# Industry 4.0-driven Final Quality Causal Analysis and Prediction from Raw Matter Characteristics

Amin Khodamoradi [1], Paulo Figueiras[1], Luis Lourenço[1], André Grilo[1], Bruno Rêga[1], Ruben Costa[1], Ricardo Jardim-Gonçalves[1]

[1] *CTS, UNINOVA, Caparica, Portugal*

**Abstract**

Concepts such as Manufacturing Quality Optimisation and Zero-Defect Manufacturing have become growing concerns for academia and industry over the past decade, especially in the context of Industry 4.0. Achieving near zero-defect quality in manufacturing processes heavily relies on optimizing operating parameters in manufacturing systems. The European Commission-funded i4Q project provides a set of solutions for manufacturers to optimize their operations and improve product quality, based on data gathered throughout the production processes within their manufacturing facilities. A use case within the i4Q project addresses MQO and ZDM within the operations of RiaStone, a stoneware ceramics factory situated in Ilhavo, Aveiro, Portugal. This paper discusses the necessary steps for manufacturing quality improvement at RiaStone, from data cleaning and integration to the application of manufacturing Data Analytics and Machine Learning solutions towards near zero-defect manufacturing and describes the methods and techniques that can be exploited to achieve higher production quality via a system that requires finding the key shop floor elements that influence final product quality.

**Keywords**

Industry 4.0, Artificial Intelligence, Manufacturing Data Analytics, Zero-defect Manufacturing, Manufacturing Quality Optimisation

## 1. Introduction

Despite the considerable advancements in techniques, equipment, and automation in process manufacturing industries, under the Industry 4.0 umbrella, over the past few decades, there is still a noticeable gap between discrete and process manufacturing [1], but bridging the existing gap between discrete and process industries, particularly in terms of production effectiveness and quality analysis, remains a challenge [2]. Analysis and improvement of product quality in process manufacturing have become growing concerns for academia and industry [3], since achieving optimal product quality heavily relies on optimizing operational parameters across manufacturing processes. In the quest for optimal product quality, the fields of Manufacturing Quality Optimization (MQO) and Zero-Defect Manufacturing (ZDM) leverage cutting-edge methodologies, such as the development of sophisticated data-driven machine learning (ML) models. Therefore, the concept of In-Process Quality Improvement (IPQI) has emerged as a new branch of quality science research dedicated to developing methodologies and applications to enhance product quality in manufacturing systems [4].

One of the use cases of the European Commission-funded i4Q project [5] focuses on MQO, ZDM and IPQI under the scope of RiaStone, a stoneware ceramics factory that produces tableware for the IKEA group, located in Ilhavo, Aveiro, Portugal [6]. An urgent challenge at RiaStone is to improve its Defect Rate (DR), which is the percentage of final products that present defects in relation to the

Proceedings of the 12th International Conference on Interoperability for Enterprise Systems and Applications (I-ESA), Innovations in Manufacturing Workshop, April 10th–12th, 2024, Crete, Greece
EMAIL: a.khodamoradi@uninova.pt (A. 1); paf@uninova.pt (A. 2); lcl@uninova.pt (A. 3); a.grilo@uninova.pt (A. 4); b.rega@uninova.pt (A. 5); rddc@uninova.pt (A. 6); rg@uninova.pt (A. 7)
ORCID: 0000-0002-2700-5384 (A. 1); 0000-0002-6068-1982 (A. 2); 0000-0003-3419-3640 (A. 3); 0000-0002-5168-4786 (A. 4); 0000-0003-3507-1442 (A. 5); 0000-0002-6142-1840 (A. 6); 0000-0002-3703-6854 (A. 7)

overall final product throughput. Hence, to ease the comprehension, the authors define the Overall Production Effectiveness (OPE) as the inverse percentage of DR, i.e.:

$$OPE = 100\% - DR \quad (1)$$

Presently, RiaStone has an OPE of around 92%, and its main goal is to achieve an OPE of at least 98%. Such an ambitious goal requires new approaches to promote innovative defect management and production control methods, namely in-line inspection technologies and integration of Information and Communication Technologies, as well as tools for autonomous, automatic, and smart decision-making. There are several challenges and inefficiencies in the processes that are impacting the OPE metric and causing significant levels of product rejection which are not detected by quality control and inspection.

