# Benchmarking and Prediction of Entities Performance on Manufacturing Processes through MEA, Robust XGBoost and SHAP Analysis

Eugénio M. Rocha
*Center for Research & Development in Mathematics and Applications*
*Department of Mathematics, University of Aveiro*
Aveiro, Portugal
0000-0003-3628-6795

Ângela F. Brochado
*PhD Student of Industrial Engineering and Management*
*Department of Economics, Management, Industrial Engineering and Tourism*
Aveiro, Portugal
0000-0002-1222-9529

Beatriz Rato
*Master Student of Mathematics and Applications*
*Department of Mathematics, University of Aveiro*
Aveiro, Portugal

Joana Meneses
*Technology and Quality Department*
*Vista Alegre Atlantis, S.A.*
Ílhavo, Portugal

***Abstract*—Determining the reasons for process variability of manufacturing processes is generically quite demanding. In the era of big data and Industry 4.0, data-driven root cause analysis (RCA) techniques are required to support the identification of such reasons. However, an important issue with classical RCA methods is their sensibility to data perturbations. In fact, adversarial data perturbation is currently one of the hot topics in the literature. Such sensibility phenomena requires the implementation of robust RCA approaches. Here, methods of operational research (multi-directional efficiency analysis), machine learning (eXtreme Gradient Boosting), and game theory (Shapley values) are merged, to obtain a robust approach that can (1) benchmark entities acting on a manufacturing process, (2) determine the importance level of process variables regarding an entity belonging to the (in)efficient group, and (3) predict the performance of the entity's future work sessions. A use case at Vista Alegre Atlantis S.A., a Portuguese leader company that manufactures porcelain tableware, high-quality glass and crystal, is analysed to show the methodology's success.**

***Index Terms*—root cause analysis, multi-directional efficiency analysis, SHAP analysis, performance prediction, porcelain industry**

## I. Introduction

The employment of preventive/corrective actions targeting process improvement is a frequent activity performed in any company that aims to be competitive. During the last years, across the literature, it has been proven that companies with a continuous improvement culture show higher efficiency levels and performance ratings (e.g., see [1]). This mindset was brought by the Lean management philosophy of waste reduction [2] and Six Sigma statistical analysis for process control [3], with its most known problem-solving methodology: Define-Measure-Analyse-Improve-Control (DMAIC).

At the production stage, the tasks of a continuous improvement (CI) team are to: (a) derive a set of actions that potentially improve the process performance (design an improvement plan), and (b) measure the effect of such actions in a precise and robust manner to evaluate the proposed improvement plan - besides several other CI tasks. Among the many techniques used in step (b), quality control tools of statistical type are commonly applied with reasonable results, particularly in the six sigma methodology. In step (a), the selection of adequate actions implies the use of some sort of root cause analysis (RCA) to identify the issues of the process under study. Nevertheless, one of the biggest issues in companies that have plenty of information about their processes is the lack of effective process mining capabilities able to determine relevant relationships between data fields, getting far more information than the common statistical data analysis. In fact, the most relevant RCA techniques for CI are still under active study, e.g., see [4]. In this work, we focus on step (a), proposing an RCA approach that merges techniques from operational research (multi-directional efficiency analysis), machine learning (eXtreme Gradient Boosting) and game theory (SHAP). The proposed work is based on a larger project of our research team that aims to implement an automated data-driven platform, based on the DMAIC methodology, for supporting the work of CI teams. As far as we know no study was found regarding performance variation prediction and (or) benchmarking in manufacturing processes, so possible limitations of the approach compared to others could not be provided. Figure 1 shows the location of this work in the global context of the DMAIC, and its connection with a complementary RCA approach, based on comparisons of metrics distributions, data entropy levels and scenarios determination; marked in the figure as RC1 [5].

To relate process performance with the data-driven RCA approach, we consider three ingredients: a series of quantifiable

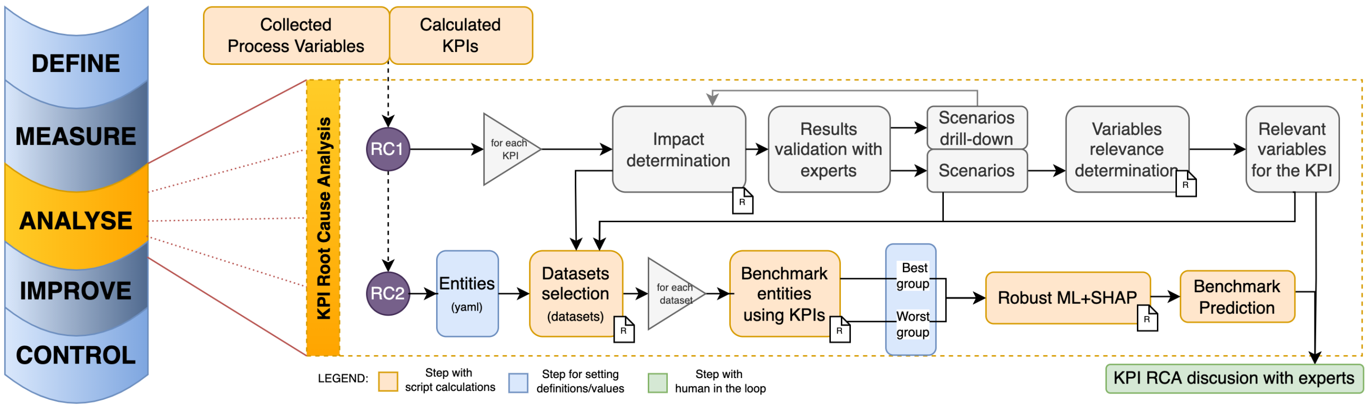


Fig. 1: Scheme representing the integration of the approach (RC2) in an automated data-driven DMAIC methodology.

indicators to be chosen *a priori* as guidelines of performance loss, a set of process variables to be linked with root causes, and a method to aggregate the information in time periods of fixed duration. In manufacturing, many key performance indicators (KPI) can be used (see [6]), as the well-known Overall Equipment Effectiveness (OEE), usually defined by

$$OEE = Quality \times Availability \times Productivity,$$

see [8]. Experts will choose the process variables according to the CI plan objectives, together with the available information in the company's data warehouse or by implementing a data collection subproject.

