Title: **Evaluating predictive capabilities in Industry 4.0 framework using Regression Models**

Author: **Shibajyoti Banerjee**
Faculty of Instrumentation and Biotechnical Systems (alumnus)
Saint Petersburg ElectroTechnical University (ETU LETI)
shbanerdzhi@stud.eltech.ru

Abstract

The recent trends and changes brought by the Fourth Industrial Revolution (I4.0) have succeeded in proposing a better future as well as ways to handle enormous amounts of entangled data generated daily. However, there is a need for visualizing and processing this information, so that it can be utilized fruitfully. To assess the performance of industrial and manufacturing equipment's overtime, prolonged monitoring of data in real-time is required. Predictive analysis is a topology that relies on prevising the future parameters of an automated process or system using computerized algorithms and previously obtained data or information and notifying about the fall in performance parameters.

This work focuses on analyzing the data obtained from a fatigue machine-based experiment, followed by applications of regression algorithms to predict and compare theoretically and predicted values.

Keywords: Predictive Regression, System Performance, Fourth Industrial Revolution, Fault Diagnosis

1. Introduction

Three Industrial Revolutions have so far prompted worldview changes in the area of assembling automation through water and steam power, large-scale manufacturing in the get-together lines, and computerization utilizing data innovation. Compared to the previous years, businesses together with analysts and approach producers worldwide have progressively pushed towards a Fourth Industrial Revolution (I 4.0). The I4.0 is described by the presentation of the Internet of things (IoT) into assembling, which empowers brilliant industrial facilities with coordinated frameworks. The subsequent intelligent processing plants can satisfy dynamic client requests with high fluctuation while coordinating human resourcefulness and automation. To help the assembling business during this transition and improve worldwide intensity, arrangement creators in a few nations have built up research and innovation plans. This shows that information has become a critical theme for scientists and businesses around the world. The fundamental object is the supposed Cyber-physical systems (CPS): Physical Elements (e.g., machines, vehicles, and workpieces), furnished with advancements, such as RFIDs, sensors, chips, telematics, or complete installed frameworks **[1]**.

1.1. The Fourth Industrial Revolution or Industry 4.0 (I 4.0)

At the yearly Davos meeting of the World Economic Forum (WEF) of 2016, there was a serious discussion on the issue of the Fourth Industrial Revolution or I4.0, and a few parts of this new stage or pattern of industrial advancement were introduced by participants, as indicated by WEF Chairman Klaus Schwab who presented the term and subject of the Fourth Industrial Revolution in Davos discussion **[2]**. The popularity of I 4.0 is further uplifted by factors such as digitization and steady correspondence, stepwise increment in proficiency, adequacy of the financial framework activity, and finally the accomplishment of self-governing conduct using brainpower integrated with autonomous control **[3]**.

1.2. Smart Manufacturing

Smart Manufacturing coordinates fabricating resources of today and tomorrow with sensors, processing stages, correspondence innovation, and information-based demonstration, control, reproduction, and predictive engineering. Smart Manufacturing aggregate CPS, IoT, Cloud Computing, administration situated registering, Artificial Intelligence (A.I.) and Data Science into one. When actualized, these covering ideas and innovations make fabricating the sign of an efficient change in industrial systems. The following elucidates the typical requirements of an intelligent manufacturing process **[4]:**

i. New materials, component and additive manufacturing, laser and net-shape manufacturing, manufacturing integration, factory automation, sensor, and software implementation.
ii. Implementing sensors, wireless technology, and data analytics. Collection of data from diverse sources, to power applications and build predictive models. Preserving and extracting past knowledge related to manufacturing.
iii. For process monitoring, quality control, and productivity analysis. Predictive engineering allows exploring the future aspects of present data to support decisions concerning future productions.
iv. Sustainable product design for manufacturing sustained manufacturing process, as well as the creation of sustainable development and processes.
v. Resource Sharing and Networking allows the sharing of vital information among businesses to gain competition.

1.3. Predictive Maintenance

The development of current strategies (e.g., Internet of Things, Sensor innovation, Artificial Intelligence, and so forth.) mirrors a change of support techniques from Reactive Maintenance (RM), Preventive Maintenance (PM) to Predictive Maintenance (Pd.M.). RM is just executed to re-establish the working condition of the device after a fault happens, and in this way will in general reason result in high responsive fix costs. PM is completed by an arranged calendar-dependent schedule or procedure emphasized to forestall breakdown, and along these lines may perform superfluous upkeep with high counteraction costs. Pd.M. permits the support recurrence to be as low as possible by forestalling impromptu RM, without bringing about expenses related to doing an excessive amount of PM **[5]**. Predictive Maintenance (Pd.M.) is a widely used technique and typically follows the algorithms of conditional monitoring, fault diagnosis, fault prognosis, and finally maintenance. It possesses the capability of predetermining areas of prospective lags, faults, and breakdowns taking into consideration factors that govern the efficiency of the system.

