Periodica Polytechnica Mechanical Engineering, 67(4), pp. 270–275, 2023

Overall Equipment Effectiveness Prediction with Multiple Linear Regression for Semi-automatic Automotive Assembly Lines

Péter Dobra[1*], János Jósvai[2]

[1] Doctoral School of Multidisciplinary Engineering Sciences, Széchenyi István University, Egyetem tér 1., H-9026 Győr, Hungary
[2] Department of Vehicle Manufacturing, Széchenyi István University, Egyetem tér 1., H-9026 Győr, Hungary
[*] Corresponding author, e-mail: dobra.peter@sze.hu

Received: 30 March 2023, Accepted: 01 September 2023, Published online: 07 September 2023

Abstract

In the field of industry, especially in the production areas, it is particularly important that the monitoring of assembly efficiency takes place in real-time mode, and that the related data-based estimation also works quickly and reliably. The Manufacturing Execution System (MES), Enterprise Resource Planning (ERP) and Customer Relationship Management (CRM) systems used by companies provide excellent support in data recording, processes, and storing. For Overall Equipment Effectiveness (OEE) data showing the efficiency of assembly lines, there is a regular need to determine expected values. This paper focuses on OEE values prediction with Multiple Linear Regression (MLR) as supervised machine learning. Many factors affecting OEE (e.g., downtimes, cycle time) are examined and analyzed in order to make a more accurate estimation. Based on real industrial data, we used four different methods to perform prediction with various machine learning algorithms, these were the cumulative, fix rolling horizon, optimal rolling horizon and combined techniques. Each method is evaluated based on similar mathematical formulas.

Keywords

OEE, machine learning, multiple linear regression, assembly line, prediction

1 Introduction

In the domain of industry, especially in the production areas, it is particularly important that the monitoring of assembly efficiency takes place in real-time mode, and that the related data-based estimation also works quickly and reliably. The efficiency of the production lines affects the operation of the entire company, therefore, it is important to predict future values as accurately as possible. Efficiency Key Performance Indicators (KPI's) have an impact on among others financial results (e.g., profits), production scheduling (e.g., assembly sequence), inventory (e.g., finished products), investments (e.g., transfer new machines), maintenance (e.g., required planned repair time) and continuous improvement (e.g., optimization of assembly processes).

Nowadays, Overall Equipment Effectiveness (OEE) is the most common efficiency metric in industrial practice. This standard indicator clearly shows current status of production and also includes different downtimes and scraps [1–3].

The aim of this paper is to predict the OEE values using machine learning. The paper is organized as follows. Section 2 focuses on the relevant scientific work regarding OEE prediction. Then, Section 3 reveales four different Multiple Linear Regression (MLR) ways such as cumulative, fix rolling horizon, optimal rolling horizon, and combined method. Section 4 concludes the paper.

2 OEE prediction with machine learning

OEE is a part of Total Productive Maintenance (TPM) concept and the basic formula for calculation is written as:

$$\mathrm{OEE} = a\ p\ q, \tag{1}$$

where a is the availability (%), p is the performance (%), q is the quality (%) [4].

Numerous systems, among others Manufacturing Execution System (MES), Enterprise Resource Planning (ERP) and Customer Relationship Management (CRM) provide assistance in the automatic recording, processing and storage of OEE values in the assembly lines [5–7].

Machine learning techniques can be divided into three categories: supervised learning, unsupervised learning and reinforcement learning [8, 9]. All types and its components

Cite this article as: Dobra, P., Jósvai, J. "Overall Equipment Effectiveness Prediction with Multiple Linear Regression for Semi-automatic Automotive Assembly Lines", Periodica Polytechnica Mechanical Engineering, 67(4), pp. 270–275, 2023. https://doi.org/10.3311/PPme.22302

can be used for OEE prediction in the field of assembly operations [10]. These estimations are aimed at the following areas:

- machine, station or tool failure;
- downtime occurrence;
- process parameters;
- products failure;
- production planning;
- type change errors, etc. [11–14].

