Understanding and Predicting Process Performance Variations of a Balanced Manufacturing Line at Bosch

Ângela F. Brochado[1][0000−0002−1222−9529]
Eugénio M. Rocha[2,3][0000−0003−3628−6795]
Carina Pimentel[1,4][0000−0003−4284−1628]

[1] Department of Economics, Management, Industrial Engineering and Tourism, University of Aveiro (UA), Portugal
[2] Department of Mathematics / UA
[3] Center for Research Development in Mathematics and Applications / UA
[4] Systems for Decision Support Research Group (GOVCOPP) / UA

Abstract. Industry 4.0 takes advantage of data-driven approaches to improve manufacturing processes. Root cause analysis (RCA) techniques are naturally required to support the identification of reasons for (in)efficiency processes. However, a limitation of classical RCA methods is their sensibility to data perturbations or outliers. Such sensibility phenomenon requires the implementation of robust RCA approaches. Here, methods of graph theory (queue directed graphs), operational research (multi-directional efficiency analysis), machine learning (extreme gradient boosting), and game theory (Shapley analysis) are merged together, in order to obtain a robust approach that is able to benchmark the workers acting on a discrete manufacturing process, determine the relevance level of process variables regarding a worker belonging to the (in)efficient group, and predict the worker performance variation into its next working session. A use case at Bosch ThermoTechnology is analysed to show the methodology's applicability.

Keywords: Discrete Manufacturing Processes · Benchmarking · Machine Learning · Root Cause Analysis

1 Introduction and Previous Work

Root Cause Analysis (RCA) can be defined as a set of methodologies for identifying the (fundamental) causes/factors of a problem, based on the analysis of deviations from a standard reference, in order to prevent its future occurrence, reduce overall process variability, and/or optimize costs. For RCA, three main types of approaches can be found in the literature: quantitative, qualitative and mixed. The most addressed and also the most used by companies' continuous improvement teams (namely in the manufacturing industry) are the mixed approaches. They comprise tools brought by the Six Sigma area, such as the 5 why's, Fault-tree Analysis, Ishikawa Diagram (also known as Fishbone Diagram), Pareto Chart, 8Ds, Failure Mode and Effects Analysis (some recent studies are [1, 2]). However, after more than 10 years of the "Industry 4.0" Era, where companies already have systems capable of extracting data from their manufacturing processes, most continuous improvement activities, namely root cause analysis, continue to be developed through conventional methods, such as manual data analysis, or the utilization of basic data mining tools [3]. These conventional methods conducted by process experts are time-consuming and provide results with high variability [4]. However, the reason why this still happens is quite simple to explain. Many companies dealing with big data still lack human resources with the expertise of applying data-driven methodologies with some level of machine learning to automate (and therefore, speed up) their RCA processes.

Hence, one of the aims of any production manager is to increase the productivity of their production processes. For such, understanding the reasons for deviations in (current) performance is fundamental to devise effective countermeasures for continuous improvement. Manufacturing production lines (MPL) that are balanced, meaning that the duration of jobs are designed to minimize bottleneck workstations and maximize operators' work time usage, are particular difficult since by construction the deviations are smoother and difficult to grasp.

In this work, we consider the problem of applying an effective and scientific valid data-driven approach to answer the following research questions:

(Q1) Is it possible to identify significant performance variations in a balanced MPL by evaluating process KPIs, metrics and/or variables associated with workers' consecutive work sessions?

(Q2) How can we determine the causes (variables) that led to high/low performance deviations?

(Q3) Is it possible to predict future performance variation based on current calculated process metrics?

In what follows, we introduce some relevant works, found in the literature, within the scope of automatic approaches to RCA for manufacturing lines. An automatic approach is understood as a sequence of algorithms/methods that follow a set of implicit/explicit rules measuring relevance, and which are capable of extracting in-depth information from real data without human intervention, giving variable relevance scores with respect to a process of study.

In fact, there are still few studies that address automatic RCA in manufacturing production problems and the majority focuses on quality problems. In [5], it was developed an approach for constructing digital cause-and-effect diagrams with quality data, where the K-means algorithm is implemented to cluster the problems and causes, and then a classification model based on a random forest is employed to classify cause text into the main cause categories. Similarly, [6] also follow the idea of constructing an automated version of a well-known lean manufacturing tool, the Value Stream Map (VSM), for multi-varieties and small-batch production, with timely on-site waste identification and automated root cause analysis. In addition, [7] developed a two-stage automatic root cause analysis (ARCA) for the phenomenon of overlap in manufacturing. The authors propose a first stage of "Problematic Moment Identification" (PMI), where relevant data is selected for the analysis by using a Exponentially Weighed Moving Averages control chart. Then factor ranking algorithms were developed and used to avoid hiding highly correlated factors and enabling information on equally probable root causes. The factor ranking algorithms are Co-Occurrences (CO), Chi-Square (CS) and Random Forest (RF). Lastly, [8] presents a big data-driven root cause analysis system including three modules of (1) Problem Identification (to describe multiple and different types of quality problems using data mining methods), Root Cause Identification (using K-Nearest Neighbor (KNN) and Neural Network (NN) classifiers to automatically predict root causes), and Permanent Corrective Action. The authors validated the approach by using data from an automobile factory.

