A classification method for making-do waste using *Machine Learning*

Método de classificação de perdas por making-do com uso de Machine Learning

Tatiana Gondim do Amaral
Gabriella Soares de Paula
Caio César Medeiros Maciel

Abstract

This research study investigates the potential use and best-fitting model for the automated making-do waste classification in construction sites, using Machine Learning techniques to reduce labor and inconsistencies of the manual method. Given the difficulty of manually analyzing a textual database of non-conformities, an automated method applying Machine Learning algorithms is proposed. A total of 8,196 records were collected from the Melius Qualidade service management platform, covering twenty-one high-end multifamily residential projects from three construction companies in Goiânia/GO, of which 3,598 were considered suitable for this research after filtering. The initial classification was done manually, followed by applying nine Machine Learning algorithms by using the Orange Data Mining software for testing and evaluation. Results indicated that grouping data by company yielded the best prediction accuracy, while the Neural Network model achieved a recall of up to 98.20%, making it the most effective. The study highlights that automation accelerates the classification process and improves precision and consistency in identifying making-do waste, significantly contributing to quality management in construction projects.

Keywords: Making-do. Machine Learning. Automation. Neural Network. Imbalanced Data.

Resumo

O estudo investiga o potencial de uso e o melhor modelo para a classificação automatizada das perdas por making-do em canteiros de obras, com técnicas de Machine Learning (ML), visando reduzir a laboriosidade e as inconsistências do método manual. Diante da dificuldade de análise manual de um banco de dados textual robusto de inconformidades, caracterizado por um volume significativo e variado, propõe-se avaliação de métodos automatizados que empregam algoritmos de ML. Para tanto, foram coletados 8.196 registros da plataforma de gestão de serviços Melius Qualidade, abrangendo vinte e um empreendimentos de alto padrão multifamiliares de três construtoras em Goiânia/GO, dos quais 3.598 foram considerados qualificados para a pesquisa. A classificação inicial foi manual, seguida pela aplicação de nove algoritmos de ML no software Orange Data Mining para teste e avaliação classificatória dos dados de entrada. Os resultados indicaram que o agrupamento dos dados por empresa possui melhor ajuste preditivo, com o modelo de Rede Neural atingindo um recall de até 98,20%, destacando-se como o mais eficaz. O estudo evidencia que a automação agiliza o processo de classificação, e aumenta a precisão e consistência na identificação das perdas por making-do, contribuindo para a otimização da gestão da qualidade em obras.

Palavras-chave: *Making-do. Machine Learning. Automação. Rede Neural. Balanceamento de dados.*

[1]Tatiana Gondim do Amaral
[1]Universidade Federal de Goiás
Goiânia – GO - Brasil

[2]Gabriella Soares de Paula
[2]Universidade Federal de Goiás
Goiânia – GO - Brasil

[3]Caio César Medeiros Maciel
[3]Universidade Federal de Goiás
Aparecida de Goiânia – GO - Brasil

Recebido em 12/03/25
Aceito em 06/09/25

AMARAL, T. G. do; PAULA, G. S. de; MACIEL, C. C. M. A classification method for making-do waste using *Machine Learning*. **Ambiente Construído**, Porto Alegre, v. 25, e146288, jan./dez. 2025.
ISSN 1678-8621 Associação Nacional de Tecnologia do Ambiente Construído.
http://dx.doi.org/10.1590/s1678-86212025000100927

Introduction

The construction sector is responsible for the infrastructural and physical development of countries and plays a key role in determining the economic equilibrium of nations, therefore, there is a constant need to develop advanced managerial methods for carrying out construction projects (Lekan *et al*., 2020).

Driven by the evolution of Construction 4.0, which entails introducing disruptive technological innovations to increase construction productivity, a movement has emerged to develop an approach aimed at improving construction process management. Disruptive technological innovations refer to the arrival of new technologies or services that fundamentally change how a market or industry operates, typically offering more accessible, simpler, and more effective solutions. These innovations can transform products, services, and processes, paving the way for new business models and consumption patterns (Lekan; Clinton; Owolabi, 2021).

Thus, researchers and practitioners are turning their attention to philosophies, methods, and tools that can enhance production performance, such as Lean philosophy. A key concept within this framework is measuring production waste, originally proposed by Ohno (1997) in his book about "seven wastes". Considering the context of the construction industry, this categorization was expanded by Koskela (2004), who proposed the eighth category of waste, known as making-do, which occurs when tasks are carried out without all the necessary resources. According to Koskela (2004), this type of waste is critical, as it can trigger other forms of waste throughout the process.

Formoso *et al*. (2011), Kalsaas (2012), Fireman and Formoso (2013), Leão, Formoso and Isatto (2014), and Saurin and Sanches (2014) reported difficulties in identifying and classifying making-do waste due to the complexity of the data collected, which varies depending on the research objectives and management practices (Pikas; Sacks; Priven, 2012; Fireman and Formoso, 2013). These data are categorized as nominal, ordinal, and textual. Nominal and ordinal data describe the project and its progress, while textual data, such as descriptions of waste, provide essential qualitative information.

The manual analysis of these data, especially textual, is labor-intensive and prone to subjective interpretations and inconsistencies, making it difficult to standardize and automate the classification process. This is an essential step for identifying the causes of waste and their impact on project performance (Medeiros, 2024). Machine Learning enables computer systems to learn to identify patterns and classify data without the need for predefined programming for each specific situation. Instead of manually defining rules embedded in a pre-written algorithm, thresholds and patterns are extracted directly from data (Flach, 2012), making the classification process faster, more dynamic, and suitable for automation.

This study addresses a knowledge gap on this topic by demonstrating that Machine Learning techniques can be applied to the making-do waste classification, offering an automated solution to perform tasks that previously required human intervention, which are more prone to biases inherent in non-automated interpretations. Furthermore, it facilitates the application of these technologies to manage other types of waste, such as repairing construction defects. In addition, this study advances the field by conducting a comparative evaluation of different ML models, allowing for the identification of those most suitable for classifying *making-do* waste among the models analyzed.

Therefore, this research study investigates the potential application of Machine Learning techniques for the automated making-do waste classification on construction sites, aiming to reduce the labor and inconsistencies compared to manual methods carried out by humans. To achieve this, different classification algorithms were tested and compared in terms of their performance, aiming to identify the models best suited to this application within the context of the construction industry.

Background and theoretical framework

Definition of making-do

Even before the development of the Toyota Production System (TPS), both Taylorism and Fordism recognized that waste in industry was not limited to material waste but also involved labor efficiency. However, Lean thinking expands this understanding by emphasizing the importance of process efficiency as a whole, aiming for the systematic elimination of all types of waste throughout the production chain (Santos; Santos, 2017).

From this perspective, Koskela (2004) defines the concept of making-do as a form of improvisation in which workers, facing a lack of adequate resources, materials, or information, strive to complete their tasks without causing significant damage. This category of waste was inspired by the work of Ronen (1992), who proposed

the concept of the complete kit, stating that no task should be started without all necessary resources. The complete kit is divided into two types: the input kit, which contains the resources required to begin the task, and the output kit, which corresponds to the final result of the preceding task.