1.4. Predictive Maintenance in Industry 4.0 and Fault Diagnosis

Pd.M. 4.0, lined up with I 4.0 standards, paints an outline for insightful Pd.M. frameworks. I 4.0 is a change in the outlook of modern procedures and items pushed by intelligent data handling, correspondence frameworks, future arranged methods, and so forth. Smart machines and production lines utilize cutting-edge innovations, for example, organizing, associated gadgets, information examination, and synthetic reasoning to arrive at a progressively proficient Pd.M. This move on Pd.M. under the setting of Industry 4.0 is termed as Pd.M. 4.0 **[6]**. The method of Pd.M. involves a thorough analysis of collected data for premature detection of devices, hardware failure, and/or crashes. The maintenance strategies are further classified into four steps:

i. *Level 1*- Visual reviews: this level of behaviors, intermittent physical reviews, and upkeep methodologies are given on investigators' skills.
ii. *Level 2*- Instrument investigations: this level of behaviors occasional examinations and up-keep systems are based on a blend of investigators' aptitude and instrument outputs.
iii. *Level 3*- Ongoing condition checking: this level behaviors nonstop constant observing of benefits and cautions are given dependent on pre-built-up rules or basic levels.
iv. *Level 4*- Pd.M. 4.0: this level behaviors nonstop real-time checking, and alarms are conveyed based on prescient strategies, for example, regression analysis.

The phases in fault diagnosis involve feature extraction and selection as the first stage where the raw deciding features are extracted; this is followed by fault classification in which supervised or semi-supervised algorithms are deployed to detect anomaly **[7], [8]**. At present, the most effective algorithms for diagnosing faults include physical, data-driven, and statistical model-based assessment methods.

2. State of the Art

Quality has become a key factor for assembling organizations. To guarantee top-notch items, broad reviews are unavoidable. Because of the high asset utilization, the advancement of elective methodologies relates to the incredible enthusiasm to a look into and mechanical application. This empowers data-driven technologies, for example, the application of AI in industry. One field of intrigue along these lines is the perceptive quality where machine learning (ML) models are utilized to foresee the normal last item quality because of recently recorded procedure parameters. ML approaches for predictive maintenance have also been mentioned **[8],** where the authors propose a hybrid ML algorithm to tackle multiple data types and formats used in the Pd.M. context.

Data is indeed a vital part of Pd.M. A good dataset is one of the most primal requirements for effective implementation of Pd.M. since it links with the outcome to be recorded for future applications and prospects. Studies like those by **[9], [10]** dedicate their attention to this domain. The author proposes a methodology for Pd.M. of data by utilizing a physical Fischertechnik model processing plant that is furnished with a few sensors, can collect, process, and store the recorded data. Integration Planning is an essential part of the manufacturing process and has been discussed taking into account production and maintenance planning with predictive analysis **[11]**. Similarly, the work done by LaCasse *et. al* **[12]** encompasses surveys related to feature set reduction for data analysis methods in context with general industrial applications, specific industrial applications, and data reductions. The study highlights prospects for feature-based data prioritization. An article by Hapuwatte and Jawahir **[13]** presents a novel structure consolidating perceptive model with Total Life Cycle (TLC) contemplations. In addition, similar to the one described before by Cho *et. al* **[8],** Bezarov *et. al* **[14]** elaborates on the abilities and efficacy of Artificial Intelligence (AI) and Deep Neural Networks (DNN) in the I4.0 era. A review by Lee *et. al* **[15]** proposes practical applications of Pd.M. in designing Quality Management Systems. The study takes into account cases provided by organizations and the type of Pd.M. system they enforce in each case.

3. **A Regression Model-Based Analysis Example**

As mentioned previously, Pd.M. systems are implemented as a part of the I4.0 paradigm; the framework helps medium and large-scale industries and organizations to keep track of performance factors, as well as in handling the vast amounts of data that are regularly produced by embedded sensing and computation. In this work, the dataset was derived from an online data repository (**TU Delft) [17]** and it details a performance record for the parameters on a fatigue-based crack growth experiment **[16]** using a combination of experiments and numerical modeling. The parameters indicate measurements that were obtained. A two-dimensional analysis problem was considered to observe how the changes in one parameter reflect in the second. Additionally, since the study is intended to conclude Pd.M. analysis, the system under consideration is assumed as Non-Linear Time-Invariant.

