SPS2022
A.H.C. Ng et al. (Eds.)

doi:10.3233/ATDE220148

A Data Scientific Approach Towards Predictive Maintenance Application in Manufacturing Industry

Xinjie DUAN [a], Adarsh VASUDEVAN [a], Ebru TURANOGLU BEKAR [a,1], Kanika GANDHI [b], and Anders SKOOGH [b]

[a] *Industrial and Materials Science, Chalmers University of Technology, Gothenberg, Sweden*
[b] *Volvo Trucks Operation, Skövde, Sweden*

Abstract. Most industries have recently started to harness the power of data to assess their performance and improve their production systems for future competitiveness and sustainability. Therefore, utilization of data for obtaining insights through data-driven approaches is invading every domain of industrial applications. Predictive maintenance (PdM) is one of the highest impacted industrial use cases in data-driven applications due to its ability to predict machine failures by implementing machine learning algorithms. This study aims to propose a systematic data scientific approach to provide valuable insights by analysing industrial alarm and event log data, which might further be used for investigation in root cause understanding and planning of necessary maintenance activities. To do that, a Cross-Industry Standard Process for Data Mining (CRISP-DM) is followed as a reference model in this study. The results are presented by first understanding the relationship between alarms and product types being processed in the selected machines by using exploratory data analysis (EDA). Along with this, the behavior of problematic alarms is identified. Afterward, a predictive analysis formulated as a multi-class classification problem is performed using various Machine Learning (ML) models to predict the category of alarm and generate rules to be used for further investigation in maintenance planning. The performance of the developed models is evaluated based on the different metrics and the decision tree model is selected with the higher accuracy score among them. As a theoretical contribution, this study presents an implementation of predictive modeling in a structured way, which uses a systematic data scientific approach based on industrial alarm and event log data. On the other hand, as a practical contribution, this study provides a set of decision rules that can act as decision support for further exploration of possible in-depth root causes through the other contextual data, and hence it gives an initial foundation towards PdM application in the case company.

Keywords. Exploratory data analysis, CRISP-DM, Machine learning, Multi-class classification, Predictive maintenance, Manufacturing

[1] Corresponding Author: Ebru Turanoglu Bekar; E-mail: ebrut@chalmers.se.

1. Introduction

Industrial digitalization is a key enabler of future competitiveness and sustainability in manufacturing companies. Therefore, manufacturing companies have grown in importance to achieve a simple goal: reducing production disturbances by minimizing unplanned downtimes and increasing efficiency of production systems with a vision of failure-free production [1]. This industrial digitalization allows for increased automation and industrial data collection through increased deployment of information and communication technologies (i.e., smart sensors) where data is collected from machines and products in order to support decision-makers for the right decision at the right time [2]. In the industrial world, a good production line is vital for the manufacturing products to ensure the timely demands of the market, thereby constituting towards the growth of the organization. The output of the production in a factory is aligned towards the proper and uninterrupted functioning of the collective machines involved in the entire production line. Any stoppages or failure of machines in the production line leads to a loss in production time and causes economic adversities for the organization [3]. With this focus, there were an increase of Predictive Maintenance (PdM) solutions by using industrial big data [4].

PdM is defined as a set of techniques designed to help determine the condition of equipment to estimate when maintenance should be performed through the implementation of smart scheduling of maintenance actions that ultimately avoid (or at least mitigate the effects of) unexpected equipment failures [5]. PdM is one of the highest impacted industrial use cases in data-driven applications due to its ability to predict machine failures by implementing Machine Learning (ML) algorithms and thereby reducing maintenance costs and increasing productivity [6]. The implementation of PdM practices is widely accepted among the manufacturing industries. However, there are some challenges in implementing PdM in real-world industrial environments such as how to structure, analyze, integrate multi-source industrial big data sets, and build robust predictive models enabling maintenance decision support [7]. Therefore, there is a need for more scalable and systematic approaches for analyzing industrial big data and investigating what type of predictive algorithms can be designed for implementation of PdM in real-world industrial environments [8]. Knowing the aforementioned challenges, this paper aims to propose a data scientific approach based on Cross Industry Process Standard for Data Mining (CRISP-DM) [9] by systematically performing a descriptive, diagnostics, and then predictive analysis using a real world industrial big data containing industrial alarms, event log data and product types. As a conclusion, it provides valuable insights to the case company, which might be used as a further investigation in root cause understanding and planning of necessary maintenance activities towards PdM implementation.

