Enhancing quality 4.0 and reducing costs in lot-release process with machine learning-based complaint prediction

Armindo Lobo, Paulo Sampaio and Paulo Novais
ALGORITMI Research Centre/LASI, University of Minho, Braga, Portugal

Abstract

Purpose – This study proposes a machine learning framework to predict customer complaints from production line tests in an automotive company's lot-release process, enhancing Quality 4.0. It aims to design and implement the framework, compare different machine learning (ML) models and evaluate a non-sampling threshold-moving approach for adjusting prediction capabilities based on product requirements.

Design/methodology/approach – This study applies the Cross-Industry Standard Process for Data Mining (CRISP-DM) and four ML models to predict customer complaints from automotive production tests. It employs cost-sensitive and threshold-moving techniques to address data imbalance, with the F1-Score and Matthews correlation coefficient assessing model performance.

Findings – The framework effectively predicts customer complaint-related tests. XGBoost outperformed the other models with an F1-Score of 72.4% and a Matthews correlation coefficient of 75%. It improves the lot-release process and cost efficiency over heuristic methods.

Practical implications – The framework has been tested on real-world data and shows promising results in improving lot-release decisions and reducing complaints and costs. It enables companies to adjust predictive models by changing only the threshold, eliminating the need for retraining.

Originality/value – To the best of our knowledge, there is limited literature on using ML to predict customer complaints for the lot-release process in an automotive company. Our proposed framework integrates ML with a non-sampling approach, demonstrating its effectiveness in predicting complaints and reducing costs, fostering Quality 4.0.

Keywords Machine learning, Quality 4.0, Lot release, Cost-sensitive learning, Imbalanced data, Threshold-moving

Paper type Research paper

1. Introduction

Industry 4.0 (I4.0) companies rely on digitisation, automation and real-time operations to improve customer service (Rojko, 2017). This movement is driven by big data and artificial intelligence (AI) tools, which are the core components of improving manufacturing processes and service quality (Escobar *et al.*, 2021). Quality 4.0 (Q4.0) addresses this challenge and can be seen as a digital transformation strategy in which quality and performance goals are the top priority (Radziwill, 2020).

The lot-release decision process in industrial operations can significantly impact efficiency and service quality. Typically governed by heuristic rules, integrating machine learning (ML) can optimise it, enhancing lot-release decision quality and customer satisfaction by minimising complaints. To do this, managing imbalanced data from diverse sources is crucial for effective implementation (Fathy *et al.*, 2021). This is the case for the studied company's software application, which currently relies on heuristic rules and must deal with heavily imbalanced datasets.

Several studies have proposed frameworks and models to improve quality control in manufacturing. Villanueva Zacarias *et al.* (2018) introduce a framework for selecting and configuring ML-based data analytics solutions, considering factors such as data quality and algorithm selection. Cho *et al.* (2022) focus on data preprocessing, using different methods to address missing values and data imbalance. This paper proposes a framework that integrates an ML model to improve the lot-release decision process, reduce quality costs and contribute to the adoption of Q4.0. This approach focusses on the last stage of the production line, gathering information from automatic production tests and repairs generated along the different production stages. Based on this information, four ML algorithms *XGBoost (XGB)*, *LightGBM (LGBM)*, *CatBoost (CatB) and Random Forest (RF)*) were conceived, tuned, evaluated and compared to classify the occurrence of a customer complaint. Two non-sampling approaches (cost-sensitive learning and threshold moving) were considered to deal with imbalanced data.

The remainder of this paper is organised as follows. Section 2 presents the related work and background. Section 3 describes the methods used to deal with imbalanced datasets, the ML algorithms considered, the evaluation metrics and how the data were collected and preprocessed. Section 4 describes the experiments carried out, and Section 5 discusses the results obtained. Finally, in Section 6 are given the main conclusions and future work directions.

2. Literature review

Improving the quality of products and services is an essential component of competitiveness for every company. I4.0 promotes the digitisation of processes to create autonomous systems and integration throughout the supply chain. This environment poses new challenges that are addressed by Q4.0. The implementation of Q4.0 can be seen as a digital transformation strategy where quality and performance are crucial (Radziwill, 2020). In this sense, customer satisfaction and product-related complaints are two of the four vital quality objectives that must be improved with Q4.0 (Dror, 2022). Compared to traditional quality, regarding quality and performance goals, Q4.0 focusses on minimising or eliminating appraisal costs by detecting problems before they occur, which has a positive impact on quality costs (Radziwill, 2020).

2.1 Related work

Over the years, the prediction of anomalies in different stages of production lines using different ML techniques has been studied thoroughly; however, this is not the case for the prediction of customer complaints (Abdelrahman and Keikhosrokiani, 2020). There is little literature on the use of ML techniques to predict them based on the results of tests carried out along production lines.

Chen and Lin (2020) address this topic in a textile company, creating a system to predict the probability of complaints about a new production order based on its inherent characteristics. As customer complaints are relatively rare, they must deal with imbalanced datasets to train the classifiers; to deal with this, they propose an

upsampling approach. To evaluate the results of the three ML classifiers tested (Decision Trees, RF and XGB), they use balanced accuracy, which is the arithmetic mean of sensitivity and specificity. In their pipeline, they also consider grid search to find the best hyperparameters. The results show that when upsampling, the grid search, the area under the roc curve (AUC) metric, and the XGB classifier were coupled, the balanced accuracy during validation was maximised and the gap between balanced accuracies during training and validation was minimised.

Yorulmuş et al. (2022) use quality data from a brake assembly line of an automobile manufacturer to develop a predictive quality model to recognise products that passed quality inspection operations without defects but are problematic. To achieve this, they considered several ML algorithms and chose the specificity and negative prediction values to compare them. The values obtained show that the gradient boost and CatB algorithms achieved the best results in detecting rare events. However, despite analysing the classification results of rare events that appear in imbalanced datasets, it is not clear what approach was used to deal with them.

The viability of using ML techniques to predict compliance quality from data of multiple processes was confirmed by Sankhye and Hu (2020). In their study, they focus on analysing data from a large-scale appliance manufacturing plant to design ML-based classification methods to predict manufacturing compliance quality according to the results of quality inspections. They compare RF and XGB, with Cohen's Kappa as the reference metric and use synthetic minority oversampling technique (SMOTE) to implement an oversampling approach to deal with the imbalanced dataset. They also analysed the impact of the feature engineering process and concluded that the results improved when certain features were created by applying prior domain knowledge to the datasets nature. Overall, the best results were obtained with XGB.

Product quality is vital for customer satisfaction and cost reduction in the automotive industry. Current research primarily targets anomaly prediction at specific production stages, overlooking predicting customer complaints during lot release, which can impact reputation and profitability. This paper addresses this gap by proposing an ML framework that uses production line data to predict customer complaints in the lot-release process, thereby promoting Q4.0 adoption and reducing quality costs. The framework incorporates algorithmic-level methods like cost-sensitive learning and threshold moving to handle imbalanced data effectively. Threshold-moving enables efficient model adjustments without retraining, offering practical and cost-effective solutions. The framework also leverages state-of-the-art tools for automatic feature creation and hyperparameter optimisation, enhancing the predictive model's robustness and performance.

