

## Article

# Forecasting Raw Material Yield in the Tanning Industry: A Machine Learning Approach

Ismael Cristofer Baierle <sup>1</sup>, Leandro Haupt <sup>2</sup>, João Carlos Furtado <sup>2</sup>, Eluza Toledo Pinheiro <sup>3</sup>  
and Miguel Afonso Sellitto <sup>3,\*</sup>

<sup>1</sup> Agroindustrial Systems and Processes Graduate Program, Universidade Federal do Rio Grande, FURG, Cel. Francisco Borges de Lima, 3005, Santo Antônio da Patrulha 95500-000, Brazil; ismaelbaierle@furg.br

<sup>2</sup> Industrial Processes and Systems Engineering Graduate Program, Universidade de Santa Cruz do Sul, UNISC, Av. Independência, 2293—Universitário, Santa Cruz do Sul 96815-900, Brazil; lhaupt@mx2.unisc.br (L.H.); jcarlosf@unisc.br (J.C.F.)

<sup>3</sup> Production and Systems Engineering Graduate Program, Universidade do Vale do Rio dos Sinos, UNISINOS, Av. Unisinos, 950—Cristo Rei, São Leopoldo 93022-000, Brazil; eluzpinheiro@unisinos.edu.br

\* Correspondence: sellitto@unisinos.br

**Abstract:** This study presents an innovative machine learning (ML) approach to predicting raw material yield in the leather tanning industry, addressing a critical challenge in production efficiency. Conducted at a tannery in southern Brazil, the research leverages historical production data to develop a predictive model. The methodology encompasses four key stages: data collection, processing, prediction, and evaluation. After rigorous analysis and refinement, the dataset was reduced from 16,046 to 555 high-quality records. Eight ML models were implemented and evaluated using Orange Data Mining software, version 3.38.0, including advanced algorithms such as Random Forest, Gradient Boosting, and neural networks. Model performance was assessed through cross-validation and comprehensive metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and Coefficient of Determination ( $R^2$ ). The AdaBoost algorithm emerged as the most accurate predictor, achieving impressive results with an MAE of 0.042, MSE of 0.003, RMSE of 0.057, and  $R^2$  of 0.331. This research demonstrates the significant potential of ML techniques in enhancing raw material yield forecasting within the tanning industry. The findings contribute to more efficient forecasting processes, aligning with Industry 4.0 principles and paving the way for data-driven decision-making in manufacturing.

**Keywords:** machine learning; digital technologies; prediction model; leather; forecasting



**Citation:** Baierle, I.C.; Haupt, L.; Furtado, J.C.; Pinheiro, E.T.; Sellitto, M.A. Forecasting Raw Material Yield in the Tanning Industry: A Machine Learning Approach. *Forecasting* **2024**, *6*, 1078–1097. <https://doi.org/10.3390/forecast6040054>

Academic Editors: Juan Ramón Traperó Arenas, Konstantinos Nikolopoulos and Vasileios Bougioukos

Received: 10 October 2024  
Revised: 9 November 2024  
Accepted: 18 November 2024  
Published: 20 November 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The industrial landscape has experienced significant changes throughout history, characterized by four distinct revolutions. Beginning with the steam-powered machines of the First Industrial Revolution and progressing to the digital systems of the Third, each era has introduced profound transformations in manufacturing processes. Presently, we find ourselves on the brink of the Fourth Industrial Revolution, known as Industry 4.0, which seeks to integrate communication and intelligence into industrial processes [1]. In this context, artificial intelligence (AI) techniques are increasingly applied in the manufacturing sector, ushering in a new era characterized by efficiency, sustainability, and innovation. AI-based manufacturing provides various advancements, including fault detection and prediction, automation, and process enhancements [2]. Such technologies are pivotal in implementing Industry 4.0 concepts, bringing intelligence to processes, and generating valuable insights from the vast amount of data produced in the current industrial era [3].

The leather industry, with its history dating back to the early days of industrialization, presents interesting opportunities for case study. Leather has been used for centuries to create a wide array of products, from shoes and clothing to material coverings and bags.

The evolution of leather processing techniques has led to significant improvements in quality over time [4]. The leather industry, particularly in Brazil, plays a crucial role in the global market. Brazil boasts the world's largest commercial cattle herd, with 244 tanning plants, 2800 leather and footwear component industries, and 120 machinery and equipment factories. The sector generates about 30,000 direct jobs, exports leather to 80 countries, and moves \$2 billion annually [5]. Despite advancements in processes, techniques, and machinery, the quality of raw materials—specifically hides—remains a challenge. The most common type, bovine hide, often suffers from neglect, as cattle are primarily raised for meat production. This neglect leads to defects that reduce the hide's quality for tanneries [4]. It is important to incorporate new technologies into the production process to address this issue.

Decision-making technologies, particularly those based on machine learning (ML) approaches, may offer significant benefits to companies in the Industry 4.0 era [3]. These technologies enable faster and better decisions, leading to more efficient production, reduced costs, and improved process management. The high integration between business environments generates large amounts of data, which can be leveraged in predictive approaches through regression techniques, neural networks, and other models that use historical data to forecast future scenarios and promote competitiveness [6].

However, there is a gap in scientific studies addressing data-driven organizational culture. While most research focuses on data infrastructure, fewer studies explore how to foster a culture that enables managers to make more precise, data-driven decisions rather than relying on intuition [7]. This gap underscores the need for solutions that improve data-driven decision-making in the industry. ML models have been widely applied in the leather industry, particularly in hide classification and defect detection. Several studies have demonstrated the effectiveness of ML techniques in these areas, achieving high accuracy rates [8–12]. However, other critical aspects of the leather industry, such as leather yield prediction, remain underexplored. Leather yield is the difference between the final and initial leather area and is a crucial indicator of a tannery's performance. The physical and chemical operations involved in tanning cause variations in leather area, either reducing or increasing yield. Leather quality and tanning process parameters influence this variability. Maximizing the use of raw materials is essential to avoid waste and losses in tanneries, making yield prediction a valuable area for ML application.

The purpose of this article is to present a method for data collection to generate a raw material yield prediction model based on machine learning and to validate this method in a leather processing company. This research bridges the gap between advanced ML techniques and practical applications in the leather industry, specifically addressing the challenge of yield prediction. The main novelty of this study lies in its application of ML techniques to predict leather yield, an area that has received limited attention in previous research. By focusing on this critical aspect of leather production, the study contributes to more efficient resource utilization and waste reduction in the tanning industry.

The research method employed is quantitative modeling, utilizing historical data from a tannery's management system to develop and evaluate ML models for yield prediction. This approach allows for a systematic analysis of various factors influencing leather yield and provides a data-driven foundation for decision-making in the production process. Through quantitative modeling, one can expect to uncover patterns and relationships in the data that are not immediately apparent, leading to more accurate yield predictions and potentially revealing insights into factors that significantly impact leather yield. These findings can inform process improvements and optimize resource allocation in tanneries.

This study addresses the urgent need for technological innovation in the leather industry, especially in raw material yield prediction, an area previously underexplored. The contemporary industrial landscape, shaped by Industry 4.0, brings challenges and opportunities for using artificial intelligence (AI) and machine learning (ML) techniques to increase efficiency, reduce waste, and optimize resource use. In a traditional and economically significant sector like leather, the impact of bovine hide quality—often compromised by

meat-focused livestock practices—presents a major obstacle to sustainable production. By applying ML models to predict leather yield, defined as the difference between the initial and final area of treated leather, this study makes a pioneering contribution toward more efficient material usage in tanneries. Its quantitative methodology, based on historical data, provides a solid foundation for decision-making and identifies previously unseen patterns and relationships, building on prior studies focused on classification and defect detection. This work stands out not only for its practical and innovative application of ML techniques but also for fostering a data-driven organizational culture, a critical aspect of digital transformation in the sector.

The article is structured into sections following the introduction. It begins with a detailed methodology section outlining the steps in data collection, processing, and model development. The methodology section is followed by a results section, presenting the outcomes of the machine learning models and their performance metrics. The discussion section interprets these results, comparing them with existing literature and exploring their implications for the leather industry. Finally, the conclusion summarizes the main findings, highlights the study's contributions to the field, and suggests directions for future research. Throughout the sections, the article focuses on the practical applications of machine learning in predicting raw material yields in the tanning industry.

## **2. Theoretical Background**

### *2.1. Industry 4.0*

Industry 4.0 represents a profound evolution in the industrial sector, marked by the integration of digital technologies and advanced automation of production processes aimed at improving efficiency, flexibility, and product customization [13]. This concept encompasses a wide range of emerging technologies, including the Internet of Things (IoT), cloud computing, additive manufacturing, and artificial intelligence (AI), all interconnected and working together to create “smart factories” [14]. Industry 4.0 proposes a production environment where machines and systems communicate and make autonomous decisions, enabling a continuous, optimized flow of data throughout every stage of the production chain. This data-driven digital approach becomes especially relevant in dynamic markets, where agility and efficiency are competitive differentiators.