2 Methodology and Mathematical Model

For the problem addressed in this work, it was necessary to implement a more adjusted methodology capable of encompassing not only quality variables but also variables related to the workers themselves, some of them calculated in a very particular way, with the support of a formal abstract structure, as will be explored in Section 3. Also, none of the RCA methodologies found in the literature adopts a robust approach. Robustness is very important to get more reliable estimates for unspecified parameters in the presence of outliers or data perturbations, for more trustworthy root cause identification and model predictions.

In particular, the approach in this work generally follows the methodology introduced recently by the authors in [9] and applied to a Ceramic Industry manufacturing, which has quite different characteristics in comparison with the ones of the manufacturing process of our use case at Bosch. The methodology steps are schematically described in Figure 1.

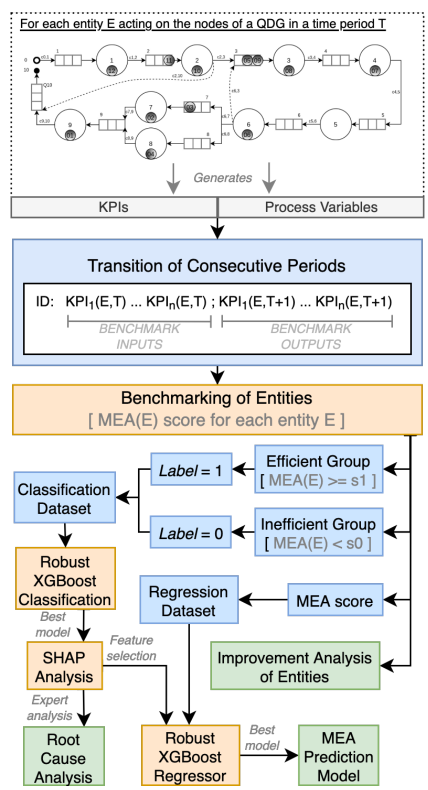

Fig. 1: A data-driven methodology for Root Cause Analysis.

The approach intends to determine the causes, i.e., the variables of workers, machines or processes of a manufacturing production line that most contribute to an entity being considered efficient or inefficient, according to the KPI values of consecutive work sessions and also predicting future performance variation scores. As displayed by the figure, the approach fuses an operational research method (multi-directional efficiency analysis), a machine learning method (extreme gradient boosting), and a game theory method (Shapley analysis), in order to obtain a robust approach for RCA. Each step of the methodology has been applied to the problem introduced in the next section, and the main results are displayed in Section 4.

First, a MPL is chosen and the notion of "entity" is defined (in this problem) as a worker, who operates on the MPL during a certain work shift. Each worker applies a set of actions to a set of workstations in the MPL. The layout and flows in the MPL can be modelled according to a formal abstract mathematical structure, called *Queue Directed Graph* (QDG) (see the bottom block of Figure 2 and [10, 11]). This mathematical structure consists of nodes (in the case of an MPL means workstations) that act on tokens (i.e., parts or products), which in turn may have to wait in queues if nodes are busy processing the previous tokens.

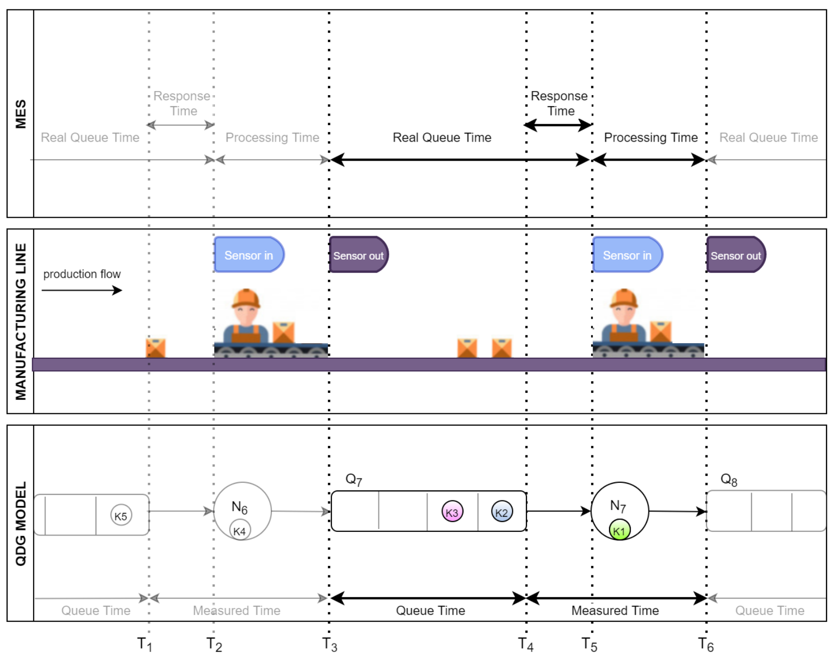

Fig. 2: Representation layers of a MPL segment - retrieved from [11].