2.2 Quality 4.0
The concept of Q4.0 has been a subject of considerable debate and discussion amongst researchers and practitioners (Oliveira et al., 2024). Q4.0 incorporates advanced technologies to improve the quality of manufacturing and services, in the context of increasing digitisation of industries (Javaid et al., 2021). It combines quality management with I4.0 to improve organisational performance, innovation and new business models (Antony et al., 2021). The critical success factors (CSFs) associated with the implementation of Q4.0 are in line with I4.0. These include investing in technology, developing the right skills, providing adequate training and knowledge, addressing cybersecurity concerns, having management support, fostering a supportive organisational culture and effectively managing resistance to change (Antony et al., 2023). This is reflected in the eight identified key ingredients for an effective implementation of Q4.0: handling big data, improving prescriptive analytics, using Q4.0 for effective vertical, horizontal and end-to-end

integration, using Q4.0 for strategic advantage, leadership in Q4.0, training in Q4.0, organisational culture for Q4.0 and, top management support for Q4.0 (Sony *et al.*, 2020). These key ingredients play a crucial role in the ability of companies to embrace Q4.0 and are aligned with the five readiness factors for Q4.0 identified by Zulfiqar *et al.* (2023): top management commitment and support, leadership, organisational culture, employee competency and presence of an ISO Quality Management System (QMS) standard. Integration of I4.0 technologies and digitisation of quality management have a substantial effect on quality technology, processes and people (LNS Research, 2017). LNS Research proposes a framework with 11 axes for Q4.0 that outlines how it can enhance existing capabilities and initiatives whilst providing a perspective on traditional quality methods. Included in these 11 dimensions are *analytics* and *data*.

The evolution of quality management has progressed from inspection to total quality management (TQM), with tools aimed at enhancing industrial processes and services (Broday, 2022). This evolution has culminated in the emergence of Q4.0, which builds upon TQM by integrating Big Data and AI (LNS Research, 2017; Escobar *et al.*, 2021). Q4.0 represents a digital transformation strategy focussed on leveraging digital tools to consistently deliver high-quality products. These tools encompass AI, Big Data, Blockchain, Deep Learning, ML, Statistics and Data Science, alongside enabling technologies such as Internet of Things (IoT), Virtual and Augmented Reality, Data Streaming, Sensors and 5G (Radziwill, 2018). Considering the evolution of quality approaches, these tools are instrumental in shaping Q4.0 as a discovery approach:

(1) Inspection: Quality assurance was based on inspection, with the use of Walter A. Shewhart's statistical process control methods.

(2) Design: Integration of quality into operations to proactively prevent quality issues, based on W. Edwards Deming's suggestions.

(3) Empowerment: Use of TQM and Six Sigma, where quality is a shared responsibility and people are empowered to participate in continuous improvement.

(4) Discovery: In an adaptive environment, quality relies on the quick identification of new data sources, root cause analysis and discovery of new knowledge (Radziwill, 2018).

It is clear that Q4.0 involves a shift from traditional quality methods to a more data-driven approach (Grandinetti *et al.*, 2020; Carvalho *et al.*, 2021). This includes the use of data analytics and ML to identify patterns and trends in quality data, which can be used to improve processes and make better decisions. This also means that Q4.0 places greater emphasis on data collection and analysis, as well as the use of digital tools to manage and track quality metrics (Thekkoote, 2022).

Amongst the technological advancements driving industrial transformation, business analytics stands out as a pivotal enabler of I4.0, playing a crucial role in this process. It is rooted in theoretical concepts like absorptive capacity, dynamic capabilities and data-driven decision-making (Duan *et al.*, 2020). Almazmomi *et al.* (2021), underscore its significance in fostering a competitive advantage, particularly through nurturing a data-driven culture and enhancing product development within I4.0. Business analytics, by providing data intelligence and expert system components, is instrumental in facilitating the successful implementation of Q4.0 within the broader context of I4.0 (Silva *et al.*, 2021). It extracts meaningful insights from vast industrial data, contributing to digital market transformation (Duan *et al.*, 2021). It is also a key element for Q4.0 alongside data, connectivity and leadership (Thekkoote, 2022). The integration of advanced analytics and big data necessitates the development of an I4.0 analytics platform, transcending mere tools and technology

(Gröger, 2018). This aligns with the view that business analytics serves as a strategic resource for gaining a competitive edge in the industrial sector during the I4.0 era. Additionally, Ehret and Wirtz (2017) highlight how the Industrial Internet of Things (IIoT) drives new business models and services, emphasising analytics, including big data and AI as enablers of innovative information and analytical services. Additionally, Fernando *et al.* (2018) underscore practical big data analytics for predicting market preferences from diverse data sources, particularly in enhancing supply chain performance, an essential aspect of industrial transformation in the I4.0 era.

Q4.0 implementation enhances customer satisfaction, product quality, service quality and competitive advantage (Antony *et al.*, 2023). I4.0 objectives mirror those of the early to mid-1990s but with two shifts: data volume surge and accelerated achievement of quality goals through emerging technologies (Radziwill, 2020). Due to rising customer expectations in the I4.0 and Q4.0 landscape challenge, companies deliver high-quality products at competitive prices (Keller *et al.*, 2014). This underscores the relevance of understanding and managing quality costs.

Schiffauerova and Thomson (2006) analysed the different models that have been proposed to quantify the Cost of Quality (CoQ), each with its unique cost or activity categories. PAF model categorises costs into Prevention, Appraisal and Failure. Crosby's model splits quality costs into conformant and non-conformant. Opportunity or Intangible cost models extend the PAF model to include opportunity costs. The Process Cost model provides a systematic approach to identifying and analysing process-related costs. Activity-Based Costing (ABC) models categorise costs into value-added and non-value-added activities. I4.0's digital technologies and data analysis drive a transformation impacting quality costs and quality management. These technologies improve customer satisfaction and reduce quality costs (Saihi *et al.*, 2021). Antony *et al.* (2023), Sony, *et al.* (2020) and Zulfiqar *et al.* (2023) highlight technology investment and employee skills as key to Q4.0. Maganga and Taifa (2023), reinforce this, identifying investment in Big Data handling, enabling technologies and human resources skills as main enablers. These costs could be offset by reduced failure costs and increased customer satisfaction, leading to greater market share (Margarida Dias *et al.*, 2021). Tools like IoT, Cyber-Physical Systems (CPS), big data and AI contribute to predictive maintenance implementation, reducing costs and preventing failures (Lee *et al.*, 2019). Q4.0 provides a management framework based on increasing customer loyalty and decreasing costs (Javaid *et al.*, 2021).

Figure 1 represents the evolution of quality costs until the Q4.0 objectives were fully achieved (right column) (DeFeo, 2018). As shown, appraisal costs are minimised or eliminated, as well as internal and external failures. In general, Q4.0 can be seen as a holistic approach to quality that uses advanced technologies to improve efficiency, reduce costs and enhance customer satisfaction.

