Improved quality control and sustainability in food production by machine learning

Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 122 (2024) 533–538

www.elsevier.com/locate/procedia

31st CIRP Conference on Life Cycle Engineering (LCE 2024)

Improved quality control and sustainability in food production by machine learning

Stefano Puttero[a,*], Elisa Verna[a], Gianfranco Genta[a], Maurizio Galetto[a]

[a]*Department of Management and Production Engineering, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy*

* Corresponding author. Tel.: +39 0110907236; *E-mail address:* stefano.puttero@polito.it

Abstract

In recent years, the food industry has faced a number of complex challenges related to both quality control and sustainability. Ensuring consumer safety and satisfaction remains a cornerstone of the food industry, supported by stringent standards that address the risks of contamination and spoilage. However, variability in raw materials, processing techniques and storage conditions are just some of the factors that affect quality in the food industry. To manage this high variability, it is essential to analyse the production process and factors that most influence food quality, aiming to predict and minimise food waste, thereby ensuring a sustainable process. This convergence of quality control and sustainability goals provides fertile ground for machine learning applications. By improving defect detection, process optimisation, resource allocation and predictive maintenance, these models help to improve product quality and reduce environmental impact. This article aims to explore the various applications of machine learning models in the food industry, where the variability of raw materials and the difficulty of controlling production and environmental factors challenge the use of traditional methods. The quality control and sustainability of an industrial corn cakes production process is used as a case study.

Keywords: Sustainable manufacturing; Quality control; Food industry; Machine learning

1. Introduction

Over the last decades, the food industry has faced several critical global challenges, both in terms of environmental sustainability and economic viability. Population growth, climate change, resource scarcity and efficient food waste management are just some of these challenges. In particular, food waste is one of the most worrying factors in the food sector. To date, it is estimated that nearly one-third of all food produced for human consumption is wasted each year, totalling about 1.3 billion tonnes worldwide [1,2]. This waste not only represents a significant economic loss, but also contributes to greenhouse gas emissions, land degradation and natural resource depletion. Accordingly, food waste management is a cornerstone of environmental conservation, food security and economic efficiency.

One of the initial approaches to mitigate food wastage is the adoption of a circular economy framework. The aim is to create symbiotic relationships among industries, wherein the by-products of one sector serve as raw materials for another, thus curtailing the volume of products that go to waste [3]. However, the transition to a circular economy requires several technological tools, including enhanced methods to predict the amount of food wasted during production. Another innovative solution that has emerged recently is the deployment of machine learning techniques in the food domain. The global food industry is undergoing a technological transformation, driven by the integration of machine learning algorithms into its operational processes. This paradigm shift holds the

10.1016/j.procir.2024.01.078

potential to revolutionize the manner in which food is produced, processed and distributed, and promises to address critical challenges such as food safety, quality control, sustainability and resource efficiency. By optimising supply chain logistics, predicting equipment failures and automating quality control measures, the industry can significantly reduce waste, lower energy consumption, and improve overall resource utilisation to meet the ever-changing consumer demands.

The paper investigates the integration of machine learning methods within the food industry, with a specific focus on minimising the prevalent waste in this sector. The objective of the paper is to develop and implement a diagnostic tool that facilitates the identification of waste-inducing factors within the production process and, subsequently, delineate effective strategies to prevent such wastage. The rest of the paper is organised into five sections. Section 2 analyses the different machine learning algorithms for data analysis and their application in the food sector. Section 3 describes the case study, while Section 4 analyses the application of machine learning to the selected production process. Section 5 presents the findings and provides a comprehensive discussion on the results obtained. Finally, Section 6 concludes the paper.

2. Machine Learning algorithms

In recent years, machine learning has evolved from a research curiosity to a useful technology with widespread commercial applications. Machine learning is defined as the set of techniques that enable computers to learn without being explicitly programmed. It is an inherently multidisciplinary field which is based on results from artificial intelligence, probability and statistics, computational complexity theory, control theory, information theory, philosophy, psychology and other different fields [4]. Such algorithms are often used to analyse large amounts of data, where traditional methods are insufficient to perform analysis and make predictions.