The remainder of the paper is structured as follows. In Section 2, related literature including data-driven studies in PdM, which use industrial alarms and event log data, is summarized. Section 3 presents the data scientific approach by describing each step. Section 4 presents the results from real-world industrial application. Finally, Section 5 concludes the paper with a summary and further research directions.

2. Related Literature

In product realization life cycle, especially under a sustainable manufacturing context, maintenance has started shifting its role from "merely values" or "just repair works" to controlling the product's condition concerning the product's physical and functional life [10]. Most companies that have established their industrial system refrain from making investments in new plants as long as the current works safely and efficiently. Maintenance, consequently, has become one of the main impacts contributing to maximizing productivity under the circumstance [11]. As [12] points out, "Depending on the specific industry, maintenance costs can represent between 15 and 60 percent of the cost of goods produced". Hence, effective maintenance has proven to be essential to many operations, not only for reducing equipment downtime, but also to optimize THE overall availability or reliability of systems by minimizing costs [13]. With the advent of digitalization, a great deal of research in the maintenance field has focused on PdM which helps companies proactively improve the availability and reliability of manufacturing systems, extending the useful lifespan of equipment, and enhancing the quality of products [14]. PdM, to an extent, averted early interventions or late remediation which respectively wasted resources and the possibilities of causing irreversible failures [15]. The recent research studies indicate that the implementation of PdM is widely accepted among the manufacturing industries and thus PdM has gained a lot of attention in manufacturing due to its ability to predict machine failures by implementing data-driven algorithms based on ML [16]. Nonetheless, performing PdM in real-world industrial environments requires a scalable and systematic approach for analyzing high-dimensional industrial big data from acquisition to deployment in a structured way. This should consist of various steps such as business understanding and formulation of analysis goals, data understanding using exploratory data analysis (EDA), pre-processing of data, designing of predictive algorithms, and evaluation of performance [17]. In this context, to answer industry (business) requirements and data-related questions, CRISP-DM, a reference model for data mining projects has been applied in some studies related to PdM by using industrial data coming from only one data source which is mainly sensors [8,7]. At the same time, the challenge is to structure and integrate heterogeneous data coming from multiple data sources for building more robust PdM solutions [18]. This study aims to use a systematic data scientific approach by integrating multi-source data including industrial alarms, event log data, and product types to investigate relationships between using exploratory data analysis and to design ML models, which enables proper planning of maintenance activities through predictive modeling.

Literature review performed based on data-driven studies using alarms and event log data indicates that most of the studies have been focused on descriptive and diagnostics analysis of alarms and event log data. According to [19], the descriptive analysis emphasizes transparency and provides a description of the historical data set while the diagnostics analysis aims in finding the root cause of the problem through performing EDA. Considering the case of an industrial alarm data, the diagnostics analysis provides details to the operators regarding what had caused or triggered the alarms, which resulted in the process deviation [20]. According to a study performed by [21], alarm management and fault detection and diagnosis disciplines have close relations however they are still separately being performed in industrial practice. Therefore, their study pointed out the benefits of using alarm data in fault detection and diagnosis. In another study done by

[22], the authors also discussed incorporation of the alarms and event log data for fault diagnosis by demonstrating examples from a real-world case study. They further highlighted the importance of this analysis for process deviation, malfunctioning of equipment, the overall performance of the production and quality of the final products. In summary, although the fault detection and diagnosis help in ensuring plant safety and product throughput [23], it is important to further expand the analysis into a predictive approach in order to create a futuristic model which might help in predicting behavior of problematic alarms to be used for further investigation. This is also highlighted in another paper as a research gap in the maintenance field that there is a need for developing generic decision models, which represent the decision making process [24]. From the knowledge obtained from the literature review, this paper contributes to research on industrial alarm data by implementing a predictive analysis using a systematic data scientific approach to develop the generic decision model that represents the decision-making for efficient maintenance planning as a core function of PdM.

3. A Data Scientific Approach based on CRISP-DM

To establish a systematic data scientific approach, the CRISP-DM is followed as a reference model in this paper. It provides a structured methodology for planning and managing the data-driven knowledge discovery process in data mining projects [25]. The CRISP-DM divides the process of data mining into six main phases which are basically business understanding, data understanding, data preparation, modeling, evaluation, and deployment [25]. The process can be iterative and the phases also develop certain dependency within each other to ensure that the data is relevant and the results of the process align with the business objective defined in the business understanding phase of the process. However, it should be noted that the CRISP-DM is taken a basis in this paper. Hence, it is adjusted for practical development and implementation according to real-world industrial application requirements.

