Application of Machine Learning for Hot Carcass Yield Prediction in Beef Cattle

Angelo Polizel Neto
Institute of Agricultural and Technological Sciences
Federal University of Rondonopolis
Rondonópolis, MT, Brazil.
polizel.neto@ufr.edu.br

Elton Fernandes dos Santos
Startup PROMEAT LTDA
São Paulo, SP, Brazil.
elton.fernandes@promeat.ai

João Pedro Lanzarini Lopes
Startup PROMEAT LTDA
São Paulo, SP, Brazil.
joao.lopes@promeat.ai

Marçal Henrique Moreira
Startup PROMEAT LTDA
São Paulo, SP, Brazil.
marcal.moreira@promeat.ai

Julio Cesar Machado Alvares
Startup PROMEAT LTDA
São Paulo, SP, Brazil.
julio.cesar@promeat.ai

Cecylyana Leite Cavalcante
Startup PROMEAT LTDA
São Paulo, SP, Brazil.
cecylyana.leite@promeat.ai

Evelyn Prestes Brito
Startup PROMEAT LTDA
São Paulo, SP, Brazil.
evelyn.brito@promeat.ai

Guilherme Roberto Matos Silva
Startup PROMEAT LTDA
São Paulo, SP, Brazil.
guilherme.silva@promeat.ai

Emerson Amaral Santos
Startup PROMEAT LTDA
São Paulo, SP, Brazil.
emerson.amaral@promeat.ai

Priscila Dias da Silva
Startup PROMEAT LTDA
São Paulo, SP, Brazil.
priscila.silva@promeat.ai

Rafael Sarto Zaratin
Startup PROMEAT LTDA
São Paulo, SP, Brazil.
rafael.zaratin@promeat.ai

***Abstract*— Hot Carcass Yield (HCY) is a critical indicator of both productive efficiency and economic value in the meat processing industry. This study presents the development and evaluation of predictive models based on Machine Learning (ML) techniques to accurately predict HCY using supervised learning algorithms trained on comprehensive historical slaughterhouse data. The dataset includes key variables such as breed, sex, age, and fat cover, which enabled the models to capture complex non-linear relationships influencing yield outcomes. Multiple ML algorithms were rigorously assessed through cross-validation to determine their predictive performance. The results reveal that the developed models achieve high accuracy and robustness, demonstrating their practical applicability for improving monitoring, management, and decision-making in industrial and zootechnical processes. This research underscores the transformative potential of Machine Learning in precision livestock farming, advancing quality control, traceability, and data-driven optimization in the meat industry.**

Keywords— Computational Intelligence; Supervised Learning Models; Productive Efficiency

I. Introduction

The pursuit of greater efficiency in the meat processing industry has driven the adoption of intelligent technologies, such as Artificial Intelligence (AI), to optimize production stages, reduce economic losses, and identify performance issues related to livestock productivity. AI aims to build systems capable of autonomously learning from large datasets and applying this knowledge to solve complex problems in a manner analogous to human reasoning [1].

Hot carcass yield (HCY) is calculated by comparing the carcass weight immediately after slaughter to the animal's live body weight, serving as a fundamental indicator in evaluating both productivity and economic return in beef production systems [2]. It has also become an important trait for producers (Santos et al., 2008). Factors such as breed, sex, age, and transportation from the farm of origin directly influence HCY, making its estimation a complex and multidimensional task [2].

In this context, the advancement of Artificial Intelligence (AI) techniques, particularly machine learning models, has proven to be a promising strategy for uncovering hidden patterns in agro-industrial datasets. These models enable predictions in scenarios that escape human perception and are capable of learning nonlinear relationships among multiple variables [3], thus contributing to accurate HCY predictions and the identification of key factors that may affect this indicator.

By replacing traditional methods based solely on average and classical statistical analysis, AI algorithms can provide more robust individual predictions, enabling more strategic decision-making along the production chain [1]. Moreover, by comparing the predicted yield with the observed yield in the processing plant, it becomes possible to diagnose operational deviations, inconsistencies in slaughter procedures, or undesired variations in the quality of processed cattle lots providing valuable insights for both producers and processors.

Therefore, this study aimed to develop and evaluate a machine learning model for predicting the hot carcass yield of cattle slaughtered at a meat processing facility, based on zootechnical and productive data. This approach seeks not only to increase the accuracy of HCY estimation but also to offer an intelligent tool to support industrial management, fostering greater standardization, traceability, and efficiency in operational processes.