In this study, an analysis of two variables has been done; '*N*' represents the number of cycles measured by the fatigue machine; while the observation variable is Cyclic_Energy in mJ which gives the output as the Cyclic energy produced by the machine per cycle (N). We conducted statistical computations using these two variables with the hopes of getting a better insight into the predictive analysis. The following are the steps that are levied for the above-mentioned analysis:

i. In the first step, the parameters of interest from the dataset were extracted a data frame or table was created. As mentioned earlier, the machine cycle counts i.e. Number of cycles (N) was taken to be the dependent variable while the Cyclic_Energy in millijoules was the independent variable.
ii. The slope and intercept values for the dataset were derived and plotted in log-scale.
iii. A box plot was created to visualize the data distribution. It was observed that most of the data lie on the outliers indicating a continued inconsistency (randomness) within the dataset.
iv. A regression plot of the data in steps of 100, 200, 1000, and a full dataset (6805 observations) was made to view its deviation and spread.
v. To implement a machine learning procedure, the model was initially trained with a ratio of 85% and 15% (ratio of training set to testing set).
vi. The next step was to run the model to assess and predict new values learned by the model using the Ordinary Least Square (OLS) method.
vii. Plots describing the comparison between original and predicted values have been depicted, and OLS statistics were computed.

4. **Observations and Results**

Two parameters as mentioned was extracted from the data set. In most cases, the data format from the available source is not always suitable for automated analysis and manipulation. Hence there is a need to pre-process the data. Additionally, the data for the dependent variable (*N*) was exponential. The dataset is a Comma Separated Value (.csv) type file bearing dimensions of 6805×2 as shown in Table 1.

Observation No.	Number of Cycles (N)	Cyclic Energy (milliJoules)
1	99.0	330.0
2	100.0	330.0
3	200.0	325.0
4	200.0	325.0
5	299.0	323.0
....	....	....
6801	342000.0	149.0
6802	342000.0	149.0
6803	342000.0	148.0
6804	342000.0	148.0
6805	342000.0	148.0

Table 1: List of Observations (6805 data points)

In this step, the slope, intercept, and plot of *Cyclic_Energy* vs *N* was made. It is to be noted that the horizontal axis is logarithmic to accumulate all the relevant data points, while the vertical axis is linear-scaled as shown in Fig. 1. From the analytics, it was also observed that the calculated slope is downwards (negative) indicating a decline in performance.

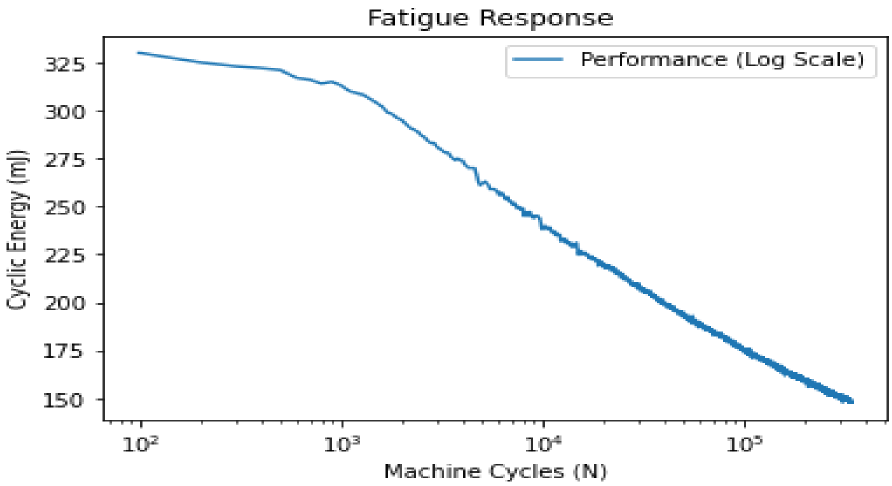

Figure 1: Cyclic_Energy (mJ) vs Number of Cycles

A Box and Whisker plot (Fig.2) was made to check the data distribution. It was observed that most of the data lie on the outliers stating a continued inconsistency (randomness) within the dataset.