The advantage of the QDG is that it is capable of representing any type of MPL with a discrete production environment, so this work can be easily extended to other use cases of discrete manufacturing. Additionally, the QDG is based on *minimal information* (MI), which is meant as the most elementary information from production line operations, enough to autonomously generate the abstract manufacturing layout and calculate metric variables, for example the so-called *AMPM* (the Average Measured Period of time a workstation is occupied processing a part), and the *AQPM* (the Average Queue Period of time a part spends at the queue of a workstation), see [11] for further details about MI and the mathematical formulas of those metrics.

Commonly, the notion of workstations' *Processing Time* is associated with time readings obtained by hardware devices on the workstations, which are further sent to the manufacturing execution system (see first block "MES" of Figure 2). This metric comprises the amount of time between the beginning of the first operation in a workstation and the moment the product leaves because all tasks, in that workstation, were completed. A QDG gives a similar metric, the so-called *Measured Time*, described in the same figure. It comprises the amount of time between the moment a part/product leaves the queue of a workstation and the moment the product leaves the workstation after all tasks have been performed. So, by setting the difference between the *Measured Time* and the *Processing Time*, it is possible to compute the amount of time a worker spent to "respond" to a part in a queue, the so-called *Part Response Time*. Then, the ***Response Time*** directly associated with MPL workers is defined as the average of Part Response Times in a time period (usually a shift), see [12].

Hence, taking into account the impact that workers' variability have on the performance of an MPL, it seems clear to study some of the variables that may cause such variability. In this work, these variables are designated as *worker-related variables*: the ***Wage***, the ***Experience Time*** (the amount of training hours invested by the company and benefited by the worker), the ***Response Time***, and the ***Delay Time***. The latest is a penalization value that measures the lapse between the planned shift start time and the time a worker effectively started working, by using a Gaussian function, defined by

$$\mathrm{DT}(t) = G_\sigma\left(t_b\right) - G_\sigma(t) \quad \text{where} \quad G_\sigma(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(t-t_b)^2}{\sigma^2}\right),$$

for a given standard deviation σ that accounts for the penalization curve spread. The reason for such penalization is the fact that earlier starting workers (artificially) increase the next workstation queue and later starting workers create a gap in the working flow, so both perturb the theoretical balancing of the line, see [26] for data-driven simulations showing this phenomena.

Hence, both *Response Time* and *Delay Time* vary along with workers and with different work sessions, while the *Experience Time* and *Wage* may differ between workers, however, both values are fixed per work session (except when certain events occur, such as pay raises, job promotions and/or job training, where there's an update).

Finally, global quality metrics are also added to the previous ones and introduced to answer all research questions, namely the percentage of *Reworks* and ***Quality***. This work, as a continuation of a previous one based on an optimization problem with a maximization function, uses the complementary values of the AMPM, AQPM and Reworks, defined here as ***cAMPM, cAQPM*** and

cRework (see [12] to assess the mathematical formulas). A complement of a variable is given as the maximum possible value of the variable minus their current value.

Further details regarding the data sets' information used in this work shall be discussed in the section that follows.

3 Data Characterization and Correlation

A chosen manufacturing production line (MPL) of Bosch Thermotechnology (Bosch TT), the business unit facility of residential hot water located in Aveiro, Portugal, is analysed through the proposed approach. At this MPL, usually 2 to 3 daily work shifts operate with a specific number of workers (shift 1 has 14 workers, shift 2 has 8 workers, shift 3 has 14 workers). The MPL is composed by 18 workstations where each worker is assigned to either 1 or 2 of them, depending on the shift, workers' experience with certain machines/processes, and even workers' availability (although not very often). In this study, the 5 chosen data sets contain information related to shift 1 on different time instances, so we will be dealing with 14 entities and 5 time periods. The *Response Time, Delay Time, cRework, cAMPM and cAQPM* are first calculated and then added as new columns to each data set, which already contain information regarding the *Experience Time*, *Wage* and *Quality* for each entity and time period. Although, each data set represents a specific time period, for confidentiality purposes the actual date is not provided here. Tables 1 and 2 provide information regarding the statistical characterization of the data sets. In the same way, the wage and experience are values changed by a offset to preserve anonymity. Notice that the results obtained by the proposed methodology are not affected, since the methods are variable translation invariant.

Although, the experts defined a set of relevant KPIs and process variables to evaluate (as described above), a data-driven approach must follow some classical pre-processing steps as looking for missing values (none in this case) and finding correlation between variables, which will perturb the interpretation of the results.

Table 1: Data sets characterization (part 1 of 2) - mean and standard deviation of the worker-related variables.

	Worker-related Variables			
Data set	*Response Time*	*Delay Time*	*Wage*	*Experience Time*
d01	1.16 ± 0.99	0.26 ± 0.14	666.21 ± 46.32	1060.71 ± 586.66
d02	0.59 ± 0.45	0.17 ± 0.13		
d03	0.59 ± 0.53	0.21 ± 0.14		
d04	0.96 ± 0.88	0.19 ± 0.15		
d05	1.06 ± 0.96	0.18 ± 0.13		
Units	seconds	per unit	euros	hours

Table 2: Data sets characterization (part 2 of 2) - mean of the shift metrics plus mean and standard deviation of the production variables.