II. THEORETICAL FRAMEWORK

A. Machine Learning Models for Predictionod

The term *Machine Learning* (ML) refers to a branch of Artificial Intelligence (AI) dedicated to the development of systems capable of autonomously learning from data, without the need for explicit programming. These algorithms are trained on structured or unstructured data, either labeled or unlabeled, to identify patterns and make predictions [1].

The goal of machine learning is to enable machines to interpret large volumes of data and extract relevant information that often remains hidden to human observation. With the exponential growth of available data, the demand for ML-based solutions has been increasing across various industrial sectors. The core idea is to allow machines to learn from data, using different computational approaches to solve complex problems more efficiently [4].

ML has become an effective tool for prediction tasks in fields such as healthcare, finance, agri-food industry, and meteorology. Its main advantage lies in the ability to model complex and nonlinear relationships between variables, even in scenarios involving multiple interdependent factors and large datasets. Among the most widely used algorithms are: Random Forest, Support Vector Machines (SVM), logistic regression, deep neural networks, and boosting techniques such as Gradient Boosting [5].

B. Model Interpretation and Decision Making

The growing adoption of machine learning (ML) models has led to significant advances in forecasting key indicators and optimizing production processes. However, the complexity of some of these models, particularly "black box" models such as deep neural networks can hinder the interpretability of their outputs, which is essential in contexts where transparency and decision justification are critical.

Model interpretability in ML refers to the ability to understand how and why a model makes certain predictions. This aspect is crucial for ensuring system reliability, facilitating validation, and supporting operational adjustments. Model interpretability can be addressed through two principal strategies: models designed for transparency by nature, such as decision trees and linear regressions; and post-hoc techniques, which aim to explain complex models retrospectively after the learning process is complete [6].

Employed decision trees to investigate how different attribute values influence the outcomes of predictive models, offering a clear view of the impact of each variable. This approach allows for strategic adjustments based on the analysis of the most influential factors affecting process performance [7].

Furthermore, interpretability techniques such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) have been widely adopted to elucidate the behavior of complex models. While often applied in educational contexts, as demonstrated by [8], these tools are equally valuable in agro-industrial environments, enabling managers to better understand the factors influencing production and to make more informed decisions.

The integration of interpretable models into the decision-making process supports in-depth analysis of production systems, contributing to the identification of bottlenecks, the improvement of operational stages, and the consolidation of trust in ML-based technologies.

C. Evaluation Metrics for Regression Models

To assess the predictive performance of machine learning models in regression tasks, such as predicting Hot Carcass Yield (HCY), it is essential to employ quantitative evaluation metrics that reflect both accuracy and reliability. The choice of appropriate metrics allows for a comprehensive understanding of how well the model generalizes unseen data and supports decision-making in practical applications.

Mean Absolute Error (MAE) measures the average magnitude of the absolute differences between predicted and observed values. It is simple to interpret and provides a direct understanding of the average error in the same unit as the target variable. In meat production systems, lower MAE values suggest precise yield estimates, which are crucial for operational efficiency and financial forecastingn [9].

Root Mean Square Error (RMSE) is a commonly used metric that measures the square root of the average of squared errors. Unlike MAE, RMSE penalizes larger errors more heavily, making it useful when outliers or large deviations are critical. In carcass yield prediction, a low RMSE suggests consistent predictions across different animals [10].

Coefficient of Determination (R^2) indicates the proportion of variance in the dependent variable (HCY) that is predictable from the independent variables. Values close to 1.0 denote a model with high explanatory power. In the context of livestock production, a high R^2 reflects the model's ability to capture complex relationships between animal traits and carcass output [11].

Mean Absolute Percentage Error (MAPE) The Mean Absolute Percentage Error (MAPE) is a statistical metric that quantifies forecast accuracy by representing the average absolute error as a percentage relative to actual values. It is particularly useful when comparing performance across different datasets or production systems. In agronomic and animal science applications, a MAPE below 10% is often considered acceptable, while values below 2% are indicative of excellent predictive capability [12].

These metrics, when interpreted together, provide a robust assessment of the model's accuracy, sensitivity to error magnitude, and practical applicability in cattle production environments. Their combined use strengthens confidence in the model's predictions and supports its integration into industrial decision-making processes.

