

---

# ENHANCING MAINTENANCE WITH A DATA-DRIVEN APPROACH

JOÃO OSTROWSKI\*, JÓZSEF MENYHART#

Department of Mechanical Engineering, Faculty of Engineering, University of Debrecen,  
Ótemető u. 2–4, H-4028 Debrecen, Hungary

\* E-mail: ostrowski.jg@gmail.com

# E-mail: jozsef.menyhart@eng.unideb.hu (Corresponding author)

Constant stream of data has been generated and stored as more devices are being connected to the internet and supported with cloud technologies. The price drop of such applications along with industry 4.0 trending, triggered an explosive growth and demand for many IT modern solutions. From an industrial point of view, sensorization practices are spreading through factories and warehouses where software is constantly adapting to provide actionable insights in a data-driven configuration. The fourth industrial revolution is empowering the manufacturers with solutions for cost reduction, which translates in competitive advantage. The sector of maintenance operations is leading in engineering innovation, from reactive to planned preventive techniques the next step in history of proactive approaches is Predictive Analytics Maintenance.

**Keywords:** *Predictive Analytics, Predictive Maintenance, data mining, sensorization, data visualization*

---

## 1. Introduction

Definition of maintenance expenses is the amount of costs involved to keep an item in good condition and/or good working order. This value analysis is important in any investment as it explains for example why home ownership can be more expensive than renting. Evaluation of maintenance costs is a routine in any industrial environment especially due to rising in automatization which translates in more equipment, systems, machinery and infrastructure. Life cycle costing techniques contribute for a big picture on understanding in more detail optimum balances between operation and maintenance, however latest applications of automatization set the space for innovations. Since Industrial Revolution the complexity of components and functionalities combined to an increasing demand, companies started relying more on skilled personnel and specialists to act fast keeping every asset in perfect functionality.

Estimating its criticality in numbers, approximately US\$ 300 billion was spent by United States annually on plant maintenance prior 2006. The total price of maintenance is a burden to numerous corpo-

rations, therefore with a persisting problem around technology found its way to address a solution. Executives and senior managers believe that the maintenance operation is solely responsible for the reliability of a manufacturing operation, being the reason why this field is in constant focus for enhancements and modernization [1].

Motivation for the research comes from the contemporary merging of knowledge fields in operation chains. The future of companies is not to be defined either as service or product providers but a mix of both [2]. When analyzing the industry opening for such transformations the Information Technology (IT) sector is the one playing a crucial role establishing the links and trends with Machine Learning capabilities, breaking the communication walls from outdated management systems delivering the power of internet connectivity. Therefore, by mixing data mining and analysis with sensor data from an industrial environment this paper proposes joining knowledges from maintenance engineering and information technology to enhance production line performance, by changing how maintenance is perceived when making use of a data-driven approach.

---

**Open Access statement.** This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited, a link to the CC License is provided, and changes – if any – are indicated. (SID\_1)

For future works the authors try to analyze and classify data with Support Vector Machine and Fuzzy logic. The basic concept based on the works of Manukovits [8–14].

## 2. Background review

### 2.1. Preventive and Predictive

Among a variety of definitions attempting to label Maintenance methodologies in this paper Predictive Maintenance is considered as a particular subset of Preventive Maintenance. While both focus on preventing operational failures of critical equipment they differ in technology and resources utilized. However, in order to detail particularities the Reliability Centered Maintenance (RCM) [3] concept is used for succinct definitions:

- Preventive Maintenance is a routine activity where components or equipment items are replaced or overhauled at a specific, predefined interval, regardless of their condition at the time;
- Predictive Maintenance is a routine activity to inspect or test for the presence of warning conditions that indicate that the item is about to fail, this based on historical data and empirical experiences. A corrective maintenance action is then scheduled to replace, repair or overhaul the item before it suffers an in-service failure.
- As a rule of thumb led by economical factors Predictive Maintenance is the right choice for a simple reason, it optimizes the life cycle from a component, by close monitoring its functioning and subjecting the part to maintenance only when it starts losing efficiency jeopardizing operation [4]. Nevertheless, there are many exceptions for this idealization, for example parts that cannot be in constant monitoring due to fieldwork. To address the case, the so called fourth maintenance revolution is finding on cloud technologies and product sensorization the easy way out.

