporting the predictive maintenance, all measurements to be monitored are presented to the deep autoencoder. When "learning", the algorithm identifies the most important features that describe the plant behavior and how the measurements depend on it under “normal“ conditions. In "online" mode, the first part of the autoencoder will then at first determine the features describing the current operating condition from the current measured values. From these features, the second part calculates the “normal“ values to be expected under the given operating conditions for each measured variable. For each measurement, statistical methods are applied to automatically check whether possible deviations between the current value and the projection are significant. This way, also large sets of measured values can be monitored automatedly, and changes in the condition of the plant that are reflected in the measured values can be detected very early and, above all, reliably with little engineering effort. This approach to identify anomalies in measured values has been successfully applied to data from various power plants (conventional power plant and wind energy plant).

Making use of available resources more efficiently

The economic pressure on power plants is rising more and more. Against the backdrop of increasing installed power, the economically efficient operation of fossil-fueled power plants in spite of a lower utilization ratio and higher stress on the components due to a more flexible mode of operation is becoming increasingly difficult. The funding models for renewable energies, in particular for wind energy plants, are changing. Here, too, the profitable operation of wind farms is becoming a growing challenge. Thus it is crucial to detect also creeping changes in a plant that indicate impending faults or failures involving high losses early and reliably.

Even experienced personnel cannot fulfill this task as it is becoming more and more difficult with increasing variability in the mode of operation of the plant and, moreover, has to be accomplished by less and less persons.
Therefore it is absolutely essential to consistently use the potentials of groundbreaking IT developments like e.g. “Big Data“ and ”machine learning“ to give existing software solutions for the monitoring of processes and plant components that crucial extra intelligence.
This leads to powerful new solutions that provide extremely valuable information for identifying significant deviations from the regular plant behavior promptly and without high personnel effort.

Available resources can thus be deployed considerably more efficiently, and maintenance measures can be planned in an even timelier and more targeted way.

Example: Detection of Anomalies in Power Plant Data Reduces the Fuel Input and Increases the Availability

For the turbo set of the considered 1,100 MW unit, more than 1,200 analog channels exist in the DCS. The evaluated data record comprises 5min averages over 60 months, i.e. more than 630 million averages. An autoencoder as described in the article was generated with a subset of the data. The deviations between measured and expected values were calculated for the entire data record. Statistical methods of analysis then identify measured values that show significant deviations from the normal behavior ("anomalies").

The results are displayed in heatmaps (Fig. 4). The representation is strongly compressed to provide a clear view of the large amount of data. There is one column for each measurement (1,200 columns altogether). Each line represents one period of time. The autoencoder was applied to 5 min data. In the diagram, the results are condensed to months (60 periods for five years).

Combination of autoencoder and statistical analysis detects anomalies

For each period, the algorithms counted how often the combination of deep autoencoder and statistical analysis detected a significant anomaly for the measurement. The periods are marked in colors as follows: light blue = no anomaly in this month, dark blue = continuous significant anomalies in the considered month, white = plant shutdown. The large white streak in the upper area relates to a major overhaul of the unit.

Ruling out anomalies due to overhauls

As described above, the autoencoder was trained with a subset of the operating data. At first, the amount of training roughly corresponds to the data up until a short time before the overhaul. It is visible that changes in the plant behavior due to measures in the context of the overhaul were correctly identified as anomalies. To represent the changed plant behavior after the overhaul, the autoencoder was trained a second time with expanded data containing also operating data from the period after the overhaul. The comparison of the results shows that anomalies in later plant operation were detected in both cases. Anomalies that occurred due to a changed plant behavior as a result of the overhaul work (Fig. 4 left) were selected as regular data by expanding the training data and thus, as expected, no longer identified in the expanded heatmap (Fig. 4 right).