D. Application of Machine Learning Models

The use of machine learning (ML) models has shown great promise in industrial settings, particularly in optimizing processes, enhancing operational efficiency, and improving quality control. These models are capable of processing large volumes of data and identifying patterns that directly impact production performance.

In the food sector, the potential of supervised classification was demonstrated through the application of neural networks to distinguish between normal milk and milk adulterated with cheese whey. Physicochemical variables such as pH, fat content, protein levels, and density were used to train the model. After training, the neural network was tested with previously unseen data, validating its effectiveness in the automated detection of fraud [13].

In animal production, the Random Forest algorithm was employed to predict the market value of Wagyu beef carcasses using both phenotypic characteristics and image-derived data. The model achieved a coefficient of determination (R^2) of 0.86, highlighting its strong ability to model nonlinear relationships among multiple variables, with direct implications for fair pricing and optimization of the commercialization process [14].

Additionally, ML models have been used for early disease diagnosis, classification of cattle breeds based on morphological traits, and identification of pathogenic microorganisms. The potential of algorithms in the development of automated systems for early disease detection was demonstrated in [15]. Classifiers were employed to distinguish cattle breeds using morphometric measurements [16], while computational models were applied for detecting bacteria harmful to animal health [17].

In a study focused on herd health, several algorithms such as logistic regression, k-nearest neighbors (KNN), Classification and Regression Trees (CART), and Random Forest were compared to predict infections caused by *Fasciola hepatica*. The use of variables such as age, sex, and carcass conformation, combined with normalization techniques, contributed to the increased accuracy of ML-based models compared to traditional methods [18].

These applications demonstrate how ML has become an indispensable tool for decision-making in the meat industry, enabling the prediction of zootechnical indicators, process optimization, and assurance of product quality.

E. Hot Carcass Yield (HCY)

Hot carcass yield (HCY) is a widely used indicator in the beef cattle industry to measure production efficiency. This parameter is defined as the percentage ratio between the carcass weight immediately after slaughter and the animal's live weight. Its application is relevant from both technical and economic perspectives, serving as a key criterion in animal performance evaluation and commercial negotiations [19][20].

Several factors influence HCY, including breed, sex, age, diet, production system, pre-slaughter management, and the animals' health status [21]. The interaction among these factors can significantly affect body composition and, consequently, the yield obtained at slaughter.

In the meat industry, HCY plays a strategic role in the zootechnical analysis of batches and in the definition of bonuses and pricing based on performance. The ability to accurately estimate this indicator from previously known variables has motivated the application of statistical modeling and machine learning techniques for predictive purposes [14][18].

The use of predictive models to estimate HCY contributes to reducing uncertainty in the production process, enabling more efficient resource allocation, better logistical planning, and more accurate decisions regarding carcass destination [22].

III. METHODOLOGY

A. Data Source and Description

The dataset used in this study was obtained from operational records of a beef cattle producer located in the state of Mato Grosso, Brazil, covering the period from January to October 2024 and comprising 23,469 animals. Each row in the dataset represents a slaughtered animal and includes zootechnical and production-related information. Table 1 summarizes the main variables included in the dataset, along with their types and descriptions.

Table 1. Description of variables included in the dataset used for predicting hot carcass yield.

Variable	Type	Description
SisBov	Categorical	Unique animal identifier
Farm of origin	Categorical	Location where the animal was raised
Responsible company	Categorical	Company responsible for the slaughter
Breed	Categorical	Animal breed
Sex	Categorical	Animal sex
Fat finishing	Categorical	Visual classification of fat cover
Live weight (Farm weight)	Numerical	Animal weight recorded at the farm
Hot carcass weight	Numerical	Animal weight immediately after slaughter
Slaughter lot ID	Categorical	Identifier of the slaughter lot to which the animal belongs

The variables were initially standardized and normalized to perform comparative analyses. Data processing included cleaning textual strings, type conversion, and attribute categorization

B. Features Engineering and Clustering

Derived variables were created and transformations were performed based on zootechnical literature. Table 2 presents the main derived variables, their types, and respective purposes.

Table 2. Key derived variables used in the study, including variable type and intended application.