Machine learning relies on different algorithms to interpret and learn from data. There is no single type of algorithm that is best for solving a problem, but the type of algorithm used depends on the type of problem to be solved, the number of variables and the type of model. Fig. 1 shows a classification of the existing machine learning algorithms according to Mahesh [5]. To analyse large amounts of data and make predictions, the most commonly used machine learning algorithms are the supervised learning algorithms. These algorithms learn from labelled training data and make predictions based on that data [6]. Linear regression and ridge regression are some examples of supervised learning models (see Section 4). Decision trees are another important family of supervised learning algorithms. These are recursive algorithms that partition the data into subgroups based on the most informative features, resulting in a tree-like model of decisions [7]. The most commonly used decision tree model is the Random Forest (see Section 4.3).

Unsupervised and semi-supervised learning algorithms are an alternative strategy to supervised learning for data analysis. In unsupervised learning, the algorithm is given input data without explicit labels. The main goal is to find patterns, structures or relationships in the data. On the other hand, semi-supervised learning uses both labelled and unlabelled data for training. The idea is to use the small amount of labelled data to drive learning, while using the large amount of unlabelled data to improve generalisation and performance [8].

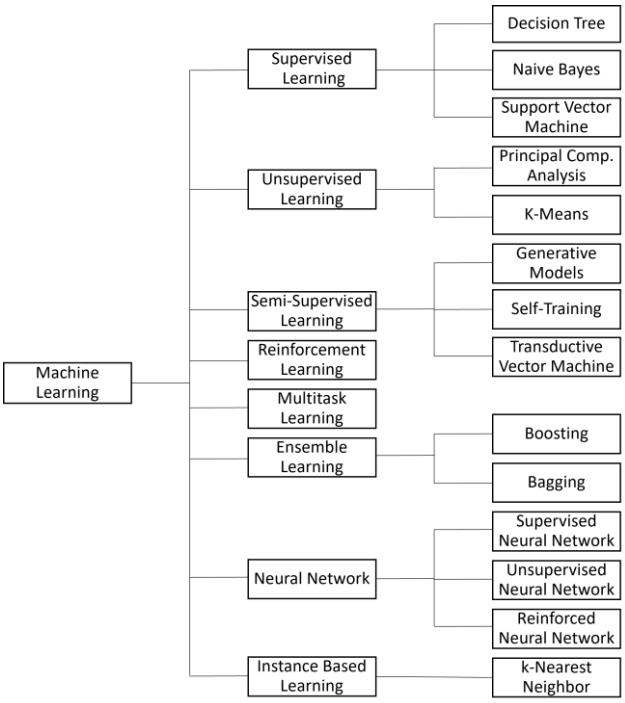

Fig. 1. Machine learning algorithms classification. Adapted from [5].

Another widely used category of algorithms is the ensemble algorithm, of which the gradient boosting algorithm is the most commonly adopted (see Section 4.4). Ensembles are based on the principle that combining the results from several models can produce better results than those of each individual model [9]. It is worth noting that among the various machine learning algorithms, there are the neural networks, which are part of the branch of machine learning called deep learning, which allow information to be extracted from images [10].

2.1. Machine learning in food industry

Among many applications, machine learning techniques are gaining widespread adoption in the food industry for a variety of purposes, including but not limited to improving quality, promoting sustainability, and optimizing operations. By analysing historical sales data, real-time stock levels and consumer behaviour, machine learning algorithms can make accurate predictions about fluctuations in demand. This predictive capability enables retailers and food suppliers to optimise inventory management, reduce overstocking and ensure that perishable products are sold or distributed before they spoil. In addition, machine learning can overcome the limitations of traditional forecasting models. According to Gobble et al. [11], classical statistical methods are not suitable for high-dimensional databases and can even cause delays in operations if the right tools are not used. Furthermore, these traditional models have their own assumptions and predefined underlying relationships between dependent and independent

variables. If these assumptions are violated, the models can lead to incorrect estimation of accident probabilities.

One of the biggest problems in the food industry is often the inability to store products for long periods of time due to spoilage [12]. Machine learning can be used to monitor and control factors such as temperature, humidity and shelf life, helping to maintain food quality and safety throughout the supply chain. By minimising food waste through proactive decision-making, machine learning not only helps to save money, but also promotes environmental sustainability by reducing the carbon footprint associated with food production and disposal [12]. However, despite their potential, machine learning algorithms are rarely used for production planning in the real food industry. Instead, simple judgement based on experience is often used [13]. In this research, an empirical dataset collected from a real food industry is used to evaluate the application of these advanced data analysis techniques in the food environment.