The proposed data scientific approach differs from the CRISP-DM with a stronger emphasis on EDA, which allows integration of domain experts' knowledge (i.e., technical understanding of data acquisition, machines and production processes) into data understanding and preparation phases. This supports overcoming some challenges such as high dimensionality, variety (diverse structures of data), and difficulties to integrate and analyze multi-source industrial big data sets [18]. In the literature, an extension of CRISP-DM methodology was presented in order to overcome domain-specific difficulties in obtaining and processing data especially for engineering applications in production [26]. Therefore, we believe that the proposed approach can be considered as a significant contribution in implementing the CRISP-DM methodology in a real-world industrial context. It should be also noted that the deployment phase of CRISP-DM is not covered in the proposed data scientific approach. Instead, this phase is recommended to be handled by the company using the insights and results obtained during the implementation of the proposed approach and thus the company can even further analyze the capabilities of the solution to make the best use of it in their real production environment.

Figure 1 demonstrates an overview of the tasks for each phase of the data scientific approach based on CRISP-DM. As it is seen from Figure 1, *Business understanding* phase is the core step, which determines the success of data mining. It includes the

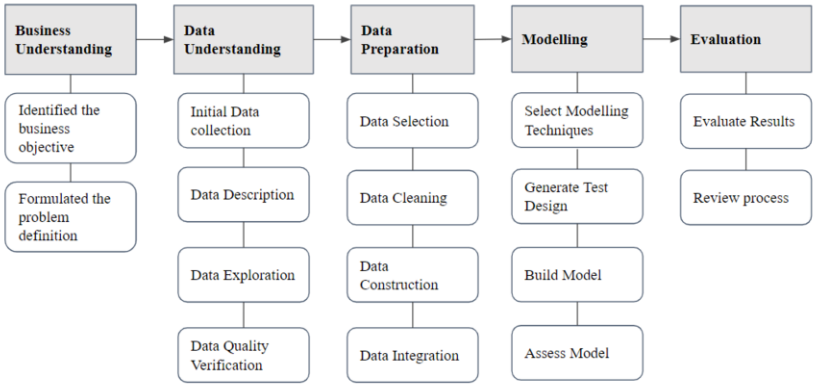


Figure 1. An overview of the data scientific approach based on CRISP-DM.

formulation of the project goal from a business perspective, analysis of objectives, and success criteria based on the industrial case study is carried out in this step. During *data understanding* phase, the study is conducted to understand if the available data of the process is sufficient to achieve the business objective or if additional data is required. This phase includes sub-steps namely, collection of initial data, description, and exploration of data, and verification of data quality [27]. In order to explore data, EDA is used to search for patterns and trends from high volume of data, and thus visualization techniques play an important part in this quest [28]. It is a promising analysis to pave industrial alarm and event analysis to extract hidden patterns and reveal undiscovered relationship without making assumptions [22]. The next phase, *data preparation*, comprises tasks involved to generate a data set that is used to build the model. This iterative process is likely to be performed multiple times and not in a prescribed order. During the data preparation phase, gap analysis, data transformation, and cleaning are performed to improve the quality of data, before feeding the data for model generation. This phase is crucial since the performance of the model depends on the data prepared in this phase. The modeling phase comprises of the selection of various modelling techniques and applying them using the prepared data. Once the modelling technique is selected and applied, the model is then tested for analysing the performance. In *modeling phase*, model's parameters are identified and tuned for its performance improvement. In some cases, there is a back and forth between data preparation and modeling according to the model performance. Finally, the model assessment is performed in order to summarize the review of the model and the performance comparison with the other models as well in *evaluation phase*.

4. A Real-World Industrial Application from Manufacturing

This section will give the results obtained from the different phases of data scientific based on CRISP-DM as described in the previous section. It should be noted that Tableau for EDA and Python programming language and its different libraries for data preprocessing and ML modeling are used during the implementation of CRISP-DM to real-world industrial application.