### 2.2. Sensorization supporting analytics

Sensorization is a new buzzword that became popular among product design teams. Its definition regards the extent or the trend of embedding as many sensors as possible within a device or appliance. Within data revolution, sensorization became convenient as an extra resource to generate more and more data. A great example of its most optimized use is on smartphones, each generation brings new embedded sensors and softwares are updated in order to make better use of hardware capabilities. In other words, this is a pattern among hardware/software industry where sensorization can be correlated to a continuous improvement cycle.

Experimentations on predictive technologies are continuously growing due to the easy access to data which is only possible due to the popularization of the sensorization philosophy applied on product design. Industries started experiencing a similar reality, with more connected machines and real time data flowing through advanced ERPs (Enterprise Resource Planning), making industrial management more transparent and organized [5].

Predictive Analytics embraces many statistical and computer science concepts to summarize data in order to predict the future of unknown events. The exploration of gathered data from different machine signals sets the ground to data mining and data visualization, which are the starting point of any predictive analytics implementation, incorporated for data-driven decision support [6]. The key point however, are the wide range of algorithms developed with prediction capabilities that with Industry 4.0 trending are being generalized and made available for accepting any type of data [7]. In its simplest form such algorithms could generate equations that mathematically explain a certain operation as a function of its working variables.

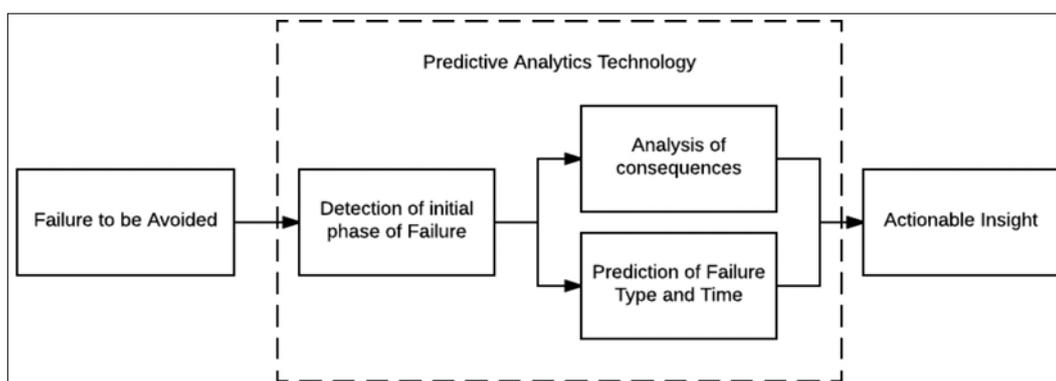


Fig. 1. Macro view of a Maintenance System using Predictive Analytics Technology

2.3. Analyzing the Predictive Analytics Maintenance model

Putting innovations within the maintenance framework generates a new process flow which is represented in Fig. 1 in addition to a common Predictive Maintenance model.

For the purpose of this paper the spotlight is on the techniques applied for optimized detection of initial phase of failure and how the outputs generated are used for more complex predictive analysis by means of machine learning. For clarification, “Failure to be Avoided” will also be described since it plays an important role as a primary input.

- Failure to be Avoided: This step is usually covered by FMEA studies managed by quality and maintenance departments in order to breakdown and specify failure modes followed by root causes and actions to be taken in each case. The information gathered by this process is crucial and delivers initial historical data used for modelling consequences analysis and failure predictions.
- Detection of initial phase of Failure: Using dynamic data being generated in real time by sensors in production, thresholds could be set for keeping values inside tolerances designed using statistical quality control, therefore whenever the values fall outside the projected, but still in a safety zone preventing complete failure, engineers are warned and asked for taking actions. Besides keeping the process in a statistical standard, strategical visualizations are also provided from mining the data, allowing quality analysts to perceive patterns and different behaviors on machines which also becomes actionable insights leading to continuous improvement. Visualizations and data mining is the main focus explored on the study case presented on this paper.