### *2.2. Artificial Intelligence*

Artificial Intelligence (AI) is one of the main technologies driving Industry 4.0, enabling systems to learn and adapt from historical and real-time data. Within the scope of AI, machine learning (ML) stands out for its ability to analyze and predict through algorithms that learn from data, identifying complex patterns and relationships that would otherwise escape human observation [15]. ML facilitates the application of predictive and prescriptive models to optimize processes, anticipate problems, and provide valuable input for strategic decisions. In the industrial context, ML plays a crucial role in predictive maintenance, quality control, and production process optimization [16].

### *2.3. ML-Based Predictive Models*

ML-based predictive models represent a practical application of machine learning, where algorithms are trained to predict future outcomes based on historical data. These predictive models are used across various sectors to improve efficiency and reduce costs, notably in predictive maintenance in manufacturing, demand forecasting in retail, and quality monitoring in the pharmaceutical industry [17]. In industrial processes, these models are especially valuable as they allow managers to identify trends and anticipate issues, making it possible to adjust operations proactively. ML for forecasting becomes even more relevant when applied to raw material management, where the precision of predictions is essential to minimize waste and optimize resource use [18].

## 2.4. Raw Material Forecasting

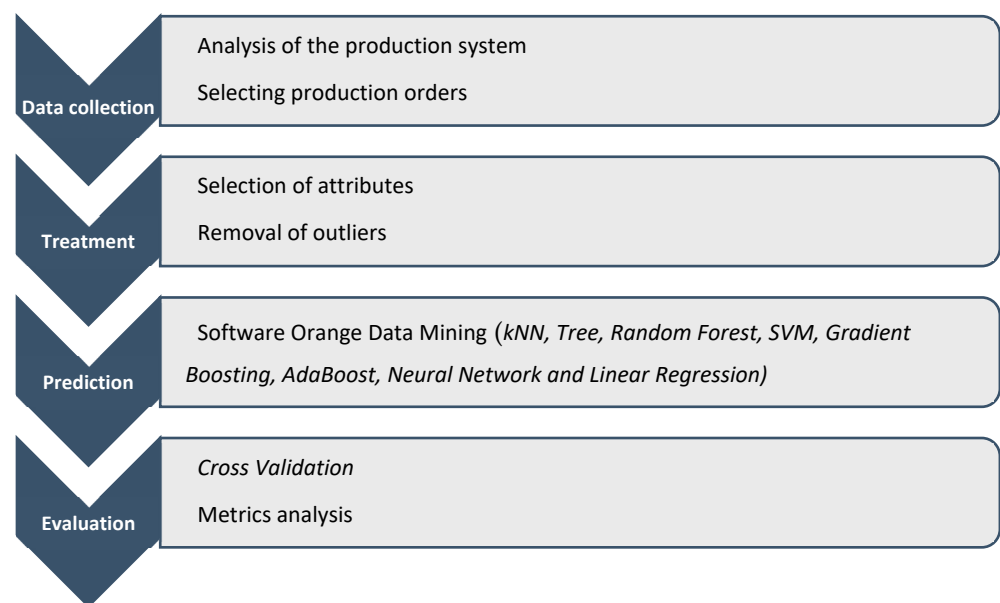
Specifically, raw material forecasting through ML is an emerging application that has gained importance in industries such as tanning, where yield control—the relationship between the final and initial area of processed leather—directly impacts production efficiency and sustainability. In this context, ML-based predictive models help calculate leather yield accurately, considering variables such as material quality, process parameters, and historical data [19]. This approach reduces waste and promotes more efficient resource use, aligning with the data-driven and sustainable production principles of Industry 4.0.

By connecting these concepts, we see that Industry 4.0 and AI converge to transform traditional processes into smarter, more autonomous systems. The use of ML to develop predictive models is a practical, direct application of these concepts, providing a proactive, data-driven approach to raw material management [20,21]. Rather than relying on reactive analyses, where managers respond to problems only after they occur, the predictive analysis enabled by ML allows challenges to be anticipated and resources to be optimized from the start, promoting a more efficient and sustainable production cycle [22].

## 3. Materials and Methods

### *Initial Quantitative Modeling and Computational Resources*

The research initially follows a quantitative approach using historical data to develop and evaluate machine learning models for predicting raw material yield. To first study the process through foundational steps, the methodology is structured into four main phases: data collection, data processing, prediction, and evaluation, as shown in Figure 1. This systematic progression allows for a thorough understanding and establishes a basis that can later be validated through practical application, ensuring robust analysis and reliable results.



**Figure 1.** Initial methodological steps.

In the initial phase, data was gathered from a tannery utilizing a specialized management system designed for such operations. It requires a thorough examination of the system's structure and its interrelated components. The system records Production Orders (POs) alongside operational data and quality control parameters. It soon became apparent that the dataset needed refinement, focusing on a subset of POs featuring more consistent operations. During the data processing phase, a detailed review of relevant operations and parameters was conducted, with outlier records eliminated through the Interquartile Range (IQR) method [23].

Following this reduction, the number of POs dropped significantly from 16,046 to 555. For the prediction phase, the Orange Data Mining software [24] version 3.30.0 was selected for its comprehensive set of built-in machine learning models and user-friendly interface. Orange operates as a visual programming tool for data analysis, machine learning (ML), and data mining. It uses modular components known as widgets, which cover tasks ranging from data visualization and preprocessing to model evaluation and predictive analysis. The software supports users of varying expertise; its graphical interface allows workflows to be built by linking widgets, while advanced users can access Python libraries for more customized data manipulation.

The ML models available in Orange for regression tasks were used, including k-Nearest Neighbors (kNN), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), Gradient Boosting (GB), AdaBoost (AB), Neural Network (NN), and Linear Regression (LR). Validation of these models was conducted using cross-validation [25] and key statistical metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and the Coefficient of Determination ( $R^2$ ). These metrics, described below, provide insights into model accuracy and performance. The software has been configured to use comma as decimal point.

- Linear Regression is a simple ML model that determines a linear function to solve problems. It performs poorly with nonlinear issues and is prone to overfitting [26]. Figure 2a displays the parameters used in Orange for the LR model.

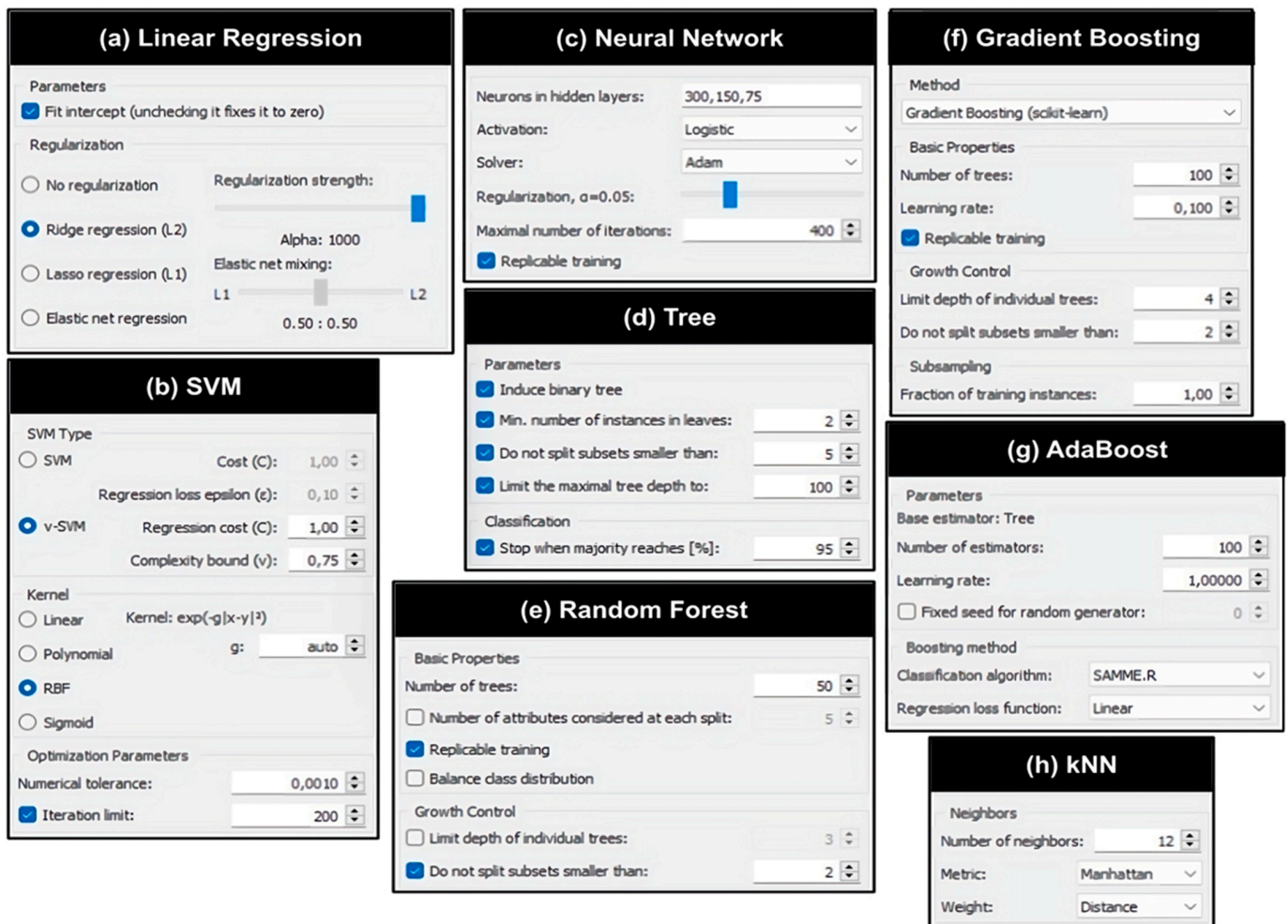


Figure 2. Configuration of parameters used in each model.