Variable	Type	Description
HCY	Numerical	Hot carcass yield: ratio between carcass weight and live weight
CWP	Numerical	Carcass weight divided by 15, used for auxiliary categorization
CWC	Categorical	Categorical range of carcass weight: 0 to 3 based on weight intervals (kg)
FA	Categorical	Numerical grouping of fat finishing on an ordinal scale (1 to 4)
BRG	Categorical	Textual grouping of breed into Nelore, Angus or Others
BRG_N	Categorical	Numerical encoding of BRG: Nelore=1, Angus=2, Others=3
AGE	Categorical	Age grouping based on dental stage (0 to 2)
SEX_B	Categorical	Binary encoding of sex: Male=0, Female=1

A brief description of the variables derived used in this study is presented below:

- **HCY**: The ratio between the hot carcass weight and the live weight of the animal, expressed as a percentage, calculated as:

HCY (%)=(Live WeightHot Carcass Weight)×100

- **PAC** An auxiliary variable obtained by dividing the carcass weight by 15, used for normalization and comparative analysis.
- **PCA**: A categorical classification of carcass weight into four ranges:
- - 0: Carcass weight up to 270 kg
- - 1: Carcass weight between 271 kg and 300 kg
- - 2: Carcass weight between 301 kg and 330 kg
- - 3: Carcass weight above 330 kg
- **AA (Fat Finishing)**: An ordinal scale derived from textual values related to fat cover classification.
- **RA(Breed Grouping)**: Simplified textual categorization of animal breed into Nelore, Angus, and Others.
- **RAN**: Numerical coding of the RA variable, where Nelore = 1, Angus = 2, and Others = 3.
- **IA (Age Grouping)**: Classification of animals based on dentition stage, divided into three age groups.
- **SA (Animal sexy)**: A binary categorical variable representing animal sex, where Male = 0 and Female = 1.

These derived variables were created to enable the application of machine learning algorithms using structured and categorical data, thereby enhancing the model's ability to capture relevant patterns for predicting hot carcass yield.

C. *Applied Groupings*

The variable related to fat finishing, originally presented in textual form with various notations (e.g., "1", "2-", "2=", "2+", "3", among others), was converted into an ordinal numerical scale ranging from 1 to 4. This standardization simplifies the classification of the degree of visual fat cover observed in slaughtered animals, facilitating its use in predictive models. Table 3 presents the applied groupings and the correspondence between the original textual values and the standardized numerical scale adopted in this study.

Table 3 – Ordinal scale for fat finishing: correspondence between the original textual classifications and the standardized numerical values used in this study.

Original Value	AA group
1, 2-, 2=	1
2+, 3	2
3=	3
3+, 4, 5	4

This mapping contributes to the homogenization of the variable and enables better integration with supervised modeling algorithms [23][24].

The classification of the animals' breed was reorganized into three main categories, as detailed in Table 4, aiming to simplify the analysis and enhance predictive modeling.

Table 4 – Breed categories of the animals used in the study, with respective grouped classifications.

Original Breed	RA Group	RAN Code
Nelore	Nelore	1
Angus, Aberdeen	Angus	2
Braford, Guzerá etc.	Outros	3

This categorization considered, in addition to the statistical representativeness of the breeds in the dataset, the phenotypic characteristics associated with productive and zootechnical performance observed in different genetic groups [25][26].

The animals' age variable was estimated based on dentition assessment; a procedure widely used in beef cattle production to indicate the stage of development and maturity of the animals [26]. For modeling purposes, dentition was categorized into ranges that reflect different levels of zootechnical maturity, as presented in Table 5.

Table 5 — Age categories of animals based on dentition assessment, representing different stages of zootechnical development.

Dentition (pairs of teeth)	Age Group	Interpretation
0 and 2 teeth	0	Very young animals
4 teeth	1	Young animals in transition
6 and 8 teeth	2	More mature animals / finishing stage

D. *Transformations and Normalization*

In addition to the groupings and transformations, the continuous numerical variable Farm Weight was standardized to ensure that its scale would not disproportionately influence the modeling process. Standardization was performed using the z-score method, which transforms the data to a distribution with zero mean and unit standard deviation [24].

This transformation was incorporated into a pipeline that also includes one-hot encoding for the categorical variables Sex (SA), Breed (RA), Fat Finishing (AA), and Age.

Standardization is a recommended practice in the literature for supervised learning models, preventing variables with different magnitudes from dominating the learning process [24].

These transformations enabled the dataset to be organized into a format suitable for the application of machine learning algorithms, whose methodological description is presented in the next section.

