Predictive quality model for customer defects

Anabela Costa Silva
Department of Production and Systems, ALGORITMI, LASI, Universidade do Minho, Braga, Portugal
José Machado
Informatics Department, ALGORITMI, LASI, Universidade do Minho, Braga, Portugal, and
Paulo Sampaio
Department of Production and Systems, ALGORITMI, LASI, Universidade do Minho, Braga, Portugal

Received 19 September 2023
Revised 24 January 2024
9 March 2024
Accepted 29 March 2024

Abstract

Purpose – In the context of the journey toward digital transformation and the realization of a fully connected factory, concepts such as data science, artificial intelligence (AI), machine learning (ML) and even predictive models emerge as indispensable pillars. Given the relevance of these topics, the present study focused on the analysis of customer complaint data, employing ML techniques to anticipate complaint accountability. The primary objective was to enhance data accessibility, harnessing the potential of ML models to optimize the complaint handling process and thereby positively contribute to data-driven decision-making. This approach aimed not only to reduce the number of units to be analyzed and customer response time but also to underscore the pressing need for a paradigm shift in quality management. The application of AI techniques sought to enhance not only the efficiency of the complaint handling process and data accessibility but also to demonstrate how the integration of these innovative approaches could profoundly transform the way quality is conceived and managed within organizations.

Design/methodology/approach – To conduct this study, real customer complaint data from an automotive company was utilized. Our main objective was to highlight the importance of artificial intelligence (AI) techniques in the context of quality. To achieve this, we adopted a methodology consisting of 10 distinct phases: business analysis and understanding; project plan definition; sample definition; data exploration; data processing and pre-processing; feature selection; acquisition of predictive models; evaluation of the models; presentation of the results; and implementation. This methodology was adapted from data mining methodologies referenced in the literature, taking into account the specific reality of the company under study. This ensured that the obtained results were applicable and replicable across different fields, thereby strengthening the relevance and generalizability of our research findings.

Findings – The achieved results not only demonstrated the ability of ML models to predict complaint accountability with an accuracy of 64%, but also underscored the significance of the adopted approach within the context of Quality 4.0 (Q4.0). This study served as a proof of concept in complaint analysis, enabling process automation and the development of a guide applicable across various areas of the company. The successful integration of AI techniques and Q4.0 principles highlighted the pressing need to apply concepts of digitization and artificial intelligence in quality management. Furthermore, it emphasized the critical importance of data, its organization, analysis and availability in driving digital transformation and enhancing operational efficiency across all company domains. In summary, this work not only showcased the advancements achieved through ML application but also emphasized the pivotal role of data and digitization in the ongoing evolution of Quality 4.0.

Originality/value – This study presents a significant contribution by exploring complaint data within the organization, an area lacking investigation in real-world contexts, particularly focusing on practical applications. The development of standardized processes for data handling and the application of predictions for classification models not only demonstrated the viability of this approach but also provided a valuable proof of concept for the company. Most importantly, this work was designed to be replicable in other areas of

The TQM Journal
Vol. 36 No. 9, 2024
pp. 155-174
Emerald Publishing Limited
1754-2731
DOI 10.1108/TQM-09-2023-0302

the factory, serving as a fundamental basis for the company's data scientists. Until then, limited data access and lack of automation in its treatment and analysis represented significant challenges. In the context of Quality 4.0, this study highlights not only the immediate advantages for decision-making and predicting complaint outcomes but also the long-term benefits, including clearer and standardized processes, data-driven decision-making and improved analysis time. Thus, this study not only underscores the importance of data and the application of AI techniques in the era of quality but also fills a knowledge gap by providing an innovative and replicable approach to complaint analysis within the organization. In terms of originality, this article stands out for addressing an underexplored area and providing a tangible and applicable solution for the company, highlighting the intrinsic value of aligning quality with AI and digitization.

Keywords Quality 4.0, Industry 4.0, Artificial intelligence, Machine learning and customer complaints
Paper type Case study

Introduction

Currently, we are experiencing the fourth industrial revolution, aiming to achieve an even higher level of operational efficiency, productivity and automation (Lu, 2017). Factors such as the use of reliable data and real-time communication have become key aspects in achieving the agile objectives of I4.0. With the development of increasingly advanced technologies like Cyber-Physical Systems (CPS) and the Industrial Internet of Things (IIoT), I4.0 is emerging to meet the needs of smart factories, where concepts like data science, artificial intelligence (AI), machine learning (ML), deep learning (DL) and even predictive models cannot be disregarded. AI explicitly focuses on making intelligent devices think and act like human beings.

In an industrial context, AI can be defined as the ability of machines to understand/interpret, learn from data and make "intelligent" decisions based on knowledge and patterns extracted from the data. In turn, ML uses concepts already employed in statistics and algorithms to work with data. The significance of ML continues to gain prominence in industrial production, representing an opportunity to prevent, predict and prescribe configurations to achieve gains in productivity, quality, energy consumption and cost reduction. Alongside the rapid changes in the industrial field, there is no longer room for decisions based on intuition, making it imperative that these decisions be based on knowledge. This underscores the importance of forecasting, highlighting patterns capable of identifying trends and simplifying analysis and decision-making processes.

With the fourth industrial revolution comes the concept of Quality 4.0 (Q4.0), transitioning from a corrective paradigm to a predictive paradigm in companies, integrating quality tools and methodologies with technology. Q4.0 initiatives can help assess supply chain risk continuously or decide whether corrective action should be taken. Additionally, these initiatives can contribute to improving cybersecurity, as documentation and benchmarking processes can help the organization detect anomalies and understand expected performance to more effectively flag potential attacks.

Based on the aforementioned themes, AI and Q4.0, and their strong interconnection, it was decided to showcase the potential of integrating them through a case study. This work focused on analyzing customer complaint data from an automotive company, aiming to demonstrate the applicability of AI concepts in the new vision of Quality. Thus, it proved relevant to develop a predictive model, for classification problems, that would help identify the responsibility for customer faults based on a set of historical data. The development of this work has allowed highlighting the principles of Quality 4.0, where there is a significant emphasis on data analysis and process automation, all done using AI.

This automotive company is a benchmark in the market, prioritizing high-quality standards and placing great importance on customer satisfaction. Given this, and considering that customer complaints are a concern that incurs substantial costs associated with non-conformities and delays in customer responses, it was crucial to focus the analysis on these complaints. To meet the needs identified by various stakeholders, a case study methodology was employed, defining 10 distinct steps.

Despite the meticulous efforts expended in this study, it is crucial to acknowledge its limitations. The primary limitation lies in the dependence on the data available on the company's internal platform, which may influence the generalization of results. Furthermore, the specific conditions of the automotive sector may not be entirely extrapolatable to other industries. Other possible limitations include the accuracy and representativeness of the historical data used. Readers must understand these constraints as they impact the extent of the applicability of the results of this study to broader contexts.

To fulfill the proposed objective, this work is organized as follows: initially, a literature review is presented, highlighting the themes of Q4.0, I4.0 and AI. Next, the methodology adopted in the work and the case study are presented, including the topic's structure, a description of the sample and the methods applied. Subsequently, the main results obtained are presented, concluding with final considerations, limitations and suggestions for future work.