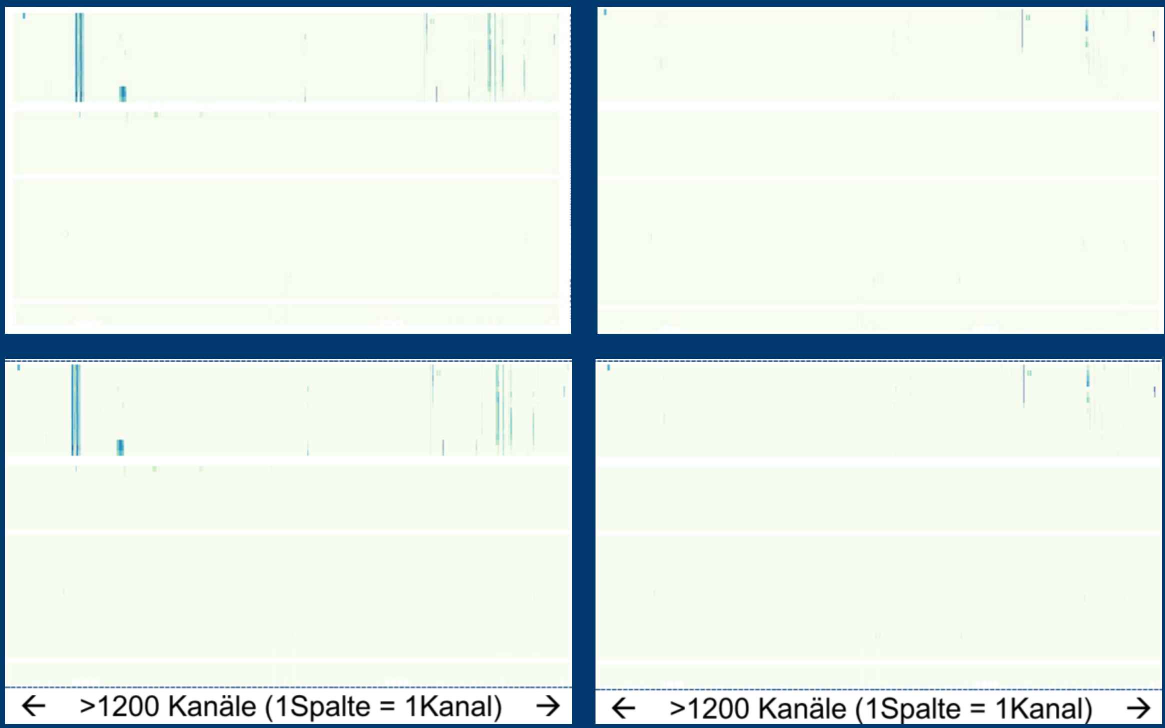


Fig. 4 Heatmap of the anomalies in a turbo set. Left: Model with ca. 50% learning period. Right: Model with ca. 75% learning period (incl. a small subset of the data after the overhaul)

Preventing false alarms, reliably identifying significant changes

Fig. 5 exclusively shows only those measured values for which anomalies occurred in the considered period of time. This was the case for only few of the more than 1,200 measurements altogether. In fact, this is to be expected for the turbo set of a modern plant. The image shows that the outlined method is able to prevent false alarms and yet to reliably detect significant changes. For instance, in June 2016 the anomaly in the measurement LBS20CT001 (red in Fig. 5 left) could be attributed to a malfunction in an LP extraction due to a defect expansion joint, which can also be observed very clearly in the chronological sequence of the extraction temperatures (Fig. 5 right).

The method is able to reliably identify a manageable quantity of suspicious values out of the large amount of measured variables, as shown in Fig. 4 and Fig. 5. Moreover, samples have shown that the revealed anomalies really correspond to faults or other conditions of the plant that deviate from the normal mode of operation.