Nowadays, due to the showcased benefits in production management operations, many companies detain a manufacturing execution system (MES) [9], gathering shop-floor data. Depending on the digitalization level of the company, we assume the information required to apply the approach of figure 1 is already available, can be extracted from the MES and calculated by a system with a mathematical formula [7], or manually acquired and transformed to the digital form.

Figure 2 presents the scheme of the proposed approach, which shall be explained in further detail in Section II, and then applied to the use case of a Portuguese porcelain and crystal manufacturing company in Section III. If we take a first look at this figure, a key notion of the approach is the definition of *Entities* that make actions on the process, directly affecting process performance. Although being defined as a generic notion, for the use case purpose, we consider that Entities are tuples on the form of:

$$Entity = (workTeam, workShift, partReference),$$

where $workTeam$ is the current team operating on the process, $workShift$ is the current shift (e.g., morning, afternoon, night), and the $partReference$ is the identification number of the part in production. Then, the main focus questions, in the light of the scientific challenges and applicability of the approach to real use cases, are the following:

(R1) Is it possible to have a valid multi-objective performance evaluation about the degree of improvement of KPIs for consecutive time periods of actions made by Entities?

(R2) Is it possible to know which process variables/KPIs are most relevant to explain higher or lower performance variation?

(R3) Is it possible to robustly predict the future performance of an entity based on his current achieved performance?

Supporting the answers to the above questions is the core target of this research.

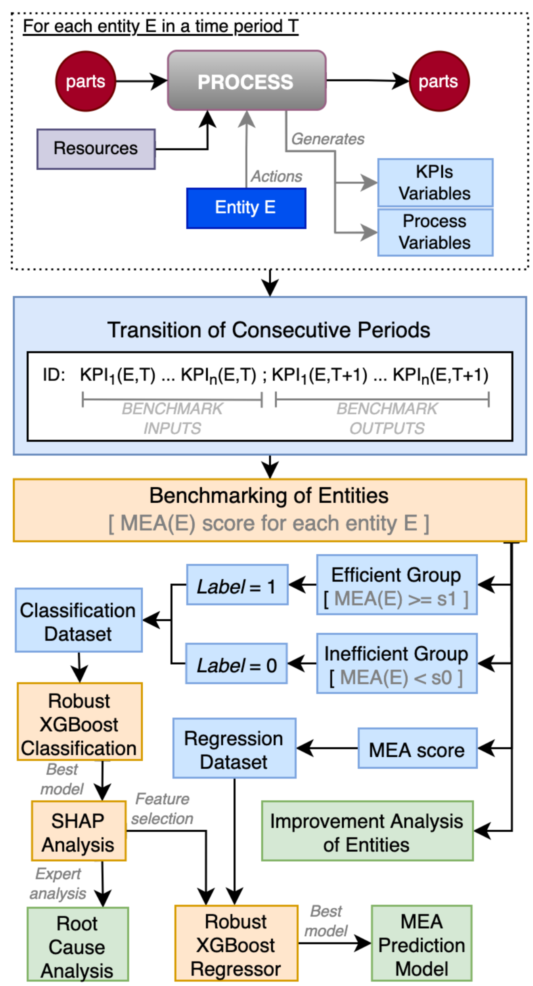


Fig. 2: Methodology scheme (color legend of figure 1).

## II. Methodology

The first step in the methodology is to define the *physical problem* (see the dashed box of figure 2). Generically, a process transforms a set of inputs (parts), and converts them into outputs. In the vast majority of manufacturing processes, the transformation requires resources, which are elements that don't directly affect the parts flow, however are essential for machines and inner processes to work reliably. Thus, they will not be considered in our model but in the form of indirect process variables. Additionally, in not fully automated systems, manufacturing processes need human intervention (individual workers or shift teams operating machines in the system), which are a part of the entities' tuples.

The three green boxes of the methodology scheme represent the main outcomes associated with questions (Q1)-(Q3) to help experts analyse the causes of process degradation conditions.

### A. Multi-Directional Efficiency Analysis

Proposed by [10] 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 [11]–[13]). 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.

Because DEA is restricted to the radial or proportional contractions of inputs (or expansions of outputs) [14], in a situation where inputs have been consumed in a certain manner, savings will be obtained in equal proportion. Having such limitation, a range of procedural concerns arise: the homogeneity of the units under evaluation, the chosen input/output set, and the measurement and weights attributed to the variables selected [15]. Here is where MEA matters, allowing to separately determine the efficiency/slack for each factor, presuming that all other input factors remain unchanged, admitting diverse combinations of earns and losses for final superior performance [16].