Advancements in quality management, though significant, bring new challenges. Digitisation of manufacturing processes alters communication, consumption patterns and value creation, influencing market strategies and product life cycles (Paritala *et al.*, 2017). Organisations must comprehend and adapt to these changes, analysing their impact on quality management (Corti *et al.*, 2021; Antony *et al.*, 2023). This ensures effective utilisation of Q4.0 to enhance customer satisfaction, enterprise efficiency and competitiveness (Liu and Gu, 2023). Q4.0, which is related to the digitalisation of quality work in the context of I4.0, is a relatively new and evolving area. Corti *et al.* (2021) and Ranjith Kumar *et al.* (2021) highlight this transition challenge and opportunities, with Corti providing a comprehensive framework for Q4.0 adoption and Kumar proposing a conceptual framework for quality in the digital transformation context. Margarida Dias *et al.* (2021) stress the need for a unified Q4.0 definition. Sureshchandar (2022) underscores the critical role of traditional quality elements such as leadership, customer focus and data-driven decision-making in the digital
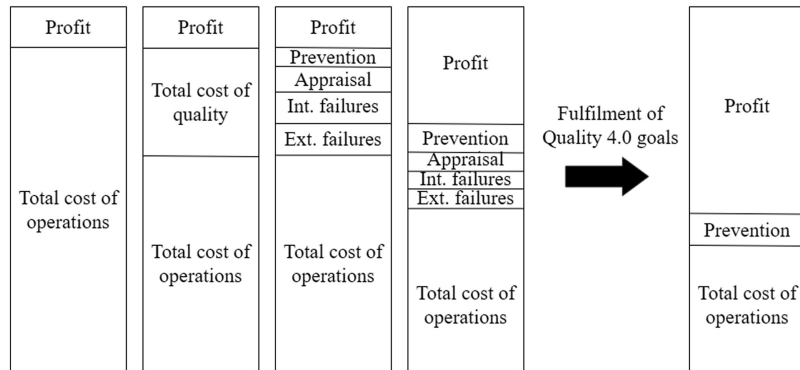
Source(s): Figure by authors adapted from Defeo (2018)

transformation journey, which aligns with the 11 axes of Q4.0 proposed by LNS Research (2017). These authors collectively suggest that whilst digitalisation presents new challenges, it also offers significant potential for enhancing quality management.

3. Materials and methods
This section details the materials and methods used in this study. The methodology, proposed framework, data preparation, data exploration, ML models used and evaluation metrics are described below.

3.1 Lot-release challenge
An automotive company is committed to minimising customer complaints, reducing costs and improving overall quality on its way to Q4.0. During different stages of its production line, different tests are performed. The products are packed in pallets that the company defines as a lot. The decision to release the lot is made in the last stage before shipping the products. This can have a great impact on customer satisfaction. Currently, the company uses a software application that manages this process by applying a set of rules that were defined heuristically. The lot is locked by default, and it is the software system that decides to release them by applying a set of rules that were manually defined by quality engineers based on their perception and knowledge. The actual rules are related to the following.

(1) Total repairs by part number, station and type of defect

(2) Faults detected in critical stations

(3) Part number blocked by quality team

After unlocking the lot, risk analysis is performed to decide whether it can be sent to customers.

Taking into account this manual system and given the vast amount of data that needs to be processed, an efficient method of handling it is crucial. Another factor to consider is the promotion of Q4.0 adoption in an effective manner. With these considerations in mind, the proposed framework incorporates an ML model to address these problems (Sarker, 2021). This will allow companies to improve their process, reduce costs and easily adjust their predictive models efficiently. Figure 2 reflects this by adding an ML module that collects information from tests and repairs at different stages to predict the probability of customer
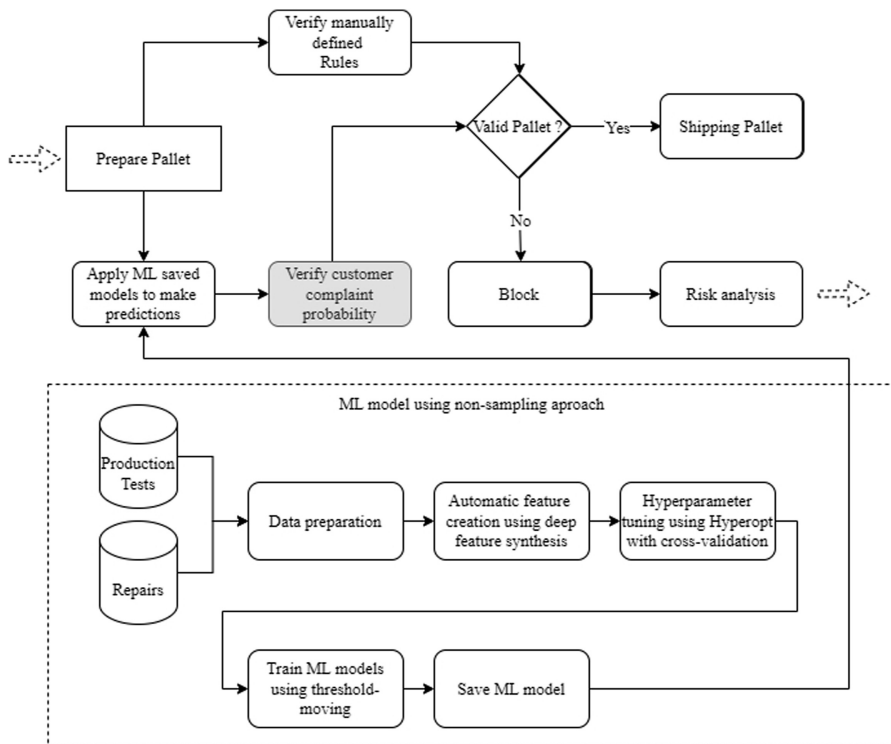
Figure 2.
Framework to improve
the lot-release decision
process

Source(s): Figure by authors

complaints before the lot is shipped to customers. In this phase, it is proposed to maintain existing manual rules to promote better decisions. The framework has been tested with real-world data and has shown promising results in terms of its ability to predict customer complaints and its efficiency. Based on the ML results and with continuous training, a more automated process can be evaluated that relies only on the ML model.

The development and evaluation of this ML module followed the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology (Wirth and Hipp, 2000). This is widely recognised as the standard for implementing data mining projects (Schröer *et al.*, 2021). This structured approach breaks down the life cycle of a data mining project into six phases:

(1) Business Understanding: Define project objectives and requirements from a business perspective and formulate a data mining problem definition.

(2) Data Understanding: Collect and familiarise with data, identifying quality issues and forming hypotheses for hidden information.

(3) Data Preparation: Construct the final dataset from raw data, including selection, transformation and cleaning tasks.

(4) Modelling: Select and apply modelling techniques, calibrating parameters for optimal performance. Certain techniques have specific data format requirements, which may require revisiting the data preparation phase.

(5) Evaluation: Assess model quality and effectiveness, ensuring alignment with business goals before deployment.

(6) Deployment: Structure and present knowledge for customer utilisation, varying in complexity from report generation to recurring data mining procedures.

Figure 3 illustrates these phases and the approach adopted in the studied company, where management and team commitment played a crucial role in navigating each phase.