- Support Vector Machine is a linear model utilized for classification and regression. Similar to LR, SVM aims to find a function that separates data, with the ability to create hyperplanes, which helps reduce overfitting and improves performance with complex data relationships [27]. Figure 2b shows the parameters used in Orange for the SVM model.
- Neural Network is a supervised learning model for evaluation and classification [28], structured to mimic the human brain. It excels at recognizing nonlinear, incomplete, noisy, and contradictory patterns. A typical NN consists of at least three layers (input, hidden, and output), with neurons connected across layers, where learning occurs by adjusting the weights between neurons [29]. Figure 2c depicts the parameters used in Orange for the NN model.
- A Decision Tree is a flowchart-like structure of decisions and their possible outcomes, starting with a root node that branches into new nodes until reaching a leaf node, which represents a result [30]. Figure 2d illustrates the parameters used in Orange for the DT model.
- Random Forest generates multiple individual Decision Trees, where a majority vote among the trees determines the outcome. Some trees may yield incorrect predictions, but the model's ensemble nature strengthens its accuracy [31]. Figure 2e shows the parameters used in Orange for the RF model.
- Gradient Boosting constructs multiple small Decision Trees, using optimization algorithms to improve each tree based on the performance of the previous one. The final model is a combination of all trees, generating more accurate predictions [32]. Figure 2f outlines the parameters used in Orange for the GB model.
- AdaBoost is an ensemble learning technique that builds models sequentially, where each model focuses on correcting the errors of its predecessor, leading to a more robust and accurate final model [33]. Figure 2g demonstrates the parameters used in Orange for the AB model.
- K-Nearest Neighbor is a supervised learning model for classification and regression that predicts outcomes based on the similarity between data points. It assumes that similar data points are grouped together and relies on distance measurements. However, kNN becomes less efficient with larger datasets and is sensitive to outliers [34]. Figure 2h highlights the parameters used in Orange for the kNN model.

The company currently uses Antara software, developed by SystemHaus [35], a specialized tool designed to meet the unique management needs of tanneries. This software supports various organizational processes by offering essential modules that enhance decision-making and operational efficiency. Although it provides a comprehensive repository of historical data, its current application is limited to reactive analysis. In this setup, managers and specialists review data only after events occur, interpreting and analyzing records to address specific issues. While this approach allows for problem-solving post-occurrence, it misses opportunities to anticipate and mitigate potential risks in advance.

Machine learning (ML) can help shift from reactive to proactive analysis by uncovering complex patterns and relationships within historical data, enabling predictive insights that highlight issues and suggest solutions ahead of time. This proactive approach is central to Industry 4.0 and the broader digital transformation movement, where data-driven methods enhance decision-making and operational foresight. To explore ML's potential for proactive analysis, the company implemented and compared eight ML-based models to identify which one delivered the most accurate predictions for tannery operations. By establishing a standardized method for generating predictive models and analyzing output parameters, the team could determine which model best forecasted and diagnosed future issues, setting the stage for more efficient, forward-thinking management. In Section 4, we present how the normalized method was established to generate predictive models and analyze the output parameters.

#### 4. Results

The study was conducted at a tannery located in southern Brazil, specializing in the production of leather for furniture upholstery. Southern Brazil accounts for approximately 30% of Brazil's leather exports, and Brazil is the world's third-largest exporter of tanned leather. The company processes approximately 70,000 hides per month, with an average yield of 22 square feet per hide. The tannery's production system is organized into three main stages: (i) wet-end, involving chemical processes to prepare the hide for tanning; (ii) tanning, which is the core process that transforms the hide into leather; and (iii) finishing, which includes the final treatments to enhance the leather's appearance and properties. The tannery's management system begins its process upon the receipt of hides, which may be in a raw state or at a more advanced processing stage. Once the hides are received, a batch is generated in the system. This batch serves as a comprehensive record, capturing critical details such as:

- The number of hides included;
- The total weight or area of the hides;
- The condition of the leather;
- The type and origin of the leather;
- The class of leather, among other attributes.

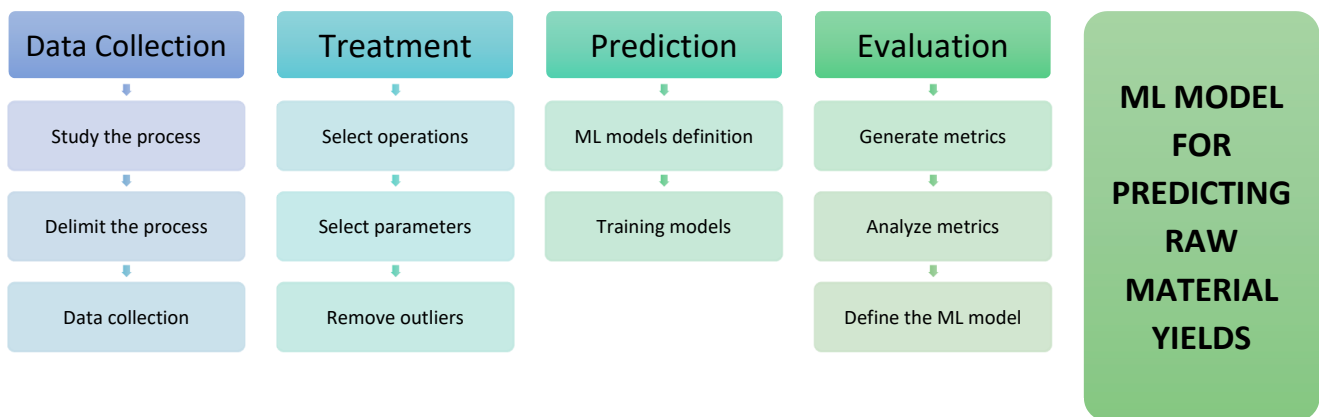
This information provides a thorough overview of the raw materials, ensuring that every batch is fully traceable and properly categorized. From these batches, the system creates volumes, which are essentially virtual records corresponding to physical pallets of hides. These volumes act as a digital representation of the pallets, offering flexibility in managing inventory. The pallets can either be broken down into smaller volumes (for example, splitting a large pallet into more manageable sections) or grouped into larger volumes if necessary. This dynamic tracking of volumes allows for better control over raw material flow throughout the production process.

These pallets of hides (represented by volumes) are then used as raw materials for Production Orders (POs). A PO refers to the formal instruction to begin processing these raw materials. Once a PO is completed, one or more new volumes are generated. These latest volumes can serve two purposes: they may be used as inputs for further production orders, entering into subsequent stages of the leather manufacturing process, or they may be designated for sale as finished products. The management system, with its detailed tracking of batches and volumes, ensures that materials are closely monitored from receipt through production and sale, providing critical oversight at each stage of the tannery process.

After gaining a thorough understanding of these operational particularities and how raw materials are handled and transformed, the method for developing prediction models was established. These models were built to enhance the system's capabilities by forecasting future needs, issues, or outcomes based on the rich dataset generated by the meticulous tracking of batches, volumes, and production outcomes. This predictive approach offers significant improvements in planning, efficiency, and decision-making within the tannery.

##### 4.1. Building Prediction Models

The Orange software enabled the creation and evaluation of a model for predicting leather yield based on operational and process parameters. This artifact is a file containing a trained model in Orange, which can be integrated into other systems, including the tannery's management software, to forecast leather yield. Figure 3 illustrates the method developed through systematic steps and highlights the challenges encountered during the process.



**Figure 3.** Method developed to generate the prediction model.

The methodology is flexible and can be applied not only to Production Orders (POs) but also to other products across various industries. Standardization is key to building reliable ML models, as data collection and processing must follow a consistent approach.

The developed method consists of the following essential steps:

- **Data Collection:** The production process must be carefully studied to understand the relationships between different stages and to establish a standardized flow of operations. Once the process is clearly defined, data from the management system should be collected. This data will serve as the foundation for subsequent analysis and treatment.
- **Data Treatment:** In this phase, the selected operations and parameters are examined. Relevant data are filtered, while missing, constant, or outlier data are either removed or corrected where possible. Additional information about the raw material, if available, can also be incorporated to enhance model accuracy.
- **Prediction:** Once the data has been collected and processed, Machine Learning (ML) models are defined and trained using the refined dataset. This phase involves setting the parameters that will allow the models to learn patterns from the historical data.
- **Evaluation:** The performance of the ML models is evaluated using well-known metrics. For regression models aimed at prediction, the recommended metrics include Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination ( $R^2$ ). Evaluation forms a continuous cycle where models are trained, and their metrics are assessed. If the results are unsatisfactory, adjustments are made to the model parameters, and the evaluation is repeated until the metrics indicate an optimal model.

