Journal of Dairy Research

cambridge.org/dar

Predicting the body weight of crossbred Holstein × Zebu dairy cows using multivariate adaptive regression splines algorithm

Ignacio Vázquez-Martínez[1,2], Cem Tirink[3], Fernando Casanova-Lugo[4], Dixan Pozo-Leyva[4], Daniel Mota-Rojas[5], Murat Baitugelovich Kalmagambetov[6], Rashit Uskenov[7], Ömer Gülboy[8], Ricardo A. Garcia-Herrera[1] and Alfonso J. Chay-Canul[1]

[1]División Académica de Ciencias Agropecuarias, Universidad Juárez Autónoma de Tabasco, Villaher-mosa, Tabasco, México; [2]Benemérita Universidad Autónoma de Puebla, Complejo Regional Norte, Tetela de Ocampo, Puebla, México; [3]Faculty of Agriculture, Department of Animal Science, Igdir University, TR76000, Igdir, Türkiye; [4]Tecnológico Nacional de México, Instituto Tecnológico de la Zona Maya, Othón P. Blanco, Quintana Roo, México; [5]Neurophysiology, Behavior and Animal Welfare Assessment, Department of Animal Production and Agriculture (DPAA), Universidad Autónoma Metropolitana Xochimilco Campus, Mexico City 04960, Mexico; [6]Aktobe Agricultural Experimental Station, Aktobe, Republic of Kazakhstan; [7]Agronomic Faculty, S. Seifullin Kazakh Agrotechnical University, Z10P6B8, 62 Zhenis av., Astana, Kazakhstan and [8]Faculty of Agriculture, Department of Animals Science, Ondokuz Mayis University, TR55139, Samsun, Türkiye

Research Article

Cite this article: Vázquez-Martínez I *et al* (2024). Predicting the body weight of crossbred Holstein × Zebu dairy cows using multivariate adaptive regression splines algorithm. *Journal of Dairy Research* **91**, 267–272. https://doi.org/10.1017/S0022029924000578

Received: 14 December 2023
Revised: 4 May 2024
Accepted: 16 May 2024

Keywords:
Body measurements; body weight; crossbred; multivariate adaptive regression splines (MARS); tropical cows

Corresponding author:
Cem Tirink;
Email: cem.tirink@gmail.com

Abstract

This study aimed to estimate live body weight from body measurements for Holstein × Zebu dairy cows ($n = 156$) reared under conditions of humid tropics in Mexico using multivariate adaptive regression splines algorithm (MARS) with several train-test proportions. The body measurements included withers height, rump height, hip width, heart girth, body length and diagonal body length. The data were divided into 65:35, 70:30 and 80:20 split data for training and testing sets, respectively. The MARS algorithm was used to construct a prediction model, which predicted the body weight from the body measurements of the test dataset. The results emphasized that the MARS algorithm had an explanation rate for 80:20 train and test set of 0.836 and 0.711, respectively, with minimum Akaike information criterion values. This indicates that it is a reliable way of predicting body weight from body measurements. The results suggest that body weight prediction can be performed with the MARS algorithm in a reliable way, therefore, this algorithm may be a useful tool for animal breeders and researchers in the development of feeding and selection-aimed approaches.

Providing regular and effective information on the body weight (BW) of animals is very important for sustainable animal husbandry and breeding. Accurate BW determination or estimation enables more precise calculation of the ideal feed allocation and will make it easier to decide drug doses, for example, and identify the most appropriate slaughter time and likely marketing price (Tırınk *et al*., 2023a). Unfortunately BW is rarely measured by small farmers due to the lack of weighscales (Lukuyu *et al*., 2016; Tebug *et al*., 2018). Of the various methods for measuring or estimating BW, the weighscale, although it is the most accurate method, is less preferred by producers because it is cumbersome, slow, expensive to implement and stressful for the animals (Wangchuk *et al*., 2018). On the other hand, visual measurement techniques such as image analysis require mathematical models to predict features such as BW and are currently only really applicable in research studies (Stajnko *et al*., 2008; Altay and Delialioğlu, 2022; Coşkun *et al*., 2023a). Therefore, there is a need for developing other practical methods that are low price and easy for small farmers to apply in practice (Dingwell *et al*., 2006; Oliveira *et al*., 2013). Alternative methods do exist based on biometric measures such as withers height (WH), heart girth (HG), hip width (HW), rump height (RH), and body length (BL: Heinrichs *et al*., 1992; Dingwell *et al*., 2006; Lesosky *et al*., 2012; Bretschneider *et al*., 2014; Lukuyu *et al*., 2016; Herrera-López *et al*., 2018; Putra *et al*., 2020).