3.2 Data exploration

Production tests and repair databases of an automotive company were used as data sources for this study, considering the period between January 2019 and April 2020. Millions of automatic tests are produced every day; to handle them, the company stores the data in a *Hadoop* cluster. The initial dataset has 2,076,529 records and 40 features, 31 related to production tests and 9 related to repairs. Table 1 lists some of them:
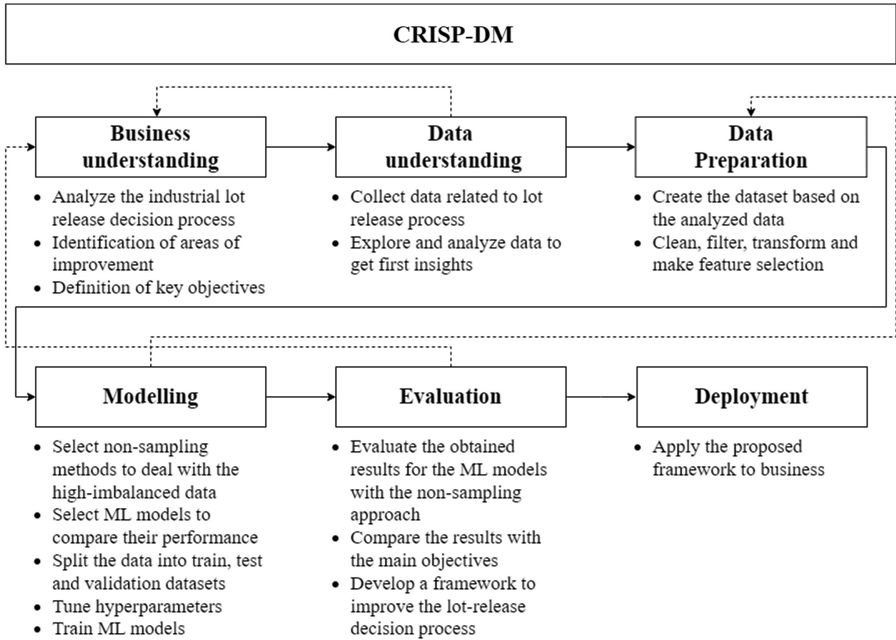

Figure 3.
CRISP-DM methodology used to implement the project

Source(s): Figure by authors

Field id	Description
product	Product ID number (type of product)
serial	Unique ID of a product
stationid	Unique station ID
gof status	Result of test sequence (GOF- "Good or Fail")
cause classid	Type of defect fixed in repair
cause_flawid	Flaw identification fixed in repair
Source(s): Table by authors	

Table 1.
Some features of interest in the original dataset

3.3 Data preparation
The following constraints were implemented to ensure a representative dataset that includes tests related to complaints and tests without complaints in the period analysed.

(1) Select the top 10 products with more complaints.

(2) Select a subset of 2 million tests without complaints.

To improve the performance of the ML models, feature engineering was carried out based on the insights gained from the data exploration. Throughout this process, some features were manually created, such as *hascomplaint*, classified as the target to determine whether a test is related to a complaint or not, *hasrepair*, which identifies whether a test has a repair or not and *srepair*, which identifies a specific type of repaired flaw. Most of the new features were created automatically using *Featuretools,* which implements the concept of deep feature synthesis (Kanter and Veeramachaneni, 2015).

In the data cleaning process, the rows with missing values were dropped. Spearman's rank correlation coefficient was used to evaluate correlation amongst features, excluding the target and dropping highly correlated ones to avoid multicollinearity.

After this operation, the final dataset comprises 1,552,324 records and 44 features, including 15 new aggregation and transformation features that were automatically created. The distribution of the two classes is as follows: 1.7% come from the class of tests related to customer complaints, and 98.3% come from the other class.

3.4 Evaluation metrics
Due to the need to deal with imbalanced data when predicting customer complaints based on repair and production tests, classification is assessed with a confusion matrix (He and Garcia, 2009). Tailored performance metrics are essential for effectively addressing imbalanced datasets. Precision, Recall, F1-Score and Matthews Correlation Coefficient (MCC) are recognised as suitable measures for evaluating model performance on such datasets (Chicco and Jurman, 2020; Bhadani *et al.*, 2023). These metrics are calculated according to the four categories of confusion metrics: True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN).

Precision is the fraction of relevant results.

$$Precision = \frac{True\,Positive}{(True\,Positive + False\,Positive)} \tag{1}$$

Recall is the fraction of positive labels correctly identified by the model.

$$Recall = \frac{True\,Positive}{(True\,Positive + False\,Negative)} \tag{2}$$

The F1-Score is the harmonic mean of precision and recall. It provides the model's ability in identifying both classes.

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{3}$$

The MCC assesses the correlation between predicted and observed binary classes in classification, yielding a high score when both positive and negative predictions are accurate.

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)}} \tag{4}$$

3.5 Machine learning algorithms selected

Given the problem's imbalanced nature, three gradient-boosting (XGB, LGBM, CatB) and a bagging (RF) algorithm were studied. These algorithms are effective in handling imbalanced datasets (Shumaly *et al.*, 2020). Specifically, XGB, LGBM and CatB are efficient, accurate and have a large set of hyperparameters that can be tuned (Bentéjac *et al.*, 2021). All selected algorithms have shown good results using a non-sampling approach to deal with imbalanced datasets (Johnson and Khoshgoftaar, 2022).

3.5.1 Gradient boosting. Boosting is an ensemble learning technique that consists of sequentially applying a number of weak learners (models that perform marginally better than random guessing) to create a strong learner (Sagi and Rokach, 2018). A boosting algorithm assigns varying weights to the output of its estimator and optimises a loss function. In gradient boosting, which combines the gradient descent algorithm and the boosting method, predictors are created consecutively rather than independently, where each tree corrects the error caused by the prior tree (Daoud, 2019).

XGB, LGBM and CatB are all advanced gradient-boosting algorithm implementations. XGB employs innovative regularisation techniques for controlling overfitting, enhancing scalability and speed whilst conserving computational resources (Chen and Guestrin, 2016). LGBM introduces *Gradient-based One Side Sampling* and *Exclusive Feature Bundling* techniques, enhancing efficiency and suitability for large-scale data processing (Ke *et al.*, 2017). CatB introduces an innovative approach for handling categorical features and implementation of ordered boosting, improving its effectiveness and performance (Prokhorenkova *et al.*, 2018).

3.5.2 Bagging. Bagging or bootstrap aggregation is an ensemble learning method. In bagging, some samples from the original dataset are randomly generated by replacement, which means that each row can be selected more than once (Breiman, 2001). After multiple data samples are generated, the outcomes generated by these weak models, which are trained independently, are evaluated, and based on the type of task (regression or classification), the average or majority of those predictions result in a more accurate estimate. RF is an extended implementation of bagging. Unlike bagging, where all features are taken into account when splitting a node, in RF, only a subset of the total features is randomly chosen, and the best split feature of the subset is used to split each node in a tree. It is robust against noisy data and outliers, making it less prone to overfitting (Breiman, 2001).

3.6 Cost-sensitive learning and thresholding

In addressing skewed class distributions, two primary strategies are employed: data-level and algorithm-level methods. The latter adjusts the learning algorithm to better manage imbalanced data, incorporating techniques like cost-sensitive learning and thresholding (Haixiang *et al.*, 2017).

Cost-sensitive learning employs a cost matrix where the penalties for misclassification are strategically set higher than those for correct classification. This can be operationalised by adjusting the "class weight" parameter in the selected algorithms to reflect the varying importance of each class. Thresholding involves establishing a decision boundary to predict class membership. For imbalanced datasets, the default threshold often proves inadequate, potentially skewing results. To counter this, the threshold is fine-tuned during training to enhance model performance. To perform this threshold analysis, the Precision-Recall Curve, which illustrates the trade-off between precision and recall across different thresholds, was considered (Davis and Goadrich, 2006; Lobo *et al.*, 2023).