The final step is the selection of the best model based on its performance metrics, comparing the results across different models to choose the most accurate predictor of leather yield. By following these steps, the method not only enhances predictive accuracy but also ensures a structured approach that can be adapted to different production orders and industries.

#### 4.1.1. Data Collection

The management system enables comprehensive traceability of the production process, allowing for a clear understanding of what happens to raw materials from the moment they arrive in batches to their final sale. This traceability is bidirectional, meaning that it is not only possible to track the raw materials used in any given Production Order (PO) but also to identify which batches contributed to a specific PO.

However, the study identified a challenge with traceability when conducted at the batch level rather than by individual units. As production progresses and volume operations (such as splitting or grouping volumes) are performed, the level of confidence in the traceability data tends to decline. For instance, when two volumes originating from



different batches are combined, the resulting volume will be composed of 50% of each batch. Over time, as volumes are further processed and manipulated, the percentage contribution of each original batch diminishes. In extreme cases, volumes may end up with insignificant percentages from several batches, complicating the traceability of raw materials.

Due to the complexity and variability introduced by this process, the study chose to disregard raw material traceability data for the POs under analysis. Instead, the focus shifted to other factors. In terms of leather yield, this study defined yield as the difference between the initial and final area of the leather during processing. In the tannery under analysis, there is no specific area measurement for individual pieces of leather. Instead, the area is measured for a set of leathers, collectively referred to as a volume. A volume contains data on the number of leathers it holds as well as their total area, which allows the system to compute the average area per hide.

Even though raw material traceability was not included in the study and POs can receive raw material from multiple volumes with varying quantities and measurements, key statistical metrics—such as the minimum, average, maximum, and standard deviation of the average leather areas in each volume—were still considered.

The scope of the study was further narrowed by selecting a sample of POs that followed a more standardized and uniform process to ensure consistency and accuracy. The first filtering criterion applied was based on the type of PO. This type of classification defines both the initial and final state of the leather involved in that specific PO. Leather in this context can be classified into one of five states, ensuring that the study focused on comparable operations and outcomes across different POs:

1. Raw: Leather freshly taken from animals may or may not be salted. Salted leather has undergone a salting process to prevent the skin from rotting.
2. Tripe: Leather that has been split into two parts: the upper part called top grain and the lower part called split leather.
3. Pre-Tanned (Wet-White): Leather tanned with synthetic and vegetable tannins, free of chromium.
4. Crust: Leather that has been tanned, dyed, and dried but is not yet finished.
5. Finished: Leather that has undergone various surface treatments, giving it properties of durability, stability, and beauty with different finishes.

Figure 4 shows the states of leather (rectangles) related to the types of OPs, which define the operations required to transform the state of the leather according to product specifications, along with the number of OPs completed for each type. The type “Crust to Finished” was chosen to obtain the maximum number of samples, resulting in 4051 completed OPs.

However, even when limited to a single type of Production Order (PO), significant variation in processes was observed due to the diverse range of products requiring specific operations. The data were filtered based on the technical specification that had the highest number of occurrences, ensuring that all selected POs shared the same product type and operational flow. This filtering process yielded 889 POs, all related to automotive upholstery leather. During further analysis, it was discovered that some POs contained zeroed initial and final quantities or showed discrepancies between these quantities. These inconsistencies were likely caused by human error, and as a result, the problematic POs were discarded, leaving 873 POs for analysis.

Despite the application of these filters, substantial variability in operations has remained across the POs. The variability was attributed to versioning in the technical specification, with certain operations being added or removed in different versions. A comparative analysis was conducted between the various versions of the technical specification, focusing on the operations involved in each. The study revealed that a greater level of standardization began from version 12 onwards. By applying this version criterion, the number of POs was further reduced to 730, which were selected for the next phase of processing.

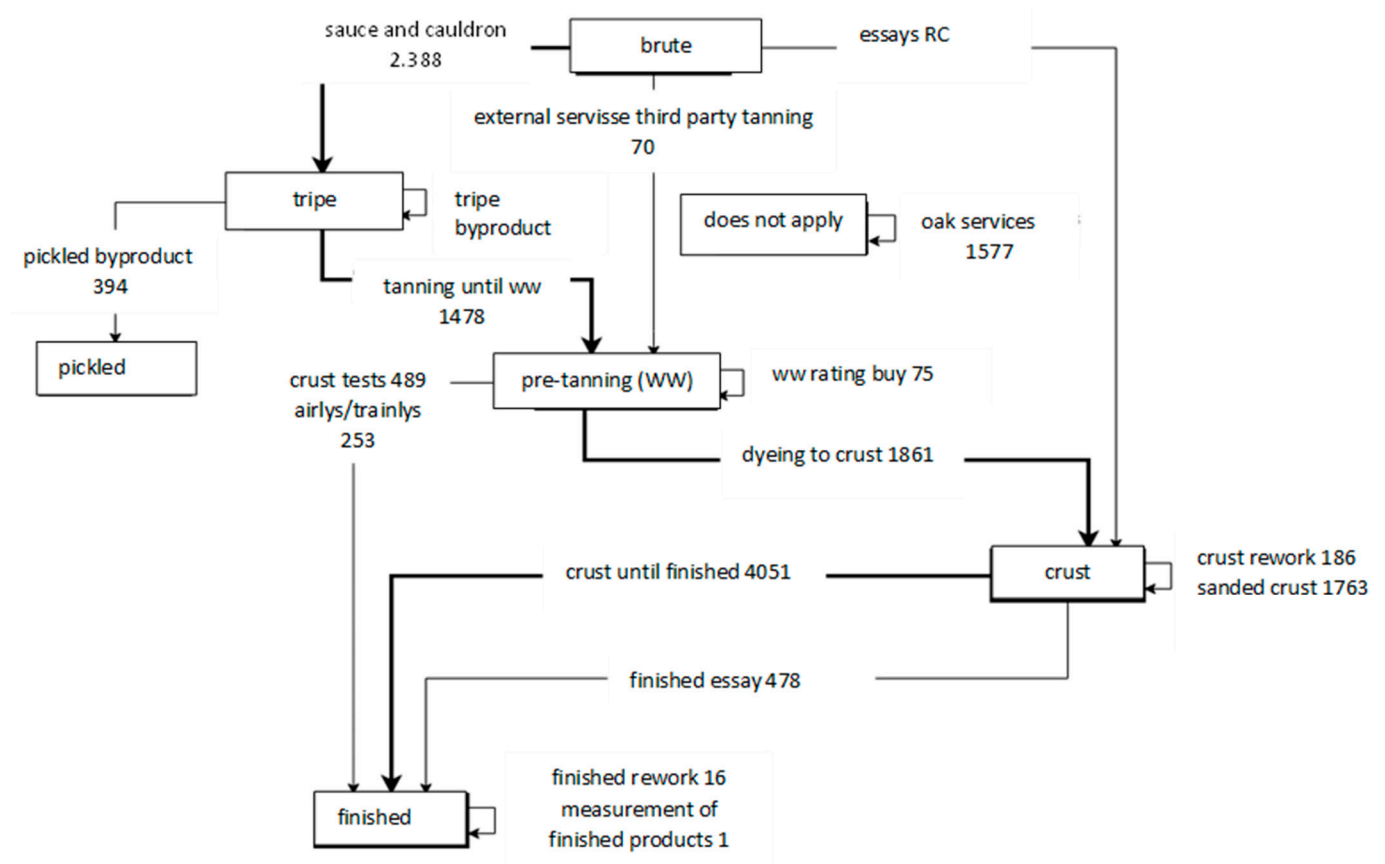
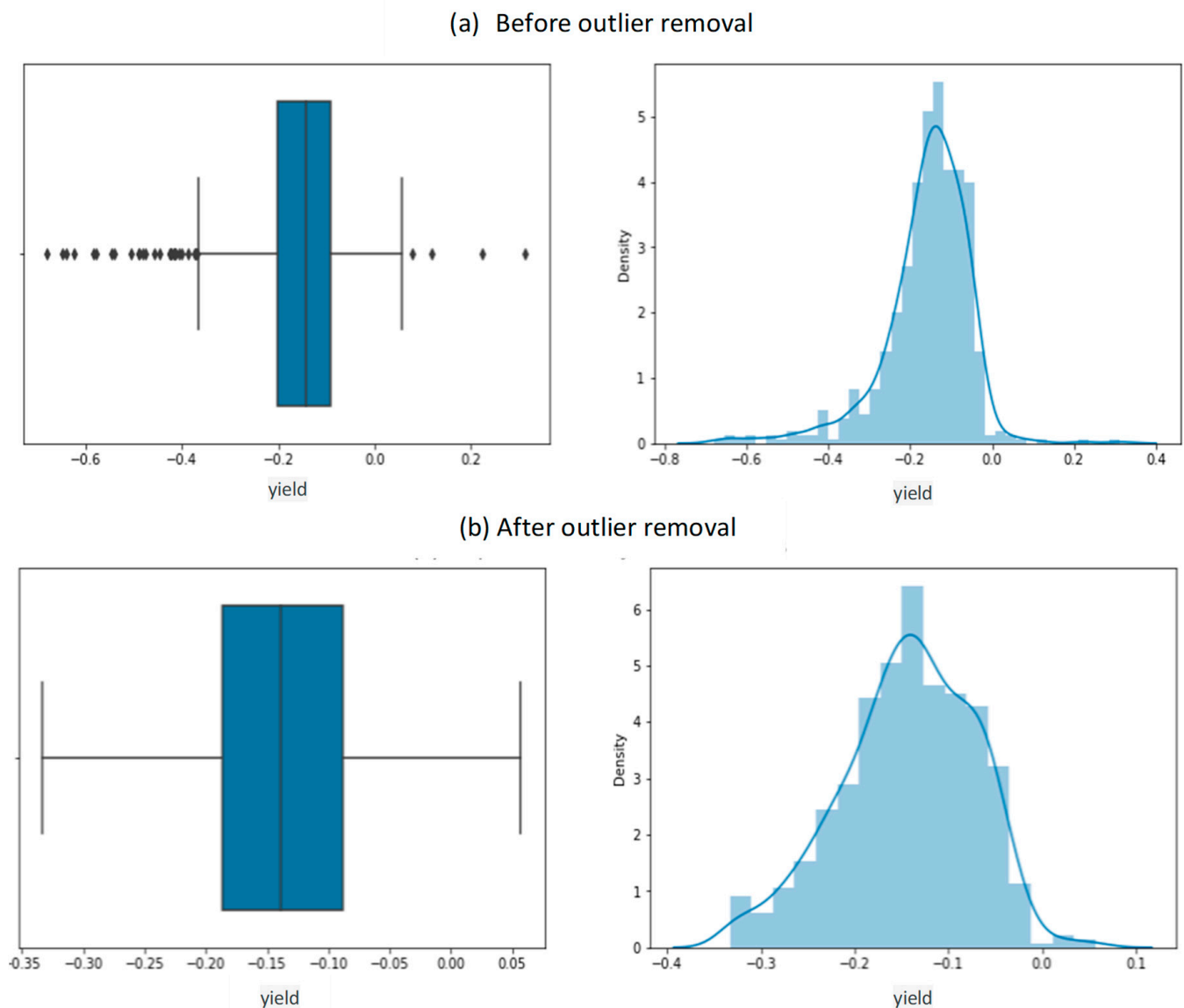