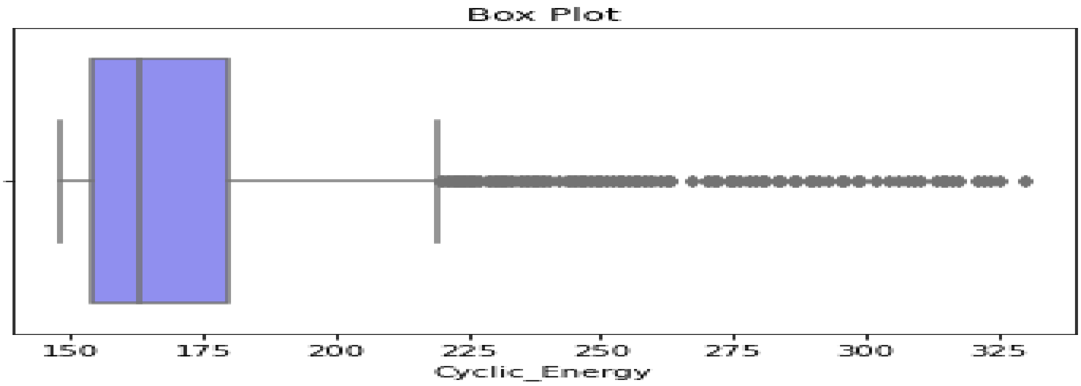

Figure 2: Box-Whisker spread Analysis

As a test for linearity of data which is an essential condition for predictive analytics. a regression plot was drawn taking into account samples of 50 (Fig 3a) 200 (Fig 3b), 1000 (Fig 3c), and complete dataset, i.e. 6805 samples (Fig 3d). As the number of observations increased, the linearity gradually disappears; and the output becomes non-linear. This is as depicted below.