4. Experiments

To conduct the study, four ML models were conceived, tuned and evaluated. The dataset was randomly split using the common ratio of 80% for training, 10% for validation and 10% for testing (Alazba *et al.*, 2023). Finding optimal hyperparameters is crucial for ML algorithm performance, but manual search can be time-consuming. Whilst Grid Search and Random Search are common methods, Bayesian optimisation with the Hyperopt library is more efficient and effective in terms of both accuracy and time (Putatunda and Rama, 2018). Each algorithm has specific hyperparameters (Chen and Guestrin, 2016; Ke *et al.*, 2017; Pedregosa *et al.*, 2011; Prokhorenkova *et al.*, 2018). To improve the F1-Score and control overfitting the following parameters were considered.

(1) "n_estimators": Specifies the number of trees or boosting stages. More trees can improve learning ability but may lead to overfitting.

(2) "max_depth": Controls overfitting by specifying the maximum depth of a tree.

(3) "learning_rate": Controls the weighting of new trees. Lower rates may improve performance but extend training time.

(4) "colsample_bytree": Controls column subsampling to prevent overfitting.

(5) "min_samples_leaf": Specifies the minimum number of samples required at a leaf node, affecting noise capture.

(6) "min_samples_split": Defines the minimum number of samples required for node splitting to control overfitting.

To tune and find the best hyperparameters for these models, cross-validation was used together with Hyperopt. Table 2 depicts the hyperparameter searching space.

This study was implemented using *Python*, version 3.7, including some libraries such as *Pandas*, *Numpy*, *MatPlotlib*, *Hyperopt* and *Featuretools*. All ML models were conceived using *Scikit-Learn*.

5. Results and discussion

Based on the best hyperparameters found for each model, which are shown in Table 3, four ML models were analysed.

Cost-sensitive learning and thresholding methods were used to deal with an imbalanced dataset of production and repair tests to predict customer complaints. The results obtained for each model based on this approach are presented in Table 4. To evaluate the results, the F1-Score and MCC were the main metrics considered.

Analysing the results, CatB had the best Recall at 66.9% and FN count, whilst XGB excelled in other metrics, including a 72.4% F1-Score and 75.0% MCC. These results were

Hyerparameter	RF	XGB	LGBM	CatB
n_estimators	{100,200,300}	{70,80,90,100}	{70,80,90,100}	{70,80,90,100}
max_depth	{10,20,30}	{3,4,5}	{3,4,5}	{3,4,5}
learning_rate	–	[0.1,0.3]	[0.1,0.3]	[0.1,0.3]
colsample_bytree	–	{0.5,0.6,0.7,0.8,,0.9,1}	{0.5,0.6,0.7,0.8,,0.9,1}	–
min_samples_leaf	{1,2,4}	–	–	–
in_samples_split	{2,4,6}	–	–	–

Note(s): RF: Random Forest; XGB: XGBoosst; LGBM: LightGBM; CatB: CatBoost
Source(s): Table by authors

Table 2.
Hyperparameter
searching space

achieved through cost-sensitive learning, adjusting the class weight parameter and threshold moving. The Precision-Recall curve was used to find the optimal threshold for the best F1-Score (Brownlee, 2020). Optimising this value results in the ideal balance between precision and recall. Figure 4 illustrates this process for the XGB model, showing the trade-off between increasing Precision and decreasing Recall. The F1-Score identifies the threshold that achieves the best balance between precision and recall.

Despite limited research on predicting customer complaints from production and repair data for industrial processes, this study leverages insights from related fields. Prior studies underscored the role of ML and AI in the automotive industry (Fernández-López et al., 2022), the significance of ML in predicting and analysing customer complaints (Alarifi et al., 2023) and improving quality prediction (Jung et al., 2021). The obtained results of CatB and XGB

Hyerparameter	RF	XGB	LGBM	CatB
n_estimators	100	100	100	90
max_depth	30	5	3	5
learning_rate	–	0.2200	0.1168	0.1298
colsample_bytree	–	0.5	0.7	–
min_samples_leaf	1	–	–	–
min_samples_split	2	–	–	–

Table 3.
Best hyperparameters found using hyperopt

Note(s): RF Random Forest; XGB: XGBoost; LGBM: LightGBM; CatB: CatBoost
Source(s): Table by authors

Model	Precision	Recall	F1-score	MCC	FP	FN
CatB	66.7%	67.1%	66.9%	66.6%	545	536
XGB	99.5%	56.9%	72.4%	75.0%	5	703
LGBM	45.9%	46.2%	46.0%	45.5%	888	877
RF	94.1%	41.2%	57.3%	62.1%	42	958

Table 4.
Obtained results for conceived models

Note(s): RF Random Forest; XGB: XGBoost; LGBM: LightGBM; CatB: CatBoost
MCC- Matthews Correlation Coefficient; FP: False Positives; FN: False Negatives
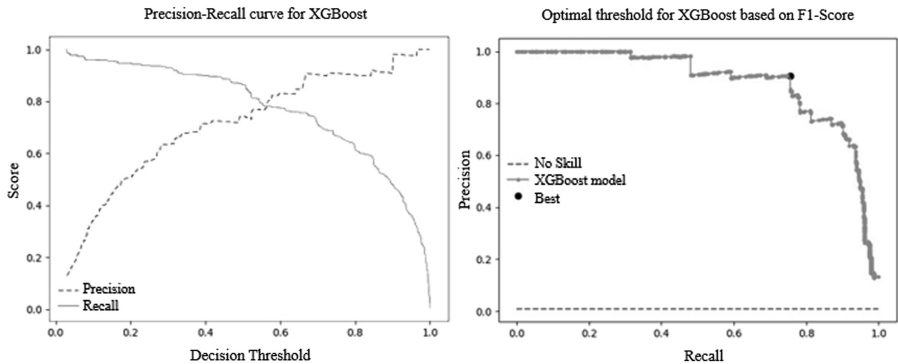**Source(s):** Table by authors

Figure 4.
Optimal threshold determination using Precision-Recall Curve with F1-Score

Source(s): Figure by authors

align with previous research, demonstrating their efficacy in tackling similar problems (Chen and Lin, 2020; Yorulmuş *et al.*, 2022). Furthermore, we found that LGBM and RF yielded lower F1-Score and MCC values compared to CatB and XGB. Our findings also demonstrate the potential of cost-sensitive learning and thresholding methods to enhance the performance of these models (Petrides and Verbeke, 2022; Coussement, 2014).

The results show that the system can predict a significant number of potential complaints based on production and repair tests. The F1-Score and MCC values indicate a balanced precision and recall, reflecting good-quality classification. In the automotive industry, recall values help reduce 0 km defects, preventing the shipment of faulty products to customers. High precision minimises false positives, reducing rework and waste and enhancing production process efficiency and profitability. The F1-Score reflects the balance between recall and precision. Furthermore, this approach that uses threshold moving allows the company to adjust its threshold according to specific needs, prioritising Precision to reduce FP or Recall to reduce FN.

These results show that incorporating an ML model into the lot-release process, as suggested by the proposed framework, can reduce customer complaints, related quality costs and enhance efficiency.

6. Conclusion

The transition to Q4.0 requires the integration of IoT, ML and Big Data to improve efficiency, productivity and innovation. Customer complaints can significantly impact automotive companies. Proactive management and anticipation of issues before shipping can boost customer satisfaction, cost savings and increase profitability.