Literature review

Industry 4.0

To drive innovation in manufacturing by taking advantage of evolved automation and IoT, the German government proposed a new economic policy based on high-tech strategies, this concept gave rise to a new industrial revolution – Industry 4.0 – representing the fourth industrial revolution (Misra *et al.*, 2016). I4.0 is characterized by huge technological advancement and underpinned by fundamental advances that are mainly based on cyber physical systems, big data, cloud computing, collaboration systems and intelligent robots. In this transformation, sensors, machines, parts and information technologies systems are connected along the value chain. These connected systems can interact with each other using internet-based protocols and analyze data to predict failures, self-configuration and adaptability to changes. As noted by Schmidt *et al.* (2015), I4.0 is the overlap of several technological developments that embrace both products and processes associated with the so-called cyber physical systems that describe the merging of the digital with the physical workflow. The impact of I4.0 is huge, not only for industries that will be much more automated and efficient, but also for the whole society, with many current human activities that will be automated and replaced by machines. The introduction of I4.0 will change both products and the whole system in terms of processes, operations and services in many ways. It is expected that I 4.0 will have consequences for employment management, enabling the creation of new business models. This should have a great effect on the market, effectively affecting the entire product life cycle, providing a new way of producing a business and allowing the improvement of processes, thus contributing to the increase of the company's competitiveness (Pereira and Romero, 2017).

Quality

The element of quality is increasingly regarded as a strategic and differentiating characteristic in an ever more globalized world, particularly within the realm of innovation and the development of new products, with the aim of meeting the needs and expectations of customers (Sampaio *et al.*, 2009).

The history of modern quality spans over two centuries, encompassing diverse cultures, continents and historical events. Various philosophies and individuals associated with a quality movement have contributed to what is now perceived as the management and engineering of quality. It can be asserted that the history of quality commenced with the Industrial Revolution and the proliferation of mass production (Kolb and Hoover, 2012). Even at that time, there was a concern for the quality of products, signifying the assurance that all manufactured products exhibited similar characteristics and were free of defects. It was in this context that the concept of a "quality inspector" emerged, responsible for inspecting each product. The concept of job specialization was also introduced, giving rise to the first factories in the United States and Europe, thereby altering the traditional model of production and dividing the former roles of craftsmen and merchants into new roles of workers and production supervisors.

The developed factory system ensured product quality by relying on the skill of workers, complemented by sporadic audits. Over time, in the pursuit of greater efficiency, new techniques and philosophies emerged to contribute to quality improvement. Consequently, quality management increasingly assumed a crucial role within production processes, as it contributed to ensuring the reliability of products/services in accordance with customer requirements. Despite these advancements and the gradual emergence of quality tools, along with various perspectives from different authors, defining quality remains a complex task without consensus in the literature. Thus, there are varied definitions for what constitutes quality.

Despite the diversity of existing definitions for the concept of quality, it is understood that it plays an extremely important role in meeting customer requirements. In this sense, coupled with the rapid advancement of technologies, the traditional meaning of quality has assumed a broader role today, referred to as Quality 4.0. This can be characterized as the digitization of the total quality management concept and its impact on quality technology, processes and people. It can also be defined as the application of Industry 4.0 technologies to quality (Carvalho *et al.*, 2021).

Quality 4.0. In increasingly complex and competitive industrial environments, such as the current one, quality is a pivotal factor contributing to the success of companies. Quality 4.0 is a branch of Industry 4.0 that aims to enhance quality by applying intelligent solutions and algorithms. Despite the significant attention given to the Industry 4.0 theme in recent years, there has been little investigation into Quality 4.0, including the transition of organizations to operate under this new paradigm. Industry 4.0 focuses on technology implementation, but not necessarily on how these technologies create value for different stakeholders, the changes within the organization, and how quality work will be performed. Organizations can benefit from transitioning to Q4.0, becoming more effective in cost management and resource allocation (Sisodia and Forero, 2020). Q4.0 refers to the digitization of quality and how these digital tools can impact processes and people. In this sense, Quality 4.0 combines new technologies with traditional quality methods, aiming to achieve levels of operational excellence, performance and innovation (Aldag and Eker, 2018). Thus, Q4.0 has evolved as a natural response to changes in the field of production, where in 2011, the term I4.0 seemed to define a new way of increasing the competitiveness of the German manufacturing industry. In this context, Q4.0 seeks to meet the requirements of companies, including the digitization of quality management systems and practices, as well as the adoption of digital tools to increase the efficiency and quality of products. Concurrently, it supports the digitization of quality management, encompassing not only products and technologies but also processes and people (Dovleac, 2021). As Q4.0 is still a recent phenomenon with no formal definitions, and given its perceived importance, it is considered relevant to present this concept by various authors. Radziwill (2018) defined Quality 4.0 as the pursuit of performance excellence during potentially disruptive digital transformation times. The author asserts that its application will bring about a change at the level of the traditional concept of quality regarding efficiency, effectiveness and satisfaction for continuous and adaptive learning. This will enable changing boundaries within and between organizations, the way information is shared, adding intelligence to monitoring and operations management, remote monitoring for productivity improvement, continuous assessment of supply chain risks, and aid in decision-making. According to Nenadál (2020), Quality 4.0 encompasses issues of advanced quality management in the digital era and should be viewed as a data-driven approach to managing quality requirements, highlighting design, development, production, services and company culture as factors enabling agile communication and efficient feedback. It emphasizes the crucial support of IT as a condition for the practical establishment of Q4.0. The concept of Quality 4.0 is also associated with a set of supporting tools: artificial intelligence; big data; blockchain; machine learning; enabling technologies, deep learning and data science (Arsovski, 2019). Salimova *et al.* (2020) emphasize that Q4.0 can be defined as the adaptive capability of a product at any stage of its life cycle, considering customer needs and the interests of other stakeholders throughout the value chain. Javaid *et al.* (2021) argue that Q4.0 corresponds to the increasing digitization of the industry, utilizing advanced technologies to enhance the quality of products and services. Its goal is to digitize all quality systems and subsequently improve existing quality approaches. According to Escobar *et al.* (2021), Quality 4.0 represents the fourth wave in the quality movement (1. Statistical Quality Control, 2. Total Quality Management, 3. Six Sigma, 4. Quality 4.0). This quality philosophy is built on the statistical and management foundations of the previous three philosophies. Q4.0 leverages industrial Big Data, IIoT and AI to address an entirely new range of engineering problems. Q4.0 is based on a new paradigm grounded in empirical learning, knowledge discovery, real-time data collection and analysis to enable intelligent decisions. Regarding the production area, the main objectives of Quality 4.0 are to develop defect-free processes, enhance human intelligence, increase the speed and quality of decision-making, and alleviate the subjective nature of human inspection.

Why Quality 4.0?. The development of an effective Quality 4.0 (Q4.0) strategy enables organizations to address quality issues stemming from inefficiencies such as a lack of multifunctional ownership, ineffective communication and fragmented traditional quality systems. The concept of Quality 4.0 presents an opportunity for organizations to reassess the root causes of current quality obstacles and engage in strategic planning to explore how new technologies and their advantages – such as enhanced data transparency and high-quality data perception – can be leveraged to foster a culture of excellence (Juran, 2019). Implementation of Quality 4.0 paradigms can lead to improvements in the performance of people, projects and products through the incorporation of technologies like AI, ML, automation and blockchain. From Radziwill's perspective (2018), the value propositions for Q4.0 initiatives fall into six categories: (1) enhance (or improve) human intelligence; (2) expedite the speed and quality of decision-making; (3) enhance transparency, traceability and auditability; (4) anticipate changes and adapt to new circumstances and knowledge; (5) evolve regarding relationships, organizational boundaries and the concept of trust to unveil opportunities for continuous improvement and new business models; (6) learn by fostering self-awareness and consciousness.