The most commonly used machine learning methods and algorithms in OEE and its components prediction from simple to complex are:

- Logistic Regression [15–17];
- Gaussian Naive Bayes [15, 16];
- K-Nearest Neighbor [15, 18, 19];
- Bayesian Ridge Regression [20, 21];
- Decision Tree Regression Algorithm [19, 22];
- Random Forest [15, 16, 22, 23];
- Support Vector Machine [15, 17, 19, 22, 24, 25];
- Support Vector Regression Generic Algorithm [23, 26, 27];
- Extreme Gradient Boost [15, 23, 26];
- Artificial Neural Network [19, 24, 28, 29];
- Deep Learning [23, 30];
- Combined methods [31].

2.1 Multiple Linear Regression

Among the machine learning methods discussed in the scientific literature, with numerous practical examples, MLR is not included, despite its potential for predicting the value of OEE. MLR is a statistical technique that uses several explanatory variables to estimate the outcome of a response variable. Calculation of MLR:

$$z_i = A_0 + A_1\ x_{i1} + A_2\ x_{i2} + \ldots + A_p\ x_{ip}, \tag{2}$$

where z_i is the dependent variable, A_0 is the y intercept, x_i is the explanatory variables, A_p is the slope coefficients for each explanatory variable, and for $i = n$ observations (n observation of one dependent variable and p independent variables).

This paper presents the prediction of OEE values through Multiple Linear Regression.

3 OEE prediction with multiple linear regression

When predicting the OEE, the authors followed the classic data processing and evaluation, which includes the following steps:

- Exploring the data;
- Cleaning the data;
- Data visualization;
- Building the model;
- Training the model;
- Predicting with the model;
- Evaluating the performance of the model.

Each step is presented in detail in Sections 3.1 to 3.7.

3.1 Exploring the data

Data from the seat structure hybrid assembly line of a Central European automotive supplier from the years of 2021 and 2022 were used. The original real data were extracted from the MES system and an SQL database. Four different main databases were used, such as OEE data, downtime data, products timestamp data, and quality data. In this article, eight hour (one shift) data is used as a record during machine learning.

3.2 Cleaning the data

The analyzed production data comes from a completely automatic source, so there was no need for major data cleaning. Only those items were excluded where there was no data for the entire shift, for example assembly operations were completed earlier (the workers continued production on another assembly line in that shift). It is important to note that the authors tried to model real production conditions and environment, so extreme values were not excluded.

3.3 Data visualization

When visualizing the data, the most important thing was the representation of the original OEE values, where it can be seen that the examined semi-automatic assembly line is in a slow growing phase. Fig. 1 shows the original OEE data, one data represent an 8-hours shift.

3.4 Building the model

The following key independent variables were considered for multiple linear regression: process failure downtime, break downtime, technical downtime, changeover downtime, quality reason downtime, logistics reason downtime, not planned downtime, other downtime reason, number of changeovers, average cycle time, number of assembled units, and number of scrap units. Generally, OEE forecasting models can be used for either production planning or industrial investment purposes. In this article, for the sake of clarity, the authors present the OEE forecast for

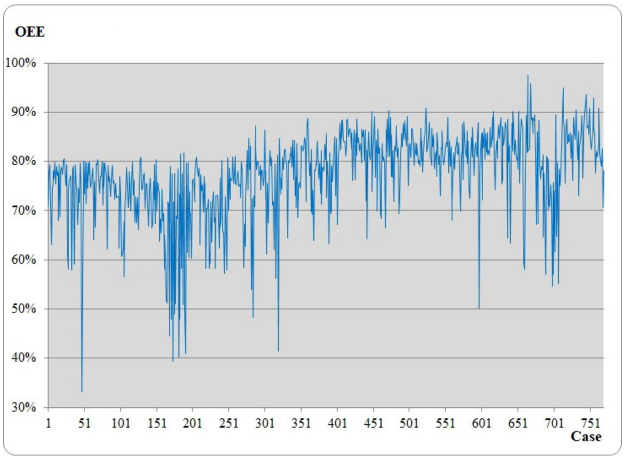

Fig. 1 OEE values at a semi-automatic assembly line

the case of an investment analysis. From a practical point of view, in the case of the industrial investment analysis, the prediction goal is to determine with the help of machine learning how the OEE values develop and, based on these, to decide whether a new assembly line is needed or whether new customer demands can be accepted.

R and R Studio program was used for the entire analysis, in addition to the rolling horizon data, cumulated data series was applied. During the rolling horizon approach, the size of the time windows can change, both for the length of the period and for the starting time, while in the cumulative approach, the cycles are counted from the first shift.