3. Case study

This paper presents a case study conducted within an Italian food company specializing in the production of organic products, specifically rice and corn cakes. The primary focus of this analysis centres on the production process of corn cakes, primarily due to the pronounced issue of waste generation.

The corn cake production process consists of 4 distinct stages. The first stage is the storage of the raw material (i.e., corn) in a temperature and humidity-controlled environment; it is essential that the corn is stored in a controlled environment to maintain the quality of the raw material. In the second stage, the corn is mixed with a mixture of water and salt to obtain the dough that will make up the corn cakes. At the end of this process, this mixture is transported through channels to tanks next to the cooking presses. In the third stage, the dough is cooked at high temperature in special presses that give the cake its typical circular shape. Once cooked, the cakes are transported to the packaging area on a vibrating conveyor. The function of the vibrating conveyor is to align the cakes so that they are ready for final packaging. The last step is to wrap the cakes in plastic film to keep them fresh for a long time.

Data were collected on waste during the regular production of corn cakes. During a period of five months, records of 1000 different batches were taken. Ten different factors influencing the occurrence of waste were identified by the tests: storage temperature (°C), storage humidity (%), production temperature (°C), production humidity (%), mixer time (s), pressing time (s), pressing temperature (°C), conveyor speed (m/s), conveyor vibration (Hz) and packing time (s). Some of these factors are related to raw material storage, such as storage humidity and storage temperature, while others are related to the production process, including conveyor speed and pressing temperature. The output variable monitored is the proportion of production waste (waste weight/input weight), given its high value during production (16%-22%). For confidentiality reasons, the real data from each day's collection could not be used and simulated data with the same trends were used. Prior to the analysis of these collected data, an outlier detection method was carried out using the InterQuartile Range (IQR) method [14].

4. Methods

Regression analysis is a statistical technique used to understand and quantify the relationship between a dependent variable and one or more independent variables. It involves finding a function, called $f(\boldsymbol{X}, \theta)$, that maps n input variables $\boldsymbol{X}$ to an output variable Y. The specific mathematical form of this function f depends on the particular problem being investigated. Typically, this model function is based on some unknown model parameter (θ) that needs to be estimated from the observed data [15]. In the following case study, $\boldsymbol{X}$ corresponds to the 10 variables listed in Section 3, while Y refers to the proportion of production waste. Four different regression models, described in the following subsections, were used to understand the relationship between $\boldsymbol{X}$ and Y: linear models, ridge regression, random forest regression, and gradient boosting regression.

Given the large number of data collected in real time, the presence of missing values, an unstructured data collection, overfitting and strong multicollinearity between the independent variables, traditional statistical methods may not be suitable, making the use of machine learning methods the only viable approach.

4.1. Linear regression model

Linear regression is a basic statistical model used to analyse and model the relationship between a dependent variable and one or more independent variables by fitting a linear equation to observed data. It is a simple but powerful way of making predictions and understanding underlying relationships in data [15]. Linear regression is a versatile tool with applications in fields ranging from finance to machine learning. Although a relatively simple model, it serves as the basis for more complex regression techniques and is a valuable tool for understanding and analysing data. The main objective of linear regression is to minimise the Residual Sum of Squares (*RSS*), defined as [15]:

$$RSS = \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2, \quad (1)$$

where n is the number of observations, y_i is the observed output, β_0 is the intercept of the model, p is the number of predictors and β_j are the coefficients of the predictors x_{ij}. The lower the RSS value, the better the model fits the relationship between the dependent and independent variables. Other parameters such as R^2, *RMSE*, residuals analysis, etc. can also be evaluated to assess the goodness of the model (see Section 5) [11]. In this study, the linear regression was used as a benchmark for the other models analysed.

