http://doi.org/10.1590/1678-992X-2023-0002
ISSN 1678-992X
Scientia Agricola
Research article

Decision trees as a tool for selecting sows in commercial herds

João Otávio Hilgemberg[1], Ines Andretta[2], Alexandre Bonadiman Mariani[2], Alisson Neimaier[3], Marcio Valk[3], Fernando Bittarello[1], Rafaela Hilgemberg[1], Cheila Roberta Lehnen[1*]

[1]Universidade Estadual de Ponta Grossa – Depto. de Zootecnia, Av. General Carlos Cavalcanti, 4748 – 84030-900 – Ponta Grossa, PR – Brasil.
[2]Universidade Federal do Rio Grande do Sul/Faculdade de Agronomia, Av. Bento Gonçalves, 9500 – 91501-970 – Porto Alegre, RS – Brasil.
[3]Universidade Federal do Rio Grande do Sul/Instituto de Matemática e Estatística, Av. Bento Gonçalves, 7712 – 91540-000 – Porto Alegre, RS – Brasil.
*Corresponding author <cheilalehnen@gmail.com>

Edited by: José Fernando Machado Menten

Received January 28, 2023
Accepted April 14, 2023

ABSTRACT: This study aimed to evaluate the use of decision trees to select sows based on the production parameters of parity order (PO) 1 sows from a commercial herd. Data were collected at a piglet production unit with a capacity of housing 5,500 sows in collective pens. Piglet production and sow culling information was collected from PO1 and PO2 sows. The period from January 2017 to March 2020 was analyzed. The correlation analysis was used to identify the influence of the production parameters on sow culling after exploring the database using the graphical analysis and descriptive statistics. The ANOVA was applied to evaluate differences in the response variables between culled and unculled sows. Two models were proposed using the decision tree method: model 1 referred to sow culling, and model 2 comprised the total number of liveborn piglets (TBA). The calculated value was close to 0, although the correlations of the production parameters with culling were statistically significant. The mean number of weaned piglets was higher for unculled sows in PO1 ($p < 0.05$). The number of weaned piglets, total number of liveborn piglets, and weaning-service interval did not differ in the unculled and culled sows in PO2 ($p > 0.05$). Using a confusion matrix as a metric tool, the decision tree method used in this study provided consistent results for this database, indicating its possible use for decision-making in sow selection.
Keywords: culling, data, machine learning, piglets, pig production

Introduction

Productive performance is one of the main selection criteria in animals reared in intensive systems. Although the productive performance of sows has improved considerably in recent years due to advances in genetic improvement, a wide variation in performance is still common (Baxter et al., 2020). Several factors are associated to low productivity and sow culling. High weight loss during lactation, feed restriction, high temperatures, long lactations, weaning-to-service intervals, diseases, injuries, reproductive failure, and irregular return to estrus are some factors that can reduce productivity and lead to the early culling of sows (Koketsu et al., 2017). Therefore, the capacity to predict which sows have high reproductive performance and high longevity to facilitate culling decisions is advantageous for producers (Iida and Koketsu, 2015) to increase productivity and reduce production costs. In this sense, studies have shown that it is possible to predict the production of piglets considering only first-parity data (Gruhot et al., 2017; Hoving et al., 2011; Iida et al., 2015; Iida and Koketsu, 2015).

Decision trees allow to integrate a large amount of information generated during the production process and to present the results in a simple and objective way thus assisting in the decision-making process on farms. Decision trees can generate knowledge representation by constructing of classifiers based on an ordered sequence of questions. The subsequent questions depend on the previous answers, and the classification is given by the answer to the last question (Kingsford and Salzberg, 2008). A decision tree is a machine learning technique capable of classifying information based on training data. A series of questions with simple answers (yes or no) is generated at each tree node. As the questions are answered, they generate other child nodes, forming an inverted tree (Kingsford and Salzberg, 2008). This method has already been applied in different production systems. Applications of decision trees in the swine industry include the areas of animal welfare (Cordeiro et al., 2018), piglet performance (Lee et al., 2019), water use (Lee et al., 2017), prevention and control of diseases (Liang et al., 2020), carcass and meat production classification (Masferrer et al., 2018), and pork pricing (Ding et al., 2010).

Therefore, predicting sow culling maintenance in the herd based on the internal reproductive performance of the production system is an exciting tool for the decision-making process. This study applied the decision tree method to production parameters of primiparous sows from a commercial herd to indicate female selection.

Material and Methods

Farm characterization

Data were obtained from a piglet production unit (PPU) in the municipality of Carambeí, Paraná State, Brazil. The PPU has a stable herd with a housing capacity of 5,500 sows and is equipped with a biosecurity system, electronic temperature, and ventilation control (evaporative plates and exhaust fans) in all sheds, as well as a farrowing unit equipped with a heated floor for the piglets. In addition, the PPU produces its replacement sows. The genetic lines used are Camborough and AG 1020 (Agroceres PIC®).

Immediately after weaning or when they are introduced into the herd, sows are submitted to the breeding sector for artificial insemination (AI), where the sows are housed in cages containing drinkers and individual feeders. Immediately after insemination, a batch is formed and sent to the gestation sector. The gestation is divided into pens with a static system to ensure that no female is introduced to the pen after the batch formation. All gestation pens are equipped with drinkers and electronic feeders. The sows are transferred to the farrowing rooms one week before the expected farrowing date. All farrowing rooms are equipped with metal farrowing crates with slatted flooring, drinkers and manual feeders, in addition to a heating system, drinkers, and feeders for the piglets.