	Shift Metrics		**Production Metrics**	
data set	*cRework*	*Quality*	*cAMPM*	*cAQPM*
d01	0.726	94.07	7.10 ± 3.70	28.13 ± 10.53
d02	1.260	97.98	7.28 ± 3.44	11.05 ± 05.80
d03	0.782	97.78	7.26 ± 3.50	12.02 ± 06.20
d04	0.813	72.59	6.64 ± 2.87	15.91 ± 07.05
d05	0.416	81.48	6.17 ± 3.21	14.35 ± 07.66
Units	percentage	percentage	seconds	seconds

Figure 3 represents the Person's correlation heatmap of the variables and KPIs, where the values can be (briefly) interpreted as measuring the strength of the linear relationship between variables (similar results where obtained by Spearman's correlation). Looking at the figure, both *Wage* and *Experience Time* have a reasonable negative correlation with the *cAMPM*. This fact is interesting, as it tells us that workers with high wages and high experience time have a lower complementary value of $AMPM$ (remember that $AMPM$ is the Average Measured Period of time a workstation is occupied processing a product, so the lower the complementary value, the higher is the measured time on the workstation). Such may seem against common sense, but the reason relies on the fact that, for production efficiency, the most experienced workers are allocated to the most complex workstations, with higher processing times and problematic jobs. Hence, to eliminate this bias, the *cAMPM* variable has been removed from the model, and it will not be used from this point further in the analysis.

4 Main Results

4.1 Benchmark of performance variation between consecutive work sessions

Proposed by [13] as a derivative of the well-known data envelopment analysis (DEA) methodology, multi-directional efficiency analysis (MEA) is a non-parametric approach that has been widely used nowadays (some applications are [14, 15, 12]). This refined approach aims to provide further insights about the potential improvement for each factor involved in the model, to make a more efficient and cost-based plan to either maximize efficiency or minimize inefficiencies. For advantages of using MEA over DEA, see [16–18].

The benchmark done here (in the output orientation version) over the tuples $(worker, day)$ can be briefly interpreted as the best ranked tuples being the ones who are capable of maximizing the KPIs of the next work session, when compared with their KPIs of the current work session, if somehow it was possible to normalize and compare all input variables between workers.

The MEA algorithm was applied to analyze the variation of performance of 14 workers from shift 1 on 5 work sessions (the data sets). Both shift metrics

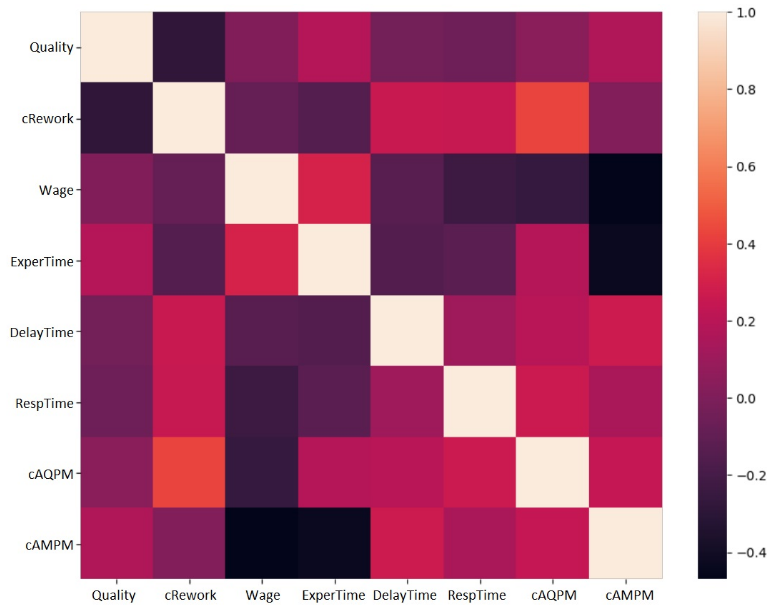

Fig. 3: Variables' correlation matrix/heatmap.

presented in Table 2 were selected as inputs (the values at t) and outputs for the MEA model (the values at $t+1$) to compute the efficiency scores. From the results, the average MEA score of the 14 workers in the transition from d03 to d04 was the lowest recorded, being near 0.4. The transition from d04 to d05 obtained a score of 0.92. On the contrary, transitions from d01 to d02 and from d02 to d03 were the transitions with the best performance improvements, attaining a 1.0 score. This represents a first level of root cause analysis, as it identifies a significant performance variation in the data. With the above information, experts are now able to explore these situations identified as significant performance variations and derive good practices for future improvement plans.

Based on this first outcome, benchmarked results were split into 2 groups and labelled: (i) the so-called *Efficient Group* (G_+) was labelled as "1", and it is composed of workers with a MEA score equal to or bigger than a defined threshold $s \in]0,1[$; (ii) the *Inefficient Group* (G_-) which was labelled as "0", and it contains all entities with MEA scores below the s threshold. For our problem, s was chosen to be the median of the obtained scores. These groups define the so-called *Classification data set*.