3. Methodology

Data from 25 sensors at a single point in time, stationary data, were extracted for a total of 136 machines. Figure 2 shows a header of first ten entries from the working dataset. Not all the sensors are presented in the figure for a clear visualization. Note that this is a historical data from a certain period of time from a manufacturer, therefore along with the variable “Machine\_ID” and the 25 values sensors, the attribute “Failure” indicates whether the machine has failed or not during the time period from which data corresponds. The sensor are anonymized due to data confidentiality terms, for that reason it is not possible to know what each one of them represents or what it measures.

To get a better grasp from the dataset, many types of data exploration were performed over it. It is a good practice to understand the characteristics of the data and comprehend some general patterns from the variables. In terms of softwares, Rapidminer and Python were utilized for data extraction, load and exploration (ETL), general data handling and visualizations as well as Tableau for deploying more interactive and deeper visualizations.

3.1. Visualizations for insights

Starting with assessing the quality of working data, which means looking for missing values, discrepancies between values in this case corresponding wrong signals and measurements from the sensors. This initial analysis is done for checking missing values and outliers in the working data by visualizing the distributions of average sensor values followed by standard deviations according Fig. 3. For a big picture about how many machines were in failure in this data set the visualization in Fig. 4 was generated.

Row No.	Machine_ID	Failure	Sensor_1	Sensor_2	Sensor_3	Sensor_4	Sensor_22	Sensor_23	Sensor_24	Sensor_25
1	M_0001	no	2.633	0.918	4.229	13.792	1.145	2.049	2.795	2.266
2	M_0002	yes	9.244	22.732	15.307	8.553	0.735	2.599	2.779	2.780
3	M_0003	no	3.183	28.526	8.735	4.898	0.893	0.213	0.139	0.099
4	M_0004	yes	2.785	2.642	5.857	7.417	0.764	0.427	0.188	0.096
5	M_0005	yes	4.889	14.415	23.903	12.985	2.200	2.984	3.017	3.110
6	M_0006	no	0.784	5.498	15.232	23.477	2.499	2.801	2.931	3.007
7	M_0007	no	6.214	26.377	13.063	6.312	0.717	0.169	0.127	0.097
8	M_0008	yes	2.590	16.258	18.281	3.524	2.690	2.582	3.028	2.555
9	M_0009	no	10.395	24.494	15.682	3.451	0.733	0.203	0.031	0.195
10	M_0010	yes	0.348	1.800	6.922	30.658	2.822	1.015	1.200	1.715