4.2. Ridge regression model

Ridge regression, also known as L2 regularisation, is a linear regression technique used to solve the problem of multicollinearity and overfitting in traditional linear regression models. It adds a penalty term to the linear regression cost function to make the model accept smaller coefficient values [16]. In linear regression, the goal is to find the coefficients that

minimise the sum of the squared differences between the predicted and actual target values. However, this can lead to overfitting in the presence of multicollinearity if the independent variables are highly correlated [15]. Ridge regression solves this problem by regularising the cost function of linear regression, which encourages the model to have smaller coefficient values [16]. According to the definition of RSS in Eq. (1), the ridge regression aims to minimise the loss function defined as:

$$L = RSS + \lambda \sum_{j=1}^{p} \beta_j^2, \quad (2)$$

where L is the loss function that ridge regression aims to minimise, λ is a tuning parameter that determines the strength of the regularisation (if λ=0, the ridge regression becomes simple linear regression) and the term $\sum_{j=1}^{p} \beta_j^2$ is the sum of the squares of the coefficients of the predictors in the ridge regression model. Coefficients that are too large are discouraged by the regularisation term, which effectively reduces them to zero without actually setting any coefficients to zero [16]. As a result, the likelihood of overfitting the training set in the model is reduced. This regression model is particularly useful when it is necessary to stabilise the estimated coefficients in high-dimensional data sets.

4.3. Random forest regression model

Tree models are a class of machine learning algorithms that make predictions by constructing a tree structure in which data is divided into subsets based on a set of decision rules [7]. These algorithms aim to maximise the information gain within the resulting child nodes by determining the best feature and split point at each node. For the regression models, the common splitting criterion is the Mean Square Error (*MSE*) [7].

There are different types of tree models, of which random forest is one of the most widely used. A random forest is a method that combines multiple decision trees to improve prediction accuracy and reduce overfitting [17]. It works by training a series of decision trees and aggregating their predictions. The predictions from each tree are combined by bagging to produce a final prediction. In regression tasks, this bagging involves averaging the scores from the different trees.

4.4. Gradient boosting regression model

Gradient boosting constructs additive regression models that successively add new models to produce a more accurate estimate of the response variable [9,18]. The main idea of this technique is to build new base models that have the highest possible correlation with the negative gradient of the loss function [18]. This learning process leads to a subsequent error correction in the case where the loss function considered is the traditional quadratic error loss. However, alternative loss functions can be used to provide a clearer understanding of the problem, and the choice of the best loss function depends on the particular case study [9].

The high degree of flexibility of gradient boosting allows it to be easily tailored to any specific data-driven activity. It adds a great deal of freedom to model design, making the selection of the best loss function a trial-and-error process [9]. In addition, it has demonstrated significant effectiveness in a range of machine learning and data mining challenges, in addition to practical applications.

5. Results

A portion of the collected dataset has been utilized to train multiple machine learning algorithms, while the residual data points have been employed for testing the models. In detail, 70% of the observations were included in the training set used for parameter estimation and validation. The remaining observations (30%) were reserved for testing and were used to assess the predictive accuracy of the fitted model. Accuracy was assessed using the coefficient of determination R^2, the Root Mean Square Error (*RMSE*), and the Mean Absolute Error (*MAE*). Table 1 shows the values obtained for the different models for the training and test phases.

Table 1. R^2, *RMSE* and *MAE* for the training and test sets for each model.

	Training			Test		
Model	R^2	*RMSE*	*MAE*	R^2	*RMSE*	*MAE*
Linear	0.788	0.010	0.008	0.777	0.010	0.008
Ridge	0.788	0.010	0.008	0.781	0.010	0.008
Random forest	0.964	0.004	0.003	0.732	0.011	0.009
Gradient boosting	0.862	0.008	0.006	0.751	0.010	0.008

By looking at the R^2 values for the test phase, it can be seen that the model that performs best is the ridge regression. This is due to the ability of the ridge regression to reduce the multicollinearity and overfitting that characterize the linear regression model. In fact, some of the independent variables analysed in the case study were highly correlated (e.g., temperature and humidity).

The decrease in R^2 from the training phase to the test phase is another important aspect worth mentioning. This is often due to the larger size of the training phase, where overfitting problems can occur (thus introducing noise and random variation into the modelling). This implies a very good performance of the model in the training phase, but a poor generalisation of the results in the test phase with unseen data. This phenomenon is much more pronounced for more complex models, i.e. models that are able to capture complex interactions in the data. A clear example is the random forest, which is an ensemble method that combines several decision trees to capture these complex, hidden relationships in the data. Table 1 shows that although the random forest is the best model in the training phase, overfitting causes a significant reduction in the R^2 value in the test phase.