Q4.0 initiatives can assist in the continuous assessment of supply chain risks or deciding on corrective actions. They can also contribute to improving cybersecurity – documentation processes and benchmarking can help organizations detect anomalies and understand expected performance more effectively, signaling potential attacks. According to Lyle (2017), fully automated or semi-automated quality data collection enables organizations to enhance efficiency in quality control. Real-time automated analysis provides swift responses when trends, out-of-spec values and variations are identified and appropriately addressed with statistical process control tools. Moreover, centralized data visibility allows all stakeholders to be involved from start to finish, contributing to the improvement of the entire supply chain, increasing production and reducing costs.

To summarize the reasons for adopting Q4.0, the following advantages, commonly described in the literature, are highlighted:

(1) Enhancing the quality and speed of decision-making

(2) Improving transparency, traceability and auditability conditions
(3) Anticipating and predicting changes, revealing deviations in datasets
(4) Facilitating adaptation to new circumstances and knowledge
(5) Identifying opportunities for continuous improvement and new business models
(6) Contributing to the reduction of rework
(7) Clarifying processes
(8) Making data-driven decisions (rather than solely relying on employee knowledge)
(9) Enhancing productive time and minimizing errors through process automation
(10) Contributing to higher quality of services and products

Industries recognize that Quality 4.0 will generate considerable value, necessitating comprehensive planning and implementation programs. Due to various quality issues, industries are beginning to build and incorporate a quality management model to enhance product properties. Machine operations are automatically adapted to unintended variations, such as environmental factors, to achieve high and stable product attributes by capitalizing on continuous sensor inputs. Small and multinational companies will quickly reach their production operations through digitization and integration of design and production processes using Q4.0 (Javaid *et al.*, 2021). However, the strategic and operational interconnection of I4.0 and Q4.0 remains an underexplored topic in the literature. According to Breteau (2022), companies must adopt a new approach to quality and harness the power of data. Data takes center stage in the era of Q4.0, especially concerning the rapid collection of large volumes of data and their efficient processing. Current best practices in quality are based on data, innovative technology and processing methods, enabling new ways of accessing information about production and distribution. According to the Boston Consulting Group's report "Quality 4.0 Takes More Than Technology," cited by Bolton (2019), about two-thirds of companies believe that next-generation quality technology involves predictive analytics, digital twins, simulation testing and built-in sensors. Companies that master the challenges associated with adopting Quality 4.0 will be able to see tangible benefits in all areas of the value chain. With Q4.0, companies can monitor processes, collect real-time data and apply analytics to predict quality issues and maintenance needs. Digital tools also enable people to perform their tasks faster, more efficiently and at lower costs. The digitization of quality will not happen overnight, and, in fact, the integration of Quality 4.0 will require companies to adopt a structured approach to both technological needs and the human element. Digital transformation accompanied by advanced quality practices will give companies a competitive advantage, creating operational efficiencies and improving products and services. But creating a culture of quality will be the more elusive goal, requiring sustained investment across the classic trinity of people, processes and technology.

Artificial intelligence

In 1956, the concept of AI was first proposed at the Dartmouth University seminar in the United States. AI plays a crucial role in the era of the fourth industrial revolution, where intelligent systems and technologies establish an active link between the physical and digital worlds. AI encompasses the science and engineering of creating intelligent machines capable of reasoning, learning, knowledge acquisition, communication, perception, planning and object manipulation. Its advantages include the use of sophisticated algorithms to "learn" from extensive data, applying gained knowledge in practical industry contexts and offering significant productivity improvements through fast and accurate data analysis (Darko *et al.*, 2020).

Lee *et al.* (2019) define AI as a cognitive science that enables humans to explore intelligent ways of modeling reasoning detection processes. It is a systematic discipline facilitating the consistent development and implementation of algorithms, as described by Liu *et al.* (2020). AI has the capacity to learn, reason, perceive and make independent decisions, replacing human labor and enhancing work and production efficiency through data processing and technical analysis.

In the journey towards a fully connected factory within the digital transformation landscape, essential concepts such as data science, artificial intelligence, machine learning, deep learning and predictive models cannot be overlooked. AI, machine learning and predictive models play a pivotal role in industrial contexts, focusing on making devices think and act intelligently. AI in industry refers to machines' ability to understand, interpret, learn from data and make informed decisions based on insights and data patterns.

Machine learning – categorized into supervised learning, unsupervised learning and reinforcement learning (Lee *et al.*, 2018) – is integral to achieving productivity, quality, energy consumption and cost savings in industrial production. The importance of ML lies in its ability to move organizations towards predictive analytics, away from exclusive dependence on descriptive analysis.

Various ML methods, such as decision trees, classification rules, logistic regression, inductive logic programming, support vector machine, Bayesian methods and artificial neural networks, have been developed by researchers (Pessoa, 2018). These methods contribute to predictive modeling, enhancing decision-making efficiency and identifying opportunities for improvement.

Performance metrics are crucial for evaluating classifiers, ensuring result reliability. Commonly referred metrics include the contingency matrix, accuracy, precision, specificity, recall, F1 Score and receiver operating characteristic (ROC) curve analysis.

Given the data-centric nature of ML, the necessity arises to provide pre-processed, high-quality data for optimal results. Feature Selection (FS) techniques are crucial in identifying relevant characteristics, removing redundant, irrelevant or noisy features, speeding up data mining algorithms and improving predictive accuracy. Cai *et al.* (2018) emphasize that Feature Selection is an effective means to address high-dimensional data challenges in ML and DM, reducing computation time and enhancing model understanding.

Methodology

The methodology employed in this study was carefully crafted based on three established data mining methodologies: CRISP-DM, SEMMA and P3TQ. CRISP-DM, resulting from collaboration between DaimlerChryrler, SPSS and NRC, establishes a comprehensive framework consisting of six fundamental phases for DM projects, covering from business understanding to the implementation of obtained results (Chapman *et al.*, 2000). SEMMA, developed by the SAS Institute, focuses on DM projects, offering a structure divided into five essential phases for the data analysis process (Azevedo and Santos, 2008). On the other hand, P3TQ, also known as Catalyst, proposed by Dorian Pyle, presents two models that assist in identifying business problems and exploring data based on these identified issues (Pyle, 2003).

Although not a completely new methodology, the approach adopted in this study was specifically developed to meet the company's data demands, aiming to be replicable across various areas and serve as a foundation for data scientists. The company recognizes the potential of data analysis and is increasingly investing in this area, understanding that standardized processes facilitate team work and improve operational efficiency.