This research aims to model the OPE of the final ceramic products, based on the collected data from shop floor data. The main research question would be: *How can one correlate raw matter properties and composition, in this case from the glazing liquid, with the final OPE?*

## 2. Related Work

MQO and ZDM stand at the forefront of enhancing production processes in vital industries. The relentless pursuit of higher-quality products in traditional large-scale process industries, driven by excess production capacity and market competition, has spurred significant advancements. Incremental improvements in product quality have been shown to yield substantial profits for the process industry [7] [8]. As the industry shifts towards solving multi-objective optimization problems, the focus is on methodologies that genuinely capture the complexity of product quality evaluation by addressing multiple conflicting indicators. By proposing innovative approaches like multi-stage and multi-model modelling frameworks, the MQO state-of-the-art strives to provide adaptive, accurate, and efficient solutions, exemplifying the industry's commitment to continual improvement [9].

As an entry point for MQO and ZDM state-of-the-art, the authors of [10] present a review on several works that apply innovative ML solutions for defect detection and prediction. These techniques range from classification or regression to clustering analyses and are applied to several type of defect detection and prediction, such as in the case of voltage control to mitigate droplet jetting defects or the classification and prediction of geometric deviation defects in CAD models. The authors of another recent review, [11], describe the usage of deep learning techniques for detecting defects in manufacturing scenarios. The authors of [12] and [13] present several ML-based defect detection approaches in industrial settings. In the former work, the study was carried out at a transmission axle assembly factory, in which a regression model was developed to simulate the typical vibration patterns of axles, enabling the identification of anomalies through the assessment of deviations between new products and the model, whereas in the latter work, the case study was conducted at a heavy-duty vehicle manufacturing factory with advanced manufacturing technology and automation levels, and applied the Cross Industry Standard Process of Data Mining (CRISP-DM) model to detect defects in the final product. As a final example, the authors of [14] present a process to introduce ML in plastic moulding processes towards MQO and ZDM. Specifically, this approach enabled the prediction and notification of process quality deterioration, resulting in fewer non-compliant parts being produced, consequently enhancing productivity while simultaneously lowering costs and minimizing environmental impact.

## 3. Research Methodology

In a well-defined ML problem, if enough data is available, it is possible to distinguish key features, and their combinations, which may have an impact on the target (the object of the problem). In the proposed research work, the target is defined as the hourly production quality which is given as the ratio between products within accepted quality thresholds and the total number of products, whether they have defects or not, i.e. the OPE.

The chosen research methodology is the CRISP-DM model [15], which splits into six phases: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation and Deployment. The methodology starts by understanding the business, production, and data-related processes in the factory, followed by the collection and storage of raw data. Afterwards, the collected data is cleaned and integrated and finally, ML methods are applied to achieve a prediction model that can predict the final product quality depending on the raw-matter composition and properties.

## 3.1. Business Understanding

The process starts by understanding the production process in the factory which includes studying the role and task of each part, the relation of data sets processes, and the interpretation of data values in each record of data. Namely, RiaStone has two main challenges related to OPE: First, density, and composition variations of incoming ceramic prime matters, produce volumetric mass density differences in post-pressing raw greenware, directly affecting the quality levels in finished stoneware products; Second, glazing and painting raw liquid matters, made at RiaStone to colour the produced tableware before the ceramic firing process, present high fluctuations in density and temperature parameters, also affecting the final OPE directly. This work tackled specifically the second challenge, related to the glazing process.