Figure 4. Flowchart of the relationship between leather condition and type of PO.

4.1.2. Treatment

The analysis of the orders (OPs) revealed that many contained a significant number of unexecuted operations. In these cases, most tasks were consolidated within a single OP, while the remaining operations were dispersed across others. OPs with fewer than 15 completed operations were discarded, reducing the total to 697 OPs. However, analyzing only the operations was insufficient. It became necessary to assess the associated parameters, particularly the availability of actual production data, which is critical for accurate analysis. Some parameter records were missing due to either operator error or system unavailability. OPs with 25 or fewer actual records were excluded, further refining the dataset to 611 Ops to ensure the reliability of the dataset.

In addition, outliers—OPs that deviated significantly from the norm and could adversely affect the performance of machine learning models—were identified and removed. The Interquartile Range (IQR) method was employed to detect these outliers based on the yield data of the OPs. After applying the IQR rule, the dataset was reduced to 555 OPs. Figure 5a presents the box plot and histogram of the yield data before outlier removal, while Figure 5b shows the same graphs after applying the IQR rule.



**Figure 5.** Removal of outliers before and after using the IQR rule.

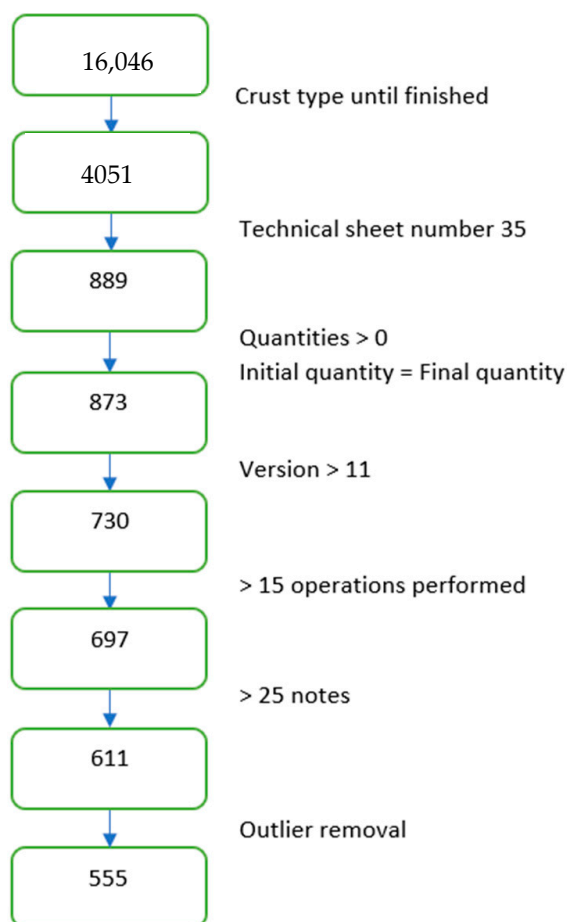
The box plot indicates that 50% of the samples had yields ranging between approximately  $-8\%$  and  $-18\%$ , with the lower and upper bounds around  $5\%$  and  $33\%$ , respectively. Data points falling outside these limits were classified as outliers and discarded. After their removal, the histogram reveals a distribution closer to a normal or Gaussian curve, indicating a more balanced and reliable dataset for further analysis.

Figure 6 shows the filtering of the POs described.

The subsequent step focused on eliminating operations that either lacked parameters, had no actual production values, exhibited constant values, or were deemed irrelevant to the yield problem. This refinement ensured that only meaningful data contributing to the analysis remained. Table 1 provides an overview of the operations and parameters retained after this filtering process.

Crust leather serves as the raw material input for this type of operation. Initially, the leather undergoes the “Crust Leather Vacuuming” process, where a vacuum machine removes any surface residue or dust, preparing it for the subsequent “Pre-Base” operation. In this step, a resin is applied to the leather to enhance adhesion for the paint used in the “RCM Paint” operation.

### NUMBER OF PRODUCTION ORDERS.



**Figure 6.** Filtering the samples based on the filters applied.

Following this, the “Rotary Press Engraving” operation uses a specialized machine to apply pressure, imprinting a design on the leather. This design may serve to conceal imperfections, provide a more natural appearance, or create a specific pattern. However, the pressure applied during this process alters the leather’s softness. The “Pre-Finished Softening” operation is performed using a softening machine to restore the leather’s original pliability, followed by a final cleaning step, the “Finished Leather Vacuuming” operation.

In the subsequent “RCM Gloss” and “Gloss Spraying” operations, a protective and aesthetic layer—often a wax—is applied to enhance the leather’s shine, color, and durability. Finally, a second softening process, the “Finished Softening” operation, is conducted to adjust the leather’s softness to meet the specific requirements of the finished product.

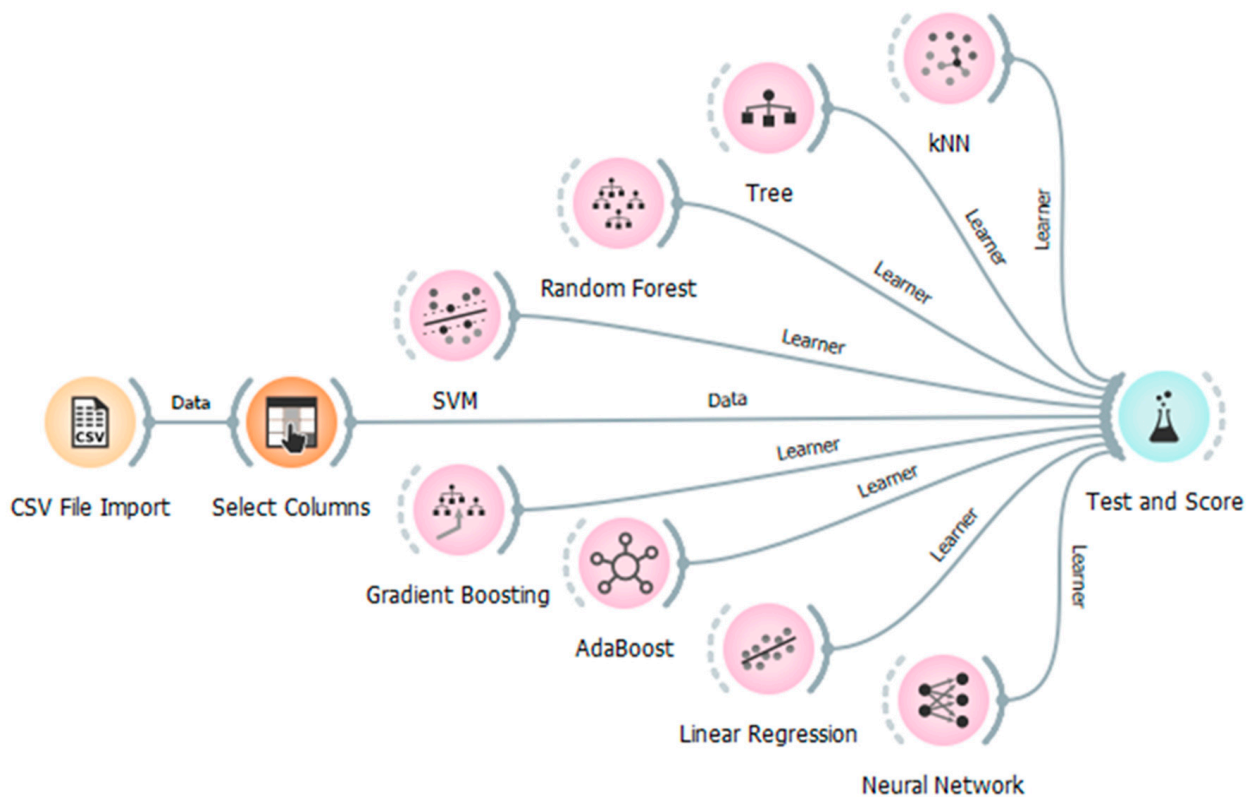
Data processing was primarily conducted manually. The parameters were first analyzed individually alongside their respective datasets, revealing numerous instances of typographical errors, which were corrected where possible. Additionally, it was observed that multiple values existed for the same operation within a single order. This variation arises from interruptions during the operation, such as operator breaks or machine maintenance, after which different parameters may be applied upon resumption. To address this, minimum, average, and maximum statistics were collected for each operation and parameter, ensuring that variations across multiple records for the same operation and parameter could be accounted for in the machine learning models.