To address this, our study focusses on improving the lot-release decision process by predicting customer complaints for an automotive company transitioning to Q4.0. We propose a framework with an ML model that gathers production line data to predict complaints. Four ML models were developed, tuned, evaluated and compared. To deal with imbalanced datasets, cost-sensitive learning and threshold-moving approaches were considered. The F1-Score and MCC were used as the main metrics to evaluate the results.

The results highlight ML's potential in refining the lot-release decision process, predicting customer complaints and reducing quality costs. XGB outperformed other algorithms with an F1-Score of 72.4% and an MCC of 75%, balancing recall and precision effectively. Compared to the heuristic rule-based system, the framework detected a significant number of tests related to customer complaints, showcasing ML's benefits. The proposed framework, using a non-sampling, threshold-moving approach, allows the prioritisation of false negatives or false positives reduction according to company-specific needs. This enables efficient model adjustment by modifying the threshold, avoiding retraining. It improves quality performance, reduces false negatives or positives and increases operational efficiency by saving time, resources and costs. The study reaffirms the importance of big data handling, improved prescriptive analytics, a change-receptive culture and top management commitment to Q4.0. We believe that our study contributes to the field of Q4.0 by underscoring the potential of ML in transforming industry practices and enhancing quality control in the era of digital transformation.

From a theoretical perspective, this study enriches the understanding of how ML can be applied to improve quality control in the context of Q4.0. It also provides insights into the importance of model evaluation metrics in the context of predicting customer complaints. Additionally, it contributes to the theoretical knowledge of handling imbalanced datasets using a non-sampling approach.

From a managerial perspective, these findings can guide strategic ML and big data investment decisions. They also stress fostering a change-receptive culture, securing top

management's Q4.0 commitment and highlighting the need for resource allocation and employee upskilling.

Leveraging a data-driven approach, a key aspect of Q4.0, the proposed framework holds several implications for the automotive industry. Contribute to reducing 0 km defects and customer complaints by preemptively identifying quality issues and enabling proactive quality control measures. This proactive approach may lead to decreased warranty and recall costs. Ultimately, it can contribute to cost savings, improved customer satisfaction and enhanced brand reputation, helping the Q4.0 transition.

Whilst the proposed framework has demonstrated promising results with real-world data, future research should replicate these findings in larger datasets for better generalisation. Exploration of alternative ML algorithms, including deep learning models, and an expanded hyperparameter search space is warranted. Additionally, resampling methods should be considered for addressing imbalanced datasets. Subsequent evaluation should comprehensively assess the framework's performance, potentially leading to the exclusive adoption of the ML approach and elimination of manual rules. Further research into the economic implications, contingent upon financial data availability, could elucidate potential cost savings and efficiency gains. The absence of financial data at this phase constrained our ability to evaluate this aspect.

Overall, this study contributes to the understanding of how advanced technologies, such as ML, can be used to improve lot-release decision processes and enhance Q4.0 adoption. The proposed framework incorporates an ML model with a non-sampling approach into the lot-release decision process and demonstrates its effectiveness in predicting customer complaints and reducing quality costs. The insights gained from this study are valuable for companies seeking to evolve towards Q4.0.

References

Abdelrahman, O. and Keikhosrokiani, P. (2020), "Assembly line anomaly detection and root cause analysis using machine learning", *IEEE Access*, Vol. 8, pp. 189661-189672, doi: 10.1109/ACCESS.2020.3029826.

Alarifi, G., Farjana Rahman, M., Mohammad, H. and Shamim Hossain, M. (2023), "Prediction and analysis of customer complaints using machine learning techniques", *International Journal of E-Business Research*, Vol. 19 No. 1, pp. 1-25, doi: 10.4018/IJEBR.319716.

Alazba, A., Aljamaan, H., Alshayeb, M. and Alshayeb, M. (2023), "Deep learning approaches for bad smell detection: a systematic literature review", *Empirical Software Engineering*, Vol. 28 No. 3, p. 77, doi: 10.1007/s10664-023-10312-z.

Almazmomi, N., Ilmudeen, A. and Qaffas, A.A. (2021), "The impact of business analytics capability on data-driven culture and exploration: achieving a competitive advantage", *Benchmarking: An International Journal*, Vol. 29 No. 4, pp. 1264-1283, doi: 10.1108/BIJ-01-2021-0021.

Antony, J., McDermott, O. and Sony, M. (2021), "Quality 4.0 conceptualisation and theoretical understanding: a global exploratory qualitative study", *TQM Journal*, Vol. 0 No. 5, pp. 1169-1188, doi: 10.1108/TQM-07-2021-0215.

Antony, J., McDermott, O., Sony, M., Toner, A., Bhat, S., Cudney, E.A. and Doulatabadi, M. (2023), "Benefits, challenges, critical success factors and motivations of Quality 4.0 – a qualitative global study", *Total Quality Management and Business Excellence*, Vol. 34 No. 7-8, doi: 10.1080/14783363.2022.2113737.

Bentéjac, C., Csörgő, A. and Martínez-Muñoz, G. (2021), "A comparative analysis of gradient boosting algorithms", *Artificial Intelligence Review*, Vol. 54 No. 3, pp. 1937-1967, Springer Science and Business Media B.V., doi: 10.1007/S10462-020-09896-5/TABLES/12.

Bhadani, R., Chen, Z. and An, L. (2023), "Attention-based graph neural network for label propagation in single-cell omics", *Genes*, Vol. 14 No. 2, p. 506, doi: 10.3390/genes14020506.

Breiman, L. and DeJong, G. (2001), "Random forests", *Machine Learning*, Vol. 45 No. October 2001, pp. 45-76, doi: 10.1023/A:1010933404324.

Broday, E.E. (2022), "The evolution of quality: from inspection to quality 4.0", *International Journal of Quality and Service Sciences*, Vol. 14 No. 3, pp. 368-382, doi: 10.1108/IJQSS-09-2021-0121.

Brownlee, J. (2020), *Probability for Machine Learning: Discover How to Harness Uncertainty with Python*, Machine Learning Mastery, available at: https://books.google.pt/books?id=uU2xDwAAQBAJ

Carvalho, A.V., Enrique, D.V., Chouchene, A. and Charrua-Santos, F. (2021), "Quality 4.0: an overview", *Procedia Computer Science*, Vol. 181, pp. 341-346, doi: 10.1016/J.PROCS.2021.01.176.

Chen, T. and Guestrin, C. (2016), "XGBoost: a scalable tree boosting system", *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM*, New York, NY, pp. 785-794, doi: 10.1145/2939672.2939785.

Chen, S.H. and Lin, W.H. (2020), "Determination of an optimal pipeline for imbalanced classification: predicting potential customer complaints to a textile manufacturer", *International Journal of Industrial Engineering : Theory Applications and Practice*, Vol. 27 No. 5, pp. 810-823.

Chicco, D. and Jurman, G. (2020), "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation", *BMC Genomics*, Vol. 21 No. 1, pp. 1-13, doi: 10.1186/S12864-019-6413-7/TABLES/5.

Cho, E., Chang, T.W. and Hwang, G. (2022), "Data preprocessing combination to improve the performance of quality classification in the manufacturing process", *Electronics*, Vol. 11 No. 3, p. 477, doi: 10.3390/ELECTRONICS11030477.

Corti, D., Masiero, S. and Gladysz, B. (2021), "Impact of industry 4.0 on quality management: identification of main challenges towards a quality 4.0 approach", *2021 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, pp. 1-8, doi: 10.1109/ICE/ITMC52061.2021.9570206.