Causes of making-do waste

Koskela (2000), Sommer (2010), and Fireman and Formoso (2013) identified prerequisites, categories of waste, and impacts related to the concept of making-do. Koskela (2000) highlighted seven prerequisites, while Sommer (2010) expanded this analysis to include eight prerequisites, seven categories of waste, and six impacts. Fireman and Formoso (2013) further complemented these findings, totaling eight categories of waste and seven impact items, as shown in Figure 1.

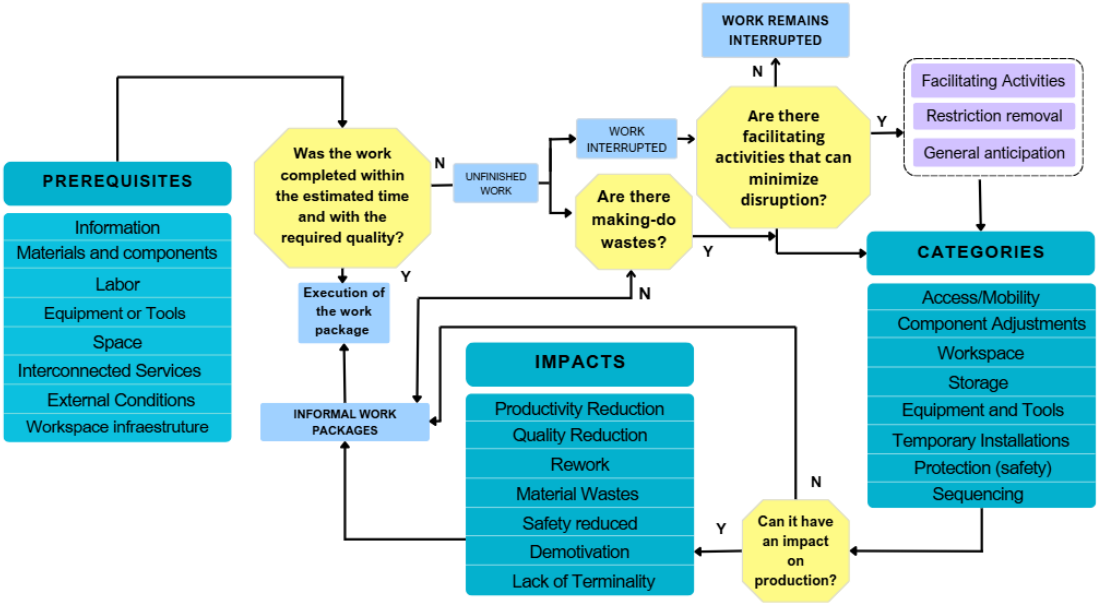
Figure 1 illustrates the classification method for making-do waste, beginning by verifying the prerequisites for task execution. When all conditions are met, the task is carried out, followed by an assessment of whether it was completed on time and with the expected quality. If not, it is determined whether waste due to improvisation (making-do) occurred and whether mitigating measures are available, such as removing constraints or rescheduling tasks. If there are no such measures, the work cannot proceed. Simultaneously, categories impacting the progress of activities, such as access, mobility, storage, tools, and protection, are analyzed. Ongoing unfinished work without mitigation can lead to low productivity, quality failures, rework, material waste, reduced safety, team demotivation, and lack of task completion.

Thus, the complexity involved in analyzing and identifying waste is clear. This difficulty underscores the need to improve the methods used by developing more precise and comprehensive analyses. The classification of the same type of waste may vary depending on the expert's or author's interpretation, highlighting the importance of a more refined and rigorous approach to understanding and qualifying making-do waste (MEDEIROS, 2024).

Machine learning fundamentals

According to Flach (2012), Machine Learning (ML) is a branch of artificial intelligence that enables computer systems to recognize patterns, make predictions, and make decisions autonomously, without the need for explicit programming that defines rules or boundaries. The same author explains that this approach allows systems to learn from the data they receive. This technique operates by analyzing a dataset where input examples (data to be processed) and output examples (expected classifications) are provided. These examples are used to learn the function that maps inputs to outputs. Once trained, a system can make predictions or classifications on new, unseen data by applying the knowledge gained during the learning process (Zhang *et al.*, 2017).

Figure 1 - **Classification method for making-do waste**



Source: adapted from Santos and Santos (2017).

Linear Regression and Logistic Regression are among the most prominent classical methods. Linear Regression is used to predict continuous values by establishing a linear relationship between independent variables and the dependent variable. By contrast, Logistic Regression is used for classification tasks, enabling the prediction of the probability of an event occurring; typically, a binary outcome (e.g., yes or no) (Amaral *et al*., 2024).

Moreover, more advanced methods, such as Deep Neural Networks and Support Vector Machines (SVM), offer greater complexity and the ability to handle nonlinear data. Deep Neural Networks consist of multiple layers of artificial neurons that enable the model to learn complex representations of the data, making them effective for tasks such as image recognition and natural language processing (Zhang *et al*., 2017). Support Vector Machines, in turn, separate data into different classes using hyperplanes and are especially useful for high-dimensional datasets (Hong, 2023). The main practical difference between these methods lies in the specific algorithms that are used to learn the function that maps inputs to outputs. The choice of algorithm and its tuning directly impact prediction accuracy, as some algorithms are better suited to certain data characteristics than others (Zhang *et al*., 2017).

Furthermore, the ability of methods to explain the parameters behind a prediction varies considerably. Some models, such as Linear Regression, offer clear and straightforward interpretations of coefficients, allowing for an understanding of how each variable influences the prediction (Medeiros, 2024). In contrast, more complex models, such as Deep Neural Networks are often considered "black boxes," in which the relationship between inputs and outputs is difficult to interpret (Gonzalez, 2018). This distinction is crucial when selecting a method, depending on the need for explanation versus predictive accuracy.

Research method

This research adopts a quantitative and empirical method with a descriptive focus, incorporating both exploratory and experimental aspects. This approach allows for the investigation of phenomena that have not yet been widely studied or understood, adopting experimental techniques to analyze cause-and-effect relationships. This method enables not only the formulation of initial hypotheses on the subject but also the manipulation of data to examine their interactions and impacts. The study was divided into four different steps: data collection; manual classification; ML-based classification; and evaluation of the results.

Data collection

A total of 8,196 non-conformance reports were extracted the Melius Quality service and materials management platform. Key features of this platform relevant to this study include the ability to standardize inspection checklists, detailed logging of occurrences with photos and descriptions, categorization by service type or responsible team, and monitoring of corrective actions taken. As these are secondary data, originally collected for operational purposes rather than research, there is a recognized limitation regarding potential inconsistencies, incomplete records, or lack of standardization in the descriptions.

The data were collected in twenty-one projects from three different construction companies that use this platform, located in Goiânia, GO. In this research study, only non-conformance reports containing a detailed and complete description, immediate correction, and corresponding corrective action were considered. Thus, out of the total data extracted from the Melius platform, 3,598 non-conformances were effectively analyzed in this study.

The three companies have a large scale of operation, being engaged in medium- to high-end projects, with approximately 30 years of experience in the market. All companies had well-established quality management systems and hold key quality certifications, such as PBQP-H, NBR ISO 9001 (ABNT, 2015a), and NBR ISO 14001 (ABNT, 2015b).