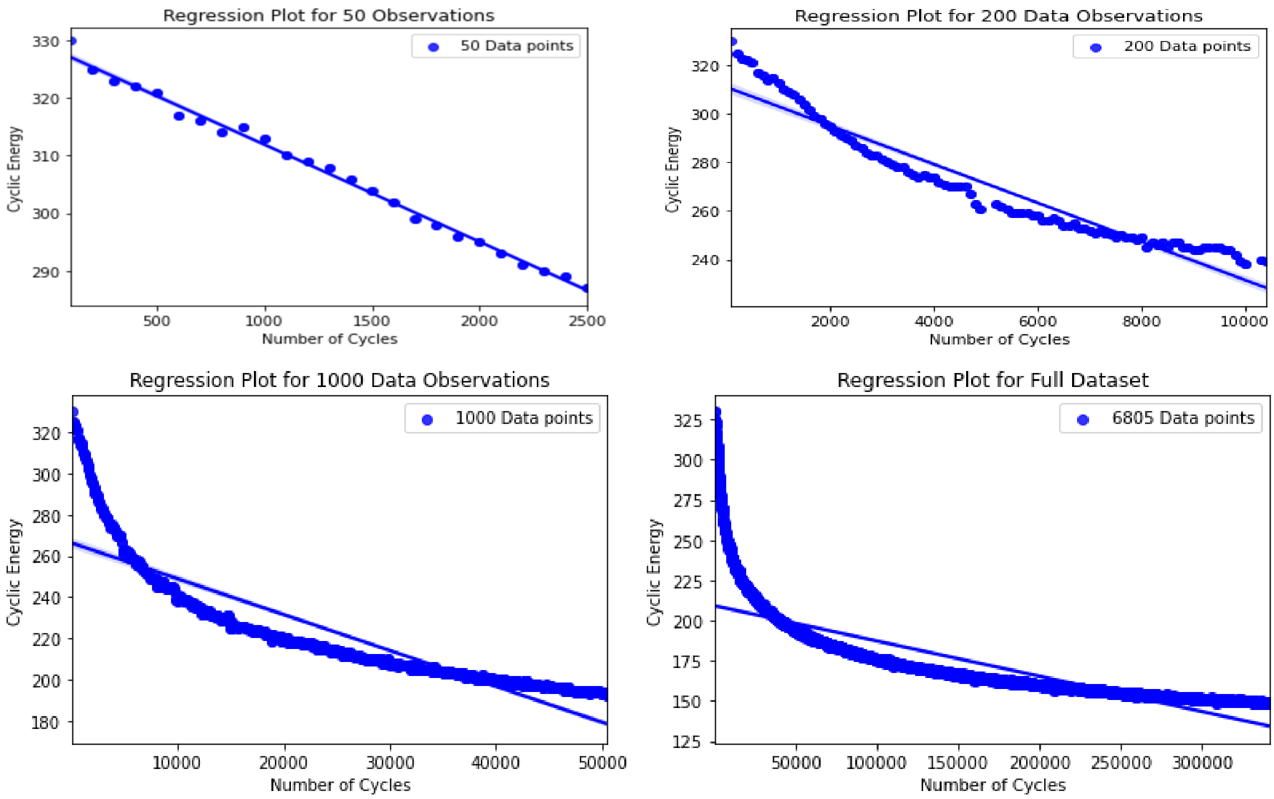

Figure 3: Loss of Functional Linearity over time

As a simple testing methodology, a predictive linear regression model was implemented. For experimental purposes, the ratio of training to testing data was taken at 85 % (~5784) and 15 % (~1021). Table 2 shows the values obtained after training with the model.

Observation Number	Actual Value (Cyclic Energy)	Predicted Value (Cyclic Energy)	Absolute Errors
2818	168.0	177.930258	9.930258
3688	161.0	168.575367	7.575367
5041	154.0	153.781585	0.218415
5878	152.0	144.557227	7.442773
6747	148.0	135.071802	12.928198
....	....	....	....
1400	185.0	193.441973	8.441973
3569	161.0	169.663145	8.663145
4886	155.0	155.304474	0.304474
5248	153.0	151.388473	1.611527
6805	238.0	206.647817	31.352183

Table 2: Predicted Value for test data (1021 values)

Fig 4 shows a bar diagram to visualize the correlations between the predicted and the original values. The graph shows a decent amount of proximity between the two variables.

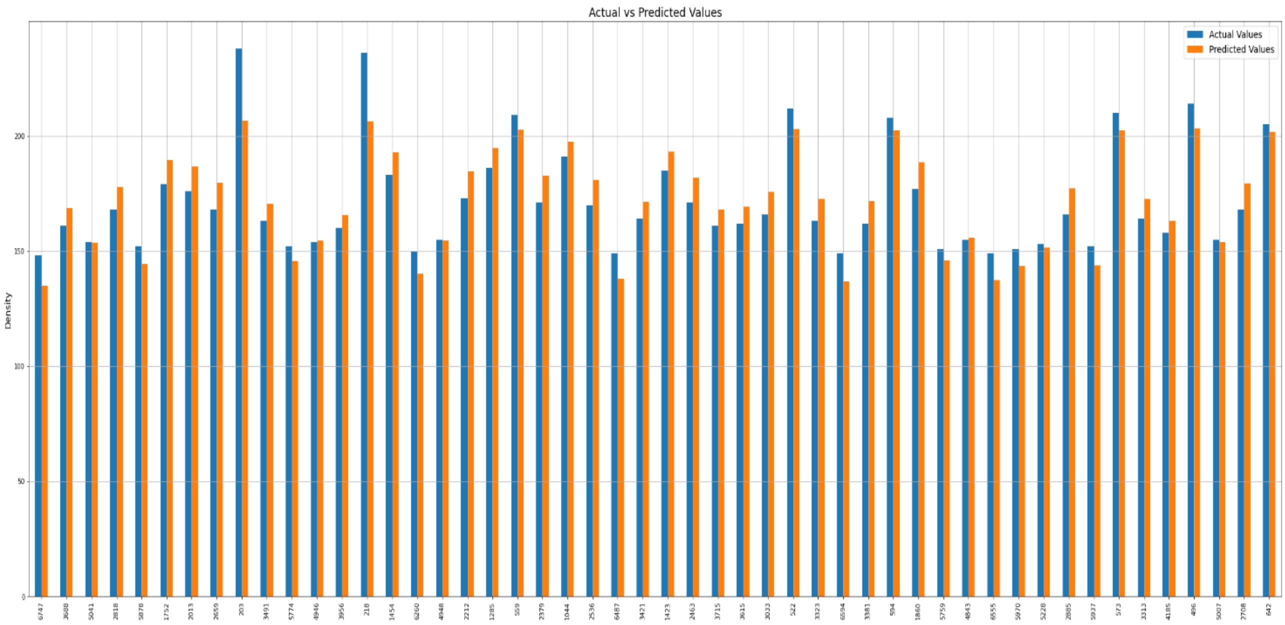

Figure 4: Original vs. Predicted value (50 Random observations)

A calculation of various statistical parameters was made to assess the viability of the results. Fig 1 indicates a negative slope of magnitude = - 0.000218 and an intercept of 209.047. Further calculations reveal a coefficient of determination $R^2 \sim 0.69$, which indicates a decent model performance. A skew value of 3.492 indicates extreme positive skewness, while kurtosis of ~ 21.630 implies heavy tails or outliers that conform to the observation in Figure 2.