The sow remains in the farrowing room until weaning, which occurs on average 21 days after farrowing, and is then transferred to the breeding sector to be again inseminated. In all production sectors, employees collect data using paper spreadsheets. These data are then sent to the office and entered the Agriness S2® farm management software.

Data collection

The data on the herd were obtained from the Agriness S2® software. The period from January 2017 to March 2020 was analyzed. The information was exported to electronic spreadsheets where each line corresponded to a sow and the variables were arranged in columns. The descriptive variables were divided into environmental, labor, date (month and time), animal (identification number, genetic line, PO, weaning-service interval - WSI, culling, type of delivery – TD, gestation length – GL, and farrowing duration – FD), and production parameters (non-productive days – NPD, TBA, number of mummified piglets – NMP, number of stillborn piglets – NSP, number of weaned piglets – NWP, piglets weaned/female/year – WFY, and mortality rate), proportion of piglets deaths at farrowing – DPart, proportion of piglets deaths at weaning – DW.

The type of delivery (TD) was established based on the information in the Agriness S2® software: dystocic, induced, normal, premature, or premature and dystocic. This information was also used to determine the types of culling, physical condition, hooves, return to estrus, false pregnancy, anestrus, low productivity, presence of mastitis, metritis, agalactia, abortion, and uterine prolapse.

Definitions

Return to estrus was defined as when the sow, after insemination and transfer to the gestation facilities, returned to manifest estrus and was again transferred to the breeding sector for new insemination (Hoving et al., 2011). In this case, only information on insemination that resulted in parturition was considered for the analysis. WSI was defined as the number of days after weaning until the first insemination. NPD was defined as the number of days when the sow was neither pregnant nor nursing (Iida and Koketsu, 2015). The number of piglets that died at farrowing or until weaning was calculated as the proportion of the number of dead piglets in relation to the number of live or weaned piglets + number of dead piglets.

Statistical analysis

The data obtained were analyzed using R 3.3.0 (RStudio Team, 2020). First, quality control and verification of biological coherence or possible typing errors of the tabulated data were performed using the descriptive statistics (minimum, maximum, mean, and standard deviation) and the graphical analysis (Table 1). The criteria for removing data from the analyses comprised information considered outside the normal pattern of the PPU, data showing biological inconsistency, and information of litters resulting from "milk mothers" sows (because it does not refer to the production

Table 1 – Mean values of variables studied for each parity order.

	Parity 1	Parity 2	Parity 3	Parity 4	Parity 5	Parity 6	Parity 7	Parity 8
Sows, n	8,869	6,213	4,105	2,612	1,420	706	313	76
AI, n	2.1 (1 - 3)	2.17 (1 - 3)	2.24 (1 - 3)	2.29 (1 - 3)	2.28 (1 - 3)	2.28 (1 - 3)	2.38 (1 - 3)	2.3 (1 - 3)
GL, days	115.1 (108 - 122)	115.2 (108 - 122)	115.3 (108 - 122)	115.2 (108 - 120)	115.2 (108 - 121)	115.3 (109 - 120)	115.5 (108 - 119)	115.8 (113 - 119)
FD, min	201.5 (10 - 1385)	227.4 (10 - 1415)	233.6 (10 - 1390)	248.8 (15 - 1411)	258.5 (10 - 1380)	256.1 (10 - 1010)	255.4 (20 - 1171)	254.1 (60 - 610)
WSI, days	6.1 (0 - 76)	4.9 (0 - 43)	5.09 (0 - 46)	4.96 (0 - 50)	5.38 (1 - 46)	5.12 (2 - 28)	4.45 (0 - 22)	-
Piglets								
TBA, n	13.62 (0 - 23)	13.00 (0 - 24)	13.56 (0 - 25)	13.98 (0 - 24)	14.12 (0 - 25)	14.06 (0 - 23)	13.77 (2 - 21)	12.87 (1 - 21)
DPart, %	0.09 (0 - 1)	0.08 (0 - 1)	0.08 (0 - 1)	0.10 (0 - 1)	0.10 (0 - 1)	0.12 (0 - 1)	0.13 (0 - 0,75)	0.11 (0 - 0,9)
Weaned, n	12.13 (0 - 20)	11.9 (0 - 20)	11.7 (0 - 20)	11.59 (0 - 20)	11.33 (0 - 20)	11.31 (0 - 19)	11.37 (0 - 20)	10.29 (0 - 15)
DW, %	0.1 (0 - 1)	0.11 (0 - 1)	0.11 (0 - 1)	0.12 (0 - 1)	0.13 (0 - 1)	0.12 (0 - 1)	0.12 (0 - 1)	0.15 (0 - 1)

AI: Artificial insemination. GL: Gestation length. FD: farrowing duration. WSI: Wean-to-service interval. TBA: Total born alive piglets. DPart: proportion of piglets deaths farrowing. DW: proportion of piglets deaths at weaning. Values in parentheses: minimum – maximum.

results of the sow itself but of the donor). After quality control and verification, the correlations between TBA, WSI, DW, DPart, AI, GL, FD, and NWP (Weaned) were calculated for PO1 (Table 2) and PO2 sows (Table 3) to assess the interaction of these variables with culling. We consider TBA as a productivity indicator of sows and NWP as a PPU productivity indicator. The ANOVA was used to evaluate the difference in TBA and NWP between culled and unculled sows individually for PO1 and PO2, adopting a level of significance of 5 % (p < 0.05).