3.5 Training the models

The data shown in Table 1 was used as a basis for the prediction models.

The selected sample period and the predicting period follow industrial practice, data for the next three months are predicted monthly based on the last three months.

3.6 Predicting with the models

During the prediction, the following four models were run and evaluated:

- Cumulative method of multiple linear regression;
- Fix rolling horizon (50 records) method of multiple linear regression;
- Optimal rolling horizon (training records changes) method of multiple linear regression;
- Combined method (cumulative and optimal rolling horizon) of multiple linear regression.

Table 1 Periods for industrial investment analysis

Sample period	Predicted period	Cycle step
150 records (3 month)	150 records (3 month)	50 records (1 month)

After running the analysis, numerous patterns and correlations were revealed thanks to the many factors taken into account. Fig. 2 shows a scatter plot example where the not planned downtime and OEE value presented as a revealed pattern.

Fig. 3 shows an example for the cumulative method of MLR where training set was the data of records 1–200 and test set was 201–350. The blue line represents the predicted values, and the red line represents the actual OEE values.

3.7 Evaluating the performance of the models

The predicted values were evaluated using three metrics: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Mean Squared Error (MSE) based on Eqs. (3) to (5):

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y}\right), \quad (3)$$

$$\text{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left(\frac{y_i - \bar{y}}{y_i}\right), \quad (4)$$

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2, \quad (5)$$

where n is the number of fitted points, y_i is the actual value, $\bar{y}$ is the predicted value.

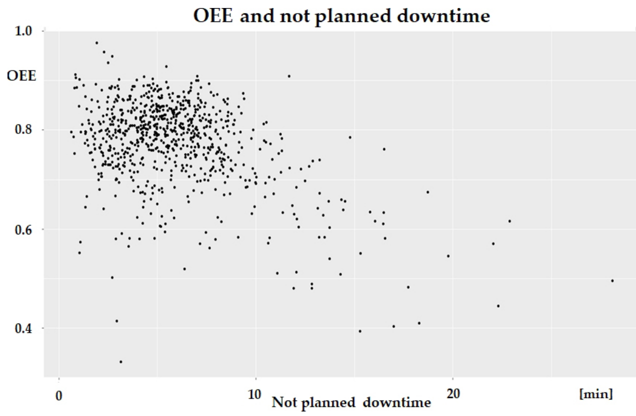

Fig. 2 Not planned downtime and OEE values as a revealed pattern

Fig. 3 Cummulative method of multiple linear regression, 1–200, 201–350

In addition, the main measure of prediction accuracy was the adjusted R-squared. Detailed evaluation results are shown in Tables 2 to 5 for each model. Based on the obtained values and metrics, a decision can be made as to whether a new investment is necessary for the given assembly line or whether an additional order can be accepted. (In reality, no new investments were made, but additional customer orders were accepted.)

Among the individual models, the combined method of MLR performed the best in terms of adjusted R-squared (0.8422), while the fixed rolling horizon method performed best in terms of the MAPE value (4.7723). Out of the many prediction variations, only one is presented in this article, but there is certainly an optimal prediction period and sampling.

4 Conclusion

This paper presented OEE prediction techniques using MLR as supervised machine learning at the domain of semi-automatic assembly lines. Prediction with real and validated industrial data analysis and compares four different ways such as cumulative, fix rolling horizon, optimal rolling horizon and combined method. Regarding industrial investments analyzis, the fixed rolling horizon

Table 2 Cumulative method of MLR

Sample period	1	1	1	1	1	1	1	1	1	1	Average
	150	200	250	300	350	400	450	500	550	600	
Predicted period	151	201	251	301	351	401	451	501	551	601	
	300	350	400	450	500	550	600	650	700	750	
MAE	0.0397	0.0301	0.0338	0.0372	0.0454	0.0437	0.0401	0.0439	0.0432	0.0485	0.0406
MAPE	6.1170	4.1910	4.4141	4.6804	5.5123	5.2550	4.8898	5.3252	5.3646	6.0414	5.1791
MSE	0.0031	0.0018	0.0022	0.0022	0.0030	0.0027	0.0023	0.0028	0.0028	0.0036	0.0026
Multiple R-squared	0.7101	0.8336	0.8225	0.8098	0.8137	0.7944	0.7935	0.7854	0.7780	0.7745	0.7916
Adjusted R-squared	0.6847	0.8229	0.8135	0.8019	0.8071	0.7880	0.7879	0.7801	0.7731	0.7699	0.7829
p-value	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16