To check the goodness of the selected model, a normality test was also performed on the residuals of the ridge regression. Using the Anderson-Darling test, the hypothesis of normality of the distribution of the residuals cannot be rejected at the 95% confidence level. This result can be readily apparent from the residual plots shown in Fig .2. In particular, Fig. 2a shows that the residuals are centred around the zero value and have a random trend above and below zero, thus showing no particular

trend. On the other hand, Fig. 2b shows the histogram of the residuals, which clearly shows the typical pattern of a normal distribution centred around zero and with constant variance.

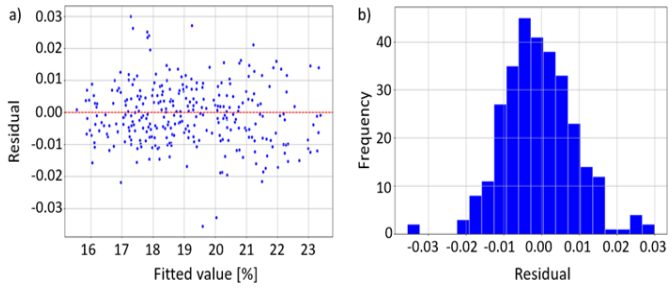

Fig. 2. Residual plots of the ridge regression model for proportion of production waste: a) scatter plot and b) histogram of the residuals.

In the context of machine learning, the permutation importance method (PIMP) is often used to define the importance of independent variables and to understand how they influence the dependent variable. The method normalises the importance measure based on a permutation test and returns significance *p*-values for each independent variable analysed [19].

The graph in Fig. 3 is obtained using this permutation method. The bar chart provides a visual representation of the importance of each independent variable: variables with larger bars have a greater impact on model performance, while variables with bars closer to zero have less impact. The x-axis shows the significance of the change in the model's performance metric due to the permutation of each independent variable, while the y-axis shows the independent variables. In particular, the values on the x-axis represent the decrease in the R^2 value when a particular variable is permuted. For instance, if the value on the x-axis for an independent variable is 0.5%, this means that the R^2 value changes by 0.5% when this characteristic is permuted, indicating that this variable is influential in the model's predictions.

As a result, a positive value indicates that the variable is important to the model and, the higher the value, the more important the variable is. The permutation importance graph in Fig. 3 shows that the most important variable is the storage humidity. A permutation of this independent variable changes the R^2 value by 2.5 percentage points.

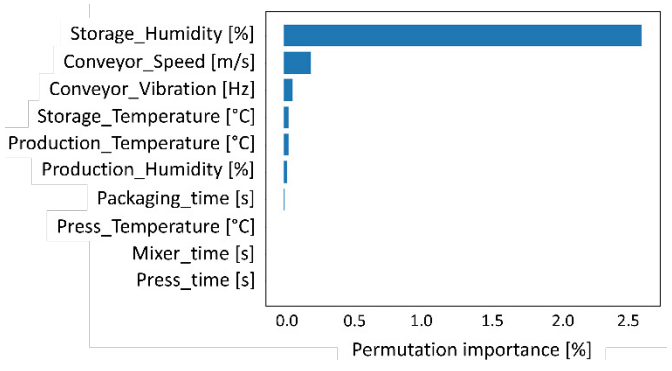

Fig. 3. Independent variable importance using the PIMP method [19].

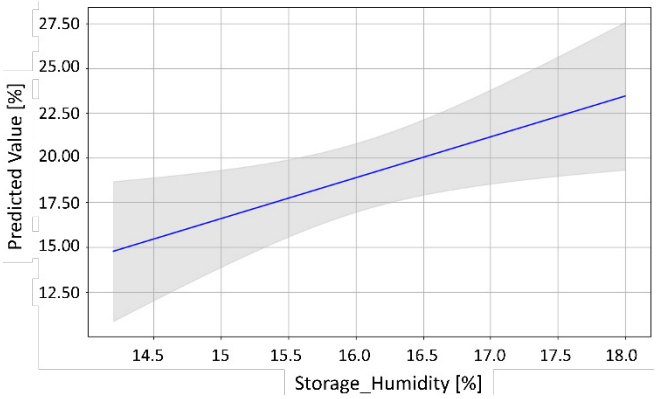

Fig. 4. Predicted proportion of waste as a function of the storage humidity.