Hence, this part of the methodology intends to benchmark a set of Entities based on their inputs and outputs. In particular, in the output-oriented version, the Entities that appear at the top of the benchmark are the ones who are capable of maximizing outputs when inputs are somehow normalized and comparable. In what follows, a description of the MEA model is presented, so as the notation used from now on.

The tuple $n = (e, T) \in \mathcal{N}$ identifies a pair Entity $e \in E$ and Time Period $T \in T$. The notation $[m]$ designates the set $\{1, ..., m\}$ for some $m \in \mathbb{N}$. Thus, any given tuple $n \in \mathcal{N}$ produces $O \in \mathbb{N}$ outputs $y_o(n)$ (with $o \in [O]$) and consumes $I \in \mathbb{N}$ inputs $x_i(n)$ (with $i \in [I]$). Looking to figure 2, the first blue box gives information about the inputs and outputs of the model, briefly explaining the notion of *Transition of Consecutive Periods*. For each tuple $n = (e, T)$ and a given set of KPIs, the inputs are just the values of the set of KPIs, determined at the time period $T$, and the outputs are the values of the same set of KPIs but at the time period $T+1$, where $T+1$ means the consecutive time period after $T$ where the same Entity acted. Hence, the MEA score represents a relative ranking, measuring the maximization of the consecutive KPI results (outputs), considering the KPI levels of the current time period (inputs).

In this mathematical model, the first $1 < D \leq I$ inputs are named discretionary inputs, i.e., variables that enter into the optimization process, because the non-discretionary inputs are variables that cannot be changed. Accordingly, the input vector is $x(n) \in \mathbb{R}^I$ and $y(n) \in \mathbb{R}^O$ is the output vector for a given (Entity, Time Period) tuple. Moreover, the data set $W = \{w(n)\}_{n \in \mathcal{N}}$ denotes the set of values $w(n) = (x(n), y(n))$ for all $n \in \mathcal{N}$.

Just to clarify what the *Transition of Consecutive Periods* means in practice, and according to the mentioned notation, we bring a small example. Suppose the set of entities is $\mathbb{E} = \{e_1, e_2\}$ and the set of time periods is $\mathbb{T} = \{2020\_02, 2020\_03\}$. For the specific tuple $n = (e_1, 2020\_02)$, the data set $W$ may contain the line:

$$w(n) = (ID_n, x(n), y(n))$$

where

$$x(n) = (OEE(e_1, 2020\_02), QF(e_1, 2020\_02)),$$
$$y(n) = (OEE(e_1, 2020\_09), QF(e_1, 2020\_09)),$$

and $QF$ are the initials of Quality Factor.

Regarding the efficiency measurement of decision making units, the Variable Returns to Scale (VRS) model has been utilized [17], being the set defined as $\Lambda = \left\{\lambda \in \mathbb{R}^{\mathbb{N}} : \sum_{n=1}^{N} \lambda_n = 1\right\}$, where $N$ is the cardinality of $\mathcal{N}$.

The MEA score for every specific observation $w(\bar{n}) = (x(\bar{n}), y(\bar{n}))$ is calculated by computing the following linear optimization programs:

$$\begin{array}{l}
\underline{\text{Problem } P_m^{\alpha}(w, \bar{n})} : \\
\min \alpha_m(\bar{n}) \ \textit{such that} \\
\sum_n \lambda_n x_m(n) \leq \alpha_m(\bar{n}), \\
\sum_n \lambda_n x_i(n) \leq x_i(\bar{n}), \ i \in [I], i \neq m, \\
\sum_n \lambda_n y_l(n) \leq y_l(\bar{n}), \ l \in [O],
\end{array}$$

$$\begin{array}{l}
\underline{\text{Problem } P_o^{\beta}(w, \bar{n})} : \\
\max \beta_o(\bar{n}) \ \textit{such that} \\
\sum_n \lambda_n x_i(n) \leq x_i(\bar{n}), \ i \in [I], \\
\sum_n \lambda_n y_s(n) \leq \beta_o(\bar{n}), s \in [O], \\
\sum_n \lambda_n y_l(n) \leq y_l(\bar{n}), l \in [O], l \neq o,
\end{array}$$

$$\begin{array}{l}
\underline{\text{Problem } P^{\gamma}(\alpha, \beta, w, \bar{n})} : \\
\max \gamma(\bar{n}) \ \textit{such that} \\
\sum_n \lambda_n x_i(n) \leq x_i(\bar{n}) - \gamma(\bar{n})(x_i(\bar{n}) - \alpha_i^*(\bar{n})), i \in [M], \\
\sum_n \lambda_n x_i(n) \leq x_i(\bar{n}), i \in [I] \setminus \{m\}, \\
\sum_n \lambda_n y_l(n) \geq y_l(\bar{n}) + \gamma(\bar{n})(\beta_l^*(\bar{n}) - y_l(\bar{n})), l \in [L],
\end{array}$$

where $\lambda \in \Lambda$, $\alpha_m^*(\bar{n})$ and $\beta_o^*(\bar{n})$ are the optimal solutions to the problems $P_m^{\alpha}(w, \bar{n})$ and $P_o^{\beta}(w, \bar{n})$, respectively. The ideal point of $(x(\bar{n}), y(\bar{n}))$ is determined by the MEA output vector $\zeta(n) \doteq (\alpha_1^*(n), ..., \alpha_d^*(n), ..., x_I(n), \beta_1^*(n), ..., , \beta_O^*(n))$. Thus, regarding the discretionary variables, from this point forward they are represented by the first indices $d$, $1 < d < I$. Thus, $i \in [D]$ indicates the discretionary inputs and $i \in [I] \setminus d$ the non-discretionary inputs. In this setting, for a certain observation $w(\bar{n}) = (x(\bar{n}), y(\bar{n}))$ the methodology consists of $(|D|+|J|+1)\times N$ linear programs, that comprises: one problem $P_d^{\alpha}(w,\bar{n})$ for each discretionary input $d\in[D]$, one problem $P_o^{\beta}(w,\bar{n})$ for each of the output dimensions $o\in[O]$, and one problem $P^{\gamma}(\alpha,\beta,w,\bar{n})$.