Table 1 presents the main activities (steps) involved, and the amount of data collected and effectively used in this investigation.

Manual processing and classification

The processing step started with manual procedures carried out by the research team, in which data were divided by company and classified according to team, phase, subphase, activities, prerequisites, categories, and impacts. During that phase, information was filtered, inconsistencies were corrected, and irrelevant data that could interfere in the analysis were removed.

In the following step, the data were processed and prepared to meet the specific requirements of the ML algorithm, in order to allow the algorithm to correctly interpret the data and accurately identify patterns. Table 2 presents a detailed organization of the data extracted from the Melius Quality platform, followed by their classification using a spreadsheet.

ML classification

Table 3 presents four essential columns for classification in the ML process.

Table 1 - **Project characteristics and data qualification**

Project	Step	Overall progress	Data collected	Qualified data	Efficiency
C-E2	Installations and finishes	99.29%	613	601	98.04%
C-E3	Plastering and subfloor, installations and finishes	92.00%	522	439	84.10%
C-E4	Completed	100.00%	600	489	81.50%
H-E3	Superstructure	42.96%	286	27	9.44%
H-E4	Superstructure	50.61%	3670	550	14.99%
H-E5	Superstructure	11.72%	274	34	12.41%
H-E6	Infrastructure	10.95%	35	25	71.43%
H-E7	Superstructure	20.52%	20	0	0.00%
H-E8	Completed	100.00%	40	0	0.00%
H-E9	Infrastructure	7.62%	1	0	0.00%
H-E10	Excavation	0.06%	4	2	50.00%
H-E11	Completed	100.00%	1253	699	55.79%
H-E12	Completed	100.00%	41	24	58.54%
H-E13	Completed	100.00%	9	7	77.78%
S-E1	Completed	100.00%	59	57	96.61%
S-E2	Completed	100.00%	154	139	90.26%
S-E3	Finishes	97.00%	323	300	92.88%
S-E4	Excavation and foundation	5.40%	78	59	75.64%
S-E5	Superstructure and masonry	52.00%	154	87	56.49%
S-E6	Excavation	0.63%	2	2	100.00%
S-E7	Completed	100.00%	58	57	98.28%
TOTAL			**8.196**	**3.598**	**43.90%**

Table 2 - **Data organization structure**

Collection						Classification						
Service	Step	Status	Criteria	Problem	Correction	Team	Step	Substep	Activities	Prerequisite	Category	Impact

Table 3 - **Example of input data**

Making-do waste	Operational procedure	Problem	Immediate correction	Merge
The lack of clear project information can lead workers to make improvised decisions.	Check the structural design for the specified information regarding slab execution.	Difficulty interpreting the structural project regarding the level at the interface between the slab and the staircase.	Meetings to align project construction decisions.	Check the structural project for the specified information regarding the slab execution. If there is any difficulty interpreting the structural project, particularly concerning the level at the interface between the slab and the staircase, hold meetings to align construction-related project decisions.

The Operational Procedure column details the specific characteristics and descriptions of a task or activity. The Problem column documents observed errors or failures, identifying deviations from the planned process. The Immediate Correction column defines actions taken promptly to address the identified problems, mitigating the impact of the failure. Finally, the Merge column combines the information from the previous three columns, presenting a consolidated version of all details related to the item. This Merge column is useful for ML processes, as it contains a complete and self-explanatory summary of each occurrence analyzed.

Next, the classification steps and method evaluation were carried out using Orange Data Mining, a free and open-source platform dedicated to machine learning and data visualization. This software allows users to intuitively create graphical representations, as well as train, validate, and compare different ML algorithms. Orange Data Mining operates with a visual structure based on modules called widgets. Each widget performs a specific function, has its own inputs and outputs, and can be interconnected with other widgets to execute tasks such as data processing and visualization (Janez *et al*., 2013).

Figure 2 illustrates the workflow of Orange Data Mining in this study, which begins with data Import through the File widget. Relevant columns are selected by using the Select Columns widget, and the Data Sampler splits the data into 80% for training and 20% for testing. The Corpus widget counts the words in the texts.

Next, the Preprocess Text widget performs cleaning and tokenization to segment the text into word units, as illustrated in Figure 3. During this process, the software automatically applies typical text mining activities such as converting uppercase letters to lowercase, removing accents, punctuation, and special characters, as well as eliminating stop words and performing basic text normalization, following the tool's default settings.

Next, Document Embedding converts the texts into vectors. Figure 2 illustrates the use of the K-Nearest Neighbors (KNN) model. However, to evaluate the best models, this study opted to include eight additional models compatible with the data format: Naive Bayes, Random Forest, Decision Tree, Logistic Regression, Neural Network, AdaBoost, Stochastic Gradient Descent (SGD), and Constant. The ML models are configured and evaluated using the Test and Score widget, which performs cross-validation and compares their performance.

Figure 2 - **Orange data mining workflow**

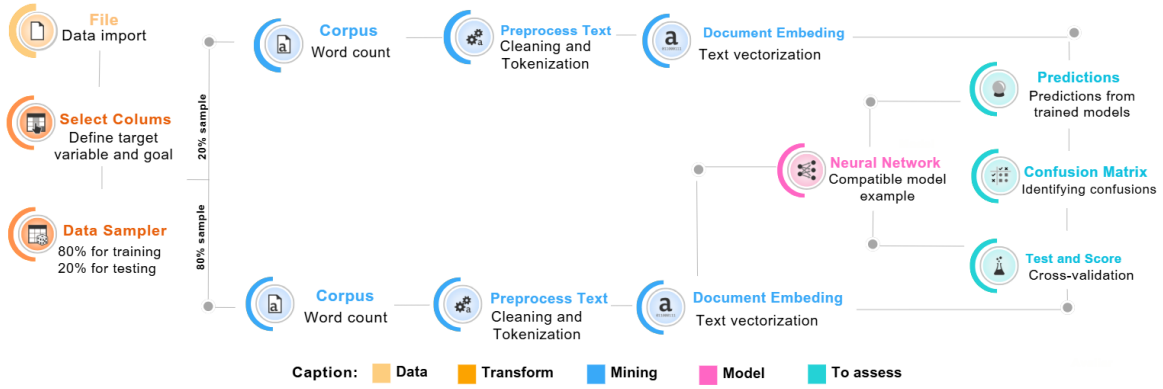
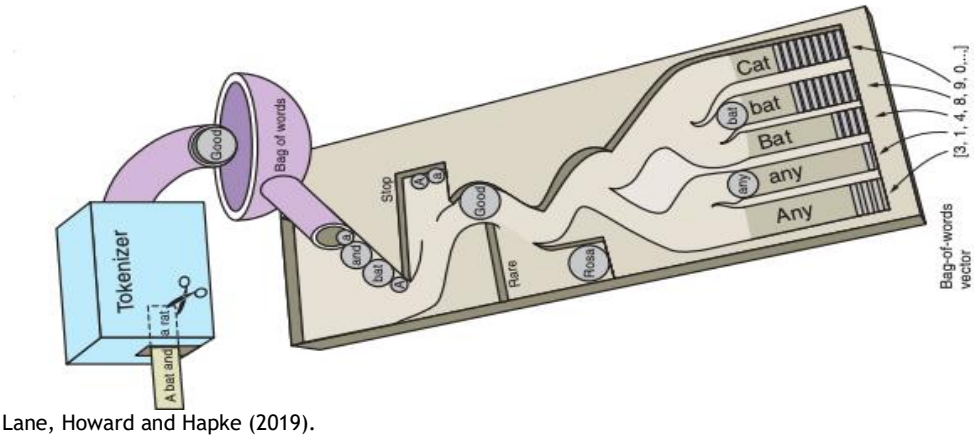


Figure 3 - **Textual data preprocessing**



Source: Lane, Howard and Hapke (2019).