Using the bootstrap method for ridge regression [16], the following variables were significant at the 95% confidence level: storage humidity, conveyor speed, conveyor vibration and packaging time. However, given the significant influence of humidity in relation to other independent variables, a deliberate focus was directed towards an in-depth examination of its effect on waste generation, along with an exploration of strategies to adjust humidity levels in alignment with the amount of waste expected. Furthermore, the storage humidity can be easily monitored and regulated by the company, making it straightforward to apply these methods.

Fig. 4 shows the expected proportion of waste as a function of storage humidity. The blue line represents the predicted proportion of waste, while the grey areas indicate the prediction intervals constructed at the 95% confidence level. As shown in Fig. 4, there is a positive correlation between the two variables and their relationship follows a linear trend according to the proposed ridge regression model. Accordingly, the higher the humidity, the more production waste.

It is worth noting that around the mean of the predictor variable, the prediction bands are relatively close to the regression line. Moving towards the extremes, the bands tend to diverge and widen. This is due to the fact that in the central zone, close to the mean, the density of the sample data is greater, thus reducing the uncertainty of the prediction. On the other hand, in the areas away from the central value, predictions are made for predictor values that are not well represented in the data sample, thus venturing into the territory of extrapolation.

As an example, a value close to the mean and a value close to the extremes of the range can be considered. For a storage humidity value of 16.0%, the identified model predicts with 95% of confidence level that the value of the proportion of waste will be between (16.00%, 21.00%). On the other hand, for an extreme humidity value of 18%, the model predicts with 95% confidence that the proportion of waste will be between (17.00%, 27.50%). This shows that for values far from the mean, the uncertainty of the prediction increases.

As a result, the implemented model can be used to predict food waste (including uncertainty) for different production scenarios and to modify waste management and production plans accordingly. This study shows how food companies can benefit from applying sophisticated analytics to historical data.

When it comes to predicting food waste with large amounts of data, machine learning techniques have proven to be more predictive than traditional statistical approaches.

6. Conclusions and future works

The efficient use of machine learning algorithms in relation to food production waste represents a potential and revolutionary strategy to address one of the most critical issues facing food companies. The need for sustainable and effective resource management is increasing as the world's population grows.

Machine learning offers several advantages for reducing food waste in the supply chain, due to its ability to handle massive data sets and make predictions based on that data. One of the key benefits of machine learning in this context is its ability to optimise production processes. By analysing historical data, algorithms can help manufacturers make informed decisions, resulting in more efficient use of resources and less waste due to overproduction or spoilage. In addition, such algorithms can make accurate predictions of gap levels based on the actual levels of the independent variables.

This study has shown the potential advantages that the food industry can derive from an examination of production data. In this very highly production context, the use of traditional statistical models is limited by the high dimensionality, complexity and correlation of these data. Therefore, the data needs to be analysed using state-of-the-art data analysis techniques such as machine learning, in order to provide more accurate estimates of production process parameters than traditional statistical models.

A case study of a company producing rice cakes was used to demonstrate how machine learning algorithms can be used as a process diagnostic tool to identify the relationship between process parameters (independent variables) and production waste (dependent variable) when large amounts of data are involved. Especially, the main problem faced by the company is excessive waste in the production process, as approximately 16%-22% of the raw material input is wasted during the corn cake production process. The aim of the paper was to identify the most influential variables and, using predictive methods based on machine learning, to identify parameter levels that would reduce this waste in production.

However, there is often a danger of using such algorithms as a 'black box', without understanding how they work and how they make decisions. There is a risk of losing the ability to interpret the data, which can be particularly problematic for the company. This can also lead to overfitting or biased data errors, resulting in incorrect final decisions. It is therefore important to define a model carefully, to understand whether or not it can be adapted to the case study under consideration, and to use it as a support tool without neglecting human intuition about the problem domain. Future work aims to use algorithms that are more specific to the case study and to avoid using the black box approach. An additional forthcoming goal is to increase the quantity and quality of data available on the production process in order to refine the predictive models defined.