For this study, real customer complaint data from an automotive company was used. Whenever a complaint about a defective product is received from a customer, the data related to that complaint is recorded, a series of fields are filled in, and the information is stored in an internal system. For this study, only complaints of the type 0 km (unit complained about by the customer) and Campo (unit complained about by the end customer) were considered. Considering the high costs and complexity associated with quality complaint analysis, the main objective was to identify variables that would minimize the impact of complaints and thus enhance operational performance. After identifying these variables and cleaning the initial sample, the next step was to obtain a predictive model that considered only the variables of interest. This model was built using AI methods to apply various techniques and evaluate which best suited the study context. Model evaluation was carried out using specific metrics to determine their effectiveness in predicting complaint responsibilities.

The methodological process was divided into 10 distinct phases, aligned with the studied methodologies and represented in Figure 1.

Business understanding and analysis: This phase, based on CRISP-DM, aimed to deeply understand the company's operational context, identifying challenges, opportunities and stakeholder needs, providing a precise analysis of relevant variables and their impacts on business outcomes.

Project plan definition: Inspired by CRISP-DM, this phase involved developing a detailed plan for project execution, establishing clear objectives, timelines and required resources to ensure a structured approach.

Sample definition: Influenced by SEMMA, this phase focused on identifying and selecting relevant data for analysis, ensuring that the sample adequately represented customer complaint processes in the automotive industry.

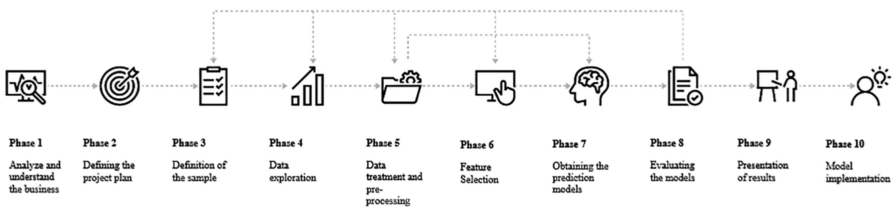
Data exploration: Using a similar approach to SEMMA, this phase involved a detailed exploratory analysis of the data, using descriptive statistics and visualizations to identify patterns and trends.

Data processing and pre-processing: This step was crucial to prepare the data for modeling, including data cleaning and handling missing values to ensure data quality and integrity.

Feature selection: Inspired by P3TQ techniques, this phase aimed to identify the most relevant variables for building the predictive model, using importance and correlation analyses to select the most significant variables.

Predictive model Acquisition: In this phase, supervised Machine Learning models were developed for complaint classification, using techniques adapted to the case study, including data splitting into training and validation sets to ensure model robustness.

Model evaluation: Model performance evaluation was conducted using specific metrics, focusing on prediction accuracy of complaint responsibilities.



Source(s): Figure by authors

Figure 1.
Methodology followed

Results presentation: Results were presented clearly and objectively, highlighting key insights obtained during the study, providing a comprehensive understanding of the findings.

Implementation: Although the complete implementation of the model was not carried out in this context, its practical application was discussed and recognized as essential for obtaining tangible benefits for the business.

Each of these phases was carefully aligned with the studied methodologies, providing a structured and scientifically grounded approach to data analysis in the company. This detailed methodology not only allowed for a deeper understanding of customer complaint processes in the automotive industry but also provided valuable insights to guide strategic decision-making and improve overall business performance. By adopting an approach based on established data mining methodologies and adapting it to the company's specific needs, it was possible to ensure the quality and relevance of the obtained results, contributing to the success of the project and the advancement of data analysis in the organization.

Case study

The data utilized in this study pertain to customer complaints from an automotive company. The customer complaint process entails numerous stages and human resources, often rendering it a time-consuming and costly endeavor for the company. Delays in responding to customers result in penalties. One bottleneck in this process is the analysis of NOK (non-compliant) units, which can be exceedingly complex and time-consuming. In this context, following discussions with representatives from various departments, the necessity to analyze customer complaint data and predict responsibility for a complaint was identified.

Customer complaints are initiated when equipment is identified as non-compliant, leading to a Field or 0 km complaint. When handling such complaints, response deadlines defined in customer procedures must be strictly adhered to. Any delays must be communicated in advance to customers, accompanied by appropriate justifications. Data related to non-conformities are immediately entered into the system, providing the factory with information about occurrences at customer facilities. This information has been available since 2004 and is stored in the company's proprietary system, also having a replica in SAP. Technicians responsible for the analysis regularly consult this data as it contains information reported by the customer, such as reported defects, the number of defects found, city of occurrence, etc. Analyzing the information available in this database aids in identifying both the problem and the assignment of responsibility – firstly, to determine if the reported defect is valid and subsequently, to address internal responsibilities. However, analyzing this information is not straightforward due to its complexity, driven by both the number of variables and the presence of noise. It is emphasized that the analysis of this information is crucial for identifying the path to resolving the problem and, consequently, reducing analysis time and customer response time.

In summary, the complaint process begins with the reception of the NOK unit, followed by an analysis of the unit and verification of evidence to assess responsibility for the complaint. If the complaint is attributed to the company, internal responsibility may lie with the supplier, development, production, or logistics. Subsequently, it is necessary to identify the root cause of the problem and define containment actions to prevent a recurrence. Considering the time required for analysis and response to a complaint, optimizing this process is crucial, as it contributes to reducing the number of analyses, customer response time, people involved and associated costs.

Therefore, efforts were made to identify key variables for analysis. Upon identifying these variables, technicians could focus only on those of interest. Subsequently, based on the preceding information, a predictive model was constructed to assess responsibility for a customer's complaint. The initial dataset considered comprised 144,036 records and 133 variables. After initial processing, a reduction in sample size was recorded. The treatment performed is described in Table 1.

Subsequently, feature selection (FS) techniques, including the filtering method, wrapper method and embedded method, were employed. The application of these techniques, in an automated manner, identifies a subset of variables that substantiate the client's consent, thereby facilitating a more streamlined analysis. It is worth noting that, prior to the application of these techniques, a correlation analysis among variables was conducted, and those exhibiting a Pearson correlation coefficient greater than or equal to 0.80 were excluded from the analysis. The outcomes are depicted in Figure 2, showcasing the characteristics selected by the distinct techniques.

Then, supervised ML algorithms were applied, specifically to address classification problems. The models aimed to predict the liability of a customer complaint, considering different combinations of features obtained in the previous phase. The data were split into training and validation sets, with an 80/20% split, respectively. It is important to highlight that the models were applied to all variants shown in Figure 2; however, in this work, only these three versions are presented, as the results for these cases were most suitable for this case study.

The results for each model and their respective metrics are presented in Figure 3. In this case, commonly used metrics in classification problems were employed, including precision, accuracy, recall, F1 and ROC curve.

The ROC curves obtained for some of the models of case 1 – without using FS are presented in Figure 4. The ROC curves obtained for some of the models in case 1 – without the use of FS are shown in Figure 4. By analyzing the ROC curve, and taking into account that the closer the ROC curve is to the upper left corner, the better it is. the performance of the model, as it indicates a high rate of true positives and a low rate of false positives, it appears that decision tree is the method with the best performance, compared to the remaining three presented.

When analyzing the aforementioned results, the following models were chosen: decision tree, XGBoost and LightGBM. These models exhibit metrics with higher values compared to others; thus, they will be utilized for applying the test dataset. Concerning the variants of applied FS techniques, there were generally no significant differences between usage and non-usage. It is noteworthy that, before employing the FS techniques, an initial cleaning and filtering of non-relevant variables for analysis were conducted. This underscores the importance of analysis, data cleansing, as well as the significance of possessing expertise in the study area.