Finally, the Predictions widget is used to generate predictions from the trained models. It takes both the processed data and the trained models as input and outputs the predicted classifications for each instance. Additionally, to represent the correlation between the reference data and the classified data, a confusion matrix can be employed (Prina; Trentin, 2020), which is used during the evaluation step.

Relevant performance metrics

To address imbalanced data, it is important to select relevant performance metrics, considering that a high number of true positives (cases that genuinely belong to the positive class and are correctly classified as such) can mask the presence of many false positives (instances where the model classifies an example as positive even though it belongs to the negative class), and vice versa. A recommended approach for model evaluation is to use the following metrics: Area Under the ROC Curve (AUC); recall; and precision. The AUC measures the model's ability to distinguish between classes and is robust to variations in data distribution. An AUC of 1.0 indicates a perfect classifier, and an AUC above 0.5 is considered acceptable (Maione, 2020).

In this study, recall was adopted as the primary metric for evaluating the models and input data, while AUC was used as a criterion to establish the minimum acceptable performance threshold. Recall measures the model's ability to correctly identify examples belonging to the positive class and is calculated as the ratio of true positives to the total actual positives. Precision, on the other hand, represents the proportion of examples classified as positive that truly belong to the positive class, thus indicating the accuracy of the model's positive predictions.

Model analysis

Using the dataset previously tested by the ML models, the data were imported into Microsoft Power BI to generate interactive graphical analyses. For data visualization, three types of charts were selected: clustered column chart; area chart; and stacked area chart, to provide a clear and effective visual interpretation of the analyzed information.

The study used a dataset segmented by companies, identified by the letters H, C, and S, with 1,368 (38.02%), 1,529 (42.5%), and 701 (19.48%) records, respectively. Additionally, a dataset labeled U was considered, representing the consolidation of records from these different companies. This approach analyzed both the individual performance of each company and the results obtained from combining the data into a single group.

Results and discussion

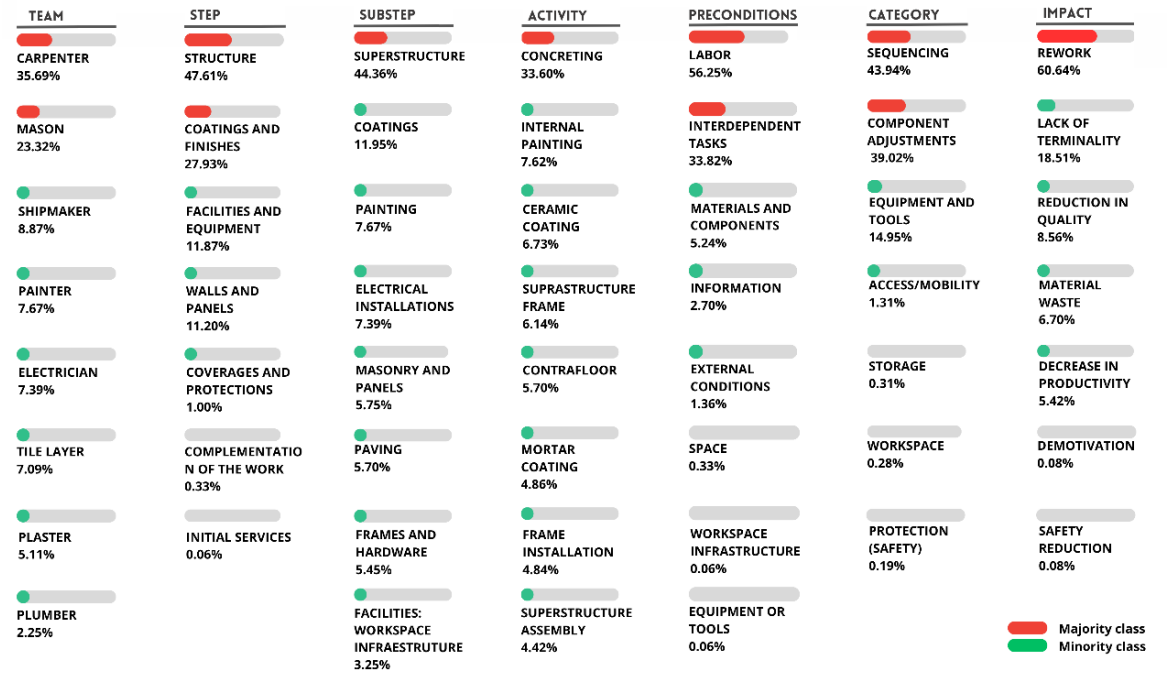
Input and output data quality

Figure 4 presents the results obtained from the manual classification of the data recorded in the database. This step required considerable effort, particularly due to the need to read, interpret, and individually categorize each report. It was not possible to automate this process using spreadsheet formulas, as each description of non-conformity required a line of reasoning based on the specific context of the occurrence. The information was not standardized and, in many cases, involved subjective descriptions or those open to multiple interpretations, which required careful human analysis to ensure accurate data classification.

Analyzing the distribution of classes, a statistically significant imbalance is clear, characterized by the predominance of certain classes and the underrepresentation of others. This imbalance is particularly notable in the input variable "impact," where the highest concentration of reports falls under the class "rework," accounting for 60.64%. In contrast, other classes, such as lack of completion, quality reduction, and material waste, show significantly lower frequencies, at 18.61%, 8.56%, and 6.70%, respectively.

This scenario characterizes a data imbalance problem, in which one class contains a significantly larger number of samples compared to the other six possible classes. This imbalance can lead the classification model to exhibit a bias toward the majority class, resulting in a high accuracy rate for that class but poor performance for the minority classes (Maione, 2020). This happens because algorithms tend to learn dominant patterns from the majority data, while ignoring or misclassifying underrepresented classes. As a result, the model may show seemingly high accuracy metrics that do not truly reflect its ability to accurately predict all classes, especially the minority ones. This undermines the overall effectiveness of the classification, leading to skewed results and impairing the correct identification of important events associated with the minority classes.

Figure 4 - **Class frequency by input data**



Overall performance

Figure 5 shows that the best performances were achieved by companies H and C, with recall rates of 98.20% and 97.70%, respectively. On the other hand, group U, which represents the data unification, and company S showed slightly lower recall rates, at 96.50% and 95.00%, respectively. These results indicate that segmenting the data by company led to better classification performance, highlighting the importance of considering the specific characteristics of each group to enhance the model's accuracy. Each company may have its own patterns of data entry, vocabulary used to describe non-conformities, service execution methods, and different team practices, all of which directly influence the distribution and characteristics of the data. By segmenting the data, the model works with more homogeneous datasets, reducing internal variability and making it easier to identify relevant patterns, thereby increasing the accuracy and reliability of predictions.