Table 3 Fix rolling horizon method of Multiple Linear Regression

Sample period	1	51	101	151	201	251	301	351	401	451	Average
	150	200	250	300	350	400	450	500	550	600	
Predicted period	151	201	251	301	351	401	451	501	551	601	
	300	350	400	450	500	550	600	650	700	750	
MAE	0.0397	0.0325	0.0386	0.0376	0.0467	0.0393	0.0334	0.0288	0.0326	0.0394	0.0369
MAPE	6.1170	4.4428	4.9991	4.7655	5.6922	4.7222	4.0832	3.5102	4.2038	5.1869	4.7723
MSE	0.0031	0.0034	0.0037	0.0023	0.0032	0.0022	0.0019	0.0013	0.0019	0.0030	0.0026
Multiple R-squared	0.7101	0.8325	0.8532	0.8784	0.8078	0.7133	0.6880	0.6257	0.5789	0.6858	0.7374
Adjusted R-squared	0.6847	0.8178	0.8404	0.8677	0.7955	0.6927	0.6656	0.6017	0.5486	0.6632	0.7178
p-value	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16

Table 4 Optimal rolling horizon method of Multiple Linear Regression

Sample period	1	51	101	151	151	151	151	151	151	151	Average
	150	200	250	300	300	300	300	300	300	300	
Predicted period	151	201	251	301	351	401	451	501	551	601	
	300	350	400	450	500	550	600	650	700	750	
MAE	0.0397	0.0325	0.0386	0.0376	0.0420	0.0446	0.0462	0.0546	0.0547	0.0606	0.0451
MAPE	6.1170	4.4428	4.9991	4.7655	5.1415	5.3832	5.6589	6.6440	6.7539	7.4597	5.7366
MSE	0.0031	0.0034	0.0037	0.0023	0.0027	0.0029	0.0030	0.0041	0.0042	0.0051	0.0034
Multiple R-squared	0.7101	0.8325	0.8532	0.8784	0.8784	0.8784	0.8784	0.8784	0.8784	0.8784	0.8545
Adjusted R-squared	0.6847	0.8178	0.8404	0.8677	0.8677	0.8677	0.8677	0.8677	0.8677	0.8677	0.8417
p-value	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16

Table 5 Combined method of Multiple Linear Regression

Sample period	1 150	1 200	101 250	151 300	151 300	151 300	151 300	151 300	151 300	151 300	Average
Predicted period	151 300	201 350	251 400	301 450	351 500	401 550	451 600	501 650	551 700	601 750	
MAE	0.0397	0.0301	0.0386	0.0376	0.0420	0.0446	0.0462	0.0546	0.0547	0.0606	0.0449
MAPE	6.1170	4.1910	4.9991	4.7655	5.1415	5.3832	5.6589	6.6440	6.7539	7.4597	5.7114
MSE	0.0031	0.0018	0.0037	0.0023	0.0027	0.0029	0.0030	0.0041	0.0042	0.0051	0.0033
Multiple R-squared	0.7101	0.8336	0.8532	0.8784	0.8784	0.8784	0.8784	0.8784	0.8784	0.8784	0.8546
Adjusted R-squared	0.6847	0.8229	0.8404	0.8677	0.8677	0.8677	0.8677	0.8677	0.8677	0.8677	0.8422
p-value	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16

approach performed best among the four methods according to the MAPE, while the combined method provide the best according to the adjusted R-squared.

The authors plan to further compare the mentioned four methods in the field of production planning. Future research could be the application of other machine learning methods in the prediction of OEE.

References

[1] Muchiri, P., Pintelon, L. "Performance measurement using overall equipment effectiveness (OEE): literature review and practical application discussion", International Journal of Production Research, 46(13), pp. 3517–3535, 2008. https://doi.org/10.1080/00207540601142645

[2] Fam, S.-F., Loh, S. L., Haslinda, M., Yanto, H., Khoo, L. M. S., Yong, D. H. Y. "Overall Equipment Effectiveness (OEE) Enhancement in Manufacture of Electronic Components and Boards Industry through Total Productive Maintenance Practices", MATEC Web of Conferences, 150, 05037, 2018. https://doi.org/10.1051/matecconf/201815005037

[3] Gujar, A. J., Kambale, N. M., Maner, S. D., Joshi, S. S., Chandne, S. G., Chavare, A. A. "A case study for Overall Equipment Effectiveness improved in manufacturing industry", International Research Journal of Engineering and Technology, 6(3), pp. 4841–4844, 2019.