4.1. Business Understanding

Understanding the business problem and defining a business objective is the first phase in CRISP-DM as described above. The case company that has been collaborated in this study had addressed a problem of having machine stoppages during the production in one of its manufacturing plant, for which the cause is unknown. The machines in the manufacturing plant had been equipped with sensors and various alarms incorporated into each machine for determining the irregularities occurring in the production process based on the triggering of alarms. Therefore, the business objective for this application was formulated to explore and investigate the underlying cause of these machine stoppages in the manufacturing plant. Therefore, the company aims to find the cause, which triggers the alarms in two selected machines and thereby devising plans to tackle these alarms, which results in reducing the overall production halt duration. Since the machines understudy for this paper perform a process on Part A1 and Part A2, the focus is mainly to find the Part A1-Part A2 combination that triggers more alarms. In short, the business objective is to reduce the production halt time by investigating the cause for the occurrence of alarms, performing an EDA, and even building an ML model to predict the category of the alarms to be used for further investigation in maintenance planning.

4.2. Data understanding and preparation

The data was acquired from the selected two machines, namely Machine A and Machine B, where a process of assembling Part A1 and Part A2 takes place. The reason for choosing these two machines is because they are critical machines in the assembly since they have very narrow limits for quality checks. It should be noted that due to the limited pages, only examples of the analysis done for Machine A are presented in this paper, however, more information is found in a master thesis study conducted by the first two authors of this paper [29].

The data was collected from two sources, first was from the process data, which contains different part numbers, measured values from various positions of the parts during assembly (shown as Pos 10, Pos 12, and Pos 13 in Table 1), the corresponding date and time, and other variables. Within the process data, there are different combinations of Part A1 and Part A2. The other data is coming from manufacturing execution systems (MES). The event-log data is stored in MES [30] which mainly includes information for event log i.e., alarm ID, and timestamps of alarms. After an initial exploration of process and MES data, it was noted that they had to be merged for performing the required data analysis. The one possible way to merge these data sets is to find a common variable like date and time so that the merged data set contains all the recordings in the process and MES data corresponding to each matching date and time respectively. After further exploration, it was also found that some of the attributes are redundant meaning that they do not make any impact on the analysis, thus they were removed from the merged data set. To improve the quality of data, various tasks were conducted. For instance, the data was checked if there are any missing values since the missing values affect the performance of the final prediction model, therefore it is mandatory to treat and fix them in the data. The alarm data from MES contained some missing values. Therefore, they were removed from the data using a function in Python programming language. Apart from the missing values, the data also had irrelevant formats for some attributes i.e., time

and date, therefore the format of these attributes was corrected as well. Table 1 gives a sample from the merged data set after various pre-processing tasks which were done to ensure proper data analysis.

Table 1. A sample from the merged data set.

Engine Number	Alarm ID	Part A1 Number	Part A2 Number	Pos 10	Pos 12	Pos 13	Start Time	End Time
17776	230	9571	1066	38	52	38	2020-09-04 04:23:26	2020-09-04 04:25:01
17776	2	9571	1066	38	52	38	2020-09-04 04:27:22	2020-09-04 04:30:27
10350	230	9611	1081	38	NaN	38	2020-09-07 22:04:56	2020-09-07 22:05:30

4.3. Exploratory data analysis (EDA)

EDA was performed to find which product variants or combination of those variants trigger more alarms. To do that, a heat map was used to visualize the relationship between Alarm IDs and different product variants and their combinations for Machine A as illustrated in Figure 2, only the types of alarms that happened three times or above during the period were presented. The darker color in Figure 2 demonstrates that more alarms are generated with the associated product variant. It can be concluded that the combination of the number of Part A1 9571 and the number of Part A2 1066 has the majority of alarms among all generated alarms, where Alarm ID 2 is the dominating factor. Additionally, alarm ID 230 and alarm ID 211 are also proven as the most problematic alarms that cause the majority of production halt. Considering relatively mass production of different combinations, which might be a reason-giving rise to the alarms, the original number of the amount of production was taken into account and calculated as percentages. According to those percentages as given in the same figure, the number of Part A1 9651 is ranking the highest with a percentage of 4.85 while the number of Part A1 9611 has less alarm in general. As a result, the numbers of Part A1 9651 and 9571 have clearly more alarms than the others, and especially the number of Part A1 9571 needs more attention since it produces most of the products. This analysis gives an important finding to the company to investigate why the specified part numbers and their combinations are triggering more alarms and how to reduce the effect on production halt by conducting an advanced root cause analysis.