The decision trees were built using the method described by Kirchner et al. (2004) and the C4.5 algorithm. The data were divided into two random groups. To train the decision tree algorithm, the database was randomly divided into a training and a testing group. Employing the model generated from the training group, predictions were made on the testing group, which was subsequently analyzed using a confusion matrix. The confusion matrix is organized as follows: true positive (TP) when the classifier correctly predicts a positive outcome and hits; false positive (FP) when the classifier incorrectly predicts a positive outcome and fault; true negative (TN) when the classifier correctly predicts a negative outcome and hits, and false negative (FN) when the classifier incorrectly predicts a negative outcome and fault (Cordeiro et al., 2018). Based on the confusion matrix used to assess the analytical capacity, the precision of the decision tree classification, accuracy, sensitivity and specificity metrics were calculated using the formulas: Sensitivity = $TP/(TP+FN)$, Specificity = $TN/(FP+TN)$, and Accuracy = $(TP+TN)/(TP+FP+TN+FN)$ (Kirchner et al., 2004).

Productivity

In the first analysis (Model 1), we used decision trees to classify sows according to first-parity productivity. A total of 6,441 primiparous sows were used. The following variables were evaluated: TBA, WSI, DW, DPart, AI, GL, FD, NWP, and TD. The sow population was first divided into two groups based on first-parity productivity and the classification method was then applied. The decision tree algorithm was trained in two different scenarios using the variables mentioned above to first identify the 25 % most productive sows (denominated excellent) and, second, to identify the 50 % most productive sows (denominated good).

Culling of sows

In the second analysis (Model 2), we sought to study the effects and understand the behavior of the culling criteria adopted by the farm and that somehow may be present in the database. First- and second-parity sows were considered in this model. A total of 5,717 sows were used. The following variables were used to build the decision tree: TBA, WSI, DW, DPart, AI, GL, FD, NWP, and TD.

Table 2 – Correlations between variables and culled sows considering only data on the first and second farrowing of sows.

	GL	FD	TBA	DPart	NWP	DW	WSI	CULLED
AI	0.13***	– 0.02 ns	– 0.03**	– 0.03**	– 0.01 ns	0.0 ns	– 0.08***	– 0.09***
GL		– 0.01 ns	– 0.16***	– 0.01 ns	0.02 ns	– 0.04***	– 0.05***	– 0.05***
FD			0.05***	0.19***	– 0.10***	0.1***	0.03*	0.04**
TBA				– 0.36***	0.20***	– 0.01 ns	0.04***	0.10***
DPart					– 0.25***	0.11***	0.03*	0.08***
NWP						– 0.73***	– 0.02*	– 0.07***
DW							0.04***	0.13***
WSI								0.04 ns

AI: artificial insemination. GL: gestation length. FD: farrowing duration. TBA: Total born alive. DPart: proportion of piglets deaths at farrowing. NWP (Weaned): number of weaned piglets. DW: proportion of piglets deaths at weaning. WSI: weaning-service interval. Culled: total culled sows. *p < 0.10; **p < 0.05; ***p < 0.01; ns: not significant.

Table 3 – Correlations between variables (data of second farrowing sows) and culled (second parity order -PO2 and total).

	GL	FD	TBA	DPart	NWP	DW	WSI	CULLED PO2	CULLED
AI	0.14***	– 0.02*	– 0.02 ns	– 0.02 ns	0.01 ns	– 0.02*	– 0.01 ns	0.0 ns	– 0.04***
GL		– 0.05***	– 0.22***	– 0.07***	0.07***	– 0.1***	– 0.02 ns	0.0 ns	– 0.05***
FD			0.13***	0.16***	– 0.08***	0.08***	0.0 ns	0.02 ns	– 0.01 ns
TBA				– 0.23***	0.13***	0.03**	0.02 ns	0.03*	– 0.09***
DPart					– 0.20***	0.08***	– 0.01 ns	– 0.01 ns	0.07***
NWP						– 0.73***	– 0.05***	0.01 ns	– 0.09***
DW							0.06***	0.02 ns	0.03**
WSI								-	0.05***
CULLED PO2									0.16***

AI: artificial insemination. GL: gestation length. FD: farrowing duration. TBA: Total born alive. DPart: proportion of piglets deaths at farrowing. NWP (Weaned): number of weaned piglets. DW: proportion of piglets deaths at weaning. WSI: weaning-service interval. Culled: total culled sows. Culled PO2: culled of second parity order - PO2 sows *p < 0.10; **p < 0.05; ***p < 0.01; ns: not significant.

Results

The analyses are limited to PO8 because of the need for more information for higher parities; thus, WSI was not considered for PO8. The descriptive statistics of the complete database are arranged according to sow parity (Table 1). To better understand the data behavior, the population was divided into culled and unculled sows. Considering PO1 sows, the mean NWP in the unculled group differs from that in the culled group ($p < 0.05$), while the mean TBA in the unculled and culled groups do not differ ($p > 0.05$). Considering PO2, the mean NWP, TBA, and WSI in the unculled and culled group do not differ ($p > 0.05$) (Table 4).

All PO1 sows were previously ranked for best production. First, 25 % of sows had ≥ 16 TBA (excellent sows), and 50 % of subsequent sows of the ranking were considered good sows ≥ 14 TBA productivity. To measure sow productivity, the classifier started with the variable number of live piglets per parturition using two groups of sows based on productivity: excellent (≥ 16 TBA) (Figure 1) and good (≥ 14 TBA) (Figure 2). The model for the excellent and good groups selected only the TBA variable based on information from PO1 sows for classification, demonstrating that TBA is the variable that best explains the sow productivity. . The model for excellent sows correctly classified the outcome, showing 86 % sensitivity, 74 % specificity, 86 % accuracy, 74 % prevalence, and a kappa index of 58 % (Figure 1). The model for good sows correctly classified the outcome with 73 % sensitivity, 88 % specificity, 80 % accuracy, 49 % prevalence, and a kappa index of 61 % (Figure 2).