[4] Nakajima, S. "Introduction to TPM: Total Productive Maintenance", Productivity Press, 1988. ISBN 0915299232

[5] Singh, R., Shah, D. B., Gohil, A. M., Shah, M. H. "Overall Equipment Effectiveness (OEE) calculation - automation through hardware & software development", Procedia Engineering, 51, pp. 579–584, 2013. https://doi.org/10.1016/j.proeng.2013.01.082

[6] Mantravadi, S., Møller, C. "An overview of next-generation Manufacturing Execution system: How important is MES for Industry 4.0?", Procedia Manufacturing, 30, pp. 588–595, 2019. https://doi.org/10.1016/j.promfg.2019.02.083

[7] Iannone, R., Nenni, M. E. "Managing OEE to optimize factory performance", In: Schiraldi, M. (ed.) Operations Management, IntechOpen, 2013, pp. 31–50. ISBN 978-953-51-1013-2 https://doi.org/10.5772/55322

[8] Smola, A., Vishwanathan, S. V. N. "Intorduction to Machine Learning", Cambridge University Press, 2008. ISBN 0-521-82583-0

[9] Fahle, S., Prinz, C., Kuhlenkötter, B. "Systematic review on machine learning (ML) methods for manufacturing processes – Identifying artificial intelligence (AI) methods for field apploication", Procedia CIRP, 93, pp. 413–418, 2020. https://doi.org/10.1016/j.procir.2020.04.109

[10] Wuest, T., Weimer, D., Irgens, C., Thoben, K.-D. "Machine learning in manufacturing: advantages, challenges, and applications", Production & Manufacturing Research, 4(1), pp. 23–45, 2016. https://doi.org/10.1080/21693277.2016.1192517

[11] Lieber, D., Stolpe, M., Konrad, B., Deuse, J., Morik, K. "Quality prediction in interlinked manufacturing processes based on supervised & unsupervised machine learning", Procedia CIRP, 7, pp. 193–198, 2013. https://doi.org/10.1016/j.procir.2013.05.033

[12] Carvajal Soto, J. A., Tavakolizadeh, F., Gyulai, D. "An online machine learning framework for early detection of product failures in an Industry 4.0 context", International Journal of Computer Integrated Manufacturing, 32(4–5), pp. 452–465, 2019. https://doi.org/10.1080/0951192X.2019.1571238

[13] Espadinha-Cruz, P., Godina, R., Rodrigues, E. M. G. "A review of data mining applications in semiconductor manufacturing", Processes, 9(2), 305, 2021. https://doi.org/10.3390/pr9020305

[14] Harding, J. A., Shahbaz, M., Srinivas, Kusiak, A. "Data mining in manufacturing: A review", Journal of Manufacturing Science and Engineering, 128(4), pp. 969–976, 2006. https://doi.org/10.1115/1.2194554

[15] Peres, R. S., Barata, J., Leitao, P., Garcia, G. "Multistage quality control using machine learning in the automotive industry", IEEE Access, 7, pp. 79908–79916, 2019. https://doi.org/10.1109/ACCESS.2019.2923405

[16] Syafrudin, M., Alfian, G., Fitriyani, N. L., Rhee, J. "Performance analysis of IoT-based sensor, Big Data processin, and Machine Learning model for real-time monitoring system in automotive manufacturing", Sensors, 18(9), 2946, 2018.
https://doi.org/10.3390/s18092946

[17] Imane, M., Aoula, E.-S., Achouyab, E. H. "Using Support Vector Regression to predict the Overall Equipment Effectiveness indicator", In: 2022 International Conference on Intelligent Systems and Computer Vision (ISCV), Fez, Morocco, 2022, pp. 1–5. ISBN 978-1-6654-9559-2
https://doi.org/10.1109/ISCV54655.2022.9806111

[18] Abdelrahman, O., Keikhosrokiani, P. "Assembly line anomaly detection and root cause analysis using machine learning", IEEE Access, 8, pp. 189661–189672, 2020.
https://doi.org/10.1109/ACCESS.2020.3029826