From here, the next step will be to determine the factors/variables that most influenced or can best explain the MEA scores, measuring the production line's ability to improve between work sessions.

4.2 Determination and analysis of variables' relevance to explain workers' performance variation

Researchers and industrial engineers have built up a wealthy literature on classification and regression models and their applications to real-life industrial cases (the literature is quite extensive). In particular in root cause analysis, one of the recent approaches it to fit a machine learning model and to use feature importance (FI) to get the factor/causes relevance. However, FI is a explainable characteristic of the particular fitted model, which may not represent correctly the problem.

In fact, a key issue in the classical FI approach in machine learning is that FI is generally not stable to small perturbation of the features' data because the used model are not robust to perturbations. The *robustness problem* has been studied thoroughly during the last decades, with a fast paced development of robust approaches in some contexts (e.g., see [19–21]). The great purpose of constructing a robust ML model is to get more reliable estimates for unspecified parameters in the presence of outliers, so the outlined root causes, model predictions are also more valid and trustworthy. Sometimes, those robust models attain a worst performance then not robust ones, but by construction they are far more reliable for FI. For the above reasons, the work from [22] combining a robust ML models with eXtreme Gradient Boosting (XGBoost) has been studied and employed in the methodology. XGBoost has been greatly recognized in the well-known Kaggle competitions due to its great performance and fast response to classification/regression predictive modelling problems, for structured or tabular data sets (some recent examples of its effectiveness are [23, 24]). In our case, after hyper-parameters optimization, the best model can be select as a good representation of a function mapping features into the Efficient/Inefficient Groups classes. Because we focus on getting variable relevance overfitting is a desired situation and was promoted, because it means that our model best characterizes the current relation between inputs (features) and the output (label).

For explaining the model results, allowing a sort of root cause analysis, we use the so-called SHapleyAdditive exPlanation (SHAP). This method is a game-theoretic approach proposed in [25], which aims to analyse complex models when there is a set of features that work as inputs and which produce a set of outputs (or predictions). The goal is to explain the predictions by computing the contribution of each feature, in the form of a value denominated the **Shapley value**. The SHAP value provides insight into how to fairly distribute the prediction among the features. Therefore, it gives a powerful measure of the importance of each individual feature in a model. The larger the SHAP value, the bigger the importance of such feature to the model explanation.

Following the steps of the proposed methodology of Figure 1, a grid hyper-parameter optimization of the (robust) XGBoost classifier was performed, and the best model metrics are described in the Table 3. Notice that this is a binary classification problem which is slightly imbalanced (27.8% vs 72.2%). These assessment metrics are considered good enough to assume the relevance of the SHAP analysis.

Table 3: Results metrics of the XGBoost.

	Precision	*Recall*	*F1-score*	*Support*
G_-	1.00	0.93	0.97	15
G_+	0.98	1.00	0.99	41
macro avg	0.99	0.97	0.98	56
weighted avg	0.98	0.98	0.98	56

At the beginning of the methodology, we address SHAP analysis when applied on a robust machine learning method as a RCA approach for the difference between the Efficient Group (G_+) and the Inefficient Group (G_-) of workers. Figure 4 shows the SHAP performance variance relevance plot. All seven variables are sorted in descending order based on their relevance to attain the specific classification class by the model. The red colour represents a high value of the variable for a specific observation, while blue represents a low value of the variable.

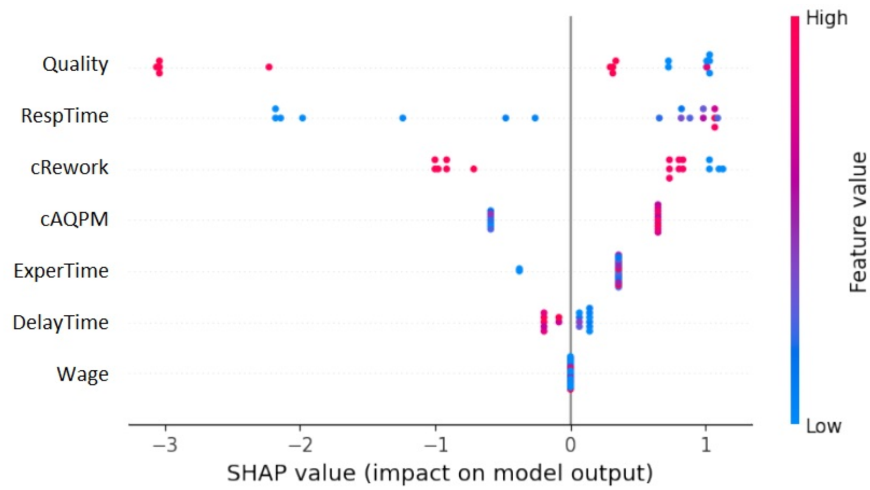

Fig. 4: Variables relevance plot computed with XGBoost+SHAP.