4.4. Modeling

The modeling phase is a systematic search for a model that meets the business objectives as defined earlier, which is efficient in predicting with minimum error and is most suitable to be used considering the properties and characteristics of the given data. After performing the EDA as explained in the previous section, it was evident that the alarm IDs such as 2, 230, and 211 are the most problematic ones. Therefore, the modeling is designed to predict these three alarms by using ML. It should be noted that the designed ML models use the historical data containing the recordings when these alarms occurred. Hence, in this study, it is only possible to predict the alarm IDs explained above instead of their occurrences in a certain period. This is because of the scope of given historical data and some data quality limitations as well. Modeling phase consists of four sub-steps, namely selecting the modeling techniques (ML algorithms), generating test

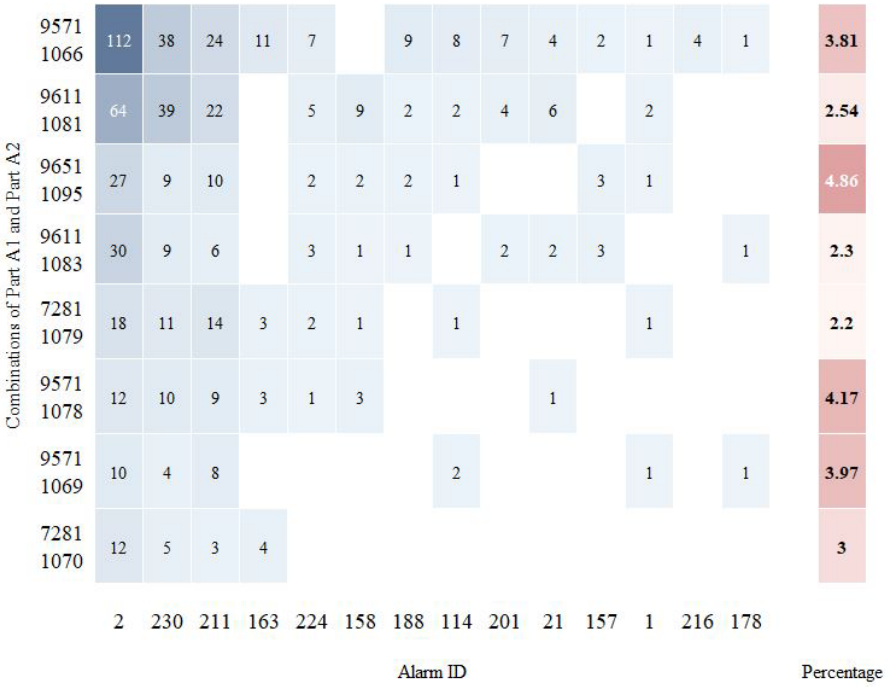


Figure 2. Relationship between Alarm ID and combinations of Part A1 and Part A2 for Machine A.

design, building the model, and assessing the performance of the model. Because the problem requires a multi-class classification approach [31] for the modeling, different ML algorithms were selected according to fitting for the problem, which are, Decision Tree, Random Forest, K-Nearest Neighbours, and XGBoost. The detailed background information regarding these algorithms can be found in an edited book [32]). These ML algorithms were used to perform the modeling and the final selection was done based on some performance metrics under different strategies.

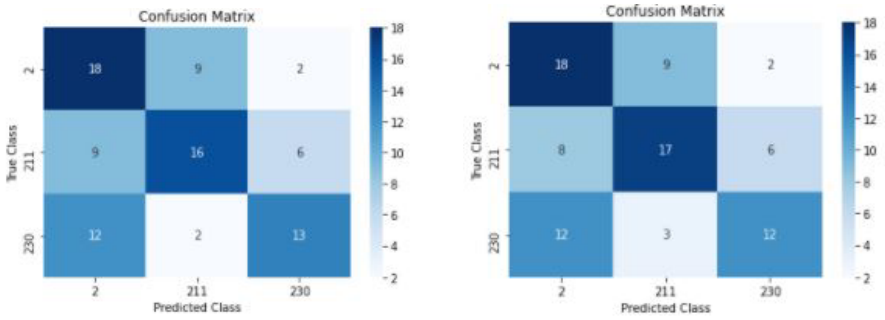
The second step in the modeling phase is to generate a test design. It is important to generate a procedure, which is used to check the model's quality and perform the validation of the model. Therefore, a test design was generated in which the data used in the modeling was split into training and testing data sets. The model was built on the training data set whereas the model's quality was tested using the test data set. Here the splitting criteria are 70 percent of the data set was used as the training data to train the model and the rest of the data set (30 percent) were used as the test data to test the model performance. After this partition of the data set, the training data was used for building previously selected lists of ML algorithms. In the building model phase, while the input variables or another word independent variables were determined such as Part A1 number, Part A2 number, their position and measured moment values, the output variables or another word dependent variables were determined alarm IDs (i.e., 2, 211, and 230). The missing values in the data set had been imputed using five different strategies, namely strategy 1 represents imputing missing values with zero; strategy 2 represents imputing missing values with median; strategy 3 represents imputing missing values with mode; strategy 4 represents imputing missing value with the nearest neighbor; and lastly strat-