Acknowledgements

This paper is part of the project NODES which has received funding from the MUR – M4C2 1.5 of PNRR funded by the European Union - NextGenerationEU (Grant agreement no. ECS00000036).

References

[1] Parfitt J, Barthel M, Macnaughton S. Food waste within food supply chains: quantification and potential for change to 2050. Philos Trans R Soc B Biol Sci 2010;365:3065–81. https://doi.org/10.1098/rstb.2010.0126.
[2] Corrado S, Caldeira C, Eriksson M, Hanssen OJ, Hauser H-E, van Holsteijn F, et al. Food waste accounting methodologies: Challenges, opportunities, and further advancements. Glob Food Sec 2019;20:93–100. https://doi.org/10.1016/j.gfs.2019.01.002.
[3] Lieder M, Rashid A. Towards circular economy implementation: a comprehensive review in context of manufacturing industry. J Clean Prod 2016;115:36–51. https://doi.org/10.1016/j.jclepro.2015.12.042.
[4] Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. Science (80-) 2015;349:255–60. https://doi.org/10.1126/science.aaa8415.
[5] Mahesh B. Machine Learning Algorithms - A Review. Int J Sci Res 2020;9:381–6.
[6] Caruana R, Niculescu-Mizil A. An empirical comparison of supervised learning algorithms. Proc. 23rd Int. Conf. Mach. Learn. - ICML '06, New York, New York, USA: ACM Press; 2006, p. 161–8. https://doi.org/10.1145/1143844.1143865.
[7] Chang L-Y, Chen W-C. Data mining of tree-based models to analyze freeway accident frequency. J Safety Res 2005;36:365–75. https://doi.org/10.1016/j.jsr.2005.06.013.
[8] Fritzke B. Growing cell structures—A self-organizing network for unsupervised and supervised learning. Neural Networks 1994;7:1441–60. https://doi.org/10.1016/0893-6080(94)90091-4.
[9] Natekin A, Knoll A. Gradient boosting machines, a tutorial. Front Neurorobot 2013;7. https://doi.org/10.3389/fnbot.2013.00021.
[10] Zhao Z-Q, Zheng P, Xu S-T, Wu X. Object Detection With Deep Learning: A Review. IEEE Trans Neural Networks Learn Syst 2019;30:3212–32. https://doi.org/10.1109/TNNLS.2018.2876865.
[11] Gobble MM. Big Data: The Next Big Thing in Innovation. Res Manag 2013;56:64–7. https://doi.org/10.5437/08956308X5601005.
[12] Tsakanikas P, Karnavas A, Panagou EZ, Nychas G-J. A machine learning workflow for raw food spectroscopic classification in a future industry. Sci Rep 2020;10:11212. https://doi.org/10.1038/s41598-020-68156-2.
[13] Garre A, Ruiz MC, Hontoria E. Application of Machine Learning to support production planning of a food industry in the context of waste generation under uncertainty. Oper Res Perspect 2020;7:100147. https://doi.org/10.1016/j.orp.2020.100147.
[14] Barbato G, Barini EM, Genta G, Levi R. Features and performance of some outlier detection methods. Http://DxDoiOrg/101080/026647632010545119 2011;38:2133–49. https://doi.org/10.1080/02664763.2010.545119.
[15] Draper NR, Smith H. Applied Regression Analysis. Wiley; 1998. https://doi.org/10.1002/9781118625590.
[16] McDonald GC. Ridge regression. WIREs Comput Stat 2009;1:93–100. https://doi.org/10.1002/wics.14.
[17] Biau G, Scornet E. A random forest guided tour. TEST 2016;25:197–227. https://doi.org/10.1007/s11749-016-0481-7.
[18] Friedman JH. Stochastic gradient boosting. Comput Stat Data Anal 2002;38:367–78. https://doi.org/10.1016/S0167-9473(01)00065-2.
[19] Altmann A, Tolosi L, Sander O, Lengauer T. Permutation importance: a corrected feature importance measure. Bioinformatics 2010;26:1340–7.