The stacked area chart presented in Figure 6 illustrates the overall performance trends of the models for each input data set. Among the models analyzed, the Neural Network achieved the best overall performance, with recall rates of 98.20% for the Stage and Impact classes, and 97.70% for the Prerequisite class. Next, the SGD model stood out for its consistent performance, reaching recall rates of 98.20% for the Team class, 97.80% for Step, and 96.70% for both Substep and Prerequisite. The kNN model also showed notable results, with recall rates of 96.70% for Stage and 94.10% for the Prerequisite and Substep classes. Finally, the Random Forest model delivered solid performance, with recall rates of 95.70% in Prerequisite, 93.80% in Stage, and 92.30% in Team. These results demonstrate the effectiveness of these four models across various data classes, with the Neural Network showing a slightly superior overall performance.

The satisfactory performance observed in both the Neural Network and SGD models can be attributed to their learning and optimization characteristics. The comparison between models reveals that those capable of better capturing nonlinear relationships and patterns in the data tend to perform better. Neural Networks, inspired by the functioning of the human brain, are designed to recognize patterns and perform complex tasks such as classification and prediction. Their layered structure of interconnected neurons enables them to process information in a way similar to communication between biological neurons (Witten *et al*., 2017).

According to Braga, Carvalho and Ludemir (2000), the main advantage of Neural Networks lies in their ability to learn from examples by continuously adjusting the weights of the connections between neurons to improve their predictions. This constant adjustment process enables the network to adapt to different scenarios and data, which explains its strong performance across various classes in the study. On the other hand, SGD stands out for its efficiency in optimizing machine learning models, as it operates by iteratively adjusting parameters—correcting neuron weights with each example or small batch of data.

Figure 5 - **Performance by data group**

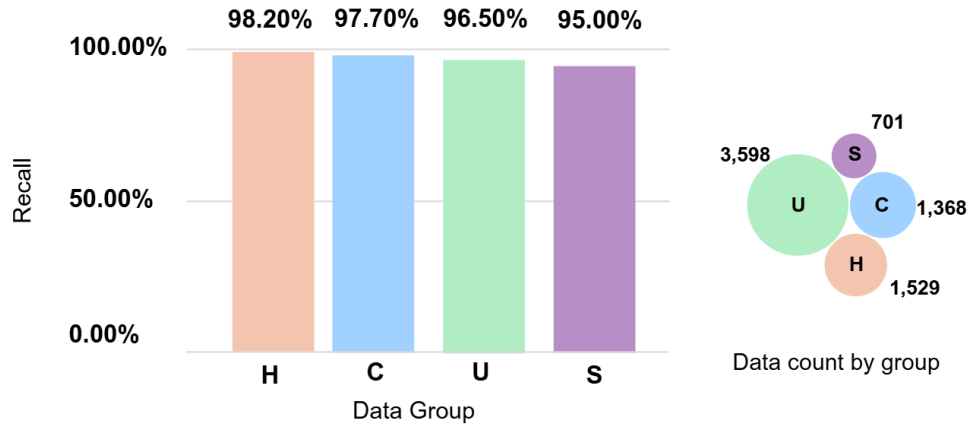
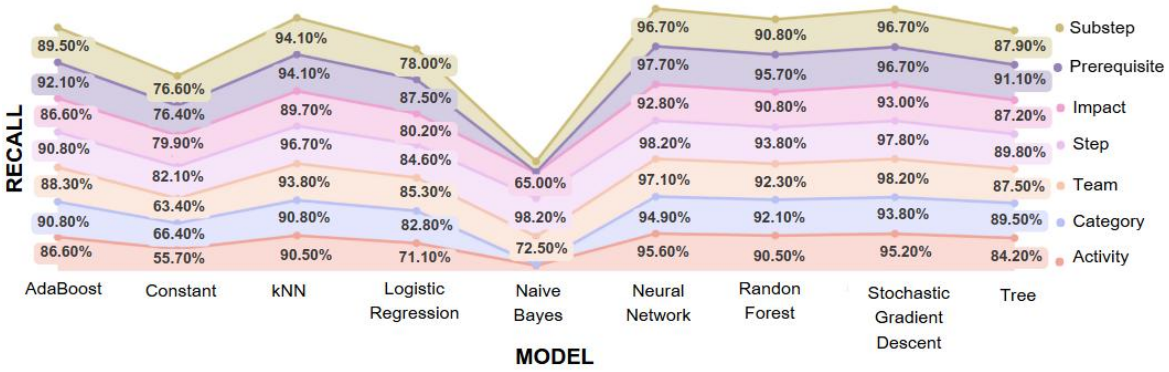


Figure 6 - **Model performances by input data**



In contrast, models such as Naive Bayes and Constant showed significantly poorer performance. The Naive Bayes model achieved a recall of 98.20% for only one type of input data, related to the Step class. Meanwhile, the Constant model reached a recall of 82.10%, highlighting its inefficiency compared to the other tested models. These results emphasize the limitations of these two models in the context of the classification analyzed.

The lower performance of the Naive Bayes and Constant models can be explained by their inherent characteristics. The Naive Bayes algorithm makes two simplifying assumptions: that features are independent of each other and that all features have equal importance (Bužic; Dobša, 2018). However, in this study, the input data features are interdependent and have varying degrees of importance. This simplification likely led to a loss of model accuracy, as it was unable to correctly capture interactions between variables, resulting in lower recall.

Moreover, the Constant model inherently produces very limited results as it always predicts the same value regardless of the input data, using the most common class from the training data for classification. This simplistic approach prevents the model from adapting to the variations and complexities of the data.

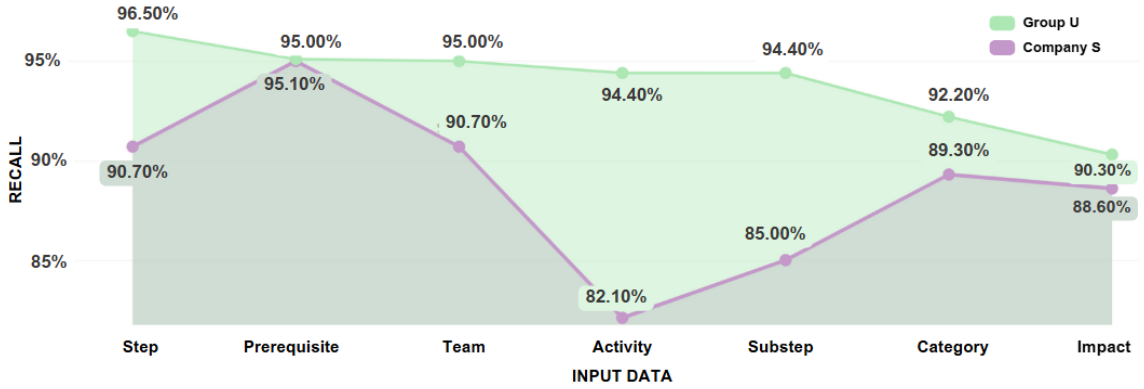
Acceptable recall by data group

The minimum recall threshold for each data group was defined based on the minimum acceptable performance according to the AUC metric, which considers values above 0.5 as an appropriate reference (Maione, 2020). Based on the results obtained, Table 2 establishes the expected minimum recall thresholds per group, defined according to their respective AUC values. This threshold serves as an evaluation criterion to determine whether the model demonstrates satisfactory performance on the given data set.