Therefore for a given data set $W=\{w(n)\}_{n\in\mathcal{N}}$, the **MEA score** of $n\in\mathcal{N}$ is given by the expression

$$MEA(n)=\frac{\frac{1}{\gamma^*(n)}-\frac{1}{D}\sum_{i=1}^{D}\frac{x_i(n)-\alpha_i^*(n)}{x_i(n)}}{\frac{1}{\gamma^*(n)}+\frac{1}{O}\sum_{o=1}^{O}\frac{\beta_o^*(n)-y_o(n)}{y_o(n)}}, \qquad (1)$$

where $\alpha_i^*(n)$, $\beta_o^*(n)$ and $\gamma^*(n)$ are the optimal solutions of $P_i^{\alpha}(w,n)$, $P_o^{\beta}(w,n)$ and $P^{\gamma}(w,n,\alpha^*,\beta^*)$, respectively. Thus, the MEA score is obtained by the directional contribution of each input and output variable. As a matter of fact, for the input $i\in[I]$ the contribution in $w(\bar{n})$ is given by

$$\text{mEff}_i(n)=\frac{x_i(n)-\gamma(n)(x_i(n)-\alpha_i^*(n))}{x_i(n)}\chi_{[D]}(i), \qquad (2)$$

where $\chi_{[D]}$ is the characteristic function of the set $[D]$, which means $\chi_{[D]}(i)=1$, if $i\in[D]$ and $\chi_{[D]}(i)=0$ if $i\notin[D]$. For the outputs $o\in[O]$ the contribution is given by:

$$\text{mEff}_o(n)=\frac{y_o(n)}{y_o(n)+\gamma(n)(\beta_o^*(n)-y_o(n))}. \qquad (3)$$

Additionally, this methodology can also provide information about individual inefficiencies. Hence, following the ideas of [14], for a data set $W=\{w(n)\}_{n\in\mathcal{N}}$, the inefficiency index for each input $i\in[I]$ and tuple $n\in N$ is given by

$$\text{mIneff}_i(n)=\frac{\sum_{n=1}^{N}\gamma(n)(x_i(n)-\alpha_i^*(n))}{\sum_{n=1}^{N}x_i(n)}. \qquad (4)$$

From MEA benchmark study, it is possible to obtain three types of information: (i) the so-called *Efficient Group* composed of Entities with a MEA score equal to or bigger than a defined threshold $s1\in]0,1]$; (ii) the *Inefficient Group* of Entities with an MEA score less than a fixed $s0\in[0,1[$ threshold; and (iii) the MEA score results for all Entities in the data set $W$. Notice that Entities with MEA scores between $[s0,s1[$ are not considered in (i) or (ii). Afterwards, (i) and (ii) are labelled with the values 1 and 0, respectively, to build and define the so-called *Classification data set*. The *Regression data set* is constructed with the information of (iii) but both data sets are enriched with the process variables and the KPIs of the time period $T$, behaving as machine learning features.

### *B. Robust XGBoost Classification for the SHAP Analysis*

From the *Classification data set*, the next step is to build a robust classification model, picking out the best model at the end of the train/test phase. To build both classification and regression models (this last one shall be discussed in the section that follows), the eXtreme Gradient Boosting (XGBoost) package for Python is used. 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 [18], and [19]). 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.

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 [20], which aims to analyse complex models when there's a set of features that work as inputs and 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 about how to fairly distribute the prediction among the features. Therefore, it gives a powerful measure about 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. Also, the interaction effect between model features can also be computed using SHAP.

A key issue in the classical feature importance (FI) approach in machine learning is that FI is generally not stable to small perturbation of the features' data. 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 [21]–[23]). 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 and model predictions are also more valid and trustworthy. Usually, those robust models attain smaller performance than not robust ones, but by construction, they are far more reliable for FI. For the above reasons, the work from [24] combining robust ML models with XGBoost has been studied and employed in the methodology.

Together, all these outcomes aim to provide process engineers with reliable and robust root cause analysis, relative to process performances, one of the main outputs from figure 2.

### *C. Robust Regression Model for Entity Performance Prediction*

Researchers and industrial engineers have built up a wealthy literature on regression models and their applications to real-life cases. Regression [25] provides an estimation about the relationship between a set of dependent and independent variables for two big purposes: (1) to infer causal relationships between these variables to enable the identification of the root causes of a problem and (2) to predict future events based on the variables information, a usual practice in machine learning. Nevertheless, the major pitfall of some existent regression approaches is their lack of robustness to data outliers and influential observations (common features of training sets). Here, the aim of having a robust regression model for the MEA score is to advise and warn manufacturing line managers about future entities performance fluctuations based on the most important features identified by the previous classifier + SHAP approach and taking the robust XGBoost model with the best optimized parameters for the fixed evaluation metric.