Milk production systems in the tropical regions of Mexico typically use crosses from *Bos taurus* and *Bos indicus* breeds and forage as the sole source for main feed, maintenance and milk production (Magaña *et al*., 2006; Rojo-Rubio *et al*., 2009; Román-Ponce *et al*., 2013), supplying around 20% of the milk consumed in the country (Magaña *et al*., 2006; Rojo-Rubio *et al*., 2009; Román-Ponce *et al*., 2013). Some studies have evaluated the relationship between biometric measures and BW in cross-bred cattle (Reis *et al*., 2008; Mota *et al*., 2013; Oliveira *et al*., 2013; Franco *et al*., 2017) as well as buffalo (Ramos-Zapata *et al*., 2023; Cruz-Tamayo *et al*., 2024), but models have not yet been developed for animals of

© The Author(s), 2024. Published by Cambridge University Press on behalf of Hannah Dairy Research Foundation

this type under conditions of the humid tropics of Mexico, nor have models that are available been evaluated for local applicability.

Studies in the last two decades were developed for predicting BW using multiple linear regression analysis, however, these regression analyses are often inadequate for prediction because of non-linearity (Ruchay *et al.*, 2021). Various machine-learning approaches have been performed to calculate BW of cattle, sheep, camels and goats. Common features of these studies report the potential of various machine learning algorithms to predict linear or nonlinear relationships between BW and biometric traits accurately and reliably (Ruchay *et al.*, 2021). However, studies reporting the prediction of BW of tropical dairy cows through machine-learning methods are limited. Therefore, this research targeted calculation of the relationship between BW and biometric measures in Holstein × Zebu crossbred cows through MARS algorithm.

Material and methods

Data recording, study site, animals and handling

The data of BW and biometric measurements were recorded in 157 crossbreed dairy cows (Holstein × Zebu). The age of the cows ranged between 3 and 6 years and the cows grazed paddocks of star grass (*Cynodon nlemfuensis*) and humidicola grass (*Brachiaria humidicola*), without supplementation. The data were collected in the commercial farm 'Rancho la Esperanza', located at 17°36′27″N, 93°11′35″W; 120 masl and 10 km of the road Juárez-Reforma, in the municipality of Juarez, Chiapas, in southern Mexico.

Biometric measurements were expressed in cm and recorded as described by Oliveira *et al.* (2013) and Bretschneider *et al.* (2014). These were: heart girth (HG), withers height (WH), rump height (RH), hip width (HW), body length (BL) and diagonal body length (DBL). We used a flexible fibre tape glass (Truper®) and a big caliper of 65 cm (Haglof®, Sweden). The animals were weighed on a scale fixed platform with a capacity of 2000 kg and accuracy of 1 kg.

Statistical analysis

The multivariate adaptive regression splines (MARS) algorithm is a non-parametric regression procedure that assists in a more applicable explanation of linear, nonlinear and interaction results among all variables examined within a cause-and-effect relationship. The most important advantage of this algorithm is that it does not necessarily need to meet the assumptions that the classical regression approach requires (Eyduran *et al.*, 2019; Akin *et al.*, 2020; Coşkun *et al.*, 2023b; Tırınk *et al.*, 2023b). This procedure generates the basic functions according to the stepwise procedures, considering all possible interaction effects among candidate knots and explanatory variables (Arthur *et al.*, 2020). The initial procedure is called the forward pass stage, and the next procedure is named the backward pass stage. In the forward pass stage, the algorithm starts with an intersection for the initial model and iteratively incorporates the initial models combined with the least training error to develop the model. Generally, this process characteristically provides an over-fitted model that influences extreme entanglement (Friedman, 1991; Eyduran *et al.*, 2019). Besides being predominantly worthy, the model constructed from the forward transition process may be weak for the dataset prior to the unstable constructed model, requiring overfitting difficulty with regard to generalization capability. The primary model that will identify the minimum quantity of the estimate model is eliminated in the last process, which is carried out to resolve the overfitting difficulty (Zaborski *et al.*, 2019; Arthur *et al.*, 2020; Faraz *et al.*, 2021). The equation of the MARS procedure carried out to estimate BW from explanatory variables can be given as:

$$\hat{y} = \beta_0 + \sum_{m=1}^{M} \beta_m \prod_{k=1}^{K_m} h_{km}(X_{v(k,m)}) \quad (1)$$

Where: $\hat{y}$ expresses the expected BW value, β_0 expresses the intercept of the model, β_m is the basis functions coefficient, K_m expresses the interaction order limit parameter, h_{km} $(X_{v(k,m)})$ term is expresses the basis function of the prediction model and v(k,m) is an indicator of the explanatory variables in the mth component of the kth product. Basic functions which can decrease the model performances achieved after aforementioned two procedure are eradicated by means of the generalized cross-validation error (GCV), whose equation is given below (Eyduran *et al.*, 2019; Zaborski *et al.*, 2019; Çanga and Boğa, 2022):

$$GCV(\lambda) = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\left[1 - \frac{M(\lambda)}{n}\right]^2} \quad (2)$$

where, n expresses the training set's sample size, y_i expresses the observed value of BW, $\hat{y}_i$ expresses the predicted value of BW and M(λ) is the penalty term that will resolve the complexity of the model containing the λ terms.

At the beginning of the MARS procedure, the multicollinearity relationship between the explanatory variables must be tested to ensure lack of conflict. For this, the data were divided into proportions (80:20, 70:30 and 65:35) for training and test sets, respectively. In the training process, the 10-fold cross-validation procedure was used to choose the best MARS model among tested 180 MARS model (degree = 1:6 and nprune = 2:38). The criteria of the goodness of fit, whose equations are given below, was used to compare the performances of the models obtained from the MARS algorithm for the train and test sets at the different proportions (Grzesiak and Zaborski, 2012; Eyduran *et al.*, 2019; Olfaz *et al.*, 2019; Zaborski *et al.*, 2019; Tırınk *et al.*, 2023a, 2023b).

1. Pearson correlation coefficient (r):

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad (3)$$

2. Root-mean-square error (RMSE):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad (4)$$

3. Standard deviation ratio (SDR):

$$SD_{ratio} = \frac{S_m}{S_d} \quad (5)$$

https://doi.org/10.1017/S0022029924000578 Published online by Cambridge University Press

4. Performance Index (PI)

$$PI = \frac{rRMSE}{1 + r} \quad (6)$$

5. Global relative approximation error (RAE):

$$RAE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \widehat{y_i})^2}{\sum_{i=1}^{n} y_i^2}} \quad (7)$$

6. Mean absolute percentage error (MAPE):

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \widehat{y_i}}{y_i} \right| \times 100 \quad (8)$$

7. Akaike information criterion (AIC):

$$\begin{cases} AIC = n.\ln\left[\frac{1}{n}\sum_{i=1}^{n} (y_i - y_{ip})^2\right] + 2k, & \text{if } n/k > 40 \\ AIC_c = AIC + \frac{2k(k+1)}{n-k-1} & \text{otherwise} \end{cases} \quad (9)$$

where, n represents that training set's sample size, k is the number of explanatory variables in the model, y_i is the observed value of BW, $\widehat{y_i}$ is the expected value of BW, s_d is the standard deviation of BW, and s_m is the standard deviation representing the errors of the optimal model. Goodness-of-fit criteria were carried out to define the best model for smallest AIC, RMSE, SD_{ratio}, MAPE, RAE, PI, and CV values for both sets, and the utmost R^2 and r values for whole models (Tatliyer, 2020).