Table 2 - **Minimum recall threshold by data group**

GROUP	MÍN AUC≥0.5 (Maione, 2020)	MIN LIMIT OF *RECALL*
H	97.2%	90.30%
C	97.5%	92.8%
S	91.4%	85.7%
U	92.8%	91.9%

Figure 7 - **Recall by input data from company S and group U**



Critical performances

When analyzing the performance of the models based on the different input data from company S and group U, it was necessary to understand the underlying causes influencing these results. Figure 7 illustrates the recall values for the various datasets, highlighting that group U achieved the best performance, particularly in the categories Step, Prerequisite, and Team, with recall rates of 96.50%, 95.10%, and 95.00%, respectively. In contrast, company S showed lower recall rates, with Prerequisite at 95.00%, and both Step and Team at 90.70%

The greater sample variability in group U has a statistically significant impact on classification efficiency, as it reduces learning bias and enhances the model's ability to identify complex patterns in the data. This increase may therefore be associated with the clustering of minority classes, which raised their frequency and improved the model's generalization, mitigating the effects of class imbalance.

In contrast, the smaller sample size of company S (Figure 5) contributes to its lower performance compared to companies with larger datasets. Each company's sample size directly impacts the model's performance, as fewer records—combined with significant class imbalance—provide the model with fewer learning examples. As a result, the model tends to perform worse for companies with smaller datasets, leading to lower recall values.

Confusion matrix

To specifically understand the prediction errors, the confusion matrix was used as an analytical tool. As shown in Table 4, this matrix displays the correct classifications along the main diagonal, while the other values represent the areas where the model fails, highlighting the points with the greatest confusion in the predictions.

The sample was selected from the data of company S, which showed the poorest performance. This choice enables a more in-depth analysis of the factors that contributed to the low performance, focusing on critical errors. The chosen target variable was Step, as it had the lowest recall among the input data analyzed, indicating greater difficulty for the model in classifying it correctly. The Neural Network, SGD, and kNN models were selected due to their stronger learning capabilities, especially in contexts with imbalanced data. This approach allows the sample to be narrowed down in order to specifically investigate the most relevant errors.

Table 5 illustrates the results distributed in the confusion matrix. The analysis of the models reveals that the Neural Network achieved the highest number of correct classifications in the Coatings and Finishes class (256 correct) and Structure class (121 correct). SGD also stood out in these two classes, with 260 correct in Coatings and Finishes and 121 correct in Structure. Meanwhile, kNN recorded 240 correct in Coatings and Finishes and 118 correct in Structure, which was the model with the lowest number of correct classifications in these categories compared to the others.

Table 4 - **Confusion matrix from a binary classification model**

Real/Prediction	Prediction	
Real	TP True positive	FP False Positivo
	FN False Negative	TN True Negative

Table 5 - **Confusion matrix for the target variable Step**

-	Real/Prediction	Coverings and protection	Complementation of the work	Structure	Installations and equipment	Walls and panels	Coatings and finishes	∑
NeuralNetwork	Coverings and protection	2	0	0	0	0	0	2
	Complementation of the work	0	8	0	0	0	0	8
	Structure	0	0	121	0	0	0	121
	Installations and equipment	0	0	0	95	0	0	95
	Walls and panels	0	0	0	0	75	0	75
	Coatings and finishes	0	0	0	0	4	256	260
	∑	2	8	121	95	561	256	
SGD	Coverings and protection	2	0	0	0	0	0	2
	Complementation of the work	0	8	0	0	0	0	8
	Structure	0	0	121	0	0	0	121
	Installations and equipment	0	0	0	94	0	1	95
	Walls and panels	0	0	0	0	69	6	75
	Coatings and finishes	0	0	0	0	0	260	260
	∑	2	8	121	94	69	267	561
kNN	Coverings and protection	1	0	0	0	0	1	2
	Complementation of the work	0	6	0	0	1	1	8
	Structure	0	0	118	2	0	1	121
	Installations and equipment	0	2	1	87	2	3	95
	Walls and panels	0	0	2	4	66	3	75
	Coatings and finishes	0	0	1	4	15	240	260
	∑	1	8	122	97	84	249	561

Notes: Green: correct predictions. Yellow: low-frequency errors. Orange: moderate-frequency errors. Red: high-frequency errors.

Regarding the errors, the Neural Network presented 4 exclusive errors in Coatings and Finishes, confusing them with Walls and Panels. SGD had 6 confusions between Walls and Panels and Coatings and Finishes, as well as a single confusion between Installations and Equipment and Coatings and Finishes. Meanwhile, kNN showed a broader distribution of errors, which was the most prone to confuse Coatings and Finishes with Walls and Panels, occurring 15 times. Additionally, kNN also presented 4 confusions involving Coatings and Finishes, Walls and Panels, and Installations and Equipment. This distribution of correct and incorrect classifications demonstrates that, although the models perform well in the main classes, confusion between related stages such as Coatings and Finishes and Walls and Panels is a common source of error, with kNN showing the greatest sensitivity to these overlaps.

Following this analysis, the Venn Diagram widget was Applied, which is a tool that performs the sample intersection of predefined datasets. Each circle in the diagram represents a set, and the overlapping areas indicate the intersections, i.e., the elements common to the other sets. This tool was linked to the confusion matrices of the three main models. Figure 8 shows the Venn diagram illustrating areas of exclusive errors (non-overlapping areas) and common errors (intersections). In this study, the primary objective of this analysis is to obtain a comparative view of how each model handles the data and to identify shared failure areas.

Analyzing the common errors among the algorithms revealed that the Neural Network and SGD do not share errors, indicating that their learning approaches are sufficiently different. This suggests that an error made by one model tends to be corrected by the other, making the combination of these two methods in a hybrid model promising for improving classification robustness. On the other hand, the intersection of four errors between the Neural Network and kNN suggests that both failed on the same classes, due to ambiguous or poorly represented features in the data, which complicates class separation. To investigate this hypothesis, the Data Table widget was applied to filter the failure data from the confusion matrix. This revealed that the analysis of shared errors between the kNN and Neural Network models uncovers a critical pattern in how information is structured and interpreted by the algorithms. Both models failed to classify Coatings and Finishes, which, upon deeper examination of the error table, was directly related to how the text merge is described.

It was observed that the four errors shared between the models arise specifically due to the inspection procedure for cleaning. The central difficulty lies in the fact that the description of the cleaning methodology is identical for all services. This prevents the models from extracting sufficiently discriminative information to correctly differentiate each service step. In the case of the Neural Network, the model tends to base its decisions on complex learning patterns, often using the last node of the network as a reference for prediction. Meanwhile, kNN performs classifications based on similarity between samples, considering the nearest neighbors. When the textual description does not provide enough variation, both models end up making the same error, as they fail to find distinctive elements to properly separate the classes.