Fig. 2. Header preview of working data set

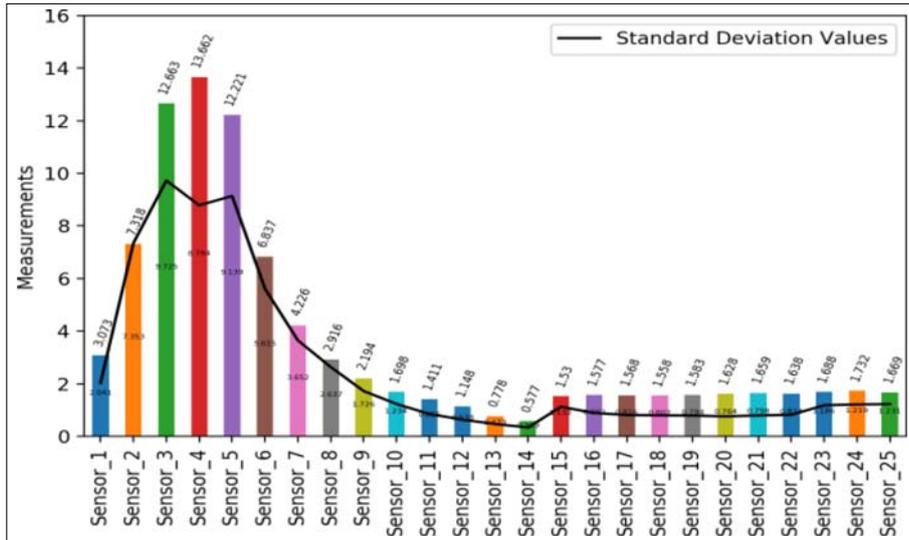


Fig. 3. Average values and standard deviation by Sensor

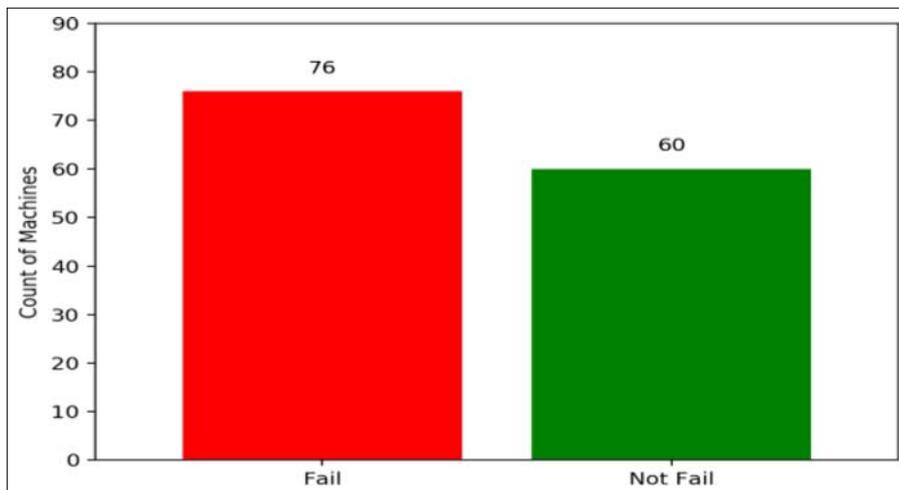


Fig. 4. Big picture on historical failure of machines analysed

3.2. Critical sensors for failure

After ensuring the data quality, keeping data outliers, a statistical correlation analysis of variables and labels was conducted, in this case how sensors correlate to machine failure. The output of this study is presented

in Fig. 5 and provides valuable information regarding the critical sensors in production. Furthermore, from the correlation matrix the top five sensors were subjected to an additional analysis. In Fig. 6 a statistical description of each sensor considered critical in operation with values for mean, standard deviation, mini-

	Failure	Sensor_1	Sensor_2	Sensor_3	Sensor_4	Sensor_5
Failure	1.000000	-0.139190	-0.277625	-0.203692	0.200298	0.368910
Sensor_1	-0.139190	1.000000	0.744845	0.436215	-0.009188	-0.344840
Sensor_2	-0.277625	0.744845	1.000000	0.673556	-0.148438	-0.628210
Sensor_3	-0.203692	0.436215	0.673556	1.000000	-0.023792	-0.656371
Sensor_4	0.200298	-0.009188	-0.148438	-0.023792	1.000000	0.052913
Sensor_5	0.368910	-0.344840	-0.628210	-0.656371	0.052913	1.000000
Sensor_6	0.451895	-0.134588	-0.429703	-0.454253	0.089500	0.696746
Sensor_7	0.436430	-0.089619	-0.305529	-0.314355	0.051629	0.531396
Sensor_8	0.420210	-0.022561	-0.175578	-0.193297	0.134703	0.403502
Sensor_9	0.375322	-0.024751	-0.206641	-0.265355	0.147908	0.424975

Fig. 5. Head of data frame with statistical correlation between variables

	Failure	Sensor_1	Sensor_2	Sensor_3	Sensor_4	Sensor_5
Failure	1.000000	-0.139190	-0.277625	-0.203692	0.200298	0.368910
Sensor_1	-0.139190	1.000000	0.744845	0.436215	-0.009188	-0.344840
Sensor_2	-0.277625	0.744845	1.000000	0.673556	-0.148438	-0.628210
Sensor_3	-0.203692	0.436215	0.673556	1.000000	-0.023792	-0.656371
Sensor_4	0.200298	-0.009188	-0.148438	-0.023792	1.000000	0.052913
Sensor_5	0.368910	-0.344840	-0.628210	-0.656371	0.052913	1.000000
Sensor_6	0.451895	-0.134588	-0.429703	-0.454253	0.089500	0.696746
Sensor_7	0.436430	-0.089619	-0.305529	-0.314355	0.051629	0.531396
Sensor_8	0.420210	-0.022561	-0.175578	-0.193297	0.134703	0.403502
Sensor_9	0.375322	-0.024751	-0.206641	-0.265355	0.147908	0.424975

Fig. 6. Statistics summary from critical sensors

imum and maximum measurements and each of quartiles. Complementing the evaluation, Fig. 7 presents a series of boxplots comparing the sensors considered as critical.