Coussement, K. (2014), "Improving customer retention management through cost-sensitive learning", *European Journal of Marketing*, Vol. 48 Nos 3/4, pp. 477-495, doi: 10.1108/ejm-03-2012-0180.

Daoud, E.A. (2019), "Comparison between XGBoost, LightGBM and CatBoost using a home credit dataset", *International Journal of Computer and Information Engineering*, Vol. 13 No. 1, pp. 6-10.

Davis, J. and Goadrich, M. (2006), "The relationship between Precision-Recall and ROC curves", *Proceedings of the 23rd International Conference on Machine Learning - ICML '06*, Vol. 148, ACM Press, New York, NY, pp. 233-240, doi: 10.1145/1143844.1143874.

DeFeo, J.A. (2018), *The Smart Factory, Industry 4. 0 and Quality*, Juran Institute, available at: https://www.youtube.com/watch?v=z4-R4YZ_Ao8&t=4s&ab_channel=Juran

Dror, S. (2022), "QFD for selecting key success factors in the implementation of quality 4.0", *Quality and Reliability Engineering International*, Vol. 38 No. 6, pp. 3216-3232, doi: 10.1002/QRE.3138.

Duan, Y., Cao, G. and Edwards, J.S. (2020), "Understanding the impact of business analytics on innovation", *European Journal of Operational Research*, Vol. 281 No. 3, pp. 673-686, doi: 10.1016/j.ejor.2018.06.021.

Duan, L., Li and Xu, D. (2021), "Data analytics in industry 4.0: a survey", *Information Systems Frontiers*. doi: 10.1007/s10796-021-10190-0.

Ehret, M. and Wirtz, J. (2017), "Unlocking value from machines: business models and the industrial internet of things", *Journal of Marketing Management*, Vol. 33 Nos 1-2, pp. 111-130, doi: 10.1080/0267257X.2016.1248041.

Escobar, C.A., McGovern, M.E. and Morales-Menendez, R. (2021), "Quality 4.0: a review of big data challenges in manufacturing", *Journal of Intelligent Manufacturing*, Vol. 32 No. 8, pp. 2319-2334, doi: 10.1007/s10845-021-01765-4.

Fathy, Y., Jaber, M. and Brintrup, A. (2021), "Learning with imbalanced data in smart manufacturing: a comparative analysis", *IEEE Access*, Vol. 9, pp. 2734-2757, doi: 10.1109/ACCESS.2020.3047838.

Fernández-López, A., Fernández-Castro, B. and García-Coego, D. (2022), *ML \& AI Application for the Automotive Industry*, Springer, Cham, pp. 79-102, doi: 10.1007/978-3-030-91006-8_4.

Fernando, Y., Chidambaram, R.R. and Sari Wahyuni-Td, I. (2018), "The impact of Big Data analytics and data security practices on service supply chain performance", *Benchmarking: An International Journal*, Vol. 25 No. 9, pp. 4009-4034, doi: 10.1108/BIJ-07-2017-0194.

Grandinetti, R., Ciasullo, M.V., Paiola, M. and Schiavone, F. (2020), "Fourth industrial revolution, digital servitization and relationship quality in Italian B2B manufacturing firms. An exploratory study", *The TQM Journal*, Vol. 32 No. 4, pp. 1754-2731, doi: 10.1108/TQM-01-2020-0006.

Gröger, C. (2018), "Building an industry 4.0 analytics platform practical challenges, approaches and future research directions", Vol. 18 No. 1, pp. 5-14, doi: 10.1007/s13222-018-0273-1.

Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H. and Bing, G. (2017), "Learning from class-imbalanced data: review of methods and applications", *Expert Systems with Applications*, Vol. 73, pp. 220-239, doi: 10.1016/j.eswa.2016.12.035.

He, H. and Garcia, E.A. (2009), "Learning from imbalanced data", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 21 No. 9, pp. 1263-1284, doi: 10.1109/TKDE.2008.239.

Javaid, M., Haleem, A., Pratap Singh, R. and Suman, R. (2021), "Significance of Quality 4.0 towards comprehensive enhancement in manufacturing sector", *Sensors International*, Vol. 2, 100109, doi: 10.1016/J.SINTL.2021.100109.

Johnson, J.M. and Khoshgoftaar, T.M. (2022), "Cost-sensitive ensemble learning for highly imbalanced classification", *Proceedings - 21st IEEE International Conference on Machine Learning and Applications, ICMLA 2022*, pp. 1427-1434, doi: 10.1109/ICMLA55696.2022.00225.

Jung, H., Jeon, J., Choi, D. and Park, A.J.Y. (2021), "Application of machine learning techniques in injection molding quality prediction: implications on sustainable manufacturing industry", *Sustainability 2021*, Vol. 13 No. 8, p. 4120, doi: 10.3390/SU13084120.

Kanter, J.M. and Veeramachaneni, K. (2015), "Deep feature synthesis: towards automating data science endeavors", *Proceedings of the 2015 IEEE International Conference on Data Science and Advanced Analytics, DSAA 2015*, IEEE. doi: 10.1109/DSAA.2015.7344858.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T. (2017), "LightGBM: a highly efficient gradient boosting decision tree", *Neural Information Processing Systems*.

Keller, M., Rosenberg, M., Brettel, M. and Friederichsen, N. (2014), "How virtualization, decentrazliation and network building change the manufacturing landscape: an industry 4.0 perspective", *International Journal of Mechanical, Aerospace, Industrial, Mechatronic and Manufacturing Engineering*, Vol. 8 No. 1, pp. 37-44.

Lee, S.M., Lee, D. and Kim, Y.S. (2019), "The quality management ecosystem for predictive maintenance in the Industry 4.0 era", *International Journal of Quality Innovation*, Vol. 5 No. 1, p. 4, doi: 10.1186/s40887-019-0029-5.

Liu, H.-C., Gu, X. and Yang, M. (2023), "From total quality management to Quality 4.0: a systematic literature review and future research agenda", *Frontiers of Engineering Management*, Vol. 10 No. 2, pp. 191-205, doi: 10.1007/s42524-022-0243-z.

LNS Research (2017), *Quality 4.0 Impact and Strategy Handbook*, LNS Research, available at: https://blog.lnsresearch.com/quality40ebook#sthash.QgBusXBV.DeI11dFc.dpbs

Lobo, A., Oliveira, P., Sampaio, P. and Novais, P. (2023), *Cost-Sensitive Learning and Threshold-Moving Approach to Improve Industrial Lots Release Process on Imbalanced Datasets*, Springer, Cham, pp. 280-290, doi: 10.1007/978-3-031-20859-1_28.

Maganga, D.P. and Taifa, I.W.R. (2023), "Quality 4.0 conceptualisation: an emerging quality management concept for manufacturing industries", *TQM Journal*, Vol. 35 No. 2, pp. 389-413, doi: 10.1108/TQM-11-2021-0328.

Margarida Dias, A., Carvalho, A.M. and Sampaio, P. (2021), "Quality 4.0: literature review analysis, definition and impacts of the digital transformation process on quality", *International Journal of Quality \& Reliability Managemen*, Vol. 39 No. 6, pp. 1312-1335, doi: 10.1108/IJQRM-07-2021-0247.