Despite computing various metrics described earlier (Accuracy, Precision, Recall, F1-Score and ROC Curve), we opted to employ the accuracy metric. Following the selection of prediction models with the best metrics, and aiming to validate their performance, a dataset

Table 1. Steps considered to obtain the number of variables and records of the final sample

	Records	Number of variables
Initial sample and attribute description	144,036	133
Check null values (remove ≥60% of null values)	144,036	79
Excluded users and single-valued variables		
Filter only field and 0 km claims	42,168	79
Identification of attributes with repeated information or that do not make sense for the analysis	42,168	42

Source(s): Table by authors

Variables	Feature Importance	Mutual information	Forward Feature Selection		Backward Feature elimination		Recursive Feature elimination		Boruta	LASSO	RIDGE
			RandomForestClassifier	KNeighborsClassifier	RandomForestClassifier	KNeighborsClassifier	DecisionTreeClassifier	RandomForestRegressor			
Claimed_Qty__Mat_			x		x	x					
Mileage			x		x			x			
idRAW_PartsList	x	x			x	x			x	x	
compl__Material	Deleted by $r >= 0.80$										
QC_Notification	x	x	x					x	x		
Part_Number	x	x				x		x			x
BusinessUnit	Deleted by $r >= 0.80$										
CD_Codegroup_text	Deleted by $r >= 0.80$										
CD_Code__Pos__1000_				x	x			x	x		
date_month	Deleted by $r >= 0.80$										
date_day	Deleted by $r >= 0.80$										
date_year	Deleted by $r >= 0.80$										
Customer				x		x					
PDCL			x	x		x					x
Complaint_Mode				x							x
Systemstatus	x	x	x	x	x	x	x	x	x		x
Product_type_formula	Deleted by $r >= 0.80$										
CuCo_text	x	x						x			
Investigat__Plant__Pos__1000_	x		x	x	x	x			x		x
CD_Code_Group__Pos__1000_			x	x		x					x
CD_Code_text	Deleted by $r >= 0.80$										
notification_date	x	x			x			x			
mileage_qty__unit			x	x							x
City		x						x		x	
Country_of_Fail					x						x
CustomerMaterial	x	x	x					x			
Customer_ProductModel	x	x		x	x	x				x	x
Family_TestSystem	x	x		x		x		x	x	x	
Resp_claim	target										

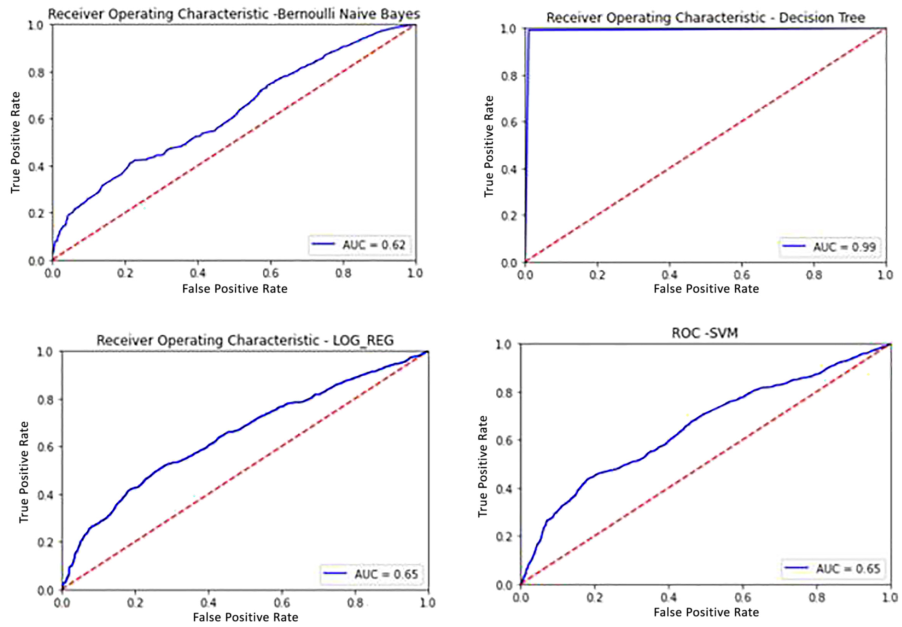
Source(s): Figure by authors

Figure 2.
Selecting features by the different methods

Without Using FS						Feature Importance						Mutual information					
Models	Accuracy	Precision	Recall	F1-Score	AUC	Models	Accuracy	Precision	Recall	F1-Score	AUC	Models	Accuracy	Precision	Recall	F1-Score	AUC
KNN	0.869220	0.880214	0.883426	0.881817	0.94	KNN	0.940598	0.940640	0.952555	0.946560	0.98	KNN	0.917833	0.920824	0.931301	0.926033	0.97
Bernoulli Naive bayes	0.565212	0.602397	0.625805	0.613878	0.62	Bernoulli Naive bayes	0.554897	0.592928	0.619150	0.605755	0.58	Bernoulli Naive bayes	0.555134	0.592411	0.623444	0.607531	0.58
GaussianNB	0.565568	0.621101	0.547231	0.581831	0.62	GaussianNB	0.549443	0.594369	0.580077	0.587136	0.61	GaussianNB	0.563552	0.592152	0.673894	0.630385	0.60
Decision Tree	0.989566	0.991821	0.989266	0.990542	0.99	Decision Tree	0.987313	0.987572	0.989480	0.988525	0.99	Decision Tree	0.990040	0.992888	0.989051	0.990966	0.99
Regressão logistica	0.603272	0.626251	0.698583	0.660442	0.65	Regressão logistica	0.563908	0.594998	0.658866	0.625306	0.63	Regressão logistica	0.591179	0.614627	0.696436	0.652979	0.63
SVM	0.615011	0.633138	0.720266	0.673898	0,65	SVM	0.573868	0.613240	0.618506	0.615861	0.64	SVM	0.578848	0.607129	0.672821	0.638289	0.64
XGBoost	0.991937	0.997399	0.987978	0.992666	1	XGBoost	0.991107	0.996748	0.987119	0.991910	1	XGBoost	0.991582	0.997182	0.987548	0.992342	1
LightGBM	0.991937	0.996538	0.988836	0.992672	1	LightGBM	0.989266	0.993318	0.989266	0.991288	1	LightGBM	0.991345	0.996535	0.987763	0.992129	1

Source(s): Figure by authors

Figure 3.
Result for ML models



Source(s): Figure by authors

Figure 4. ROC curves obtained for Bernoulli Naive Bayes, decision tree, logistic regression and SVM models

was introduced to test the validity of the already trained model (226 complaints without assigned responsibility). The utilized models were decision tree, XGBoost and LightGBM. Figure 5 presents the confusion matrix obtained for each of the models.

It was found that the model that presents the best results is XGBoost (Table 2).

The precision metric analysis revealed that 59.7% of complaints were correctly classified, although this value is not considered high; however, it is important to be attentive to cases of "false negatives and false positives." These results highlight the feasibility of reducing the analysis time for a complaint by focusing on incorrectly categorized complaints. In other words, considering the confusion matrix, for a sample of 226 complaints, the analysis of only 91 complaints would be necessary, leading to a reduction in the number of units to be examined.