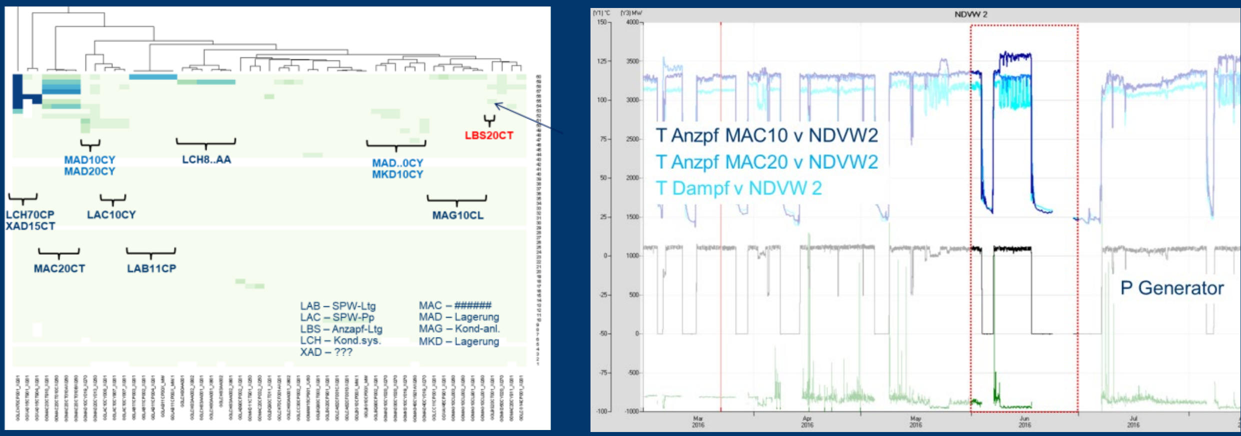


Fig. 5 Left: Exclusive view of the anomalies in a turbo set (75% learning period). Right: Detailed view of a detected anomaly

Systematic prevention of greater damages

In a review at another unit, the system showed a suddenly occurring abnormal running behavior of the turbine (Fig. 6). More detailed evaluations indicated a major turbine damage at the IP turbine and/or a bearing damage. To prevent even greater damage, the plant was shut down safely. The cooling of the turbine was initiated as early as possible, and the endoscopy was prepared. The endoscopy confirmed the IP turbine damage.

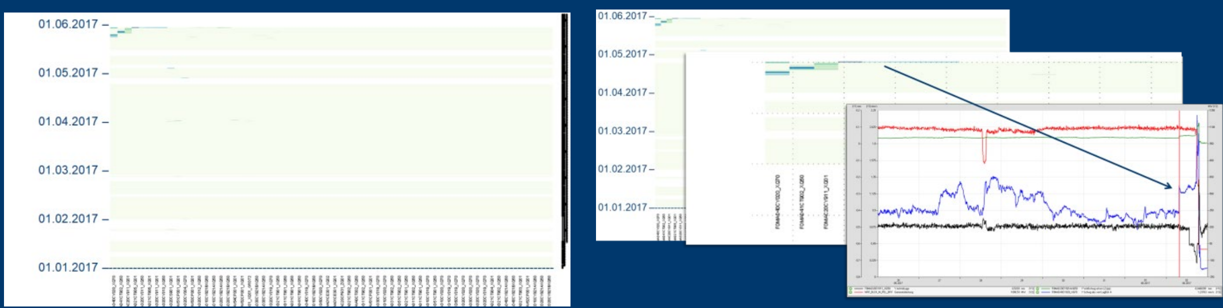


Fig. 6 left: Heatmap of the anomalies in a turbo set (1st half of 2017). Right: Detailed view of an early detected anomaly in the running behavior of the turbine

Efficient use of available resources

The tool “anomaly detection“ thus automatedly monitors entire plant sections, allowing the operating personnel to focus their attention on the areas where actual changes appear. A systematic analysis allows to react to these changes by acting predictively, efficiently deploying the available resources, in the context of predictive maintenance. This way, sources of loss are eliminated early and unplanned shutdowns are reduced. The mean specific heat requirement is thus reduced and the availability is improved.

For more information:

Phone +49 201 801-4175

STEAG Energy Services GmbH

System Technologies

Rüttenscheider Str. 1-3

45128 Essen

Germany

www.steag-systemtechnologies.com