E. *Predictive Modeling*

The supervised learning model adopted in this study was the Random Forest Regressor, implemented using the Scikit-Learn library. The objective of the modeling was to predict the hot carcass weight, from which the hot carcass yield (HCY) can be calculated using a simple formula based on the animal's live weight. The choice of Random Forest is justified by its ability to capture nonlinear relationships among attributes and its overall strong performance in multivariate regression tasks with heterogeneous data [23]. Thus, although the HCY was not directly predicted by the model, the prediction of hot carcass weight provides a reliable basis for the subsequent calculation of this index.

F. *Model Evaluation*

The dataset was split into training and testing sets, using 80% of the data for training and 20% for testing. The model's performance on the test set was evaluated based on the following metrics:

- **Mean Absolute Error (MAE):** measures the average of the absolute errors between the predictions and the actual values.
- **Root Mean Squared Error (RMSE):** measures the average of the squared errors, giving greater weight to larger errors.
- **R^2 (Coefficient of Determination):** measures the proportion of the variance in the data explained by the model.
- **Mean Absolute Percentage Error (MAPE):**measures the average percentage error between the predictions and the actual values.

These metrics are recommended for evaluating regression models as they provide different perspectives on model performance in terms of accuracy, robustness, and interpretability [10].

IV. Results and Discussion

After data preprocessing and transformation, the Random Forest Regressor was trained using 80% of the dataset, while the remaining 20% was reserved for model evaluation. The main objective was to predict the Hot Carcass Yield (HCY) of slaughtered cattle using zootechnical and production-related variables. This modeling approach allows for accurate and robust yield forecasting, which is crucial for enhancing production planning and decision-making within the meat industry. By estimating carcass performance based on live weight, the model offers a strategic tool capable of influencing both economic efficiency and operational outcomes in meat processing systems.

A. *Model Performance*

The model's performance was evaluated using widely recognized metrics in the literature for regression problems, providing a comprehensive analysis of the accuracy, robustness, and explanatory power of the generated predictions. The adopted metrics and their respective results on the test set are presented in Table 6.

Table 6 – Evaluation metrics of the Random Forest Regressor model performance on the test dataset.

Metric	Obtained Value
MAE (Mean Absolute Error)	0.0119
RMSE (Root Mean Squared Error)	0.0161
R^2 (Coefficient of Determination)	0.9497
MAPE (Mean Absolute Percentage Error)	1.28%

The results obtained indicate that the model exhibits high predictive accuracy. An R^2 value of 0.9497 demonstrates that approximately 95% of the variability in Hot Carcass Yield (HCY) is explained by the variables used, characterizing a well-fitted model [10]. Furthermore, the MAE of 0.0119 and RMSE of 0.0161 indicate that both the mean absolute error and the root mean squared error are low, meaning the average difference between predicted and actual values is minimal values close to zero are desirable for these metrics [9].

Additionally, the MAPE of 1.28% shows that the average percentage error between predicted and actual values remains below 2%, a threshold considered acceptable for regression tasks in agricultural and animal science domains [12].

These results demonstrate that the model exhibits strong generalization capability and is suitable for practical applications in predicting hot carcass yield (HCY) in cattle. The high coefficient of determination (R^2 = 0.9497) and the low mean absolute percentage error (MAPE = 1.28%) indicate a high predictive accuracy, which is particularly relevant in production systems where small variations in yield directly affect profitability. Furthermore, the low absolute error reinforces the model's ability to reliably reflect the animals' biological efficiency, expressed by the conversion of live weight into marketable carcass mass. These findings highlight the model's potential as a decision-support tool in the beef industry, contributing to the selection of animals with higher yield and improved zootechnical performance [27].

B. Dispersion of Actual vs. Predicted Values

Fig. 1 illustrates the dispersion between the actual and predicted values of Hot Carcass Yield (HCY).

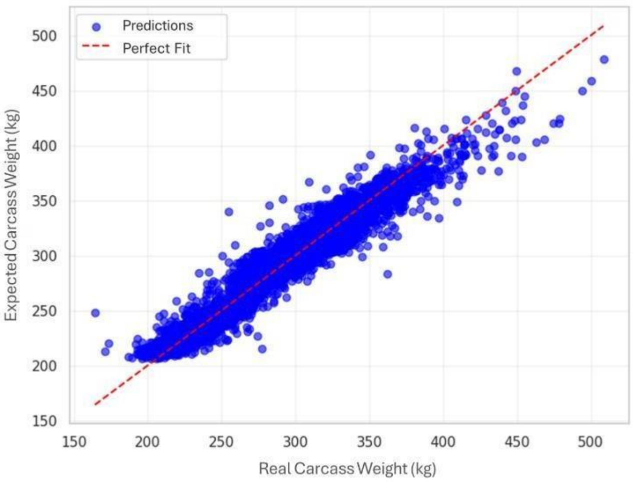

Fig. 1: Dispersion of actual versus predicted values of Hot Carcass Yield (HCY).