It was decided to study the possibility of improving the accuracy value, so two approaches were undertaken: the first related to the implementation of AutoML, and the second, using text processing techniques to handle the variable with the problem description. By applying

Decision Tree				XGBoost				LightGBM			
Predict	0	1	All	Predict	0	1	All	Predict	0	1	All
Real				Real				Real			
0	10	63	73	0	5	68	73	0	15	58	73
1	54	99	153	1	23	130	153	1	63	90	153
All	64	162	226	All	28	198	226	All	78	148	226

Source(s): Figure by authors

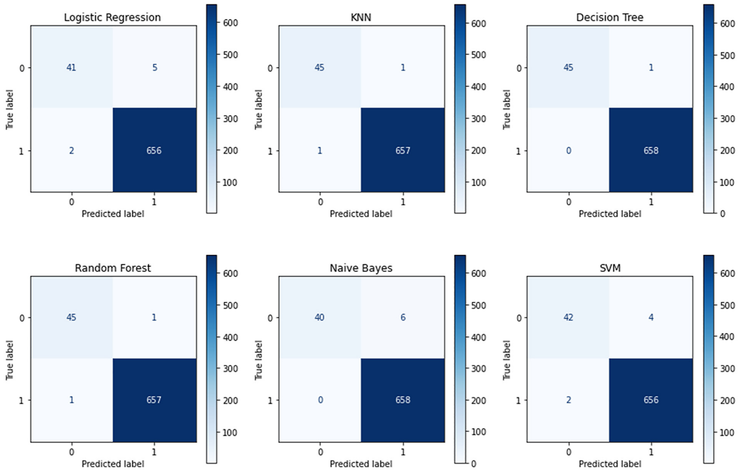
Figure 5. Confusion matrix obtained for the three models

AutoML models (TPOT, H2O and Auto-Sklearn), accuracy was improved when Auto-Sklearn was applied with a value of 64.16%, meaning that a percentage of incorrectly categorized units was 35.8% of the received units. In the second approach, translation was performed, and clustering was applied so that each data row was assigned a cluster. By applying the Elbow method, it was determined that $k = 50$. With the application of KMeans, and after each row was categorized with a certain cluster number, the model is capable of learning, and when a new description is introduced, it automatically categorizes it into a specific cluster. Traditional ML and AutoML models were applied, and it was concluded that they did not show improvements in results. After having a model capable of predicting the responsibility of a customer's received complaint, it was important to analyze internal attribution – supplier responsibility. After applying ML models, it was found that DT presented the best accuracy – 64.05% (Figure 6). Thus, it is possible to reduce the number of complaints to be analyzed and automatically forward them to the person in charge of supplier complaint management. Besides reducing the analysis time, this will avoid intermediate steps and contribute to responding to the customer in a shorter time.

Table 2.
Result of the metrics of the models

	Accuracy	Precision	Specificity	Recall	F1
Decision tree	48.23	61.11	13.70	64.71	61.86
XGBoost	59.73	65.66	6.84	84.97	74.07
LightGBM	46.46	60.81	20.55	58.82	59.80

Source(s): Table by authors



	Model	Accuracy	F1 Score	Recall	Precision
0	Logistic Regression	0.990057	0.994693	0.99696	0.992436
1	KNN	0.997159	0.998480	0.99848	0.998480
2	Decision Tree	0.998580	0.999241	1.00000	0.998483
3	Random Forest	0.997159	0.998480	0.99848	0.998480
4	Naive Bayes	0.991477	0.995461	1.00000	0.990964
5	SVM	0.991477	0.995448	0.99696	0.993939

Source(s): Figure by authors

Figure 6.
Metrics and confusion matrix applied to ML models to predict complaint responsibility with supplier data

The importance of integrating customer and supplier complaint information from the same data source was recognized (which was not the case until then). A preliminary analysis of the variables was conducted, identifying those relevant for the analysis, and the databases were merged by the complaint number. The information was made available in HIVE in a table format, making access to information simpler, eliminating the need to resort to different systems or request specific accesses and reducing unnecessary information that complicates the analysis, as variables that proved unnecessary were already excluded.

It was deemed relevant to invest part of the work in automating the data loading process. Automations were created to ensure that data from these sources is already available on a daily basis. Having a script that allows data manipulation, processing, modification, additional information integration, and storage proved advantageous in the context of Q4.0. This procedure also facilitates faster implementation of new analyses, freeing up people for other tasks.

Finally, to streamline and standardize processes, a guide for solving classification problems was presented. The aim was to extend the implementation of predictive models to other areas of the factory, being applied to other use cases accordingly.

Results

The data that served as the basis for the development of this work originated from customer complaint records, encompassing information from 2004, totaling 144,036 entries with 133 variables. Given that most variables were in string format, the identification and treatment of outliers were deemed unjustified. Nonetheless, a transformation into numerical values using ordinal encoding was necessary. Subsequently, the correlation among variables was scrutinized using Pearson's correlation coefficient, facilitating the identification of correlated variables and verifying the absence of significant correlation with the target variable. Following this, feature selection (filter, wrapper and embedded methods) techniques were applied to the dataset comprising customer complaints. These supervised techniques pinpointed the most relevant variables for customer responsibility, namely "Short_text_for_code__Pos__1000_5," "Systemstatus," "Object_part__Pos__1000_," and "Short_text_for_code__Pos__1000_4."

Moving on, the application of machine learning (ML) models ensued. Data were partitioned into training and validation sets in an 80/20 ratio, and the models were trained, exhibiting satisfactory metrics. It is noteworthy that various ML models were applied to different Feature Selection variants, yet no significant disparities in model metrics were observed. This underscores the importance of initial data analysis and cleaning, coupled with domain knowledge. Following the application of ML models and metric analysis, the top three performers were identified: decision tree, XGBoost and LightGBM.

After training and validation, the models were put to the test using a sample of complaints without assigned responsibility. XGBoost emerged as the superior model, achieving an accuracy of 59.73%. In this scenario, out of 226 complaints, only 91 (corresponding to false positives and false negatives) would require analysis, enabling technicians to focus solely on these cases and reduce the number of units under scrutiny.

Shifting focus to internal responsibility, a new model demonstrated an accuracy of 64.05% in predicting complaints related to supplier responsibility. The application of predictive models proved effective in saving time and contributing to enhancements in the customer complaint process. The significance of having information from various sources centralized and automatically updated was emphasized, streamlining the data query process for new analyses.

In the context of Industry 4.0 (Q4.0), increasingly intertwined with analytical comprehension, this work highlighted the positive contribution of data analysis, treatment and the application of AI techniques to the Q4.0 vision. The identified advantages in terms of quality encompassed decision-making support, result prediction, information extraction from data, clearer and standardized processes, data-driven decision-making and improved analysis time.

Thus, and as the results obtained were to be synthesized, a summary of the initial and final status achieved with the development of this work is presented below:

1. Standardize data sources and make them available in a single data repository

Initial status: Customer complaint data available in an internal company system and in SAP; costs associated with SAP access permissions; excessive variables for analysis.

Current status: Selection of variables to consider; merging customer complaint and supplier complaint information into a single table; table available in HIVE; no need for specific SAP transaction authorizations; information in one place, containing only necessary data; management of variables to include in the sample done in a Python script.