## III. Use Case of a Portuguese Porcelain and Crystal Manufacturing Company

The methodology presented in the previous section is general enough to be applied to numerous situations. Howbeit, to see its applicability potential, we choose a case where a continuous improvement action plan has already been applied, with success, with the aim of showing that a deep data-driven RCA approach can give further useful information than classical approaches. The following use case characterizes a crucial factory process of Vista Alegre Atlantis S.A., a Portuguese leader company that manufactures porcelain tableware, decorative pieces, high-quality glass and crystal. The company is a global reference for manufacturing one of the purest crystals in the world and for earning numerous design awards internationally. Their top priority has been closely devoted to final product quality, keeping with very high quality and product conformity standards.

After many years in the market, nowadays, the company is trying to balance excellent final product quality with excellent process quality. Manufacturing line processes have been analysed, and several improvement projects conducted. Recently due to a moderate rate of product rejections by conformation, a *Traycasting* machine (nomenclature given by the company) was intervened. The problem was identified at the $10^{th}$ week of 2019, and after a long period of problem study and planning, the quality team implemented a set of improvement actions. The Plan-Do-Check-Act (PDCA) methodology for problem-solving was followed, and the results of implementation assessed at the Check stage. Despite achieving several improvements, after final evaluation it was clear that a deeper root cause analysis of the problem data was missing to induce more effective improvement actions. Thus, according to the methodology designed in figure 2, the first step was to define the KPIs (see Table I) and the process variables (see Table II). Due to production process confidentiality, the company did not reveal the real definitions of Q1, Q3, Q8, Q10 in Table II, so a general description is given.

At Table I, $n$ refers to the number of shifts that operated in a certain week. Each shift lasts for about $ShiftTotalTime$ seconds, and the $TheoTC$ corresponds to the theoretical cycle time of the product family. A data set with all the required data for the process variables and for the KPI calculations, aggregated weekly, from week 2 of 2020 to week 6 of 2021, was extracted from the SQL system of the company.

TABLE I: KPIs mathematical formulas

| *KPI* | *Mathematical Formula* |
|---|---|
| OEE | $\sum_{i=1}^{n} \frac{[Produced - NOK]_i \cdot TheoCT_i}{ShiftTotalTime}$ |
| Quality Factor | $\sum_{i=1}^{n} \frac{[Produced - NOK]_i}{Produced_i}$ |

TABLE II: Quality process variables

| *Process Variable* | *Description* |
|---|---|
| Q1 | Issue 1 |
| Q2 | Manufacturing defect |
| Q3 | Issue 2 |
| Q4 | Crack defect |
| Q5 | Defect by presence of plaster |
| Q6 | Issue 3 |
| Q7 | Air bubble defect |
| Q8 | Issue 4 |
| Q9 | "Others" |
| Q10 | Issue 5 |
| Produced | Produced parts |
| NOK | "Not Ok" produced parts |

### A. Plugins to calculate KPIs and extract process variables

To generate the data from the company data sets, three Python plugins were implemented. Each plugin receives a specific Entity and Time Period: $(workTeam, workShift, partReference)$, $(week, year)$.

The first plugin filters the data set by Entity, making it possible to compute the KPIs (OEE and Quality Factor). The second plugin generates the process variables, also filtering the data. Afterwards, a plugin computes the number of parts that came out of production for Q1 to Q10 separately.

### B. Results discussion

The methodology was employed to the *Traycasting* process data set, and results are displayed on figures 3-9.

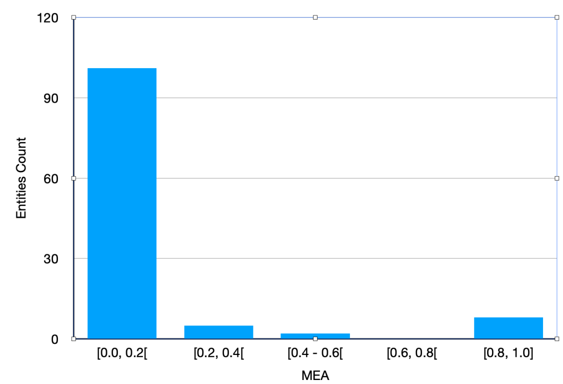


Fig. 3: MEA score distribution for all benchmarked entities.

As mentioned earlier, the MEA score is a relative benchmark value, between zero (less efficient) and one (most efficient), used in our situation to rank Entities that were able to improve the KPIs of the next session, when the KPIs of the current session were (by the algorithm) normalized to have a comparable ranking between entities. Thus, when we see a small MEA value, it means that the entity was not able to improve its performance at that point in time, i.e., it does not mean that its performance decreased. Efficient entities naturally will have more difficulty improving their performance (which is already good), since they are near to the process's upper limit of production. Nevertheless, in the scope of this work, our interest is to capture the entities that still were able to improve and then extract the main reasons for such events.

The MEA score distribution (figure 3) shows that some entities were still able to make improvements, having six entities that attain the maximum score of 1.

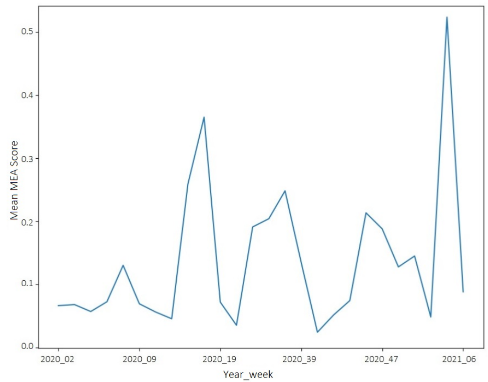


Fig. 4: Mean MEA score fluctuation for all benchmarked entities from week 2 of 2020 to week 6 of 2021; some weeks are missing since the selected products were not produced.