## 3.2. Data Understanding

Raw data collected from glazing liquid analysis records the liquid's composition in 15-minute intervals. To simplify the process only a few columns were considered, as shown in Figure 1. *'Linha'* is the line number, *'ref_atual'* is the product reference, *'cor_atual'* is the product colour, *'densidade_act'* means current density, and *'temperaturavid_act'* means current temperature. *'Data_inicio_leitura'* and *'Data_fim_leitura'*, correspond respectively to the recording beginning and ending timestamps.

| id | Linha | ref_atual | cor_atual | Data_inicio_leitura | Data_fim_leitura | densidade_act | densidade_min | densidade_max | temperaturavid_act |
|---|---|---|---|---|---|---|---|---|---|
| 887886 | 1 | TIGELA FK 16 | Cz Escuro | 2021-10-28 14:38:20.000 | 2021-10-28 14:43:20.000 | 0 | 1688 | 1692 | 4800 |
| 887885 | 2 | SOPA FK 23 | Verde Seco | 2021-10-28 14:36:23.080 | 2021-10-28 14:41:23.080 | 1673 | 1688 | 1692 | 31 |
| 887884 | 3 | SOPA FK 23 | Cz Escuro | 2021-10-28 14:34:48.020 | 2021-10-28 14:39:48.020 | 0 | 1688 | 1692 | 4800 |
| 887882 | 4 | SOBREMESA FK 20 | Cz Escuro | 2021-10-28 14:34:25.000 | 2021-10-28 14:39:25.000 | 0 | 1688 | 1692 | 4800 |
| 887878 | 1 | TIGELA FK 16 | Cz Escuro | 2021-10-28 14:33:20.000 | 2021-10-28 14:38:20.000 | 0 | 1688 | 1692 | 4800 |
| 887881 | 7 | SOBREMESA FK 20 | Azul Esc | 2021-10-28 14:33:13.030 | 2021-10-28 14:38:13.030 | 0 | 1688 | 1692 | 4800 |
| 887879 | 8 | PRATO FK 26 | Verde Seco | 2021-10-28 14:32:59.090 | 2021-10-28 14:37:59.090 | 0 | 1688 | 1692 | 4800 |

**Figure 1.** Example of glazing data

Quality data are extracted from quality control teams' reports that are manually filled by human operators. Presently, the quality ratio is measured by checking 40 random samples for every 1000 products per hour. The files are divided into eight tables corresponding to the eight production lines. The quality report for each line is divided into hourly quality records, as noticeable in Figure 2, from which *'hora'* is the shift hour, *'colaborador'* is the collaborator's name, *'referencia'* is the product reference, *'Cor'* is the product colour, *'$1^0$'* is the number of products with optimal quality, *'$2^0$'* is the number of products with second level quality, *'R'* is the number of rejected products, *'%qld'* is the hourly quality percentage (OPE), *'Frete colado'* is the number of glued products and *'defeitos/obs'* is the description of the highest occurring defect reasons during that hour.

| hora | colaborador | referência | Cor | 1ª | 2ª | R | total | % qld. | Frete colado | defeitos/obs. |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Lalimar | FK 26 | Azul Escuro | 816 | 20 | 35 | 871 | 93,7% | 96 | empeno frete colado |
| 2 | Lalimar | FK 26 | Azul Escuro | 744 | 20 | 24 | 788 | 94,4% | 72 | empeno frete colado |
| 3 | Barbara | FK 26 | Azul Escuro | 648 | 20 | 80 | 748 | 86,6% | 48 | pingos |
| 4 | Barbara | FK 26 | Azul Escuro | 600 | 20 | 188 | 808 | 74,3% | 48 | pingos/escorrido zona A |
| 5 | Barbara | FK 26 | Azul Escuro | 576 | 60 | 197 | 833 | 69,1% | 96 | pingos/escorrido zona A |
| 6 | Yeni | FK 26 | Azul Escuro | 576 | 20 | 103 | 699 | 82,4% | 96 | pingos/picos/residuos |
| 7 | Yeni | FK 26 | Azul Escuro | 744 | 20 | 128 | 892 | 83,4% | 216 | pingos/picos/residuos |
| 8 | Yeni | FK 26 | Azul Escuro | 600 | 20 | 85 | 705 | 85,1% | 216 | pingos/picos/residuos |
| total/média | | | | 5304 | 200 | 840 | 6344 | 83,6% | 888 | |