egy 5 represents without imputation (e.g., keeping the missing values as it is in the data set). The model building also includes a key step, which is the parameter settings and tuning. Hence, the parameters were set up according to the requirements of each selected ML algorithm and they were tuned for each strategy to handle missing values. The obtained results after the parameter tuning are given in Table 2. The percentages in the table represent the accuracy score for each ML algorithm with respect to the determined strategies for handling missing values.

Table 2. Accuracy scores of each ML algorithm with respect to the strategies for handling missing values.

Model	Strategy 1	Strategy 2	Strategy 3	Strategy 4	Strategy 5
K-Nearest Neighbors (KNN)	41.37%	41.38%	43.67%	45.98%	X
Decision Tree	52.87%	54.02%	54.02%	47.12%	X
Random Forest	51.72%	49.42%	50.05%	43.67%	X
XGBoost	X	X	X	X	52.87%

As it has been seen from Table 2, the decision tree algorithm with strategies 2 and 3 have similar results in terms of the highest accuracy score, 54.02%. Therefore, a further analysis was performed to select the best model between them. To do that, confusion matrices were generated for the strategy 2 and 3 in Figure 3a and 3b respectively, and the corresponding false negative rates of the prediction were compared for final selection. A confusion matrix is used to understand the performance of a classification model and it helps to summarise correct predictions of classes by generating a matrix in which true (actual) classes are given as in the rows while predicted classes are given as in the columns [33]. True class is defined as the real occurred type of alarm obtained from the historical data and the predicted class is the alarm type that the model predicts based on the input parameters given to the model. False negative is defined as the outcome when the actual class is wrongly predicted e.g., an outcome from the model is predicted as alarm ID 211 but it was actually alarm ID 2, which is considered as a false negative in this case. On the other hand, true positive is defined as the outcome, e.g. the alarm ID 211 occurred in actual and the model also predicted alarm ID 211 as the predicted class. Furthermore, false negative rate is also calculated by using the ratio of false negatives to the sum of false negatives and true positives derived from the confusion matrix [29].



(a) Imputation with median

(b) Imputation with mode

Figure 3. Confusion Matrices for strategy 2 and 3

It has been observed from Figure 3 that both strategies have similar false negative rates for the prediction of Alarm ID 2, whereas strategy 2 has slightly higher false negative rate for alarm ID 211 and strategy 3 has a higher false negative rate for alarm ID 230. In this context, strategy 2 has slightly lower overall false negative rate for all three alarms, and hence the Decision Tree algorithm with strategy 2 was chosen as the final model.

4.5. Evaluation

The Decision Tree algorithm with strategy 2 (imputing missing values with median) was selected as a final model explained in the modeling phase. It also has the capability to provide the decision rules, which can support for the decision-making process to explore the possible in-depth root causes in production and maintenance. Therefore, the extracted rules from this model are one of the important outcomes, which can help a better understanding in a visual format of how the trees propagate and facilitate the decision of the domain experts in the company. The interesting information from the rules is that the moment values in a certain position are crucial in identifying alarm IDs. Therefore, this set of rules would also support describing the crucial features in identifying root causes of problematic alarms. In this phase, this model was further evaluated to check whether the model aligns with the business objective. In a conclusion, the combination of Part A1 and Part A2 triggering more alarms was found using EDA and a model was built to predict the type of alarm. This shows that the business objective was achieved as identified in the business-understanding phase of the data scientific approach.