**Table 1.** List of selected operations and parameters.

Operation	Parameter
Crust Leather Vacuuming Vacuuming machine	Pressure (bar)
	Carpet speed (m/min)
Pre-Base RCM machine	Product quantity (g)
	Roller speed
	Carpet speed (m/min)
	Thickness (mm)
	Average oven temperature (°C)
	Viscosity Ford 6 (")
RCM Paint RCM machine	Product quantity (g)
	Roller speed
	Carpet speed (m/min)
	Viscosity Ford 4 (")
	Thickness (mm)
Rotary Press Engraving Rotopress	Average oven temperature (°C)
	Pressure (Bar)
	Roller speed
	Carpet speed (m/min)
Pre-Finishing Dampening machine	Tightness 1
	Softness to Standard
	Tightness 4
Finished Leather Vacuuming Vacuum machine	Pressure (Bar)
	Carpet speed (m/min)
RCM Gloss RCM machine	Product quantity (g)
	Roller speed
	Carpet speed (m/min)
	Viscosity Ford 4 (")
	Thickness (mm)
Gloss Spraying Painting machine	Average oven temperature (°C)
	Product quantity (g)
	Ink pressure (bar)
	Pulv. pressure (bar)
	Carousel speed (RPM)
	Carpet speed (m/min)
Finished Softening Softening machine	Viscosity Ford 4 (")
	Average oven temperature (°C)
	Squeeze 1
	Speed (m/min)
	Softness to Standard
Finished Softening Softening machine	Tightness 2.2
	Tightening 2.3
	Carpet speed (m/min)

#### 4.1.3. Prediction

Figure 7 shows the diagram generated in the Orange software.



**Figure 7.** Modeling in Orange software.

In the “CSV File Import” widget, the CSV file containing the processed leather yield data is loaded into the software. Following this, the “Select Columns” widget is used to choose the features and target columns. In this case, the target variable is the yield, while the features consist of the remaining variables. The machine learning models, along with the training and testing data, are then linked to the “Test and Score” widget, which handles the training, validation, and presentation of model statistics. A cross-validation strategy with five folds was employed for training and validation. The model parameters were optimized through trial and error, with adjustments made and error metrics monitored after each iteration. The optimal parameter sets for each model are those already displayed in the previously shown Figure 2.

#### 4.1.4. Evaluation

Table 2 summarizes the evaluation statistics of the models generated in Orange software [36].

The Mean Squared Error (MSE) measures the average squared difference between observed and predicted values, with an ideal model yielding an MSE of zero. The Root Mean Squared Error (RMSE) is the square root of the MSE, while the Mean Absolute Error (MAE) represents the average deviation of predicted values from actual values. The  $R^2$  score ranges from 0 to 1, representing the correlation between the model’s variables and its predictions; values closer to zero imply weaker correlations. To generate the statistics shown in Table 2, we used the cross-validation mode, which simulates the predictions of new objects by repeatedly dividing the original training data set into training and testing on objects [37]. The procedure is used when a validation set is unavailable or when the dataset is too small to be divided into a training and a testing set.

**Table 2.** Evaluation statistics of the generated models.

Model	MSE	RMSE	MAE	R <sup>2</sup>
AdaBoost	0.003	0.057	0.042	0.331
Random Forest	0.003	0.057	0.043	0.325
Gradient Boosting	0.003	0.058	0.044	0.308
SVM	0.004	0.065	0.050	0.127
Linear Regression	0.004	0.067	0.053	0.065
kNN	0.005	0.069	0.054	0.008
Neural Network	0.005	0.069	0.055	0.022
Tree	0.005	0.074	0.057	0.145

The AdaBoost, Random Forest, and Gradient Boosting models demonstrated the best performance across the MSE, RMSE, MAE, and R<sup>2</sup> metrics, with AdaBoost showing a slight edge in MAE and R<sup>2</sup>. Both AdaBoost and Gradient Boosting are ensemble learning models that use Decision Trees as base models, which individually performed poorly, underscoring the effectiveness of boosting techniques, where learning occurs sequentially, with each model correcting the errors of its predecessor. These models performed well due to the relatively small dataset and high variability in parameters and leather yield. The poor performance of simpler models highlights the advantage of error-correcting models like AdaBoost and Gradient Boosting. In contrast, the Neural Network model delivered average results, as it typically requires larger datasets to excel.

An MAE of 0.042 translates to an error margin of 4.2%. For instance, if the predicted yield is 8%, the actual yield is likely to fall between 3.8% (8–4.2%) and 12.2% (8% + 4.2%). In the context of yield prediction, this error is relatively high, making the model unsuitable as a decisive tool but still valuable for decision support. The elevated error could be attributed to the absence of data on raw material characteristics, unrecorded production parameters, and a lack of batch traceability.

## 5. Discussion and Future Directions

A raw material yield prediction model in an Industry 4.0 scenario can benefit organizations. As a decision-support tool, as raw materials or parameters change, the system should be able to predict the yield. This prediction will help specialists select raw materials and parameters or, with the aid of other models, suggest the best parameters for the preferred raw material to optimize yield.

Furthermore, this model could be part of an automated production system in a highly developed Industry 4.0 environment. For example, through sensors, a company's machine collects and sends real-time data to a data processing server. This server, in turn, feeds a yield prediction model with the collected data and provides real-time yield predictions. This prediction can be valuable to the manager, who can be notified when the yield falls below expectations and take appropriate action. It can also be used as input for other systems for further calculations and machine configurations with new parameters to optimize yield and raw material usage.

The performance of eight ML-based prediction models was evaluated using data collected from a management system in a real tannery. The data scope needed to be reduced due to the complexity of the tanning process and the data available in the system. A significant number of errors were identified in the data, such as unreported, missing, or incorrectly reported parameters by operators, which reduced the number of samples used in this study. Considering these challenges, the generated and analyzed models could predict leather yield with limited error.

In an Industry 4.0 context, a raw material yield prediction model offers significant advantages for organizations by serving as a decision-support tool. As raw materials

or production parameters fluctuate, the system should be capable of forecasting yield, enabling specialists to make informed decisions. This predictive capability would assist in selecting the most suitable raw materials and parameters or, with the integration of other models, suggest optimal parameter configurations to maximize yield for a given raw material.

Beyond decision support, such a model could be seamlessly integrated into an automated production system within a fully developed Industry 4.0 environment. For instance, machines equipped with sensors could continuously collect and transmit real-time data to a central processing server. This server would then feed the yield prediction model, which would generate real-time yield forecasts. These predictions could be invaluable to managers, allowing them to receive alerts when yields fall below expectations and take corrective measures promptly. Moreover, the predictions could serve as inputs for other systems, supporting further calculations and enabling machines to automatically adjust their configurations with new parameters, thereby optimizing both yield and raw material efficiency.

In this study, the performance of eight machine learning-based prediction models was evaluated using data from a real tannery's management system. However, due to the inherent complexity of the tanning process and limitations in the available data, the data scope had to be reduced. A key challenge encountered was the significant number of data errors, including unreported, missing, or inaccurately recorded parameters by operators, which led to a reduction in the number of usable samples for the analysis.