4. Concluding remarks

4.1. Insights from data mining and visualizations

Regarding overall operation performance, Fig. 4 pre-sets a rate of 55% machine failure during the time of analysis, leaving a lot of space for enhancements and confirming the motivations for this research.

Sensors 6, 7, 8, 9 and 5 were the ones considered critical to keep operations running effectively, due to

its statistical correlation with machine failure according to Fig. 5.

Comparing the behavior of sensors when resulting in failure and non failure in Fig. 7 it is clear that once in failure sensors 6, 7 and 5 had higher distribution of measurements. However, outliers (points above the box distribution) impaired a better evaluation.

4.2. Future research

As mentioned above, the first topic for future development of this research is not just excluding the outlier values but analyzing each occurrence to understand its root causes, classifying each point according to possible causes established by FMEA documentation and

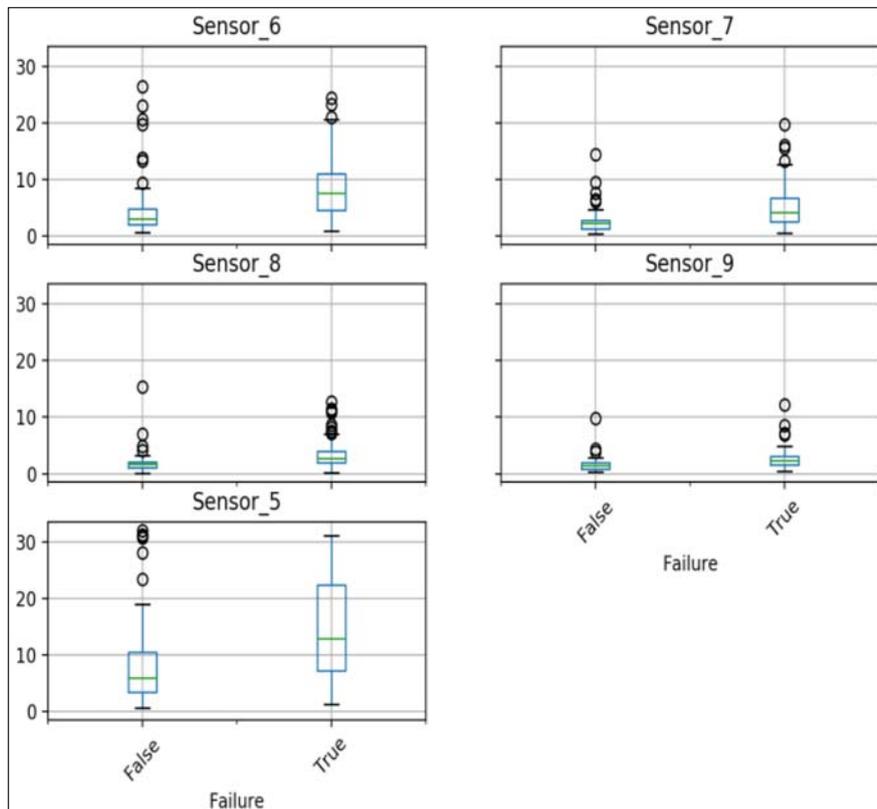


Fig. 7. Box plots comparison of critical sensors

updating the documentation in case of new findings. After studying the points they should be ignored to enhance Fig. 7 visualization providing a clearer comparison method between failure and non failure behaviors.

The relationship between variables and failure was estimated by mean of statistical correlation analysis, a further step on this study is to run a new correlation among the sensors considered critical only. This will allow a deep understanding of how one sensor affects the other considering that the final result is a machine breakdown. In a maintenance engineering perspective this will enable a straightforward understanding of failure flow.

As a future possibility the association of sensors and failure should also be evaluated by means of association rules algorithms, such as Apriori and FP-Growth. By utilizing this methodology the statistical values of Confidence and Lift could be a valuable output for working improvements.

Still approaching the maintenance engineering, Fig. 5 is a reliable input for scheduling maintenance of critical components as well as machines in a optimized predictive configuration. In addition to that, by close monitoring those critical sensors, keeping a dynamic real time analysis grouped with a threshold set up for each sensor will prevent that measurement values go outside of designed values in a very cost efficient way.