The proximity of the points to the perfect fit line indicates a strong predictive capability of the model, as recommended by [28] for the visual assessment of predictive model performance.

The homogeneous distribution of points around the fit line suggests that the model performs well across different ranges of Hot Carcass Yield (HCY), without bias toward specific subranges. This balanced dispersion, observed along the diagonal, reinforces the model's ability to capture variability in HCY across diverse animal profiles, thereby enhancing its practical applicability for carcass classification and pricing strategies.

C. Análise dos resíduos

Fig. 2 shows the model residuals, highlighting potential outliers based on a confidence interval defined by the parameter k=k

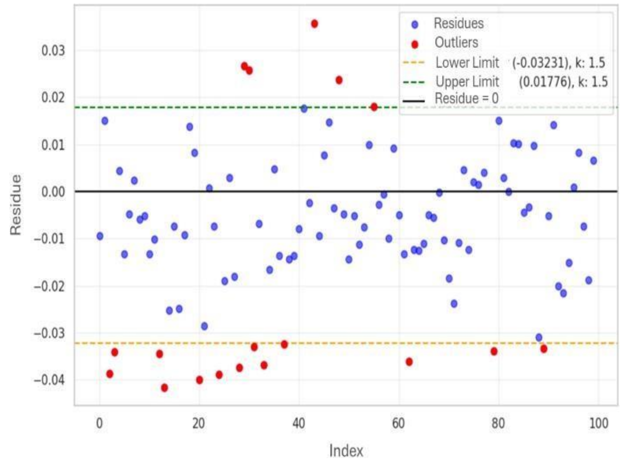

Fig. 2: Residuals of the Random Forest Regressor model for predicting hot carcass yield (HCY), highlighting outliers outside the confidence interval defined by k=k.

The residuals are mostly concentrated near zero, indicating the absence of systematic bias in the predictions. According to [29], for a good regression, the residuals should be approximately symmetrically distributed around zero and free of systematic patterns.

Although some outliers were identified in the residual analysis (Fig. 2), their presence does not compromise the overall performance of the model.

Residual analysis also contributes to assessing the robustness of HCY prediction. Since the deviations remain low and well distributed, there is evidence that the model does not exhibit systematic bias, which strengthens its reliability as a decision-support tool for industrial management.

D. Residuals Distribution

Fig. 3 illustrates the distribution of the model residuals, highlighting their concentration and spread.

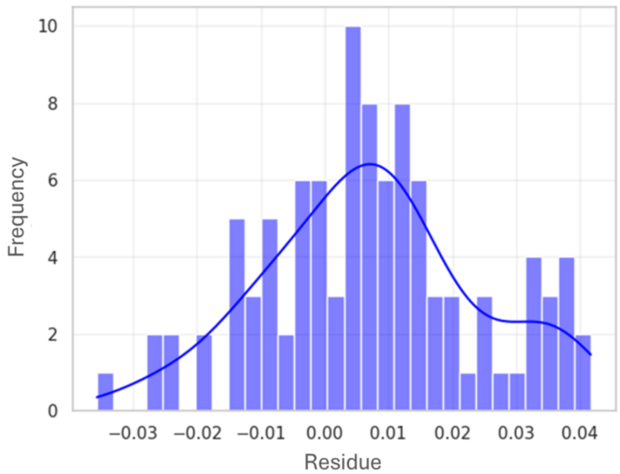

Fig. 3. Distribution of residuals from the Random Forest regression model, showing the concentration of errors around zero and the overall variance in prediction errors.

The visualization reveals that the residuals are predominantly centered around zero, indicating minimal systematic bias and supporting the model's accuracy across the dataset. This balanced distribution is essential for validating the reliability of the predictions.

The approximately symmetrical distribution, resembling a normal distribution, is a positive indication of the quality of the fit, as it meets the assumptions of residual normality [29].

The low dispersion of residuals in this sample suggests that the model maintains its predictive accuracy even in individual observations. This stability is particularly advantageous for slaughterhouses seeking consistency in carcass yield evaluation, both in operational and strategic contexts.