To predict the culling criteria of sows, we used TBA, TD, WSI, DW, DF, AI, GL, FD, and NWP as explanatory variables. Total culling was the dependent variable since this approach made the model more explanatory than culling according to parity order. Decision trees of PO1 sows consider the influence of other variables on the response variable (Figure 3). The bar chart shows that the Weaned (NWP) has the most important variable to determine the culling of a sow. According to the algorithm, only sows that did not wean any piglets or whose TD was dystocic or premature were culled. In the confusion matrix, the model was excellent in classifying sows that were not culled correctly (98 % sensitivity); however, it showed low accuracy (14 %) in classifying sows that were culled, with low confidence of the classifier (kappa = 17 %).

A decision tree that groups data from PO1 and PO2 sows indicates that sow culling is a complex decision and considers more than one variable (Figure 4). The addition of PO2 information increased the branches of the tree, which became more complex. In this respect, WSI, AI, and NWP (Weaned) were the variables that most influenced the formation of the tree. Sows that weaned, on average fewer than 9.3 piglets were culled. Based on this classification, sows with a WSI of 6 to 24 days and those submitted to more than two AI stayed in the herd. Furthermore, when WSI was less than five days, maintenance of the sow depended on a combination of the AI and NWP results. However, sows that weaned more than 14 piglets were culled. The classifier for this tree was more balanced (kappa = 46 %) and correctly classified 59 % of unculled sows and 85 % of culled sows.

To study the effect of the other variables on the groups of good (top 50 %) and excellent (top 25 %) sows,

Table 4 – Analysis of variance of TBA and NWP between culled and unculled sows individually for PO1 and PO2.

	PO1			PO2		
	Unculled	Culled	P	Unculled	Culled	P
TBA	13.62 ± 3.36	11.67 ± 5.03	ns	13.0 ± 3.87	16.2 ± 2.77	ns
NWP	12.14 ± 3.15	7.67 ± 6.66	*	11.9 ± 3.24	12.6 ± 2.88	ns
WSI				6.10 ± 6.03	4.75 ± 0.5	ns

PO: parity order. TBA: Total born alive. NWP: number of weaned piglets. WSI: weaning-service interval. Culled: total culled sows *p < 0.10; **p < 0.05; ***p < 0.01; ns: not significant.

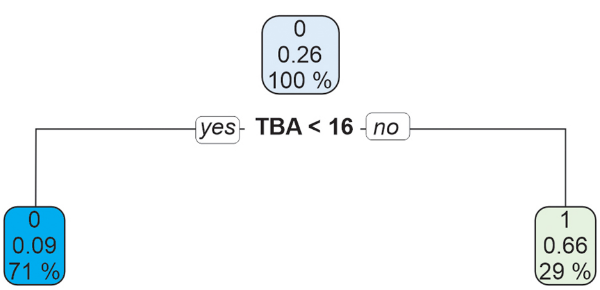

Figure 1 – Decision tree to classify the sows through the Total Born Alive (TBA) as excellent considering PO1 sows. Values in the boxes (leaves) are: - rating: 0 (poor) and 1 (good); - the average of the tree leaf: ex. 0.26 hits index and means that of the animals considered good, 100 % proportion of total samples in this leaf. Colors: lower the mean value - bluer and higher the mean value – greener. Accuracy: 0.83. Kappa: 0.58. Sensitivity: 0.86. Specificity: 0.74. Prevalence: 0.74.

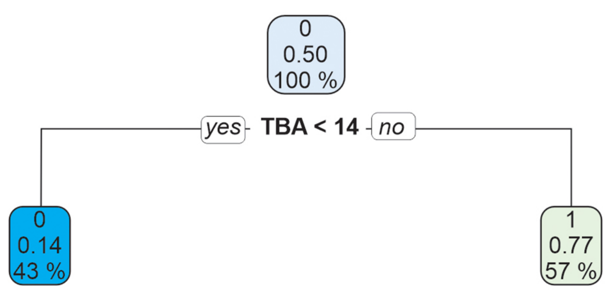

Figure 2 – Decision tree to classify the sows through the Total Born Alive (TBA) as good considering PO1 sows. Values in the boxes (leaves) are: - rating: 0 (poor) and 1 (good); - colors: lower the mean value - bluer and higher the mean value – greener; - the average of the tree leaf: ex. 0.50 hits index and means that of the animals considered good, 100 % proportion of total samples in this leaf. Accuracy: 0.80. Kappa: 0.61. Sensitivity: 0.73. Specificity: 0.88. Prevalence: 0.49.