Regarding the MEA variability along time, figure 4 displays the average MEA scores of all benchmarked Entities for each week_year. Four pronounced peaks can be found in the graphic, meaning that for several weeks straight, the average entities' performance score increased (e.g., see mean MEA score fluctuation between 2020_13 and 2020_17). As expected, after a significant improvement is quite hard to keep improving, so a non-increase or even a decrease in the variation of performance can happen. Anyway, globally, the KPI difference between $t+1$ and $t$ is around 2.053 for the OEE, and 0.704 for the quality factor, having an accumulated value in the study period of 238.52 for the OEE and 81.682 for the quality factor; which seems to be a good result.

Figures 5 and 6 provide a more particularized analysis. The first graphic displays the entities' performance for each team separately and the second one for each product. Indirectly, both give information about the *best work teams* and *best products* in terms of demonstrated performance, where abnormal performance scores may also be explored. These so-called *performance outliers* provide a second level of RCA of the process. In particular, team $T1$ attain the majority of the improvements, $T3$ had only one full efficient occurrence that seems like an outlier, and $T2$ was completely steady concerning improvement actions. In a different perspective, figure 6 shows that parts $P2$ and $P4$ are the most problematic. It is also clear that higher OEE values enable further improvements and difficult parts only allow entity improvements at the minimum values of OEE, see $P2$ and $P3$ at OEE $min$. With the above information, experts can look at these specific identified situations, dig into them, and derive good/bad practices for future improvement plans.

After a grid hyper-parameter optimization of the XGBoost classifier, the best model attained (macro and weighted) precision, recall and F1 of 89%, and accuracy of 91%. Notice that the problem is a balanced binary classification with 45% vs 55%. These evaluation metric values were considered good enough to assume that the SHAP analysis is relevant.

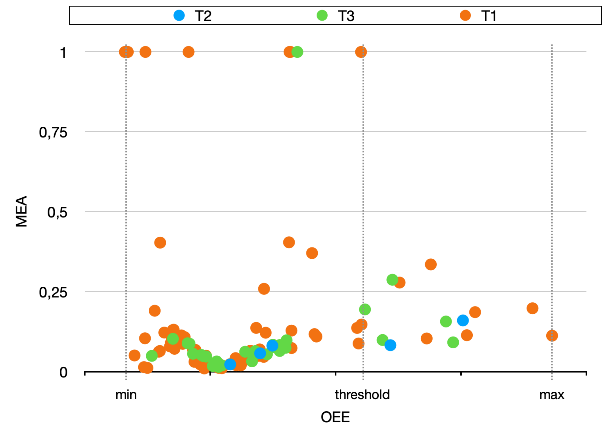


Fig. 5: Relationship between the individual MEA scores of each Work Team T1, T2, T3 and the calculated OEE.

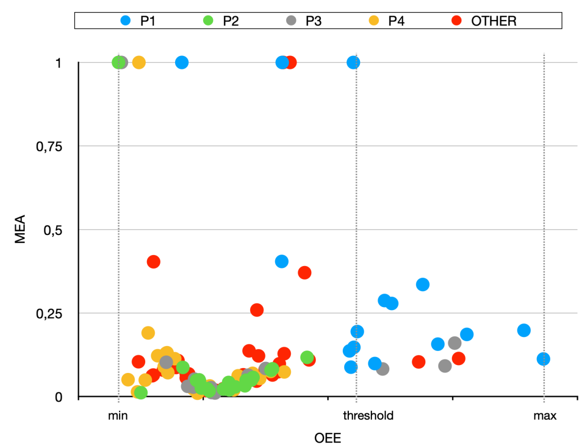


Fig. 6: Relationship between the individual MEA scores obtained for each Part Reference P1, P2, P3, P4, OTHER and the calculated OEE.

Recall that SHAP analysis will be a RCA for the difference between the Efficient Group ($G_+$) of entities and the Inefficient Group ($G_-$) of entities. Figure 7 displays the SHAP variance importance plot, drawn with all the dots of the train data. These variables are ranked according to their importance to the model in descending order. The color represents whether the value of that variable is high or low for an observation (red or blue, respectively). So, in terms of importance, a high OEE value has a positive impact in order to be in the $G_+$ group (meaning it has a positive SHAP value; see the X-axis). Focusing on metrics with a SHAP value $> 0.5$, we conclude that $G_+$ prefer to have lower values of *Quality* and *Produced*, which is also marked in the opposite way, and higher values for these variables push entities to the $G_-$ group. It is very intriguing to identify this competing behaviour between OEE and Quality. Another interesting result is the fact that process variables Q1-Q6 and Q10 do not play a significant role for an entity to be in one of the groups $G_+$ or $G_-$. So, the percentage of those product defects have similar patterns in both groups and CI actions on them would not produce significant process performance improvements.