E. Sample of Predictions

Table 7 presents a sample of 10 observed HCY values, their corresponding predictions generated by the model, and the resulting residuals (difference between observed and predicted values).

Table 7 – Sample of observed HCY values, model predictions, and residuals.

Observed	Predicted	Residual
0.532787	0.523317	0.009470
0.515766	0.530789	-0.015023
0.562642	0.524038	0.038604
0.559579	0.525440	0.034139
0.523148	0.527483	-0.004335
0.541371	0.528037	0.013334
0.527778	0.522912	0.004866
0.527381	0.529768	-0.002387
0.530108	0.524274	0.005834
0.533493	0.528401	0.005092

The mean of the observed values was 0.53, identical to the mean of the predictions, which reflects the model’s accurate average performance regarding central tendency. Residuals in the sample were close to zero, demonstrating strong alignment between predicted and observed values. Additionally, the residuals’ standard deviation was low (0.0165), highlighting the model’s consistency and low variability in predictions.

Carcass yield is known to be influenced by multiple factors, including breed, age, sex, and feeding system. These variables impact both the efficiency of body mass utilization and meat quality, indicating that production strategies aiming to improve these outcomes should incorporate such parameters for better animal selection and system optimization [30].

Predicting hot carcass yield using zootechnical variables presents an innovative and practical approach. The findings from this study illustrate that machine learning techniques like the Random Forest Regressor offer robust technical support for operational decision-making in slaughterhouses, enhancing production efficiency and traceability. By pinpointing the most influential factors on hot carcass yield, this method enables targeted zootechnical and strategic interventions to optimize herd performance from the source.

Accurate hot carcass yield prediction across different production categories holds strategic importance for producers and the meat processing industry alike. It also informs cost management, facilitates fairer payments to livestock suppliers, and supports the development of competitive retail pricing strategies [31].

V. CONCLUSION

The results of this study highlight the potential of machine learning algorithms, particularly the Random Forest Regressor, for practical applications within the beef production chain. The high predictive accuracy achieved evidenced by an R^2 close to 0.95 and a mean absolute percentage error below 2% demonstrates that Hot Carcass Yield (HCY) can be reliably estimated using routinely collected zootechnical and production variables.

This modeling approach not only supports improvements in industrial processes such as carcass classification and pricing, but also informs strategic decision-making across the supply chain, from nutritional management to genetic selection. The model’s consistent performance across different yield ranges enhances its applicability to a wide range of production contexts.

Moreover, by identifying the key factors influencing HCY, the model provides technical insights that can guide targeted interventions aimed at improving the biological efficiency of animals and increasing overall system profitability.

Future research may explore hybrid approaches or incorporate genomic data and carcass imaging through computer vision, further enhancing both the accuracy and applicability of predictive technologies in slaughterhouses and animal breeding programs.

ACKNOWLEDGMENT

We would like to thank Finep and MCTIC for their financial support, with resources from FNDCT, granted to the Inovadoc Project – Reference 0779/23 (Contract No. 03.24.0206.00) and by the National Council for Scientific and Technological Development (CNPq, process number 409694/2022-3). We also extend our gratitude to the Novapec group, represented by Arlindo Vilela, for their valuable collaboration.

REFERENCES

[1] D. P. Castro et al., 'Artificial intelligence applied to animal production,' Ciência Rural, vol. 55, 2025.

[2] D. Olmedo et al., 'Desempenho e características da carcaça de novilhos terminados em pastejo rotacionado ou em confinamento,' Arq. Bras. Med. Vet. Zootec., vol. 63, no. 2, pp. 348–355, 2011.

[3] P. Joshi, Artificial Intelligence with Python. Mumbai, India: Packt, 2017.

[4] B. Mahesh, 'Machine Learning Algorithms - A Review,' Int. J. Sci. Res., vol. 9, no. 1, pp. 381–386, 2020.

[5] L. Alves, 'Recomendação de algoritmos de aprendizado de máquina para predição de falhas de software por meio de meta-aprendizado,' M.Sc. dissertation, PUCRS, 2016.

[6] W. J. Murdoch et al., 'Interpretable machine learning: definitions, methods, and applications,' arXiv:1901.04592, 2019.

[7] J. J. D. Silva, 'Interpretabilidade de modelos de aprendizado de máquina: uma abordagem baseada em árvores de decisão,' M.Sc. dissertation, UFG, 2023.