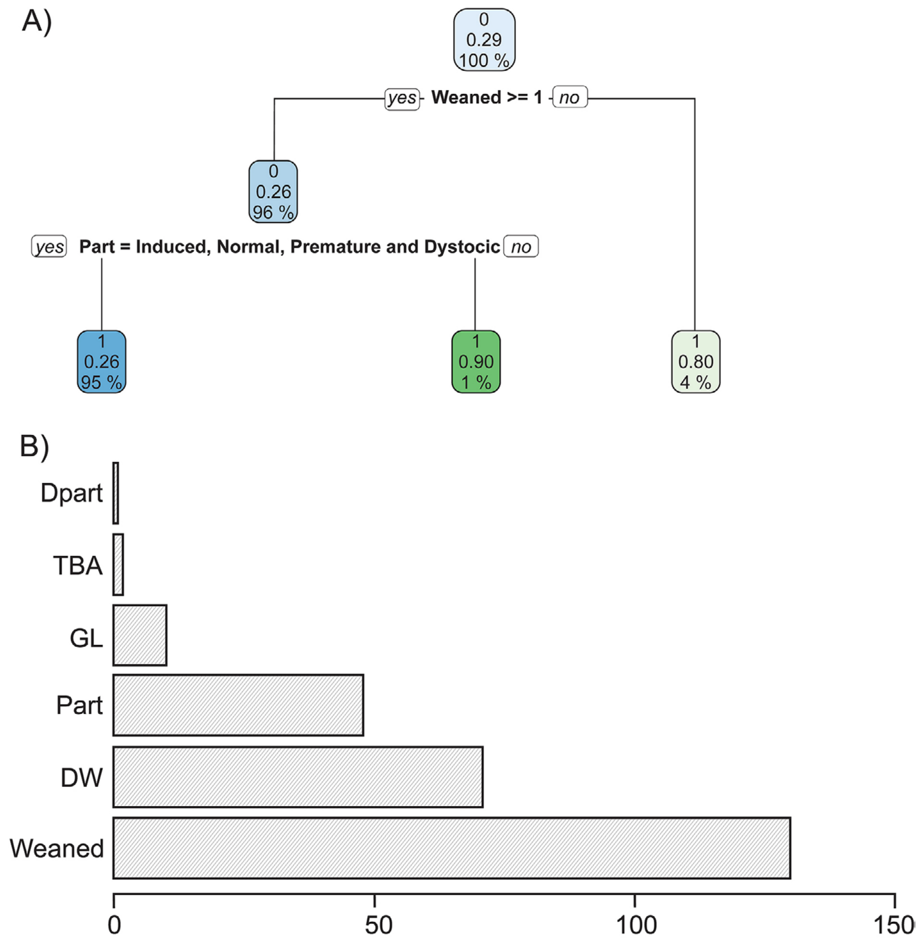

Figure 3 – A) Decision tree to classify sow culling considering PO1 data. B) Order of influence of variables on the construction of the decision tree. Sensitivity = 0.98. Specificity = 0.14. Prevalence = 0.71. Accuracy = 0.74. Kappa: 0.17. $p < 0.001$. NWP (Weaned): number of weaned piglets. DPart: proportion of piglets deaths at farrowing. TBA: Total born alive. GL: gestation length. TD (Part): types of delivery (dystocic, induced, normal, premature, or premature and dystocic). DW: proportion of piglets deaths at weaning; Values in the boxes (leaves) are: - rating: 0 (poor) and 1 (good); - colors: lower the mean value - bluer and higher the mean value – greener; - the average of the tree leaf: ex. 0.29 hits index and means that of the animals considered good, 100 % proportion of total samples in this leaf.

TBA was removed from the model since it was the variable most used in the previous models (Figure 5). It was not possible to use the classification of excellent sows. However, it was possible to predict good sows. This method was not accurate to classify sows as good. However, we can see that a long FD (between 118 and 527 minutes) and a long GL (118 days) compromise the number of liveborn piglets. In contrast to the previous classifications, model accuracy was limited to 34 % for sows classified as not good and 78 % for those classified as good (kappa = 0.12).

Discussion

The graphical characteristics of decision trees allow straightforward interpretation of the results. In addition, from an economic viewpoint, decision trees are computationally inexpensive to train, assess, and store data. However, these trees are prone to overfitting since small changes in the dataset can generate extremely different trees (Valletta et al., 2017).

Efficiency is the focus of the swine industry, which is achieved by increasing herd productivity and reducing production costs. In this respect, the aim is to identify sows that produce more and stay in the herd longer. We, therefore, applied decision trees to identify the variables that most influence the permanence of sows in a commercial herd. Considering the PO1 data, we observed that TBA was the essential variable since it was the only one in the trees for both excellent and good sows. In the tree for excellent sows (≥ 16 TBA), 29 % of the population was classified as excellent and the model proved to be accurate. The model correctly classified 86 % of not-excellent sows and 74 % of excellent sows. However, when the classifier was less demanding (≥ 14 TBA), 57 % of the population was classified as