A global analysis of the SHAP results shows that variables *Wage*, *DelayTime*, and *Experience Time* are the less relevant, whereas *Quality* and *cRework* are the most relevant. So, looking at the top relevant variables *Quality* and *cRework*, it can be said that lower values of both metrics are determinant for obtaining the respective classification. The same thing can be mentioned about lower complementary values of *Reworks*: if there is a big number of reworks in a current work session are determinant to define the classification class. For higher Quality values, the interpretation is not as obvious, so in this situation, Figure 5(left) can help clarify what is the most prominent impact it has on the model - a negative impact but in some situations it can also have a slight positive impact.

On the other hand, higher values of *Response Time* are also associated with better determinant where, again, it is not clear about lower values. Figure 5(left) indicates a (global) positive model impact through this variable.

For this work, the partial dependence plot was also computed with SHAP. However, because the outcome did not provide any significant conclusion, it was decided not to include it at the analysis.

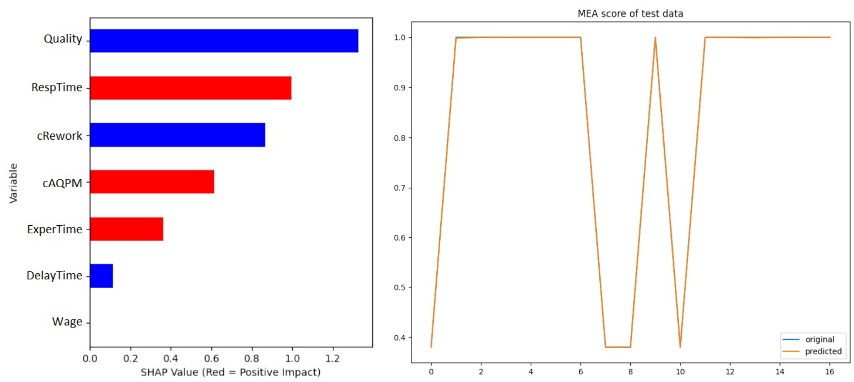

Fig. 5: (left) Variables relevance plot computed with SHAP; (right) MEA score prediction results of workers' performance for the next work session.

4.3 Prediction of the performance variation benchmark

By the end of our RCA approach, the regression data set with the MEA scores, plus the outcomes provided by the SHAP analysis, were used to train the robust XGBoost algorithm, and a regression model was obtained. This model is able to predict the entities' performance (MEA score) for the next work session based on the some variables of the current work session, in particular, it was used $(Quality, RespTime, cRework)$.

These features were selected, based on the results of an algorithm created to detect which feature (or, in this case, which set of features) could better predict the value of the MEA score. Figure 5(right) compares the MEA performance results of the test data set with the predicted outcomes of the XGBoost robust regression model. The registered RMSE metric is 0.00376, indicating that the model has a quite good fit. Just by looking at the result in Figure 5(right), the same conclusion of good fit can be taken. Thus, with this magnitude of error, managers may truly rely on the prediction model results to accurately predict future benchmark performance variations. This is quite useful for situations when the algorithm detects significant drops, so managers can try to identify the reasons or root causes for such events.

5 Conclusion and Future Work

In this work, a set of data sets from a use case problem of a discrete manufacturing process at Bosch TT were analysed and it was possible to:

1. Identify significant performance variations of 14 workers operating on a balanced MPL, between the transition of consecutive work periods;
2. Determine the causes/factors, i.e., the variables that led to high/low performance variations;
3. Build a prediction model capable of detecting future performance variations score in the MPL based on the values of the current variables produced by the 14 line workers on a shift.

With these results, experts are now able to explore specific events identified as significant performance variations and derive good practices for future improvement plans, or even identify, in time, the causes/variables that will lead to such future variation.

The followed methodology fuses techniques from operational research (MEA), machine learning (XGBoost) and game theory (SHAP) was employed. The attained XGBoost regression model registered a quite good RMSE metric. One of the limitations described in [9] is that although the model is robust by construction, its applicability depends on the set of process variables chosen. Because these variables become the model features, they may induce low values of the machine learning evaluation metrics for the classification model, the foundation for the SHAP analysis. In this work, this problem was overcome, by drawing, analysing and eliminating redundant variables detected with the correlation matrix. Another concern mentioned in [9] is that the methodology approach does not deal (automatically) with the existence of unbalanced classes.

After further real-data validations, this work is planned to be deployed to production testing as a set of micro-services, communicating with Kafka brokers connected to the manufacturing execution system.

Regarding future work, one of the main goals of the research team is the design and development of a data-driven platform to support continuous improvement activities in companies. Although in a prototype phase, this platform will follow the structure of an extensive protocol (a well-defined set of rules and steps), in order to deliver a unified continuous improvement tool that joins all areas of operation of a company, from the top management, down to the operational level. To this end, the DMAIC (Define-Measure-Analyse-Improve-Control) strategy, well-known as a data-driven improvement approach to help reduce process variation, and deriving from the Six Sigma area, will be used as the protocol foundation and integration tool.