**Figure 2.** Example of a quality report for one shift of eight hours in an inspection table

## 3.3. Data Preparation

A key challenge in the RiaStone use case is the large amount of data that needs to be collected and analysed. Both data sets presented in the subsequent section present problems in terms of data veracity and validity. The glazing data set (Figure 1) presents the following challenges: (i) several columns are deemed not relevant for this specific use case, like line velocity, stoppage times, or number of breakdowns, since they were not significant for the defect prediction model; (ii) several data records had to be cleaned since, data is collected by different sensors at different time intervals, but the density measurement occurs at 15-minute intervals, hence the data was cleaned to only account for this temporal granularity; (iii) other parameters, like the theoretical, minimum, and maximum thresholds of several measured properties (temperature, density, viscosity, etc.) presented values that were often wrong or simply zero and were removed. Regarding the quality control data set (Figure 2), each report table is separated into three shifts, with eight hours each, and has a list of the main defect types encountered in those hours. The issue here is that there is no standardization of the defect types' nomenclature; rather it is based on operators' inputs, which are not uniformized across operators, nor the number of corresponding products that possess these defect types, which entails that no specific defect tracking of individual products is realized. Furthermore, there is no certainty that the recorded quality and, consequently, defects' numbers are completely correct and in line with the production reality.

### 3.3.1. Data Cleaning and Integration

After removing columns in both datasets and handling the previous issues in the data records, the total amount of usable data records in the glazing and quality data sets are, respectively 580538 rows and 21373 rows. The data sets were stored in a relational database (based on PostGreSQL): the cleaned glazing data was stored directly, while the unstructured monitoring reports were converted into relational data via a Python script. Since there is no track-and-trace system for individual products, a data correlation analysis was performed to integrate both data sets. From such correlation analysis, it was decided that a time gap of eight and a half hours would be used to integrate the data sets, corresponding to the time between the glazing process and the quality process. Hence, a soft margin was used to integrate as much row data from two tables as possible, aligning the data from the two data sets based on common identifiers such as *'prod_id'* and *'color_id'*. The resulting dataset combines the quality and glazing data sets. Finally, other data pre-processing steps were made, such as various transformations including adding new columns (*'hour', 'date', 'end_date', 'shift'*), populating them with calculated values based on existing columns, and excluding records that do not have glazing-related defects.

## 3.4. Modelling, Feature Engineering and Feature Selection

In the feature engineering and selection step, some categorical features like defect type are encoded via the one-hot encoder algorithm. For feature selection, Extra Tree-based stepwise, variance-based, and importance-based methods were used to come up with a narrower feature subset which was more significant as input for the ML model. The result of the feature importance analysis is shown in Figure 3. The most important features are the day of the year and, most importantly, the density of the glazing liquid. The data set has a time-series nature, and its target values are numerically continuous, thus the problem considers a time-series regression problem. Different regression models (Lasso, ridge regression, linear Support Vector Regression, Decision Tree Regression (DTR), and Extra Tree) were trained and evaluated for the given quality prediction problem, from which DTR was selected, fine-tuned and adapted to the proposed problem. The models were evaluated through a 3-fold Cross Validation approach specific for time-series data and several precision metrics were reported. For training the and evaluating the models, the following algorithm was used:

1. Split dataset into features (X) and target variable (y).
2. Split data into training and testing sets using the '*train_test_split*' function from scikit-learn.
3. Define the model.
4. '*TimeSeriesSplit*' 3-fold cross-validation is used to handle time series data.
5. '*GridSearchCV*' is used to perform a grid search over the parameter grid, optimizing for negative mean squared error to find the best value for the hyperparameters.