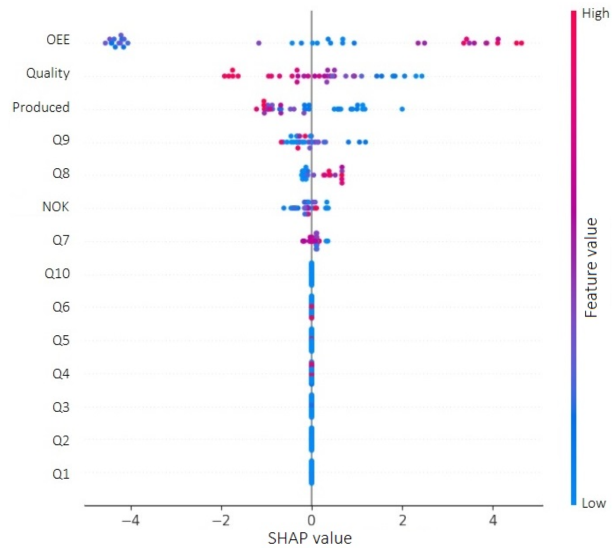


Fig. 7: Variance importance plot computed with SHAP.

Figure 8 shows in more detail the partial dependence plot of OEE and Quality regarding the groups allocation. It is clear that it displays an approximately linear with negative slope relationship between both KPIs. So, oppositely to what company engineers would assume, efficient entities are associated with a high OEE value and a medium-low Quality value.

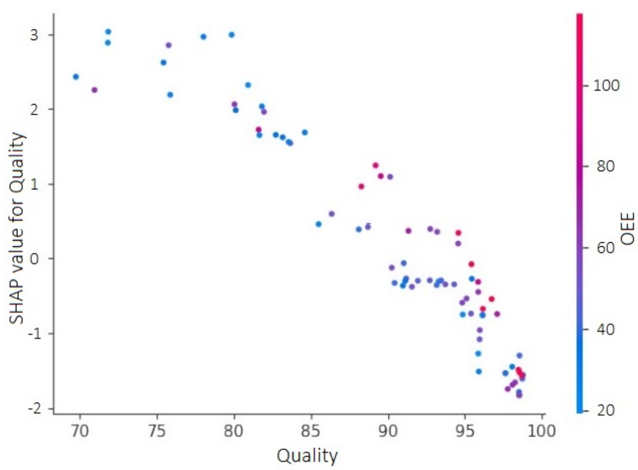


Fig. 8: Quality and OEE relationship according to SHAP analysis.

### C. Predicting Entities Performance

Another outcome of the approach described in figure 2 is the possibility of deploying a machine learning model to predict the entities' MEA score at their future working session, every time they finish the job. The features used for the regression model were only the current session values of

$$(OEE,\quad Quality,\quad Produced),$$

which are the three more relevant features in the SHAP analysis. Applying standard cross-validation of 70%-30%, figure 9 compares the MEA performance results of the test data set with the predicted outcomes of the XGBoost robust regression model. The attained RMSE metric is 0.09, which is quite good for the problem at hand. This level of error allows a good margin of confidence in the model's application, giving production managers another tool for team planning.

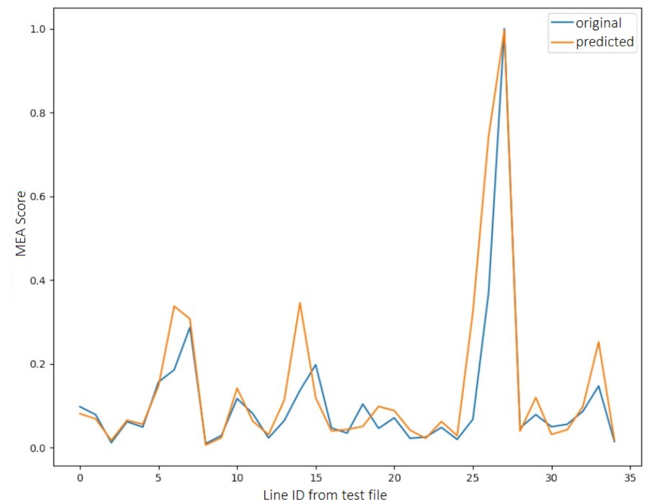


Fig. 9: MEA score prediction results of next session entities performance by the robust XGBoost regression model.

## IV. Conclusion

In this work, an innovative approach was presented, merging techniques from different research areas to derive a robust approach of root cause analysis (RCA) for manufacturing processes. A key ingredient is the notion of entities that act on processes and their relative ranking regarding their KPI performance on the next work session based on the KPI performance of the current work session. A use case in the Porcelain industry sector was explored, showing the approach. Even using a data set compromising already a continuous improvement plan with moderate success, the approach was able to identify further potential issues of improvement. A clear limitation is the fact that the applicability of the approach depends on the set of process variables chosen by the expert, since they become features that may induce low values of the machine learning evaluation metrics of the classification model (that is the base of the SHAP analysis). Another concern, is that the approach does not deal (automatically) with the possible existence of unbalanced classes, whereas in our use case such was not a problem. The above issues will be look at in future studies, where the integration and fusion of data with other root cause analysis approaches may be discussed, and the implementation of an automatic selection of (relevant) process variables may be constructed. This work is intended to be deployed to production testing as a set of micro-services, communicating with Kafka brokers connected to the manufacturing execution system.

## Acknowledgments

The first 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/04106/2022. The second author has a PhD grant in the scope of the Project Augmented Humanity (PAH) [POCI-01-0247-FEDER-046103], financed by Portugal 2020, under the Competitiveness and Internationalization Operational Program, the Lisbon Regional Operational Program, and by the European Regional Development Fund.

**Data availability statement:** The data sets used in this work are confidential information of Vista Alegre Atlantis company, so they are not publicly available. The data labels presented were changed to ensure anonymity without compromising any of the characteristics of the real data collected.