In addition to these failures, other confusions were observed due to descriptive similarity between the services. Both in the procedures and in the nonconformities, the descriptions for Coatings and Finishes and Walls and Panels exhibit similarity. This issue is exacerbated by the fact that these stages share similar materials and tools, such as mortar, leveling ruler, and face plumb, which are frequently mentioned in both services. Thus, the absence of distinct textual elements for each category contributes to the models' confusion.

Further analysis of the intersection between SGD and kNN reveals a single error, suggesting that these two methods have different weaknesses but share at least one failure in the same way. Since SGD relies on gradient optimization and kNN on the proximity between samples in feature space, this error may be associated with a region where class separability is weak. Tracing this failure in the database showed that the error is once again related to the cleaning item, reinforcing the models' difficulty in correctly distinguishing this category. The traceability of the shared errors is summarized in Table 6.

Figure 8 - **Venn diagram evaluating common errors among the main models**

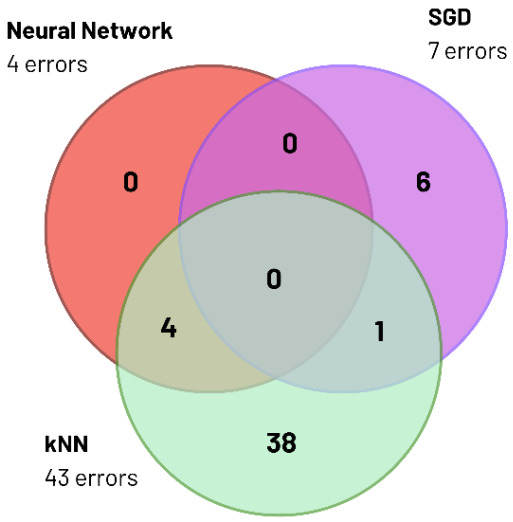


Table 6 - **Traceability of shared errors**

Inter-section	Step (Real)	Merge	Step (Prediction)	Cause
Neural Network and kNN	Coatings and finishings	Check the complete completion of the work, organization, waste segregation, and cleaning of the site, equipment, and tools. Dirtiness. Approved after concession.	Walls and panels	The description of the cleaning methodology is the same for all services.
Neural Network and kNN	Coatings and finishings	Visually inspect if the edges are sharp and use a leveling ruler to check their verticality. Also, verify the squaring. The opening of the suite 02 door is smaller than the specified measurement. Approved after concession.	Walls and panels	Descriptive similarity between the services.
SGD and kNN	Installations and equipment	Verify the full completion of the work, organization, waste segregation, and cleaning of the site, equipment, and tools. Dirtiness. Approved after concession.	Coatings and finishings	The description of the cleaning methodology is identical for all services.

Conclusions

The research work analyses the potential to apply Machine Learning techniques for the automated classification of making-do waste in construction sites, with the aim of reducing labor and inconsistencies that result from manual methods. In addition, it compares the performance of different classification algorithms and identifies the most suitable ones for this application in the context of building construction. The main theoretical contribution is concerned with how to do quantitative analysis of data on making-do waste, as pointed out below.

The classification of *making-do* waste in construction is a challenging task, especially when analyzing an imbalanced database. This study contributes by applying ML techniques to automate the classification process of this waste, achieving promising results by separating the data of each company, which allowed for more accurate performance compared to data grouping.

Using different ML algorithms, such as Neural Network and SGD, proved effective, with recalls reaching up to 98.20%, demonstrating these models' ability to handle complex and interdependent data. The results reinforce the importance of considering the particularities of projects and companies to maximize model accuracy. Data segmentation by company resulted in better performance compared to unifying records from different companies, demonstrating that for practical applications in large companies, model training needs to be tailored according to each construction company's specificities.

The study also identified weaknesses that negatively impact development, such as data imbalance and the use of databases with small sample sizes associated with low variability of examples. The models most sensitive to these limitations were Naive Bayes, which achieved a recall of 98.0% but was limited to a single type of input data, *Stage*. The Constant model did not perform acceptably under any condition, obtaining a maximum recall of 82.10% for the same variable.

It is worth noting that only 43.90% of the total collected data were qualified for the research. This low yield reflects a significant loss of information, mainly caused by the absence of adequate records, as well as incomplete or inaccurate records of nonconformities and corrective actions. The loss of information hinders precise analysis and prevents decisions from being made based on a robust dataset, which may compromise quality management and waste control on construction sites. Bearing this in mind, a possible solution could be to implement an intelligent system in the quality management software that automates the detection of potential nonconformities during inspection processes on construction sites and automatically suggests the correct corrective actions.

Furthermore, the graphical analyses generated by the Orange Data Mining visual widgets proved fundamental for evaluating the performance of the models used in this study. The confusion matrix was essential, allowing a clear visualization of prediction errors across different classes and revealing recurring patterns of confusion among them. Combined with the use of the Venn diagram, intersections of common errors were identified

among the models, showing that the Neural Network and kNN shared failures, especially in categories with similar textual descriptions, such as the Coatings and Finishes steps.

Therefore, this work advances by conducting a comparative evaluation among ML models, allowing the identification of the most suitable ones for classifying making-do waste. The models analyzed included K-Nearest Neighbors (kNN), Naive Bayes, Random Forest, Decision Tree, Logistic Regression, Neural Network, AdaBoost, Stochastic Gradient Descent (SGD), and Constant. The models selected were Neural Network and SGD. It is important to highlight, however, that model accuracy still depends on factors such as the quality (detail and clarity) and quantity of available data, as well as the structure of the texts used in the descriptions of problems and corrective actions.

For future studies, it is recommended to explore the Python Script widget of Orange Data Mining, which allows the insertion of custom algorithms for specific tasks. Creating scripts that perform automatic database balancing could be an effective solution to minimize errors caused by class imbalance, one of the main challenges identified in this study. Furthermore, the inclusion of new variables, such as the financial impact of making-do waste, is suggested to provide a more comprehensive view of the effects of this waste on the overall performance of projects.

References

AMARAL, T. G. *et al.* Modelo de classificação das perdas por making-do com uso de aprendizado de máquina. In: ENCONTRO NACIONAL DE TECNOLOGIA DO AMBIENTE CONSTRUÍDO, 20., Maceió, 2024. **Anais [...]** Maceió: ANTAC, 2024.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR ISO 14001**: sistemas de gestão ambiental: requisitos com orientações para uso. 3. ed. Rio de Janeiro, 2015b.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR ISO 9001**: sistema de gestão da qualidade: requisitos. 3. ed. Rio de Janeiro, 2015a.

BRAGA, A. P.; CARVALHO, A. C. P. L. F.; LUDEMIR, T. B. **Redes neurais artificiais**: teoria e aplicações. Rio de Janeiro: LTC, 2000.

BUŽIC, D.; DOBŠA, J. Lyrics classification using naive bayes. In: INTERNATIONAL CONVENTION ON INFORMATION AND COMMUNICATION TECHNOLOGY, ELECTRONICS AND MICROELECTRONICS, 41., Opatija, 2018. **Proceedings [...]** Opatija: IEEE, 2018.