6. Train the model with training data and the best possible hyperparameters from the previous step.
7. Make predictions on the testing data using the trained model.
8. Evaluate the model's performance by calculating the Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared ($R^2$) score using the predicted and actual target values.

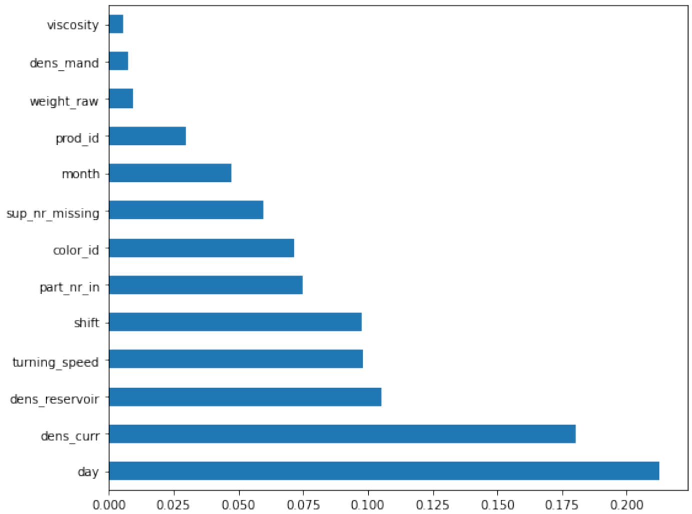


**Figure 3.** Feature importance analysis

The DTR model was chosen due to the precision metrics' results, its interpretability characteristics and the feature importance report presented in Figure 3, which shows several features with relative importance for predicting the target. Figure 4 shows the error distribution of the DTR model.

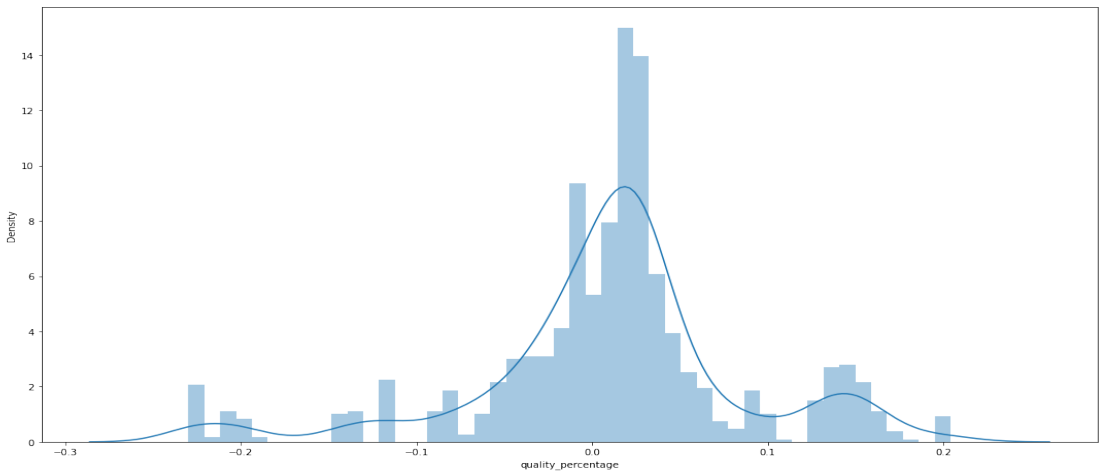


**Figure 4.** DTR regression model residual distribution

## 4. Conclusion and Discussion

The application of the DTR model had a significant impact on RiaStone's pursuit for ZDM and MQO, mainly because, from the decision tree structure that resulted from the DTR model and also the feature importance analysis shown in Figure 3, the density of the glazing liquid is crucial for the overall OPE. Hence, a simple liquid stirring and mixing procedure was added to the process, and the OPE automatically raised from 92% to at least 96% in the following weeks. Hence, *one can correlate the raw matter properties of the glazing liquid with the OPE by applying explainable regression models, such as the DTR mode to find the most important characteristics for the prediction of the OPE.*

The presented work is part of an ongoing research, and it is the authors' opinion that, although the DTR got the best results and was chosen due to the explainability potential of decision trees, other models could have better results with some fine-tuning work. In fact, these model fine-tuning procedures will continue until the end of the i4Q project.