The methodology applied to the problem addressed in this paper (see Figure 1) will pertain to the *Analyse* phase of the DMAIC, and it represents one of two possible approaches developed by one of the authors for root cause analysis of key performance indicators (KPIs) (see Figure 6, RC2 path). The RC1 approach was tested and applied to another use case presented by Bosch TT, which can be found in the following work [27].

Subsequently, it is intended to introduce in the DMAIC's *Improve* stage the PDCA cycle (Plan-Do-Check-Act), a well-known approach for continuous improvement and problem-solving [28]. Thus, the Improve phase can integrate multiple PDCA cycles, as many as necessary, to effectively produce an improvement, which must be measured based on a quantifiable metric.

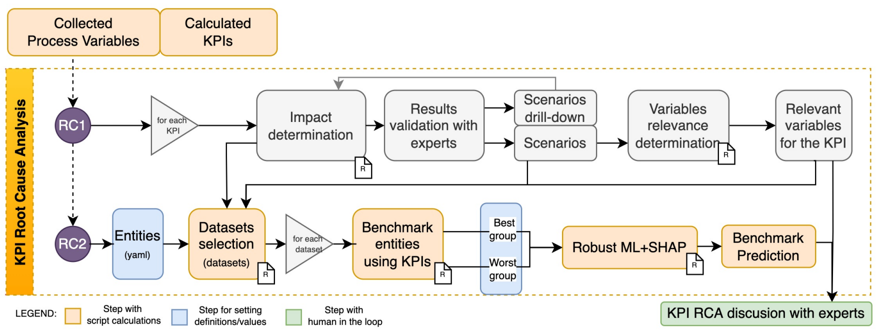

Fig. 6: KPI Root Cause Analysis approaches for the DMAIC's *Analyse* phase.

Data Availability Statement. All data sets used in the present study are confidential information of Bosch company manufacturing systems, so they are not publicly available.

Acknowledgements. The present study was developed in the scope of the Project Augmented Humanity (PAH) [POCI-01-0247-FEDER-046103], in which the first author has a PhD grant, financed by Portugal 2020, under the Competitiveness and Internationalization Operational Program, the Lisbon Regional Operational Program, and by the European Regional Development Fund. The second author was partially supported by the Center for Research and Development in Mathematics and Applications (CIDMA), through the Portuguese Foundation for Science and Technology, reference UIDB/04106/2020.

References

1. Rathi, Rajeev, Maddireddy Chandra Gopi Reddy, Annam Lakshmi Narayana, Utlapalli Lakshmi Narayana, and Md Saeedur Rahman: *Investigation and Implementation of 8D Methodology in a Manufacturing System.* In Materials Today: Proceedings 50 (January): 743–50 (2022). https://doi.org/10.1016/J.MATPR.2021.05.273
2. Zhai, G. F., X. H. Liu, J. Deng, W. Sulistiyowati, M. R. Adamy, and R. B. Jakaria: *Product Quality Control Based on Lean Manufacturing and Root Cause Analysis Methods.* In Journal of Physics: Conference Series 1402 (2): 022038 (2019). https://doi.org/10.1088/1742-6596/1402/2/022038.
3. Peças, Paulo, João Encarnação, Manuel Gambôa, Manuel Sampayo, and Diogo Jorge: *PDCA 4.0: A New Conceptual Approach for Continuous Improvement in the Industry 4.0 Paradigm.* Applied Sciences 11 (16): 7671 (2021). https://doi.org/10.3390/APP11167671.