5. Discussion

In this paper, a data scientific approach based on CRISP-DM was used for conducting a diagnostic and predictive analysis using industrial alarm and event log data coming from a real-word manufacturing company. Therefore, this approach provides some theoretical and practical contributions. From a theoretical point of view, it demonstrates an implementation of a predictive approach on industrial alarm data by using ML algorithms, which can be considered as a novel contribution in this area since most of the studies have followed descriptive and diagnostics approaches and not explored much in terms of implementing predictive modeling. Therefore, this study also fills this research gap as also mentioned in a study done by [24]. Additionally, this study also helps the company to implement the developed solution in the manufacturing plant in order to avoid the unplanned production stoppages and facilitate efficient planning of maintenance activities, which can be considered a significant practical contribution. The data scientific approach performed in this study is easily modified so that the company can follow the same structure with visualization and modeling techniques to analyze a vast amount of historical data collected for other data mining applications in manufacturing.

6. Conclusion

This paper proposes a data scientific approach based on CRISP-DM to perform the diagnostic and predictive analysis in a structured manner using industrial alarm and event

log data. The diagnostics analysis was done by performing EDA and the results from EDA provided insights into which combination of product types triggered more alarms. The performed data visualization as a part of EDA further gave insights into the behavior of the alarms occurring in the machine under study. After the problematic product type's combination was explored, this was used for predictive modeling, wherein the Decision Tree algorithm was selected to classify the alarm IDs and extract the decision rules produced from the algorithm. As a practical contribution, this study provides a set of decision rules that can act as decision support for further exploration of possible in-depth root causes through the other contextual data, and hence it provides good implications to the company towards PdM application. This study contributes to the literature by demonstrating a predictive approach in the analysis of industrial alarm data using ML and implementation of enhanced CRISP-DM methodology in a real-world industrial context. As further research, this study can be extended to address the process deviation or equipment malfunctioning, and therefore it is important to devise proper alarm management in order to achieve a good overall production performance in manufacturing.

Acknowledgement. This paper has been produced from the Master's thesis study of the first and second authors at the Chalmers University of Technology. The authors would like to thank the Production 2030 Strategic Innovation Program funded by VINNOVA for their funding of the research project PACA - Predictive Maintenance using Advanced Cluster Analysis (Grant No. 2019-00789), which this study has been conducted. Thanks to Kanika Gandhi and Sven Wilhelmsson for their guidance and support with the real-time data from a real-world manufacturing system. Thanks also to Alexander Karlsson for his support. This study has been conducted within the Production Area of Advance at the Chalmers University of Technology.

References

[1] May G, Kyriakoulis N, Apostolou K, Cho S, Grevenitis K, Kokkorikos S, et al. Predictive Maintenance Platform Based on Integrated Strategies for Increased Operating Life of Factories. In: IFIP International Conference on Advances in Production Management Systems. Springer; 2018. p. 279-87.

[2] Bokrantz J, Skoogh A, Berlin C, Wuest T, Stahre J. Smart Maintenance: a research agenda for industrial maintenance management. International Journal of Production Economics. 2020;224:107547.

[3] Mehmeti X, Mehmeti B, Sejdiu R. The equipment maintenance management in manufacturing enterprises. IFAC-PapersOnLine. 2018;51(30):800-2.

[4] Lee J, Ni J, Singh J, Jiang B, Azamfar M, Feng J. Intelligent Maintenance Systems and Predictive Manufacturing. Journal of Manufacturing Science and Engineering. 2020;142(11):110805.

[5] Ahmad R, Kamaruddin S. An overview of time-based and condition-based maintenance in industrial application. Computers & industrial engineering. 2012;63(1):135-49.

[6] Wuest T, Weimer D, Irgens C, Thoben KD. Machine learning in manufacturing: advantages, challenges, and applications. Production & Manufacturing Research. 2016;4(1):23-45.

[7] Kurrewar H, Bekar ET, Skoogh A, Nyqvist P. A Machine Learning Based Health Indicator Construction in Implementing Predictive Maintenance: A Real World Industrial Application from Manufacturing. In: IFIP International Conference on Advances in Production Management Systems. Springer; 2021. p. 599-608.

[8] Zhai S, Gehring B, Reinhart G. Enabling predictive maintenance integrated production scheduling by operation-specific health prognostics with generative deep learning. Journal of Manufacturing Systems. 2021.

[9] Schröer C, Kruse F, Gómez JM. A Systematic Literature Review on Applying CRISP-DM Process Model. Procedia Computer Science. 2021;181:526-34.