## Acknowledgements

The authors acknowledge the contribution of the European Commission-funded Horizon 2020 research project i4Q (Grant agreement ID: 958205) for the development and validation of the presented work.

## Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

## References

[1] J. Bughin, M. Chui and J. Manyika, "Clouds, big data, and smart assets: Ten tech-enabled business trends to watch," *McKinsey Quart.,* vol. 56, no. 1, p. 75–86, 2010.

[2] T. Chai, S. J. Qin and H. Wang, "Optimal operational control for complex industrial processe," *Annual Reviews in Contro,* vol. 38, pp. 81-92, 2014.

[3] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh and A. H. Byers, "Big Data: The Next Frontier for Innovation, Competition," San Francisco, CA, USA: The McKinsey Global, 1 may 2011. [Online]. Available: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation. [Accessed 12 1 2024].

[4] A. Gandomi and M. Haide, "Beyond the hype: Big data concepts, methods, and analytics," *International Journal of Information Management,* vol. 35, pp. 137-144, 2015.

[5] European Commission, "i4Q: Industrial Data Services for Quality Control in Smart Manufacturing," Community Research and Development Information Service (CORDIS), 2021. [Online]. Available: https://cordis.europa.eu/project/id/958205.

[6] I4Q Consortium, "i4Q: RiaStone Pilot," 2021. [Online]. Available: https://www.i4q-project.eu/riastone.

[7] S. Jia, Q. Yuan, W. Cai, M. Li and Z. Li, "Energy modeling method of machine-operator system for sustainable machining," *Energy Conversion and Management,* vol. 172, pp. 265-276, 2018.

[8] Q. Yang, J. Zhang and Z. Yi, "Applied Soft Computing," *Predicting molten steel endpoint temperature using a feature-weighted model optimized by mutual learning cuckoo search,* vol. 83, p. 105675, 2019.

[9] X. Yin, Z. Niu, Z. He, Z. Li and D. Lee, "An integrated computational intelligence technique based operating parameters optimization scheme for quality improvement oriented process-manufacturing system," *Computers & Industrial Engineering,* vol. 140, p. 106284, 202.

[10] L. Meng, B. McWilliams, W. Jarosinski, H.-Y. Park, Y.-G. Jung, J. Lee and J. Zhang, "Machine Learning in Additive Manufacturing: A Review," *Journal of The Minerals, Metals & Materials Society,* vol. 72, pp. 2363-2377, 2020.

[11] A. Saberironaghi, J. Ren and M. El-Gindy, "Defect Detection Methods for Industrial Products Using Deep Learning Techniques: A Review," *Algorithms,* vol. 16, no. 2, p. 95, 2023.

[12] N. Leberruyer, J. Bruch, M. Ahlskog and S. Afshar, "Enabling an AI-Based Defect Detection Approach to Facilitate Zero Defect Manufacturing," in *International Conference on Advances in Production Management Systems*, Trondheim, Norway, 2023.

[13] N. Leberruyer, J. Bruch, M. Ahlskog and S. Afshar, "Toward Zero Defect Manufacturing with the support of Artificial Intelligence—Insights from an industrial application," *Computers in Industry,* vol. 147, p. 103877, 2023.

[14] B. Silva, R. Marques, D. Faustino, P. Ilheu, T. Santos, J. Sousa and A. D. Rocha, "Enhance the Injection Molding Quality Prediction with Artificial Intelligence to Reach Zero-Defect Manufacturing," *Processes,* vol. 11, no. 1, p. 62, 2023.

[15] R. Wirth and J. Hipp, "CRISP-DM: Towards a Standard Process Model for DataMining," in *4th international conference on the practical applications of knowledge discovery and data mining*, London, UK, 2000.