4. Kim, Byunghoon, Young-Seon Jeong, Seung Hoon Tong, and Myong K. Jeong: *A Generalised Uncertain Decision Tree for Defect Classification of Multiple Wafer Maps*. International Journal of Production Research 58 (9): 2805—2821 (2020).
5. Xu, Z., and Dang, Y.: *Automated Digital Cause-and-Effect Diagrams to Assist Causal Analysis in Problem-Solving: A Data-Driven Approach*. International Journal of Production Research 58 (17): 5359–5379 (2020).
6. Wang, Hao nan, Qi qi He, Zheng Zhang, Tao Peng, and Ren zhong Tang: *Framework of Automated Value Stream Mapping for Lean Production under the Industry 4.0 Paradigm*. Journal of Zhejiang University-SCIENCE A (Applied Phisics and Engineering) 22 (5): 382–95 (2021) https://doi.org/10.1631/JZUS.A2000480
7. e Oliveira, Eduardo, Vera L. Miguéis, and José L. Borges: *On the Influence of Overlap in Automatic Root Cause Analysis in Manufacturing*. International Journal of Production Research (2021). https://doi.org/10.1080/00207543.2021.1992680
8. Ma, Qiuping, Hongyan Li, and Anders Thorstenson: *A Big Data-Driven Root Cause Analysis System: Application of Machine Learning in Quality Problem Solving*. Computers Industrial Engineering 160 (October): 107580 (2021). https://doi.org/10.1016/J.CIE.2021.107580.
9. Rocha, Eugénio M., Brochado, Ângela F., Rato, Beatriz, and Meneses, Joana.: *Benchmarking and Prediction of Entities Performance on Manufacturing Processes through MEA, Robust XGBoost and SHAP Analysis*. Manuscript submitted for publication (2022).
10. Lopes, Maria J., Rocha, Eugénio M., Georgieva, Petia, and Ferreira, Nelson: *General Model for Metrics Calculation and Behavior Prediction in the Manufacturing Industry: An Automated Machine Learning Approach*. In Handbook of Research on Applied Data Science and Artificial Intelligence in Business and Industry, (January), 263—290 (2021). https://doi.org/10.4018/978-1-7998-6985-6.CH012
11. Brochado, Ângela F., Rocha, Eugénio M., Almeida, Duarte, Sousa, Amaro de, and Moura, Ana.: *A Data-Driven Model with Minimal Information for Bottleneck Detection - Application at Bosch Thermotechnology*. Manuscript submitted for publication (2022).
12. Rocha, Eugénio M., Brochado, Ângela F., and Moura, Ana.: *Workers Benchmarking Using Multi-Directional Efficiency Analysis in a Manufacturing Production System*. In Procedia Computer Science 200, January: 1451—1460 (2022). https://doi.org/10.1016/J.PROCS.2022.01.346
13. Bogetoft, Peter, and Jens Leth Hougaard: *Efficiency Evaluations Based on Potential (Non-Proportional) Improvements*. Journal of Productivity Analysis 12 (3): 233–47 (1999).
14. Manevska-Tasevska, Gordana, Helena Hansson, Mette Asmild, and Yves Surry: *Exploring the Regional Efficiency of the Swedish Agricultural Sector during the CAP Reforms – Multi-Directional Efficiency Analysis Approach*. Land Use Policy 100, January: 104897 (2021).
15. Lei, Xiyang, Xuefei Zhang, Qianzhi Dai, and Lin Li: *Dynamic Evaluation on the Energy and Environmental Performance of China's Transportation Sector: A ZSG-MEA Window Analysis*. Environmental Science and Pollution Research 28 (9): 11454–68 (2021).
16. Asmild, Mette, and Kent Matthews: *Multi-Directional Efficiency Analysis of Efficiency Patterns in Chinese Banks 1997-2008*. European Journal of Operational Research 219 (2): 434–41 (2012).
17. Dyson, R. G., R. Allen, A. S. Camanho, V. V. Podinovski, C. S. Sarrico, and E. A. Shale: *Pitfalls and Protocols in DEA*. European Journal of Operational Research 132 (2): 245–59 (2001).

18. Hahn, G.J., M. Brandenburg, and J. Becker: *Valuing Supply Chain Performance Within and Across Manufacturing Industries: A DEA-Based Approach*. International Journal of Production Economics, June, 108203 (2021).
19. Wang, Qiushi, Zhenpo Wang, Lei Zhang, Peng Liu, and Zhaosheng Zhang: *A Novel Consistency Evaluation Method for Series-Connected Battery Systems Based on Real-World Operation Data*. In IEEE Transactions on Transportation Electrification 7 (2): 437–51 (2021).
20. Qian, Jianjun, Wai Keung Wong, Hengmin Zhang, Jin Xie, and Jian Yang: *Joint Optimal Transport with Convex Regularization for Robust Image Classification*. In IEEE Transactions on Cybernetics 52 (3): 1553–64 (2022).
21. Liu, Weidong, Xiaojun Mao, and Xin Zhang. *Fast and Robust Sparsity Learning over Networks: A Decentralized Surrogate Median Regression Approach*. In IEEE Transactions on Signal Processing 70: 797–809 (2022).
22. Chen, Hongge, Huan Zhang, Duane Boning, and Cho Jui Hsieh: *Robust Decision Trees Against Adversarial Examples*. In Proceedings of 36th International Conference on Machine Learning, ICML 2019 2019-June (February): 1911–26 (2019).
23. Dairu, Xie, and Zhang Shilong: *Machine Learning Model for Sales Forecasting by Using XGBoost*. In IEEE International Conference on Consumer Electronics and Computer Engineering, ICCECE 2021, January, 480–83 (2021).
24. Gautam, Vinay, and Prabhjot Kaur: *Malware Classification Based on Various Machine Learning Techniques*. In Proceedings of 2nd International Conference on Artificial Intelligence: Advances and Applications 141–151 (2022).
25. Shapley, L.: *A Value for N-Person Games*. In Contributions to the Theory of Games II, 307–17. Princeton University Press (1953).
26. Rocha, Eugénio M., and Lopes, Maria J.: *Bottleneck Prediction and Data-Driven Discrete-Event Simulation for a Balanced Manufacturing Line*. In Procedia Computer Science 200, January: 1145—1154 (2022). https://doi.org/10.1016/J.PROCS.2022.01.314
27. Lopes, Maria J., and Rocha, Eugénio.M.: *Impact analysis of KPI scenarios, automated best practices identification, and deviations on manufacturing processes*. Manuscript submitted for publication (2022).
28. Brochado, Ângela F., Rocha, Eugénio M., and Pimentel, Carina.: *PDCA Protocol to ensure a Data-Driven Approach for Problem-Solving*. In Proceedings of the 12th Annual International Conference on Industrial Engineering and Operations Management, Istanbul, Turkey (2022).