FIREMAN, M. C. T.; FORMOSO, C. T. Integrating production and quality control: monitoring *making-do* and unfinished work. In: ANNUAL CONFERENCE OF THE INTERNATIONAL GROUP FOR LEAN CONSTRUCTION, 21., Fortaleza, 2013. **Proceedings [...]** Fortaleza, 2013.

FLACH, P. **Machine learning**: the art and science of algorithms that make sense of data. Cambridge: Cambridge University Press, 2012.

FORMOSO, C.T. *et al.* An Exploratory study on the measurement and analysis of *making-do* in construction sites. In: In: ANNUAL CONFERENCE OF THE INTERNATIONAL GROUP FOR LEAN CONSTRUCTION, 19., Lima, 2011. **Proceedings [...]** Lima, 2011.

GONZALEZ, L. **Regressão Logística e suas Aplicações**. São Luis, 2018. 46 f. Monografia (Especialização) - Curso de Ciência da Computação, Universidade Federal do Maranhão, São Luis, 2018.

HONG, H. **Machine learning and deep learning in computational**: toxicology. Cham: Springer, 2023.

JANEZ, D. *et al.* Orange: data mining toolbox in python. **Journal of Machine Learning Research**, v. 14, p. 2349–2353, 2013.

KALSAAS, B. T. Further work of measuring workflow in construction site production. In: ANNUAL CONFERENCE OF THE INTERNATIONAL GROUP FOR LEAN CONSTRUCTION, 20., San Diego, 2012. **Proceedings [...]** San Diego, 2012.

KOSKELA, L. **An exploration towards a production theory and its application to construction**. Espoo, Finlândia: VTT, 2000. (VTT Publication, 408).

KOSKELA, L. Making-do: the eighth category of waste. In: INTERNATIONAL GROUP FOR LEAN CONSTRUCTION ANNUAL CONFERENCE, 12., Elsinore, 2004. **Proceedings [...]** Elsinore, 2004.

LANE, H.; HOWARD, C.; HAPKE, H. M. **Natural language processing in action**: understanding, analyzing and generation text with Python. Manning, 2019.

LEÃO, C. F.; FORMOSO, C. T.; ISATTO, E. L. Integrating production and quality control with the support of information technology. In: ANNUAL CONFERENCE OF THE INTERNATIONAL GROUP FOR LEAN CONSTRUCTION, 22., Oslo, 2014. **Proceedings [...]** Oslo, 2014.

LEKAN, A. *et al.* Lean thinking and industrial 4.0 approach to achieving construction 4.0 for industrialization and technological development. **Buildings**, v. 10, 221, 2020.

LEKAN, A.; CLINTON, A.; OWOLABI, J. The disruptive adaptations of construction 4.0 and industry 4.0 as a pathway to a sustainable innovation and inclusive industrial technological development. **Buildings**, v. 11, n. 3, 79, 2021.

MAIONE, C. **Balanceamento de dados com base em oversampling em dados transformados**. Goiânia, 2020. 135 f. Tese (Doutorado em Ciência da Computação em Rede) - Universidade Federal de Goiás, Goiânia, 2020.

MEDEIROS, C. C. M. **Método de classificação de perdas por *making-do* com uso de *Machine Learning***. Aparecida de Goiânia, 2024. Dissertação (Mestrado Profissional em Engenharia de Produção) – Programa de Pós-graduação em Engenharia de Produção, Universidade Federal de Goiás, Aparecida de Goiânia, 2024.

OHNO, T. **O Sistema Toyota de produção**: além da produção em larga escala. Porto Alegre: Bookman, 1997.

PIKAS, E.; SACKS, R.; PRIVEN, V. Go or No-Go decisions at the construction workface: uncertainty, perceptions of readiness, making ready and making-do. In: ANNUAL CONFERENCE OF THE INTERNATIONAL GROUP FOR LEAN CONSTRUCTION, 20., San Diego, 2012. **Proceedings [...]** San Diego, 2012.

PRINA, B.; TRENTIN, R. GMC: geração de matriz de confusão a partir de uma classificação digital de imagem do ArcGIS. In: GERAÇÃO DE MATRIZ DE CONFUSÃO A PARTIR DE UMA CLASSIFICAÇÃO, 17., João Pessoa, 2020. **Anais [...]** João Pessoa, 2020.

RONEN, B. The complete *kit* concept. **The International Journal of Production Research**, v. 30, n. 10, p. 2457-2466, 1992.

SANTOS, P. R. R.; SANTOS, D. G. Investigação de perdas devido ao trabalho inacabado e o seu impacto no tempo de ciclo dos processos construtivos. **Ambiente Construído**, Porto Alegre, v. 17, n. 2, p. 39-52, abr./jun. 2017.

SAURIN, T. A.; SANCHES, R. C. Lean Construction and resilience engineering: complementary perspectives of variability. In: ANNUAL CONFERENCE OF THE INTERNATIONAL GROUP FOR LEAN CONSTRUCTION, 22., Oslo, 2014. **Proceedings [...]** Oslo, 2014.

SOMMER, L. **Contribuições para um método de identificação de perdas por improvisação em canteiros de obras.** Porto Alegre, 2010. Dissertação (Mestrado em Engenharia) – Programa de Pós-Graduação em Engenharia Civil, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2010.

WITTEN, I. H. *et al.* **Data mining**: practical machine learning tools and techniques. 4. ed. Cambridge: Morgan Kaufmann, 2017.

ZHANG, C. *et al.* An up-to-date comparison of state-of-the-art classification algorithms. **Expert Systems with Applications**, v. 82, p. 128–150, 2017.

Declaração de Disponibilidade de Dados
Os dados de pesquisa só estão disponíveis mediante solicitação.

Tatiana Gondim do Amaral
Conceptualization, Data curation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

Escola de Engenharia Civil e Ambiental | Universidade Federal de Goiás | Av. Universitária, Quadra 86, Lote Área 1488, Setor Leste Universitário, Goiânia – GO – Brasil | CEP 74605-220 | Tel.: (62) 98168-0902 | E-mail: tatianagondim@ufg.br

Gabriella Soares de Paula
Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Validation, Visualization, Writing – original draft.

Escola de Engenharia Civil e Ambiental | Universidade Federal de Goiás | Tel.: (62) 99642-8364 | E-mail: gabriella.soares@discente.ufg.br

Caio César Medeiros Maciel
Data curation, Formal analysis, Methodology, Supervision, Validation.

Programa de Pós-Graduação em Engenharia de Produção | Universidade Federal de Goiás | Estrada Municipal, Quadra e Área Lote 04, Bairro Fazenda Santo Antônio | Aparecida de Goiânia – GO – Brasil | CEP 74971-451 | Tel.: (62) 99393-4718 | E-mail: caiocesar.eng@hotmail.com

Editores: **Carlos Torres Formoso**, **Ariovaldo Denis Granja** e **Dayana Bastos Costa**

Ambiente Construído
Revista da Associação Nacional de Tecnologia do Ambiente Construído
Av. Osvaldo Aranha, 99 - 3º andar, Centro
Porto Alegre – RS - Brasil
CEP 90035-190
Telefone: +55 (51) 3308-4084
www.seer.ufrgs.br/ambienteconstruido
www.scielo.br/ac
E-mail: ambienteconstruido@ufrgs.br