2. Identify variables of interest to assess complaint responsibility without the need for a cumbersome and in-depth analysis:

Initial status: Excess variables constituting the customer complaint database; difficulty identifying what is important for identifying the root cause of the problem.

Current status: Data cleaning – elimination of variables causing noise in the analysis (e.g. user, repeated information, missing values); application of FS techniques allowing automatic identification of variables of interest; implementation of FS methods in a script, making the process repeatable as needed.

3. Relate customer complaint data to other data that may be at the root cause of the complaint, thus analyzing possible correlations

Initial status: No analysis associating customer complaints with other data sources.

Current status: Studied key internal players in the internal responsibility of a complaint; identified three pathways: supplier, development and production; the supplier represented 40% of internal responsibility; joint analysis of customer and supplier information; construction of an ML model to predict supplier responsibility for a complaint.

4. Apply integrated quality principles with AI techniques to make processes more automated and faster

Initial status: Reception of NOK units and the need to analyze 100%; no automatic way to identify important variables regarding supplier responsibility; existence of historical data not analyzed; absence of joint analysis of customer and supplier data.

Current status: Analyzed customer and supplier complaint data; data cleaning; variable transformations; identification of important variables; implementation of FS techniques capable of systematically and automatically identifying relevant features; data integration; application of ML models capable of predicting complaint responsibility; creation of automations for data loading; presentation of data in the form of a dashboard, facilitating analysis; provision of a guide for applying to new use cases. All these developments are based on the principles of Quality 4.0, where great emphasis is given to data analysis, process automation, all done with the use of AI.

5. Support decision-making regarding complaint responsibility

Initial status: Decision based on unit analysis; time-consuming process contributing to delays in customer response.

Current status: Construction of predictive models capable of predicting complaint responsibility, as well as responsibility lying with the supplier; possibility of only looking at false negatives and false positives, thus reducing the number of units to be analyzed; decision making based on ML models.

6. Contribute to reducing analysis time and costs in the complaint process

Initial status: All units received by the customer are analyzed; the analysis process is time-consuming, which sometimes delays the entire customer response process; there are agreements with the time to respond to a complaint, and if these are not met, costs will be incurred by the company.

Current status: Obtaining predictive models allows the number of complaints to be analyzed to be reduced; with a smaller number of analyses, and immediate identification of responsibility, the process becomes faster, avoiding delays in responding to the customer, consequently avoiding costs, as well as human resources necessary for the analysis.

7. Analyze and identify the impacts of using a predictive model for customer complaints in the company

Initial status: NA.

Current status: As mentioned earlier, models based on AI techniques and using ML have been built, capable of predicting customer complaints. Various approaches were tested for accuracy optimization. However, these models have not yet been put into production, making it impossible to assess the actual gains and impacts for the company. Nevertheless, based on the results obtained from the models, optimization and automation of data loading, and how these can be presented with information extracted, it is believed that the introduction of these models and techniques in a production context will bring numerous improvements. Highlighting: reduction in the number of units to analyze; reduction in analysis time; reduction in the number of human resources allocated to analysis; reduction in the number of meetings to discuss responsibility; decision based on intelligent models; reduction in response time to the customer; alignment of information between various departments due to information access; possibility of new analysis with data loaded daily; presentation of information using dashboards, transforming data into useful and easy-to-understand information.

In terms of Quality 4.0, and considering that it is increasingly linked to analytical understanding, this work has shown that data analysis, its treatment and the application of AI techniques contribute positively to the Q4.0 vision. Thus, the advantages found with this work in terms of quality are enumerated: contribution to decision-making, result prediction, information extraction from data; clearer and standardized processes; decision based on data and improvement of analysis time.

Based on the shared information, the results obtained in this study reflect a significant advancement in the field of quality, especially in the context of digital transformation and Quality 4.0. By integrating advanced data analytics and AI techniques into the customer complaint management process, this work presents substantial contributions to research and business practice.

One of the key innovations of this study lies in the effective application of ML techniques to predict responsibilities in customer complaints, something that had not been widely explored in the automotive industry until now. The creation and implementation of predictive models enabled not only faster and more accurate analysis of complaints but also the proactive identification of potential root causes, leading to more efficient and targeted intervention.

Furthermore, the development of a detailed and structured methodology to address the specific challenges of the complaint process demonstrates a commitment to operational excellence and continuous improvement. Standardizing data, automating the data loading process, and integrating information from different sources into a single repository represent significant advances in optimizing workflows and reducing redundancies.

By providing a clear and objective overview of the steps taken, the methods employed and the results achieved, this study not only contributes to the advancement of academic knowledge in the field of quality and data analysis but also offers valuable insights for the automotive industry and other organizations seeking to enhance their quality processes and meet the demands of the digital era.

In summary, the results of this study represent a new frontier in the application of analytical and AI approaches to address complex quality challenges, demonstrating the transformative potential of these technologies when strategically and integratively applied to business processes.

Conclusions

This study represents an innovative synthesis between the principles of Quality 4.0 and the capabilities of Machine Learning, aiming to optimize the customer complaint management process. Quality 4.0 seeks to transition to data-driven decisions, promoting improvements in both processes and product quality. The application of ML techniques in this context reveals a growing potential to increase efficiency, identify improvement opportunities and assess the business in real-time.

In predicting the liability of complaints, the application of Feature Selection methods identified crucial variables. While the XGBoost model achieved an accuracy of 59.7%, it proved limited, leading to the exploration of advanced AutoML techniques. These techniques increased accuracy to 66.16%, but still indicated the need for improvements. Various factors, such as the complexity of the problem, suggest exploring more advanced ML techniques to automatically optimize model performance.

The inclusion of new variables, such as detailed data from supplier complaints, production, or measurements, can enrich the dataset, enhancing the model's generalization capability. This is especially relevant when considering the multifaceted nature of complaints and the complexity of the operational environment.

The practical implications of this study go beyond optimizing the complaint process. The proposed methodology can be extended to other cases within the company, promoting efficiency and quality in different operational contexts. The flexibility of the approach, aligned with Q4.0 principles, suggests that the proposed improvements have the potential to be adapted to various industries. Additionally, the methodology can be replicated in similar companies facing challenges in complaint management.

For future research, the continuous pursuit of improvements in model metrics is crucial. The inclusion of new data sources, along with considerations of limitations such as null variables and limited datasets, is a promising avenue to enhance prediction. Despite limitations, the conclusions of this study offer considerable practical implications, representing a versatile tool to enhance operational efficiency and product quality.

In summary, this study not only fills gaps in the understanding of complaint management but also establishes a valuable precedent for future research aiming to integrate Q4.0 principles and advanced ML techniques. By addressing the specific challenges of the company in question, it opens the door to replicating this innovative approach in similar contexts, consolidating contributions to the continuous evolution of business processes and product quality in the era of digitization.

References

Aldag, M.C. and Eker, B. (2018), "What is quality 4.0 in the era of industry 4.0?", *International Conference on Quality of Life*, November, pp. 31-34.

Arsovski, S. (2019), *Social Oriented Quality : From*, Vol. 1, pp. 397-404.

Azevedo, A. and Santos, M.F. (2008), "KDD, SEMMA and CRISP-DM: a parallel overview", *IADIS European Conference Data Mining*, pp. 182-185.