Furthermore, to bring the state of art when it comes to Machine Learning applications to industrial environments, a next step to this paper is applying and comparing efficiencies of many different predictive models focusing on prediction of machine failure based on sensor measurement trends.

## Acknowledgements

The work/publication is supported by the EFOP-3.6.1-16-2016-00022 project. The project is co-financed by the European Union and the European Social.

## References

- [1] Dabbs T. (2014), The True Cost of Maintenance. Pump & Systems
- [2] Stevenson W. J. (2015), Operations Management. McGraw-Hill, Twelfth Edition
- [3] Nowlan F. S., Heap H. F. (1978), Reliability-Centered Maintenance. United Airlines, for Office of Assistant Secretary of Defense, Washington, D.C.
- [4] Dunn S. (2016), Big Data, Predictive Analytics and Maintenance. Retrieved from: <https://www.assetivity.com.au/article/>
- [5] Lee J., Kao H., Yang, S. (2014), Service innovation and smart analytics for Industry 4.0 and big data environment. Product Services Systems and Value Creation. Proceedings of the 6th CIRP Conference on Industrial Product-Service Systems
- [6] Larose D. T., Larose C. D. (2015), Data mining and predictive analytics. John Wiley & Sons
- [7] Blackburn R., et al. (2015), A predictive analytics approach for demand forecasting in the process industry. International Transactions in Operational Research, 22(3), 407–428.
- [8] Kocsis I., Mankovits T., Vámosi A., Deák K. (2014), SVM variants used in the investigation of some engineering optimization problems. In: Rodrigues H. C., Herskovits J., Mota Soares C. M., Guedes J. M., Araújo A. L., Folgado J. O., Moleiro F., Madeira J. F. A. (szerk.) Book of Abstracts of the 4th International Conference on Engineering Optimization. Lisboa, Portugália, 2014.09.08–2014.09.11. Lisboa: IST Press, pp. 213–214. (ISBN: 978-989-96276-6-6)
- [9] Borsavölgyi T., Mankovits T., Kocsis I. (2011), SVM alkalmazása műszaki feladatok optimalizálására [Usage of SVM for optimization in technical problems]. In: Pokorádi L. (szerk.), Műszaki Tudomány az Észak-kelet Magyarországi régióban 2011. Miskolc, Magyarország, 2011.05.18. Debrecen: Debreceni Akadémiai Bizottság (DAB), pp. 499–507. (Elektronikus Műszaki Füzetek; 9.) (ISBN: 978-963-7064-25-8)
- [10] Mankovits T., Vámosi A., Kocsis I., Huri D., Kállai I., Szabó T. (2015), Shape design of axially symmetric rubber part using finite element method and support vector machines. In: Bodzás S., Mankovits T. (szerk.) Proceedings of the 3rd International Scientific Conference on Advances in Mechanical Engineering (ISCAME 2015). Debrecen, Magyarország, 2015.11.19. Debrecen: University of Debrecen Faculty of Engineering, 2015. pp. 114–119. (ISBN: 978-963-473-917-3)
- [11] Mankovits T., Szabó T., Kocsis I., Páczelt I. (2014), Optimization of the shape of axi-symmetric rubber bumpers, Strojnicki Vestnik. Journal of Mechanical Engineering, 60(1), 61–71.
- [12] Kocsis I., Mankovits T. (2013), Application of non-parametric regression in engineering optimization. Analele Universitatii din Oradea Fasciola Management si Inginerie Tehnologica / Annals of the University of Oradea Fascicle of Management and Technological Engineering, 12(22), 159–162.
- [13] Mankovits T., Kocsis I., Portik T., Szabó T., Páczelt I. (2013), Shape design of rubber part using FEM. International Review of Applied Sciences and Engineering, 4(2), 85–94.
- [14] Mankovits T., Vámosi A., Kocsis I., Huri D., Kállai I., Szabó T. (2015), Shape design of axially symmetric rubber part using finite element method and support vector machines. In: Bodzás S., Mankovits T. (szerk.) Proceedings of the 3rd International Scientific Conference on Advances in Mechanical Engineering (ISCAME 2015). Debrecen, Magyarország, 2015.11.19. Debrecen: University of Debrecen Faculty of Engineering, pp. 114–119. (ISBN: 978-963-473-917-3)