[8] F. C. Silva and A. M. Santana, 'Explicabilidade dos modelos de aprendizado de máquina no cenário educacional: uma revisão sistemática,' RENOTE, vol. 22, no. 1, 2024.

[9] C. J. Willmott and K. Matsuura, 'Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance,' Clim. Res., vol. 30, no. 1, pp. 79–82, 2005.

[10] T. Chai and R. R. Draxler, 'Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature,' Geosci. Model Dev., vol. 7, no. 3, pp. 1247–1250, 2014.

[11] D. C. Montgomery, E. A. Peck, and G. G. Vining, Introduction to Linear Regression Analysis, 5th ed. Hoboken: Wiley, 2012.

[12] R. J. Hyndman and A. B. Koehler, 'Another look at measures of forecast accuracy,' Int. J. Forecast., vol. 22, no. 4, pp. 679–688, 2006.

[13] G. F. S. Valente et al., 'Aplicação de redes neurais artificiais como teste de detecção de fraude de leite por adição de soro de queijo,' Rev. Inst. Laticínios Cândido Tostes, vol. 69, no. 6, pp. 425–432, 2014.

[14] T. Tamagawa et al., 'Prediction of carcass unit price based on image analysis and carcass grading traits using Random Forest in Japanese Black cattle,' Anim. Sci. J., vol. 95, no. 3, p. 211, 2024.

[15] S. Ghosh and R. Dasgupta, Machine Learning in Biological Sciences: Updates and Future Prospects. Singapore: Springer Nature, 2022.

[16] P.-M. Parés Casanova, I. Sinfreu Basi, and D. Villalba Mata, 'Principal component analysis of cephalic morphology to classify some Pyrenean cattle,' Anim. Genet. Resour., vol. 50, pp. 59–64, 2012.

[17] M. L. Hermann-Bank et al., 'Characterization of the bacterial gut microbiota of piglets suffering from new neonatal porcine diarrhoea,' BMC Vet. Res., vol. 11, art. 139, 2015.

[18] M. Ergin and Ö. Koşkan, 'Classification performances of several machine learning algorithms and logistic regression for Fasciola hepatica in cattle,' Pesq. Agropec. Bras., vol. 59, e03563, 2024.

[19] L. F. G. Menezes, Avaliação de novilhos das gerações avançadas do cruzamento rotativo Charolês – Nelore, M.Sc. dissertation, UFSM, 2004.

[20] P. L. de P. Rezende et al., 'Carcass characteristics of feedlot-finished nellore heifers slaughtered at different weights,' Acta Sci. Anim. Sci., vol. 41, no. 1, pp. 1–8, 2019.

[21] J. Restle et al., 'Efeito da suplementação energética sobre a carcaça de vacas de diferentes idades,' Rev. Bras. Zootec., vol. 30, no. 3, pp. 1076–1083, 2001.

[22] D. Matthews et al., 'Predicting carcass cut yields in cattle from digital images using artificial intelligence,' Meat Sci., vol. 184, 108671, 2022.

[23] S. B. Kotsiantis, 'Supervised machine learning: a review of classification techniques,' Informatica, vol. 31, no. 3, pp. 249–268, 2007.

[24] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed., Amsterdam: Elsevier, 2011.

[25] L. M. Menezes et al., 'Avaliação do desempenho de bovinos de diferentes grupos genéticos em sistema de terminação,' Rev. Bras. Zootec., vol. 35, no. 6, pp. 2343–2351, 2006.

[26] K. Euclides Filho, Melhoramento Genético Aplicado à Produção Animal. Campo Grande: Embrapa Gado de Corte, 2001.

[27] M. P. Santos et al., 'Prediction of hot carcass yield in beef cattle using fuzzy expert system,' Braz. J. Biosyst. Eng., vol. 8, no. 2, pp. 158–164, 2014.

[28] M. Kuhn and K. Johnson, Applied Predictive Modeling. New York: Springer, 2013.

[29] D. C. Montgomery, E. A. Peck, and G. G. Vining, Introduction to Linear Regression Analysis, 5th ed. Hoboken: Wiley, 2012.

[30] T. Sakowski et al., 'Genetic and Environmental Determinants of Beef Quality—A Review,' Animals, vol. 12, no. 4, p. 528, 2022.

[31] P. S. Pacheco et al., 'Carcass physical composition and meat quality of Charolais cattle of different categories,' Rev. Bras. Zootec., vol. 40, 2011.