Bolton, D. (2019), "Report: quality 4.0 adoption requires more than just emerging technology", available at: https://www.etq.com/blog/report-quality-4-0-adoption-requires-more-than-justemerging-technology/

Breteau, S. (2022), "Embarking on the quality 4.0 transformation journey", available at: https://www.sdcexec.com/software-technology/emerging-technologies/article/21940777/qima-embarking-on-the-quality-40-transformation-journey

Cai, J., Luo, J., Wang, S. and Yang, S. (2018), "Feature selection in machine learning: a new perspective", *Neurocomputing*, Vol. 300, pp. 70-79, doi: 10.1016/j.neucom.2017.11.077.

Carvalho, A.V., Enrique, D.V., Chouchene, A. and Charrua-Santos, F. (2021), "Quality 4.0: an overview", *Procedia Computer Science*, Vol. 181 No. 2019, pp. 341-346, doi: 10.1016/j.procs.2021.01.176.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. and Wirth, R. (2000), "Crisp-dm 1.0 Step-by-step data mining guide", available at: https://mineracaodedados.wordpress.com/wp-content/uploads/2012/12/crisp-dm-1-0.pdf

Darko, A., Chan, A.P.C., Adabre, M.A., Edwards, D.J., Hosseini, M.R. and Ameyaw, E.E. (2020), "Artificial intelligence in the AEC industry: scientometric analysis and visualization of research activities", *Automation in Construction*, Vol. 112 No. May 2019, 103081, doi: 10.1016/j.autcon.2020.103081.

Dovleac, R. (2021), "Knowledge management systems in Quality 4.0", *MATEC Web of Conferences*, Vol. 342, 09003, doi: 10.1051/matecconf/202134209003.

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.

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 June, 100109, doi: 10.1016/j.sintl.2021.100109.

Juran (2019), "Quality 4.0: the future of quality?", available at: https://www.juran.com/blog/quality-4-0-the-future-of-quality/

Kolb, R.R. and Hoover, M.L. (2012), "The history of quality in industry", Sandia National Laboratories, August, pp. 1-17.

Lee, J.H., Shin, J. and Realff, M.J. (2018), "Machine learning: overview of the recent progresses and implications for the process systems engineering field", *Computers and Chemical Engineering*, Vol. 114, pp. 111-121, doi: 10.1016/j.compchemeng.2017.10.008.

Lee, J., Singh, J. and Azamfar, M. (2019), "Industrial artificial intelligence", *ArXiv*.

Liu, J., Chang, H., Forrest, J.Y.L. and Yang, B. (2020), "Influence of artificial intelligence on technological innovation: evidence from the panel data of China's manufacturing sectors", *Technological Forecasting and Social Change*, Vol. 158 April, 120142, doi: 10.1016/j.techfore.2020.120142.

Lyle, M. (2017), "Revealing the value of data through quality intelligence", *Quality Magazine*, October, pp. 25-30.

Misra, G., Kumar, V., Agarwal, A. and Agarwal, K. (2016), "Internet of Things (IoT) – a technological analysis and survey on vision, concepts, challenges, innovation directions, technologies, and applications (an Upcoming or future generation computer communication system technology)", *American Journal of Electrical and Electronic Engineering*, Vol. 4 No. 1, pp. 23-32, doi: 10.12691/ajeee-4-1-4.

Pereira, A.C. and Romero, F. (2017), "A review of the meanings and the implications of the Industry 4.0 concept", *Procedia Manufacturing*, Vol. 13, pp. 1206-1214, doi: 10.1016/j.promfg.2017.09.032.

Pessoa, V. (2018), "Exploration of machine learning techniques for automatic optical inspection", in *Statistical Field Theor*, Vol. 53 No. 9, doi: 10.1017/CBO9781107415324.004.

Pyle, D. (2003), *Business Modeling and Data Mining*, Morgan Kaufmann, doi: 10.1016/B978-1-55860-653-1.X5000-3.

Radziwill, N. (2018), "Let's get digital: the many ways the fourth industrial revolution is reshaping the way we think about quality", *ArXiv*, pp. 1-10.

Salimova, T., Vatolkina, N., Makolov, V. and Anikina, N. (2020), "The perspective of quality management system development in the era of industry 4.0", *Humanities and Social Sciences Reviews*, Vol. 8 No. 4, pp. 483-495, doi: 10.18510/hssr.2020.8447.

Sampaio, P., Saraiva, P. and Rodrigues, A.G. (2009), "ISO 9001 certification research: questions, answers and approaches", *International Journal of Quality and Reliability Management*, Vol. 26 No. 1, pp. 38-58, doi: 10.1108/02656710910924161.

Sisodia, R. and Forero, D.V. (2020), "Quality 4. 0 – how to handle quality in the Industry 4. 0 Revolution", Chalmers University of Technology.

Further reading

Abbott, D. (2014), *Applied Predictive Analytics: Principles and Techniques for the Professional Data Analyst*, John Wiley & Sons, Indianapolis, Indiana.

Campbell, K.S. (1999), "Collecting information: qualitative research methods for solving workplace problems", *Technical Communication*, Vol. 46 No. 4, pp. 532-545.

Carneiro, V.H.S. (2019), "Abordagens híbridas de Machine Learning / Simulação para sistemas logísticos dinâmicos", Tese de Mestrado, Faculdade de Engenharia da Universidade do Porto, available at: https://repositorio-aberto.up.pt/bitstream/10216/121197/2/342930.pdf

Cote, C. (2021), "What is predictive analytics? 5 Examples", available at: https://online.hbs.edu/blog/post/predictive-analytics

Dias, A., Carvalho, A. and Sampaio, P. (2022), "Quality 4.0: literature review analysis, definition and impacts of the digital transformation process on quality", *International Journal of Quality and Reliability Management*, Vol. 39 No. 6, pp. 1312-1335, doi: 10.1108/ijqrm-07-2021-0247.

Eckert, K.B. and Britos, P.V. (2021), "Proposed extended analytic hierarchy process for selecting data science methodologies", *Journal of Computer Science and Technology (Argentina)*, Vol. 21 No. 1, pp. 49-58, doi: 10.24215/16666038.21.E6.

Ferreira, P.M. (2010), "Aplicação de Algoritmos de Aprendizagem Automática para a Previsão de Cancro de Mama", available at: https://www.dcc.fc.up.pt/∼ines/aulas/1516/DM1/TesePedroFerreira.pdf

Gama, J. (2000), "Combining classification algorithms", *Aicommunications*, Vol. 13 No. 2, pp. 135-136.

Kelleher, J.D., Namee, B.mac and D'Arcy, A. (2015), "Fundamentals of machine learning for predictive data analytics algorithms", Worked Examples, and Case Studies, available at: http://library1.nida.ac.th/termpaper6/sd/2554/19755.pdf

Wu, X., Kumar, V., Ross, Q.J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Yu, P.S., Zhou, Z.H., Steinbach, M., Hand, D.J. and Steinberg, D. (2008), "Top 10 algorithms in data mining", *Knowledge and Information Systems*, Vol. 14 No. 1. pp. 1-37, doi: 10.1007/s10115-007-0114-2.

Corresponding author
Anabela Costa Silva can be contacted at: anabelacs@gmail.com