Despite these challenges, the generated models were able to predict leather yield with a limited margin of error. However, the study identified several avenues for future work aimed at improving the models and reducing prediction errors. These opportunities include enhancing data quality, expanding the dataset to capture more variability in raw materials and parameters, and incorporating additional factors such as environmental conditions and machine-specific behaviors. By addressing these areas, future iterations of the model could offer even more precise predictions, further optimizing yield and resource utilization in a smart manufacturing environment.

Key improvements were identified to improve accuracy in the context of the tanning industry:

- Incorporate raw material characteristics: Factors such as the supplier, country of origin, animal type, and the presence of defects can significantly influence yield. Including these variables in the model would offer a more comprehensive understanding of how raw material quality impacts the final output.
- Consider operation execution time: The duration of each operation within the production process may affect yield. Collecting and integrating execution time into the model would provide insights into how process timing influences efficiency and product quality.
- Account for variations in production parameters: Critical parameters like temperature, speed, pressure, and humidity often fluctuate throughout the production process. However, current systems typically record only a single value, often an average, which fails to capture these real-time variations. Developing an Internet of Things (IoT) infrastructure, where sensors attached to machines continuously monitor and record parameter fluctuations, would allow for a deeper analysis of how these dynamic variables affect leather yield. This real-time data could lead to more precise predictions and better process control.
- Implement hide-level traceability: Yield models could benefit from tracking individual hides rather than relying solely on batch-level data. Variability in raw material characteristics can occur even within the same batch. Technologies such as Radio Frequency Identification (RFID), dot peening, and laser engraving, as proposed by [38], could be employed to trace each hide throughout the production process. This hide-level traceability would enable more granular data collection, improving model accuracy by accounting for intra-batch variability.



By addressing these areas, future iterations of the yield prediction model could offer more precise insights, enabling better decision-making and optimization of both raw material usage and production processes.

## 6. Conclusions

This study aimed to establish a comprehensive method for collecting, processing, predicting, and evaluating data to develop machine learning (ML) models capable of predicting leather yield based on historical data from a real tannery. Leather yield is influenced by a multitude of variables, ranging from the quality and characteristics of the raw materials to the specific parameters of the tanning process. The proposed method proved to be highly valuable for the company under study, as it not only facilitated the generation of ML models but also provided a structured approach to data management and process optimization. Some key utilities of the method may be highlighted:

- **Analyzing Current Data Collection Practices:** The method allowed for a critical examination of how data are currently gathered within the company. This step is crucial because inconsistent or incomplete data can undermine the effectiveness of predictive models. By understanding the existing data collection framework, companies can identify gaps and inefficiencies that need to be addressed.
- **Identifying Relevant and Consistent Data:** Not all data recorded in production systems are equally valuable for yield prediction. The method helped to sift through the available historical data, identifying which variables had the most consistent and relevant impact on yield. The method ensures that only high-quality data are fed into the ML models, improving their accuracy and reliability.
- **Standardizing Data Collection:** One of the most important outcomes of the method was the standardization of the data collection process. In many companies, data are collected in a fragmented or inconsistent manner, which can lead to discrepancies and errors. Standardization ensures that data are uniformly recorded across different production stages, making it easier to analyze and compare.
- **Generating Actionable Information for Decision-Making:** By processing the data and generating insights, the method provided valuable information that can be used to inform decision-making at both the operational and managerial levels. For instance, it can help select the best raw materials or adjust process parameters to optimize yield.
- **Improving Managerial Process Control:** The method also contributed to better process control. By providing a clearer understanding of how different variables affect yield, managers can make more informed decisions regarding production adjustments, ultimately improving efficiency and reducing waste.
- **Identifying Areas for Improvement in Data Control and Collection:** The implementation of the method highlighted several areas where data collection and control could be improved. These insights are crucial for companies looking to optimize their production processes and ensure that their data are reliable and actionable.
- **Creating a Robust Database:** The method facilitated the creation of a comprehensive database of information. This database serves as the foundation for generating predictive models and can be expanded over time as more data are collected, leading to increasingly accurate predictions.
- **Increasing Yield and Reducing Waste:** Ultimately, the method aims to improve yield and optimize raw material usage. By identifying inefficiencies and making data-driven decisions, companies can reduce waste, lower operational costs, and increase profitability.

Regarding challenges and initial difficulties, implementing this method may present challenges for companies that lack proper control over their data collection processes. In the case of the tannery studied, significant effort was required to prepare and process the data before it could be used to train ML models. The amount of necessary effort highlights the importance of having a well-structured data collection system in place from the outset. Moreover, the introduction of unit-level traceability—tracking individual hides throughout

the production process—was identified as a critical step toward generating more accurate and contextually relevant models for decision-making.

Regarding model performance and results, in this study, the performance of several ML algorithms was evaluated, with the AdaBoost algorithm emerging as the top performer. Using cross-validation, the AdaBoost model achieved the following metrics:

- Mean Absolute Error (MAE): 0.042
- Mean Squared Error (MSE): 0.003
- Root Mean Squared Error (RMSE): 0.057
- $R^2$  (Coefficient of Determination): 0.331

These results indicate that the AdaBoost model is currently the most effective at predicting leather yield. However, the Random Forest and Gradient Boosting models also demonstrated comparable, albeit slightly lower, performance.

Finally, regarding opportunities for improvement, despite the promising results, there is still room for enhancing the accuracy of the ML models. Several key areas for improvement were identified:

- **New Data Aggregation:** Aggregating data from additional sources or expanding the dataset to capture more variability in raw materials and process parameters could lead to more robust models.
- **Incorporating Execution Time:** The duration of each operation within the production process is another variable that could influence yield. Including this data in the model could improve its predictive power.
- **Real-Time Parameter Capture:** The current system records process parameters such as temperature, speed, pressure, and humidity as single values, often averages. However, these parameters can fluctuate during the production process, and capturing real-time variations through an IoT (Internet of Things) infrastructure would provide a more accurate representation of how these factors impact yield.
- **Hide-to-Hide Traceability:** Implementing traceability at the individual hide level, rather than relying on batch-level data, could significantly improve model accuracy. Traceability allows for a more granular understanding of how raw material characteristics vary within the same batch and how these variations affect yield. Technologies such as RFID, dot peening, and laser engraving could be employed to achieve this level of traceability.

In conclusion, the method developed in this study serves as a powerful tool for both generating ML models and enhancing data management practices within the tannery industry. Machine learning (ML) predictive models can play a crucial role in ensuring the traceability of raw materials, particularly in sectors like leather production, where material quality directly impacts yield and efficiency. By analyzing historical data on various suppliers and their associated leather batches, ML models can identify patterns and correlations that help predict which suppliers consistently provide higher-quality materials. These models can be integrated into the supply chain system to track the origin of each batch of leather, allowing managers to trace the quality of each piece back to its specific supplier. This traceability not only enhances transparency but also provides actionable insights to optimize supplier selection, improving the overall quality of raw materials entering the tanning process.

Furthermore, by utilizing ML models to predict potential quality issues before they manifest in the tanning process, manufacturers can take preventive actions to reduce waste and improve the consistency of their products. ML algorithms can detect anomalies in leather characteristics, such as thickness, texture, and moisture levels, which may indicate poor-quality hides. When these predictors are linked to specific suppliers, they allow for targeted adjustments in the procurement process, ensuring that only high-quality hides are selected. As a result, the tanning process becomes more efficient, with fewer defects and losses, aligning with the goals of sustainability and cost optimization in leather manufacturing.

**Author Contributions:** Conceptualization, L.H., J.C.F., and I.C.B.; methodology, L.H., J.C.F., E.T.P., and I.C.B.; software, L.H., and J.C.F.; validation, L.H., E.T.P., and J.C.F.; formal analysis, L.H., E.T.P., J.C.F., M.A.S., and I.C.B.; investigation, L.H., J.C.F., E.T.P., and I.C.B.; resources, L.H., J.C.F., M.A.S., E.T.P., and I.C.B.; data curation, L.H., J.C.F., M.A.S., E.T.P., and I.C.B.; writing—original draft preparation, I.C.B., and M.A.S.; writing—review and editing, M.A.S., E.T.P., and I.C.B.; visualization, M.A.S., and I.C.B.; supervision, J.C.F., and I.C.B.; project administration, J.C.F.; funding acquisition, M.A.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** The CNPq: the Brazilian Research Agency, funded this research under grant agreement No. 303496/2022-3.