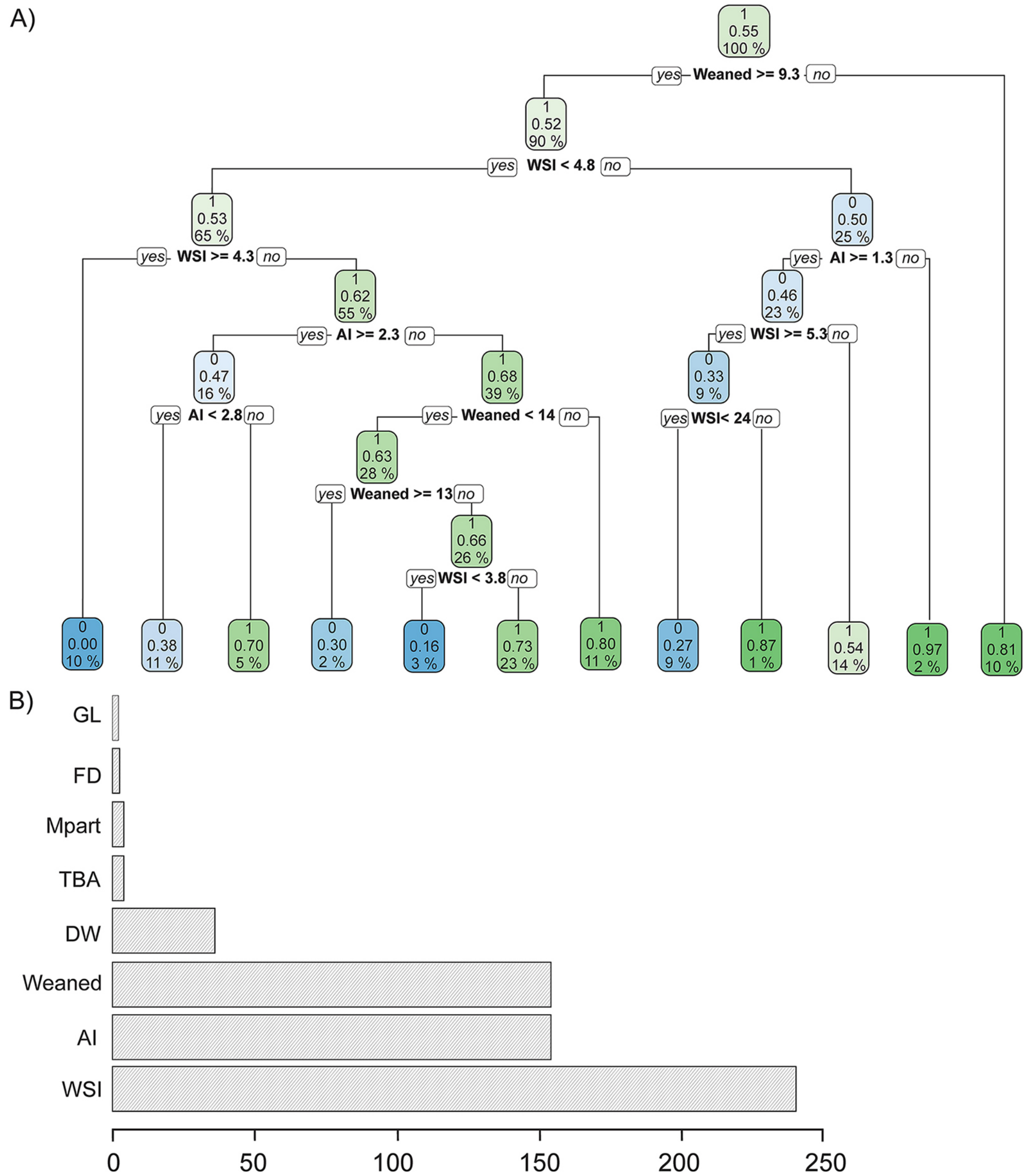

Figure 4 – A) Decision tree to classify sow culling considering PO1 and PO2 data. B) Order of influence of variables on the construction of the decision tree. Sensitivity = 0.59. Specificity = 0.85. Prevalence = 0.45. Kappa: 0.46. p < 0.001. NWP (Weaned): number of weaned piglets. WSI: weaning-service interval. AI: artificial insemination. - GL: gestation length. FD: farrowing duration. DPart: proportion of piglets deaths at farrowing. DW: proportion of piglets deaths at weaning. TBA: Total born alive. Values in the boxes (leaves) are: - rating: 0 (poor) and 1 (good); - colors: lower the mean value - bluer and higher the mean value – greener; - the average of the tree leaf: ex. 0.55 means that of the animals considered good, 100 % proportion of total samples in this leaf.

good sows and the model was more accurate, correctly classifying 73 % of not-good sows and 88 % of good sows. Therefore, the models were capable of predicting sow productivity with some accuracy based on PO1 data. The great contribution of TBA to sow classification was expected. This parameter is one of the main criteria used for sow selection on farms since it is strongly associated to profitability, mainly because more prolific sows stay longer in the herd (Gruhot et al., 2017), diluting fixed costs and thus making the activity more profitable (Zak et al., 2017). For these reasons, low productivity is a leading cause of culling first- and second-parity sows (Andersson et al., 2015).

Correlations between the TBA of culled and unculled sows were not significant (p >0.05 and 0.064, respectively) indicating that TBA is not the main variable influencing culling. Therefore, the culling criteria of sows must t consider different factors, not only TBA. We emphasize that sow selection based only on productivity can have consequences for the welfare of dams and