## References

[1] Parajuli, Manish, Christofer M. Harper, Bolivar Senior, and Neil Grigg. 2022. "Benefits and Challenges Associated with the Implementation of Lean in Transportation Agencies." Transportation Research Record 2676 (2): 186–96.

[2] Womack, James P., Daniel T. Jones, and Daniel Roos. 2007. The Machine That Changed the World. Simon and Schuster.

[3] Snee, Ronald D. 2010. "Lean Six Sigma – Getting Better All the Time." International Journal of Lean Six Sigma 1 (1): 9–29.

[4] Braglia, Marcello, Mosè Gallo, and Leonardo Marrazzini. 2021. A Lean Approach to Address Material Losses: Materials Cost Deployment (MaCD). International Journal of Advanced Manufacturing Technology 113 (1–2): 565–84.

[5] M.J.Lopes, E.M. Rocha. "Impact Analysis of KPI Scenarios, Automated Best Practices Identification, and Deviations on Manufacturing Processes." (submitted to ETFA2022).

[6] Sousa, João, João Pedro Mendonça, and José Machado. 2022. "A Generic Interface and a Framework Designed for Industrial Metrology Integration for the Internet of Things." Computers in Industry 138 (June): 103632.

[7] Zhang, Yun Yi, Zhen Zhong Hu, Jia Rui Lin, and Jian Ping Zhang. 2021. "Linking Data Model and Formula to Automate KPI Calculation for Building Performance Benchmarking." Energy Reports 7 (November): 1326–37.

[8] Hung, Yick-Hin, Leon Y.O. Li, and T.C.E. Cheng. 2022. "Uncovering Hidden Capacity in Overall Equipment Effectiveness Management." International Journal of Production Economics 248 (June): 108494.

[9] Aramja, Adil, Oualid Kamach, and Rachid Elmeziane. 2021. "Companies' Perception toward Manufacturing Execution Systems." International Journal of Electrical and Computer Engineering 11 (4): 3347–55.

[10] Bogetoft, Peter, and Jens Leth Hougaard. 1999. "Efficiency Evaluations Based on Potential (Non-Proportional) Improvements." Journal of Productivity Analysis 12 (3): 233–47.

[11] Manevska-Tasevska, Gordana, Helena Hansson, Mette Asmild, and Yves Surry. 2021. "Exploring the Regional Efficiency of the Swedish Agricultural Sector during the CAP Reforms – Multi-Directional Efficiency Analysis Approach." Land Use Policy 100 (January): 104897.

[12] Lei, Xiyang, Xuefei Zhang, Qianzhi Dai, and Lin Li. 2021. "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.

[13] Rocha, Eugénio M., Ângela F. Brochado, and Ana Moura. 2022. "Workers Benchmarking Using Multi-Directional Efficiency Analysis in a Manufacturing Production System." Procedia Computer Science 200 (January): 1451–60.

[14] Asmild, Mette, and Kent Matthews. 2012. "Multi-Directional Efficiency Analysis of Efficiency Patterns in Chinese Banks 1997-2008." European Journal of Operational Research 219 (2): 434–41.

[15] Dyson, R. G., R. Allen, A. S. Camanho, V. V. Podinovski, C. S. Sarrico, and E. A. Shale. 2001. "Pitfalls and Protocols in DEA." European Journal of Operational Research 132 (2): 245–59.

[16] Hahn, G.J., M. Brandenburg, and J. Becker. 2021. "Valuing Supply Chain Performance Within and Across Manufacturing Industries: A DEA-Based Approach." International Journal of Production Economics, June, 108203.

[17] Bogetoft, Peter, and Lars Otto. 2011. Benchmarking with DEA, SFA, and R. Edited by International series in operations research management science. Vol. 157. International Series in Operations Research Management Science. New York, NY: Springer New York.

[18] Dairu, Xie, and Zhang Shilong. 2021. "Machine Learning Model for Sales Forecasting by Using XGBoost." 2021 IEEE International Conference on Consumer Electronics and Computer Engineering, ICCECE 2021, January, 480–83.

[19] Gautam, Vinay, and Prabhjot Kaur. 2022. "Malware Classification Based on Various Machine Learning Techniques," 141–51.

[20] Shapley, L. 1953. "A Value for N-Person Games." In Contributions to the Theory of Games II, 307–17. Princeton University Press.

[21] Wang, Qiushi, Zhenpo Wang, Lei Zhang, Peng Liu, and Zhaosheng Zhang. 2021. "A Novel Consistency Evaluation Method for Series-Connected Battery Systems Based on Real-World Operation Data." IEEE Transactions on Transportation Electrification 7 (2): 437–51.

[22] Qian, Jianjun, Wai Keung Wong, Hengmin Zhang, Jin Xie, and Jian Yang. 2022. "Joint Optimal Transport with Convex Regularization for Robust Image Classification." IEEE Transactions on Cybernetics 52 (3): 1553–64.

[23] Liu, Weidong, Xiaojun Mao, and Xin Zhang. 2022. "Fast and Robust Sparsity Learning over Networks: A Decentralized Surrogate Median Regression Approach." IEEE Transactions on Signal Processing 70: 797–809.

[24] Chen, Hongge, Huan Zhang, Duane Boning, and Cho Jui Hsieh. 2019. "Robust Decision Trees Against Adversarial Examples." 36th International Conference on Machine Learning, ICML 2019 2019-June (February): 1911–26.

[25] Draper, Norman R., and Harry Smith. 2014. Applied Regression Analysis. Applied Regression Analysis. 3rd ed. Wiley.