**Data Availability Statement:** Data will be made available on request.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Heikkila, M.; Malkamo, V.; Etelaaho, P.; Kippola, T.; Koskela, M. Latency Validation Method for 3D 5G Networks' URLLC Applications. In Proceedings of the 16th European Conference on Antennas and Propagation, EuCAP 2022, Madrid, Spain, 27 March–1 April 2022. [CrossRef]
2. Angelopoulos, A.; Michailidis, E.T.; Nomikos, N.; Trakadas, P.; Hatziefremidis, A.; Voliotis, S.; Zahariadis, T. Tackling Faults in the Industry 4.0 Era—A Survey of Machine-Learning Solutions and Key Aspects. *Sensors* **2020**, *20*, 109. [CrossRef] [PubMed]
3. Goecks, L.S.; Habekost, A.F.; Coruzzolo, A.M.; Sellitto, M.A. Industry 4.0 and Smart Systems in Manufacturing: Guidelines for the Implementation of a Smart Statistical Process Control. *Appl. Syst. Innov.* **2024**, *7*, 24. [CrossRef]
4. Sivakumar, V. Towards Environmental Protection and Process Safety in Leather Processing—A Comprehensive Analysis and Review. *Process Saf. Environ. Prot.* **2022**, *163*, 703–726. [CrossRef]
5. CICB—Brazilian Tanning Industry Center. Available online: <https://www.cicb.org.br/cicb/dados-do-setor> (accessed on 13 September 2024). (In Portuguese).
6. Schaefer, J.L.; Tardio, P.R.; Baierle, I.C.; Nara, E.O.B. GIANN—A Methodology for Optimizing Competitiveness Performance Assessment Models for Small and Medium-Sized Enterprises. *Adm. Sci.* **2023**, *13*, 56. [CrossRef]
7. Furstenau, L.B.; Sott, M.K.; Kipper, L.M.; MacHado, E.L.; Lopez-Robles, J.R.; Dohan, M.S.; Cobo, M.J.; Zahid, A.; Abbasi, Q.H.; Imran, M.A. Link between Sustainability and Industry 4.0: Trends, Challenges, and New Perspectives. *IEEE Access* **2020**, *8*, 140079–140096. [CrossRef]
8. Winiarti, S.; Prahara, A.; Murinto; Ismi, D.P. Pre-Trained Convolutional Neural Network for Classification of Tanning Leather Image. *Int. J. Adv. Comput. Sci. Appl.* **2018**, *9*, 212–217. [CrossRef]
9. Pereira, R.F.; Medeiros, C.M.S.; Filho, P.P.R. Goat Leather Quality Classification Using Computer Vision and Machine Learning. In Proceedings of the 2018 International Joint Conference on Neural Networks, Rio de Janeiro, Brazil, 8–13 July 2018. [CrossRef]
10. Sousa, C.E.B.; Medeiros, C.M.S.; Pereira, R.F.; Neto, M.A.V.; Neto, A.A. Defect Detection and Quality Level Assignment in Wet Blue Goatskin. In Proceedings of the 11th International Conference on Advances in Information Technology, Bangkok, Thailand, 1–3 July 2020. [CrossRef]
11. Mohammed KM, C.; Prasad, G. Defective Texture Classification Using Optimized Neural Network Structure. *Pattern Recognit. Lett.* **2020**, *135*, 228–236. [CrossRef]
12. Tan, U.; Puntusavase, K. Decision-Making System in Tannery by Using Fuzzy Logic. *Adv. Intell. Syst. Comput.* **2021**, *1158*, 391–398. [CrossRef]
13. Schwab, K. *The Fourth Industrial Revolution*; Crown Business: New York, NY, USA, 2016.
14. Kagermann, H.; Wahlster, W.; Helbig, J. *Recommendations for Implementing the Strategic Initiative INDUSTRIE 4.0*; Final Report of the Industrie 4.0 Working Group; National Academy of Sciences: Washington, DC, USA, 2013.
15. Russell, S.; Norvig, P. *Artificial Intelligence: A Modern Approach*, 3rd ed.; Pearson Education: Boston, MA, USA, 2016.
16. Zhong, R.Y.; Xu, X.; Klotz, E.; Newman, S.T. Intelligent Manufacturing in the Context of Industry 4.0: A Review. *Engineering* **2017**, *3*, 616–630. [CrossRef]
17. Wuest, T.; Weimer, D.; Irgens, C.; Thoben, K.-D. Machine Learning in Manufacturing: Advantages, Challenges, and Applications. *Prod. Manuf. Res.* **2016**, *4*, 23–45. [CrossRef]
18. Yap, M.H.; Hew, C.S.; Lai, K.K. Machine Learning Predictive Models for Resource Optimization: An Industry 4.0 Approach. *J. Ind. Inf. Integr.* **2019**, *15*, 1–13.
19. Borges, F.S.; Santos, D.P.; Silva, G.F. Leather Yield Prediction Using Machine Learning Models: Towards a Sustainable Tanning Industry. *J. Clean. Prod.* **2021**, *289*, 125826.
20. Guan, Y.; Zhuang, X. Machine Learning Applications in the Manufacturing Industry: A Review. *J. Manuf. Sci. Eng.* **2019**, *141*, 061010.
21. Zhao, L.; Xie, S. Machine Learning-Based Prediction for Production Optimization in the Leather Industry. *Int. J. Comput. Integr. Manuf.* **2017**, *30*, 447–457.

22. Suganthi, L.; Iniyan, S. A Review on Machine Learning Models for Prediction of Leather Processing Yields. *Int. J. Ind. Eng. Technol.* **2020**, *12*, 29–42.
23. Yang, J.; Rahardja, S.; Fränti, P. Outlier Detection: How to Threshold Outlier Scores? In Proceedings of the International Conference on Artificial Intelligence, Information Processing and Cloud Computing, Sanya, China, 19–21 December 2019. [[CrossRef](#)]
24. Lemaître, G.; Nogueira, F.; Aridaschar, C.K. Imbalanced-Learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *J. Mach. Learn. Res.* **2017**, *18*, 1–5.
25. Wong, T.T. Performance Evaluation of Classification Algorithms by k-Fold and Leave-One-Out Cross Validation. *Pattern Recognit.* **2015**, *48*, 2839–2846. [[CrossRef](#)]
26. Wang, S.; Ning, Y.; Shi, H. A New Uncertain Linear Regression Model Based on Equation Deformation. *Soft Comput.* **2021**, *25*, 12817–12824. [[CrossRef](#)]
27. Du, J.; Sun, L.; Xu, K.; He, Z.; Zhang, W.; Chen, G.; Chen, X.; Reed, G.T. Nonlinear Distortion Mitigation by Machine Learning of SVM Classification for PAM-4 and PAM-8 Modulated Optical Interconnection. *J. Light. Technol.* **2018**, *36*, 650–657.
28. Baierle, I.C.; Schaefer, J.L.; Sellitto, M.A.; Fava, L.P.; Furtado, J.C.; Nara, E.O.B. MOONA Software for Survey Classification and Evaluation of Criteria to Support Decision-Making for Properties Portfolio. *Int. J. Strateg. Prop. Manag.* **2020**, *24*, 226–236. [[CrossRef](#)]
29. Babaei, H.; Mendiola, E.A.; Neelakantan, S.; Xiang, Q.; Vang, A.; Dixon, R.A.F.; Shah, D.J.; Vanderslice, P.; Choudhary, G.; Avazmohammadi, R. A Machine Learning Model to Estimate Myocardial Stiffness from EDPVR. *Sci. Rep.* **2022**, *12*, 1–17. [[CrossRef](#)]
30. Fontoura, L.C.M.M.; De Castro Lins, H.W.; Bertuleza, A.S.; D’assunção, A.G.; Neto, A.G. Synthesis of Multiband Frequency Selective Surfaces Using Machine Learning with the Decision Tree Algorithm. *IEEE Access* **2021**, *9*, 85785–85794. [[CrossRef](#)]
31. Alabadee, S.; Thanon, K. Evaluation and Implementation of Malware Classification Using Random Forest Machine Learning Algorithm. In Proceedings of the 7th International Conference on Contemporary Information Technology and Mathematics ICCITM, Mosul, Iraq, 25–26 August 2021; pp. 112–117. [[CrossRef](#)]
32. Monego, V.S.; Anochi, J.A.; de Campos Velho, H.F. South America Seasonal Precipitation Prediction by Gradient-Boosting Machine-Learning Approach. *Atmosphere* **2022**, *13*, 243. [[CrossRef](#)]
33. Kim, D.; Philen, M. Damage Classification Using Adaboost Machine Learning for Structural Health Monitoring. *Proc. SPIE* **2011**, *7981*, 659–673. [[CrossRef](#)]
34. Guo, G.; Wang, H.; Bell, D.; Bi, Y.; Greer, K. KNN Model-Based Approach in Classification. *Lect. Notes Comput. Sci.* **2003**, *2888*, 986–996. [[CrossRef](#)]
35. SystemHaus. Available online: <https://systemhaus.com.br/en/antara-erp> (accessed on 22 September 2024).
36. Demšar, J.; Curk, T.; Erjavec, A.; Gorup, Č.; Hočevar, T.; Milutinovič, M.; Možina, M.; Polajnar, M.; Toplak, M.; Starič, A.; et al. Orange: Data Mining Toolbox in Python. *J. Mach. Learn. Res.* **2013**, *14*, 2349–2353.
37. Xu, L.; Fu, H.Y.; Goodarzi, M.; Cai, C.B.; Yin, Q.B.; Wu, Y.; She, Y.B. Stochastic cross validation. *Chemom. Intell. Lab. Syst.* **2018**, *175*, 74–81. [[CrossRef](#)]
38. Thakur, M.; Tveit, G.M.; Vevle, G.; Yurt, T. A framework for traceability of hides for improved supply chain coordination. *Comput. Electron. Agric.* **2020**, *174*, 105478. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.