5. Conclusion

Variations in data gathered in real-time often become hard to analyze. However taking resort of statistical and computational benefits, using big data and analytics, industrial efficiency can surge with better performance.

This work was limited to basic regression modeling (OLS). In the near future, with the help of superior algorithms and efficient methods, smart systems are expected to handle non-linear and inhomogeneous data with better efficiency.

6. Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

7. Data Source

All data used in this work were collected from the open data repository of TU DELFT [17].

References

1. Thoben, K. D., Wiesner, S., & Wuest, T. (2017). “Industrie 4.0” and smart manufacturing-a review of research issues and application examples. *International journal of automation technology*, *11*(1), 4-16.
2. Prisecaru, P. (2016). Challenges of the fourth industrial revolution. *Knowledge Horizons. Economics*, *8*(1), 57.
3. Ślusarczyk, B. (2018). Industry 4.0: Are we ready?. *Polish Journal of Management Studies*, *17*.
4. Kusiak, A. (2018). Smart manufacturing. *International Journal of Production Research*, *56*(1-2), 508-517.
5. Ran, Y., Zhou, X., Lin, P., Wen, Y., & Deng, R. (2019). A survey of predictive maintenance: Systems, purposes, and approaches. *arXiv preprint arXiv:1912.07383*.
6. Zhang, J., Zhang, D., Yang, M., Xu, X., Liu, W., & Wen, C. (2018, October). Fault Diagnosis for Rotating Machinery with Scarce Labeled Samples: A Deep CNN Method Based on Knowledge-Transferring from Shallow Models. In *2018 International Conference on Control, Automation and Information Sciences (ICCAIS)* (pp. 482-487). IEEE.
7. Xu, G., Liu, M., Wang, J., Ma, Y., Wang, J., Li, F., & Shen, W. (2019, August). Data-Driven Fault Diagnostics and Prognostics for Predictive Maintenance: A Brief Overview. In *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)* (pp. 103-108). IEEE.
8. Cho, S., May, G., Tourkogiorgis, I., Perez, R., Lazaro, O., de la Maza, B., & Kiritsis, D. (2018, August). A hybrid machine learning approach for predictive maintenance in smart factories of the future. In *IFIP International Conference on Advances in Production Management Systems* (pp. 311-317). Springer, Cham.
9. Klein, P., & Bergmann, R. (2018). Data Generation with a Physical Model to Support Machine Learning Research for Predictive Maintenance. In *LWDA* (pp. 179-190).
10. Klein, P., & Bergmann, R. (2019). Generation of Complex Data for AI-based Predictive Maintenance Research with a Physical Factory Model. In *ICINCO (1)* (pp. 40-50).
11. Rødseth, H., Schjølberg, P., Wabner, M., & Frieß, U. (2017, September). Predictive Maintenance for Synchronizing Maintenance Planning with Production. In *International Workshop of Advanced Manufacturing and Automation* (pp. 439-446). Springer, Singapore.
12. LaCasse, P. M., Otieno, W., & Maturana, F. P. (2019). A survey of feature set reduction approaches for predictive analytics models in the connected manufacturing enterprise. *Applied Sciences*, *9*(5), 843.
13. Hapuwatte, B. M., & Jawahir, I. S. (2019). A total life cycle approach for developing predictive design methodologies to optimize product performance. *Procedia Manufacturing*, *33*, 11-18.
14. Bezobrazov, S., Anfilets, S., Dolny, R., & Yakimovich, D. (2019). Predictive Analytics For Industry 4.0. *Industry 4.0*, *4*(6), 273-276.
15. Lee, S. M., Lee, D., & Kim, Y. S. (2019). The quality management ecosystem for predictive maintenance in the Industry 4.0 era. *International Journal of Quality Innovation*, *5*(1), 4.

16. Pascoe, J. A., Zavatta, N., Troiani, E., & Alderliesten, R. C. (2020). The effect of bond-line thickness on fatigue crack growth rate in adhesively bonded joints. *Engineering Fracture Mechanics*, 106959.
17. https://data.4tu.nl/repository/uuid:1495c951-33e2-43ab-9f41-5ed3b57ba02f