[19] Vilela De Souza, B., Barros Dos Santos, S. R., De Oliveira, A. M., Givigi, S. N. "Analyzing and predicting Overall Equipment Effectiveness in manufacturing industries using machine learning", In: 2022 IEEE International Systems Conference (SysCon), Montreal, QC, Canada, 2022, pp. 1–8. ISBN 978-1-6654-3993-0
https://doi.org/10.1109/SysCon53536.2022.9773846

[20] Imane, M., Aoula, E.-S., Achouyab, E. H. "Using Bayesian Ridge Regression to predict the Overall Equipment Effectiveness performance", In: 2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Meknes, Morocco, 2022, pp. 1–4. ISBN 978-1-6654-2210-9
https://doi.org/10.1109/IRASET52964.2022.9738316

[21] Khdoudi, A., Barka, N., Masrour, T., El Hassani, I., El Mazgualdi, C. "Online prediction of automotive tempered glass quality using machine learning", Research Square, 2022.
https://doi.org/10.21203/rs.3.rs-2040065/v1

[22] Lee, J. H., Noh, S. D., Kim, H.-J., Kang, Y.-S. "Implementation of Cyber-Physical Production System for quality prediction and operation control in metal casting", Sensors, 18(5), 1428, 2018.
https://doi.org/10.3390/s18051428

[23] El Mazgualdi, C., Masrour, T., El Hassani, I., Khdoudi, A. "Machine learning for KPIs prediction: a case study of the overall equipment effectiveness within automotive industry", Soft Computing, 25(4), pp. 2891–2909, 2021.
https://doi.org/10.1007/s00500-020-05348-y

[24] Lejon, E., Kyösti, P., Lindström, J. "Machine learning for detection of anomalies in press-hardening: Selection of efficient methods", Procedia CIRP, 72, pp. 1079–1083, 2018.
https://doi.org/10.1016/j.procir.2018.03.221

[25] Wang, T., Chen, J., Gao, X., Qin, Y. "Real-time monitoring for disk laser welding based on feature selection and SVM", Applied Science, 7(9), 884, 2017.
https://doi.org/10.3390/app7090884

[26] El Mazgualdi, C., Masrour, T., El Hassani, I., Khdoudi, A. "Using machine learning for predicting efficiency in manufacturing industry", In: Ezziyyani, M. (ed.) Advanced Intelligent Systems for Sustainable Development (AI2SD'2019), Springer, 2019, pp. 750–762. ISBN 978-3-030-36670-4
https://doi.org/10.1007/978-3-030-36671-1_68

[27] Zouhri, W., Rostami, H., Homri, L., Dantan, J.-Y. "A genetic-based SVM approach for quality data classification", In: Masrour, T., Cherrafi, A., El Hassani, I. (eds.) Artificial Intelligence and Industrial Applications, Springer, 2021, pp. 15–31. ISBN 978-3-030-51185-2
https://doi.org/10.1007/978-3-030-51186-9_2

[28] Khdoudi, A., Masrour, T. "Prediction of industrial process parameters using artificial intelligence algorithms", [preprint] arXiv:1812.10537v2, 03 February 2019.
https://doi.org/10.48550/arXiv.1812.10537

[29] Khdoudi, A., Masrour, T., El Mazgualdi, C. "Using machine learning algorithms for the prediction of industrial process parameters based on produc design", In: Ezziyyani, M. (ed.) Advanced Intelligent Systems for Sustainable Development (AI2SD'2019), Springer, 2019, pp. 728–749. ISBN 978-3-030-36670-4
https://doi.org/10.1007/978-3-030-36671-1_67

[30] El Mazgualdi, C., Masrour, T., Barka, N., El Hassani, I. "A learning-based decision tool towards smart energy optimization in the manufacturing process", Systems, 10(5), 180, 2022.
https://doi.org/10.3390/systems10050180

[31] Bonada, F., Echeverria, L., Domingo, X., Anzaldi, G. "AI for improving the Overall Equipment Efficiency in manufacturing industry", In: Martínez, L. R., Rios, R. A. O., Prieto, M. D. (eds.) New trends in the use of Artificial Intelligence for the Industry 4.0, IntechOpen, 2020, pp. 1–22. ISBN 978-1-83880-141-0
https://doi.org/10.5772/intechopen.89967