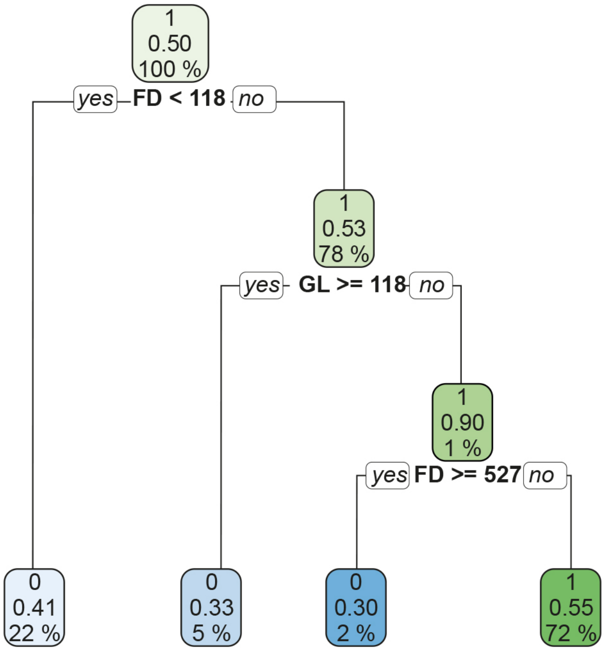

Figure 5 – Decision tree to classify sows as good (top 50 %) in terms of TBA from PO1, considering the duration of farrowing (FD) and gestation (GL). Accuracy: 0.56. Sensitivity = 0.34. Specificity = 0.78. Prevalence = 0.49. Kappa: 0.12. $p < 0.001$. GL: gestation length. FD: farrowing duration. TBA: Total born alive. Values in the boxes (leaves) are: - rating: 0 (poor) and 1 (good); - colors: lower the mean value - bluer and higher the mean value – greener; - the average of the tree leaf: ex. 0.50 means that of the animals considered good, 100 % proportion of total samples in this leaf.

their piglets, and affect the economic efficiency of pig production (Bergman et al., 2018).

In model 2 (Figures 3 and 4), we can understand which criteria influenced the culling of sows on the farm studied. Including only PO1 data, the tree contained only two variables: NWP and TD. Only 5 % of sows were culled based on these criteria: 4 % when the sows weaned less than one piglet and 1 % when farrowing was dystocic or premature. The results showed that the classifier correctly classified 98 % of unculled sows but only 14 % of culled sows. This finding suggests that, when in doubt, the sow was not culled to increase the opportunity to express its potential. However, adding second-parity information increased the number of branches in the tree composed of WSI, AI and NWP. Neither low-productivity nor high-productivity sows stayed in the herd. Low-productivity sows (≤ 9.3 weaned piglets) were culled, corresponding to 10 % of the population. Low productivity (litters with ≤ 7 piglets) affects longevity of sows since they are culled due to reproductive failures (Baxter et al., 2020). However, although desired, high productivity and large litters can compromise the productive life of sows, increasing farrowing duration and the number of stillbirths (Björkman et al., 2017; Udomchanya et al., 2019). Sows that produce an average of 12 to 14 piglets stay longer in the herd (Andersson et al., 2015). Our results showed that sows that weaned between 13 and 14 piglets stayed in the herd (2 % of the population), while those that weaned more than 14 piglets were culled (11 % of the population). These results can be attributed to the fact that the increase in the number of born piglets tends to increase mortality, mainly because of the low birth weight, viability, and performance of the piglets (Campos et al., 2012). Other associated factors are heat stress during lactation, increased mobilization and loss of body reserves, increased shoulder injuries due to limited postural change, and reduced maternal capacity decreasing longevity of sows (Andersson et al., 2015; Baxter et al., 2020).

Another observation from the decision tree is that both sows submitted to few (≤ 1.3) and many inseminations (≥ 2.8), as well as sows with high WSI (> 24 days), were culled (8 % of the population). In addition, sows with a WSI between 3.8 and 4.3 days and those that weaned fewer than 13 piglets were culled (23 % of the population). These results can be explained by reproductive failures, irregular estrus, embryo death, and failure in estrus detection, mainly because reproductive disorders are the leading cause of culling of younger sows (de Hollander et al., 2015). On the other hand, sows that weaned more than 9.3 piglets and had a WSI between 5 and 24 days stayed in the herd (9 % of the population). Ideally, the WSI should be kept between 3 and 7 days to maximize production and reduce the NPD of sows and production costs. Likewise, sows submitted to 2.3-2.8 AI were more likely to stay in the herd (11 %). Artificial insemination protocols that use two doses for 24 hours are safe since they also target sows with early ovulation or diagnostic problems (late estrus).

The decision tree method accurately classified sows according to productivity using only first-parity information. This tool could also identify the main variables that influenced the culling of sows in the current database, demonstrating their relationships and patterns. These findings show that the method can be applied on farms to understand critical factors and improve productivity. However, the present results are limited to the database of the farm studied since small changes in data composition can considerably alter the presentation of the decision trees. It is also important to emphasize that the main focus of the study was to evaluate the applicability of the method within the farm rather than reporting values that could be used for selection. Thus, further studies using a larger database and involving a larger number of production units are necessary for a more detailed analysis. Decision trees accurately represented the patterns and relationships between production variables and culling. Regarding productivity, the decision trees could identify good/excellent sows based on first-parity TBA data. As for the culling criteria, sows with a WSI longer than 5.3 days are culled earlier, and those that wean between 13 and 14 piglets stay in the herd for a longer period.

Acknowledgments

The Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), the Fundação Araucária, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for grants awarded.

Authors' Contributions

Conceptualization: Hilgemberg, JO; Andretta, I; Valk, M; Lehnen, CR. **Data curation:** Hilgemberg, JO; Neimaier, A. **Methodology:** Lehnen, CR; Andretta, I; Valk, M; Hilgemberg, JO. **Formal Analysis:** Hilgemberg, JO.; Neimaier, A; Valk, M. **Writing-original draft:** Hilgemberg, JO; Lehnen, CR; Andretta, I; Valk, M. **Writing and editing:** Hilgemberg, JO; Lehnen, CR; Andretta, I; Valk, M; Hilgemberg, R.; Mariani, A. B.; Bittarello, F.