Oliveira, D., Alvelos, H. and Rosa, M.J. (2024), "Quality 4.0: results from a systematic literature review", *The TQM Journal*, Advance online publication, doi: 10.1108/TQM-01-2023-0018.

Paritala, P.K., Manchikatla, S. and Yarlagadda, P.K.D.V. (2017), "Digital manufacturing- applications past, current, and future trends", *Procedia Engineering*, Vol. 174, pp. 982-991, The Author(s), doi: 10.1016/j.proeng.2017.01.250.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E. (2011), "Scikit-learn: machine learning in Python", *Journal of Machine Learning Research*, Vol. 12, pp. 2825-2830.

Petrides, G. and Verbeke, W. (2022), "Cost-sensitive ensemble learning: a unifying framework", *Data Mining and Knowledge Discovery*, Vol. 36 No. 1, pp. 1-28, doi: 10.1007/S10618-021-00790-4/TABLES/8.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V. and Gulin, A. (2018), "Catboost: unbiased boosting with categorical features", *Advances in Neural Information Processing Systems*, Vol. 2018 Section 4, pp. 6638-6648, Decem.

Putatunda, S. and Rama, K. (2018), "A comparative analysis of hyperopt as against other approaches for hyper-parameter optimization of XGBoost", *Proceedings of the 2018 International Conference on Signal Processing and Machine Learning - SPML '18*, ACM Press, New York, NY, pp. 6-10, doi: 10.1145/3297067.3297080.

Radziwill, N.M. (2018), "Quality 4.0: let's Get Digital - the many ways the fourth industrial revolution is reshaping the way we think about quality", October.

Radziwill, N.M. (2020), *Connected, Intelligent, Automated - the Definitive Guide to Digital Transformation and Quality 4.0*, Quality Press.

Ranjith Kumar, R., Ganesh, L. and Rajendran, C. (2021), "Quality 4.0-a review of and framework for quality management in the digital era", *International Journal of Quality \& Reliability Managemen*, Vol. 39 No. 6, pp. 1385-1411, doi: 10.1108/IJQRM-05-2021-0150.

Rojko, A. (2017), "Industry 4.0 concept: background and overview", *International Journal of Interactive Mobile Technologies (IJIM)*, Vol. 11 No. 5, p. 77, Kassel University Press, doi: 10.3991/ijim.v11i5.7072.

Sagi, O. and Rokach, L. (2018), "Ensemble learning: a survey", *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Vol. 8 No. 4, p. e1249, doi: 10.1002/WIDM.1249.

Saihi, A., Awad, M. and Ben-Daya, M. (2021), "Quality 4.0: leveraging Industry 4.0 technologies to improve quality management practices-a systematic review", *International Journal of Quality \& Reliability Management*, Vol. 40 No. 2, pp. 628-650, doi: 10.1108/IJQRM-09-2021-0305.

Sankhye, S. and Hu, G. (2020), "Machine learning methods for quality prediction in production", *Logistics*, Vol. 4 No. 4, p. 35, doi: 10.3390/logistics4040035.

Sarker, I.H. (2021), "Machine learning: algorithms, real-world applications and research directions", *SN Computer Science*, Vol. 2 No. 3, pp. 1-21, doi: 10.1007/s42979-021-00592-x.

Schiffauerova, A. and Thomson, V. (2006), "A review of research on cost of quality models and best practices", *International Journal of Quality & Reliability Management*, Vol. 23 No. 6, pp. 647-669, doi: 10.1108/02656710610672470.

Schröer, C., Kruse, F. and Gómez, J.M. (2021), "A systematic literature review on applying CRISP-DM process model", *Procedia Computer Science*, Vol. 181, pp. 526-534, doi: 10.1016/J.PROCS.2021.01.199.

Shumaly, S., Neysaryan, P. and Guo, Y. (2020), "Handling class imbalance in customer churn prediction in telecom sector using sampling techniques, bagging and boosting trees", *2020 10h International Conference on Computer and Knowledge Engineering, ICCKE 2020*, pp. 82-87, doi: 10.1109/ICCKE50421.2020.9303698.

Silva, A.J., Cortez, P., Pereira, C. and Pilastri, A. (2021), "Business analytics in industry 4.0: a systematic review", *Expert Systems*, Vol. 38 No. 7, pp. 517-539, doi: 10.1111/exsy.12741.

Sony, M., Antony, J. and Douglas, J.A. (2020), "Essential ingredients for the implementation of Quality 4.0 A narrative review of literature and future directions for research", *The TQM Journal*, Vol. 32 No. 4, pp. 779-793, doi: 10.1108/TQM-12-2019-0275.

Sureshchandar, G.S. (2022), "Quality 4.0-understanding the criticality of the dimensions using the analytic hierarchy process (AHP) technique", *International Journal of Quality \& Reliability Management*, Vol. 39 No. 6, pp. 1336-1367, doi: 10.1108/IJQRM-06-2021-0159.

Thekkoote, R. (2022), "Enabler toward successful implementation of Quality 4.0 in digital transformation era: a comprehensive review and future research agenda", *International Journal of Quality \& Reliability Management*, Vol. 39 No. 6, pp. 1368-1384, doi: 10.1108/IJQRM-07-2021-0206.

Villanueva Zacarias, A.G., Reimann, P. and Mitschang, B. (2018), "A framework to guide the selection and configuration of machine-learning-based data analytics solutions in manufacturing", *Procedia CIRP*, Vol. 72, pp. 153-158, doi: 10.1016/J.PROCIR.2018.03.215.

Wirth, R. and Hipp, J. (2000), "CRISP-DM: towards a standard process model for data mining".

Yorulmuş, M.H., Bolat, H.B. and Bahadır, Ç. (2022), "Predictive quality defect detection using machine learning algorithms: a case study from automobile industry", *Lecture Notes in Networks and Systems*, Vol. 308 August 2022, pp. 263-270, doi: 10.1007/978-3-030-85577-2_31.

Zulfiqar, M., Antony, J., Swarnakar, V., Jayaraman, R. and McDermott, O. (2023), "A readiness assessment of Quality 4.0 in packaging companies: an empirical investigation", *Total Quality Management and Business Excellence*, Vol. 34 Nos 11-12, pp. 1334-1352, doi: 10.1080/14783363.2023.2170223.

About the authors

Armindo Lobo is Data Scientist at DTx - Digital Transformation Colab and Ph.D. Candidate specialising in AI and Quality 4.0. Armindo holds a degree in Systems Engineering from the University of Minho and completed post-graduate studies at Universidad Complutense de Madrid. He has extensive experience in the tech industry, having assumed various leadership and technical roles at Primavera BSS. He was Chief Executive Officer (CEO) and Co-founder of Primecog, as well as Chairman of the General Meeting of Primavera BSS. Armindo Lobo is the corresponding author and can be contacted at: lobo.armindo@gmail.com

Paulo Sampaio is Associate Professor with Habilitation at the School of Engineering of the University of Minho, Integrated Researcher of the ALGORITMI Research Centre/LASI and Coordinator of the Research Group on Quality and Organizational Excellence. His research topics are related to quality and organizational excellence.

Paulo Novais is Full Professor of Computer Science at the Department of Informatics, the School of Engineering, the University of Minho (Portugal) and Researcher at the ALGORITMI Centre. He is Coordinator of the Portuguese Intelligent Systems Associate Laboratory (LASI). His main research objective is to make systems a little more smart, reliable and sensitive to human presence and interaction.