[10] Takata S, Kirnura F, van Houten FJ, Westkamper E, Shpitalni M, Ceglarek D, et al. Maintenance: changing role in life cycle management. CIRP annals. 2004;53(2):643-55.
[11] Nakagawa T. Maintenance theory of reliability. Springer Science & Business Media; 2006.
[12] Mobley RK. An introduction to predictive maintenance. Elsevier; 2002.
[13] Ran Y, Zhou X, Lin P, Wen Y, Deng R. A survey of predictive maintenance: Systems, purposes and approaches. arXiv preprint arXiv:191207383. 2019.
[14] He Y, Gu C, Chen Z, Han X. Integrated predictive maintenance strategy for manufacturing systems by combining quality control and mission reliability analysis. International Journal of Production Research. 2017;55(19):5841-62.
[15] Jimenez JJM, Schwartz S, Vingerhoeds R, Grabot B, Salaün M. Towards multi-model approaches to predictive maintenance: A systematic literature survey on diagnostics and prognostics. Journal of Manufacturing Systems. 2020;56:539-57.
[16] Carvalho TP, Soares FA, Vita R, Francisco RdP, Basto JP, Alcalá SG. A systematic literature review of machine learning methods applied to predictive maintenance. Computers & Industrial Engineering. 2019;137:106024.
[17] Bekar ET, Nyqvist P, Skoogh A. An intelligent approach for data pre-processing and analysis in predictive maintenance with an industrial case study. Advances in Mechanical Engineering. 2020;12(5):1687814020919207.
[18] Yan J, Meng Y, Lu L, Li L. Industrial big data in an industry 4.0 environment: Challenges, schemes, and applications for predictive maintenance. IEEE Access. 2017;5:23484-91.
[19] Gröger C. Building an industry 4.0 analytics platform. Datenbank-Spektrum. 2018;18(1):5-14.
[20] Vásquez JW, Travé-Massuyès L, Subias A, Jimenez F, Agudelo C. Alarm management based on diagnosis. IFAC-PapersOnLine. 2016;49(5):126-31.
[21] Lucke M, Chioua M, Grimholt C, Hollender M, Thornhill NF. Integration of alarm design in fault detection and diagnosis through alarm-range normalization. Control Engineering Practice. 2020;98:104388.
[22] Bezerra A, Silva I, Guedes LA, Silva D, Leitão G, Saito K. Extracting value from industrial alarms and events: A data-driven approach based on exploratory data analysis. Sensors. 2019;19(12):2772.
[23] Mahadevan S, Shah SL. Fault detection and diagnosis in process data using one-class support vector machines. Journal of process control. 2009;19(10):1627-39.
[24] Bousdekis A, Lepenioti K, Apostolou D, Mentzas G. Decision Making in Predictive Maintenance: Literature Review and Research Agenda for Industry 4.0. IFAC-PapersOnLine. 2019;52(13):607-12.
[25] Wirth R, Hipp J. CRISP-DM: Towards a standard process model for data mining. In: Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining. vol. 1. Springer-Verlag London, UK; 2000. .
[26] Huber S, Wiemer H, Schneider D, Ihlenfeldt S. DMME: Data mining methodology for engineering applications–a holistic extension to the CRISP-DM model. Procedia Cirp. 2019;79:403-8.
[27] Chapman P, Clinton J, Kerber R, Khabaza T, Reinartz T, Shearer C, et al. CRISP-DM 1.0: Step-by-step data mining guide. SPSS inc. 2000;9:13.
[28] Skiena SS. The data science design manual. Springer; 2017.
[29] Adarsh, Xinjie. A systematic data science approach towards predictive maintenance application in manufacturing industry @ONLINE; 2021. https://hdl.handle.net/20.500.12380/302632.
[30] Chand S, Davis J. What is smart manufacturing. Time Magazine Wrapper. 2010;7:28-33.
[31] Susto GA, Schirru A, Pampuri S, McLoone S, Beghi A. Machine learning for predictive maintenance: A multiple classifier approach. IEEE Transactions on Industrial Informatics. 2014;11(3):812-20.
[32] Han J, Pei J, Kamber M. Data mining: concepts and techniques. Elsevier; 2011.
[33] Luque A, Carrasco A, Martín A, de las Heras A. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. Pattern Recognition. 2019;91:216-31.