References

Andersson E, Frössling J, Engblom L, Algers B, Gunnarsson S. 2015. Impact of litter size on sow stayability in Swedish commercial piglet producing herds. Acta Veterinaria Scandinavica 58: 31. https://doi.org/10.1186/s13028-016-0213-8

Baxter EM, Schmitt O, Pedersen LJ. 2020. Managing the litter from hyperprolific sows. p. 71-106. In: Farmer, C. ed. The suckling and weaned piglet. Wageningen Academic Publishers, Wageningen, The Netherlands. https://doi.org/10.3920/978-90-8686-894-0_3

Bergman P, Gröhn YT, Rajala-Schultz P, Virtala AM, Oliviero C, Peltoniemi O, et al. 2018. Sow removal in commercial herds: Patterns and animal level factors in Finland. Preventive Veterinary Medicine 159: 30-39. https://doi.org/10.1016/J.PREVETMED.2018.08.010

Björkman S, Oliviero C, Rajala-Schultz PJ, Soede NM, Peltoniemi OAT. 2017. The effect of litter size, parity and farrowing duration on placenta expulsion and retention in sows. Theriogenology 92: 36-44. https://doi.org/10.1016/j.theriogenology.2017.01.003

Campos PHRF, Silva BAN, Donzele JL, Oliveira RFM, Knol EF. 2012. Effects of sow nutrition during gestation on within-litter birth weight variation: a review. Animal 6: 797-806. https://doi.org/10.1017/S1751731111002242

Cordeiro AFS, Nääs IA, Leitão FS, Almeida ACM, Moura DJ. 2018. Use of vocalization to identify sex, age, and distress in pig production. Biosystems Engineering 173: 57-63. https://doi.org/10.1016/j.biosystemseng.2018.03.007

Ding L, Meng J, Yang Z. 2010. An early warning system of pork price in China based on decision tree. Social Science Research Project of the Ministry of Education in China, Beijing, China. https://doi.org/10.1109/ICEEE.2010.5660808

Gruhot TR, Calderón Díaz JA, Baas TJ, Stalder KJ. 2017. Using first and second parity number born alive information to estimate later reproductive performance in sows. Livestock Science 196: 22-27. https://doi.org/10.1016/j.livsci.2016.12.009

de Hollander CA, Knol EF, Heuven HCM, van Grevenhof EM. 2015. Interval from last insemination to culling. II. Culling reasons from practice and the correlation with longevity. Livestock Science 181: 25-30. https://doi.org/10.1016/J.LIVSCI.2015.09.018

Hoving LL, Soede NM, Graat EAM, Feitsma H, Kemp B. 2011. Reproductive performance of second parity sows: relations with subsequent reproduction. Livestock Science 140: 124-130. https://doi.org/10.1016/j.livsci.2011.02.019

Iida R, Koketsu Y. 2015. Number of pigs born alive in parity 1 sows associated with lifetime performance and removal hazard in high- or low-performing herds in Japan. Preventive Veterinary Medicine 121: 108-114. https://doi.org/10.1016/j.prevetmed.2015.06.012

Iida R, Piñeiro C, Koketsu Y. 2015. High lifetime and reproductive performance of sows on southern European Union commercial farms can be predicted by high numbers of pigs born alive in parity one. Journal of Animal Science 93: 2501-2508. https://doi.org/10.2527/jas.2014-8781

Kingsford C, Salzberg SL. 2008. What are decision trees? Nature Biotechnology 26: 1011-1013. https://doi.org/10.1038/nbt0908-1011

Kirchner K, Tölle KH, Krieter J. 2004. Decision tree technique applied to pig farming datasets. Livestock Production Science 90: 191-200. https://doi.org/10.1016/j.livprodsci.2004.04.003

Koketsu Y, Tani S, Iida R. 2017. Factors for improving reproductive performance of sows and herd productivity in commercial breeding herds. Porcine Health Management 3: 1. https://doi.org/10.1186/s40813-016-0049-7

Lee W, Ham Y, Ban TW, Jo O. 2019. Analysis of growth performance in swine based on machine learning. IEEE Access 7: 161716-161724. https://doi.org/10.1109/ACCESS.2019.2951522

Lee W, Ryu J, Ban TW, Kim, SH, Choi H. 2017. Prediction of water usage in pig farm based on machine learning. Information and Communication Engineering 21: 1560–1566. https://doi.org/10.6109/jkiice.2017.21.8.1560

Liang R, Lu Y, Qu X, Su Q, Li C, Xia S, et al. 2020. Prediction for global African swine fever outbreaks based on a combination of random forest algorithms and meteorological data. Transboundary and Emerging Diseases 67: 935–946. https://doi.org/10.1111/tbed.13424

Masferrer G, Carreras R, Marti-Puig P, Serra M. 2018. Sorting hams using bagged decision trees in a commercial pig slaughterhouse. Frontiers in Artificial Intelligence and Applications 308: 84-88. https://doi.org/10.3233/978-1-61499-918-8-84

Udomchanya J, Suwannutsiri A, Sripantabut K, Pruchayakul P, Juthamanee P, Nuntapaitoon M, et al. 2019. Association between the incidence of stillbirths and expulsion interval, piglet birth weight, litter size and carbetocin administration in hyper-prolific sows. Livestock Science 227: 128-134. https://doi.org/10.1016/j.livsci.2019.07.013

Valletta JJ, Torney C, Kings M, Thornton A, Madden J. 2017. Applications of machine learning in animal behaviour studies. Animal Behaviour 124: 203-220. https://doi.org/10.1016/j.anbehav.2016.12.005

Zak LJ, Gaustad AH, Bolarin A, Broekhuijse MLWJ, Walling GA, Knol EF. 2017. Genetic control of complex traits, with a focus on reproduction in pigs. Molecular Reproduction and Development 84: 1004-1011. https://doi.org/10.1002/mrd.22875