Descriptive statistical evaluation was carried out with the R software (R Core Team, 2018). To tabulate descriptive statistics for whole variables we used the 'psych' package (Revelle, 2017). To show the relationship between explanatory and response variables, Pearson's correlation coefficient was determined by using the 'performance analytics' package (Peterson and Carl, 2020). To test the multicollinearity problem, the function of the variance inflation factor was used with the 'car' package in R software (Fox and Weisberg, 2019). The MARS algorithm was carried out with the 'caret' package for different proportions (Kuhn, 2022). To show the performances of the 'made for all' models, the 'ehaGoF' package was used (Eyduran, 2020).

Results

Table 1 shows the calculated descriptive statistics for the data. The coefficient of variation (CV %) was calculated to be lower than 30% for all traits, meaning that the measured data were reliable for the data analysis.

Pearson's correlation analysis was performed to present the relationship between explanatory and response variables. The estimated Pearson correlation coefficients and their significance levels are given in Figure 1. All correlation coefficients in Figure 1 were determined to be statistically significant. The greatest correlation coefficient was determined between BW and HG. In addition, HW and WH had higher coefficients with BW than with each other. The lowest correlation coefficients were determined between BW and BL (0.51), DBL (0.47) and RH (0.37), respectively. Although there is a relatively high correlation coefficient among the other variables, it is still possible to discuss a positive relationship.

Table 1. Descriptive statistics for all variables recorded in 157 crossbred dairy cows (Holstein × Zebu)

Variables	Mean ± Standard error	Minimum	Maximum	CV (%)
BW	520.26 ± 5.10	366	704	12.23
HG	191.31 ± 0.67	170	209	4.37
WH	144.33 ± 0.46	133	163	3.99
RH	140.88 ± 0.56	130	195	4.94
HW	54.02 ± 0.26	44	65,5	6.09
BL	91.97 ± 0.52	65	105	7.04
DBL	109.56 ± 0.51	95	128	5.80

BW, Body weight; HG, Heart girth; WH, Withers height; RH, Rump height; HW, Hip width; BL, Body length; DBL, diagonal body length.

Before implementing the MARS algorithm, the multicollinearity problem was assessed. For this aim, the variance inflation factors existing between explanatory variables were determined. Values were 2.18, 2.16, 1.46, 1.80, 1.42 and 1.40 for HG, WH, RH, HW, BL and DBL, respectively. Since all were below 10 there was no multicollinearity problem that would cause overfitting and hence the MARS algorithm could be applied.

To compare the models obtained through the MARS algorithm, model comparison criteria were applied for different proportions of the training and test sets as shown in Table 2. The outcomes of the model evaluation criteria showed that the greatest analytical model power was obtained for the 80:20 training/test proportion, which had the lowest AIC values for both sets. Also, R^2 and r values were determined for this model to be 0.836 and 0.711 for the training and test sets, respectively. In addition, relative importance values were calculated for all proportions, as shown in Table 3. HG had the biggest effective variable for determining the BW for all proportions.

The best predictive model was provided by the 80:20 proportion. Equation (10) shows that the BW can be described with the five basis functions in the MARS prediction model.

$$BW = 476.671 + \begin{cases} if\ HG < 187{:}(HG - 187) \times 6.183 \\ if\ HG > 187{:}(HG - 187) \times 4.249 \\ +(HW - 50) \times 5.929 + (BDL - 116) \times 5.472. \end{cases} \quad (10)$$

According to this, the first term of the selected best MARS prediction model was an intercept that had a coefficient of 476.671. In the second term, HG was determined with the cutpoint of 187 cm and negative coefficient of −6.183. The third term (HG-187) had a cutpoint of 187 cm with a coefficient of 4.249. The fourth term and the third basis function were for HW, with a cutpoint of 50 cm with a coefficient of 5.929. For changes of HW of 50 cm, the effective fourth term on body weight was affected by 5.929. The fifth term was for DBL, again, with cutpoint 116 cm with a coefficient of 5.472.

Discussion

The dearth of studies for predicting BW from biometric measurements in Holstein × Zebu crossbred cattle is a challenge for

https://doi.org/10.1017/S0022029924000578 Published online by Cambridge University Press

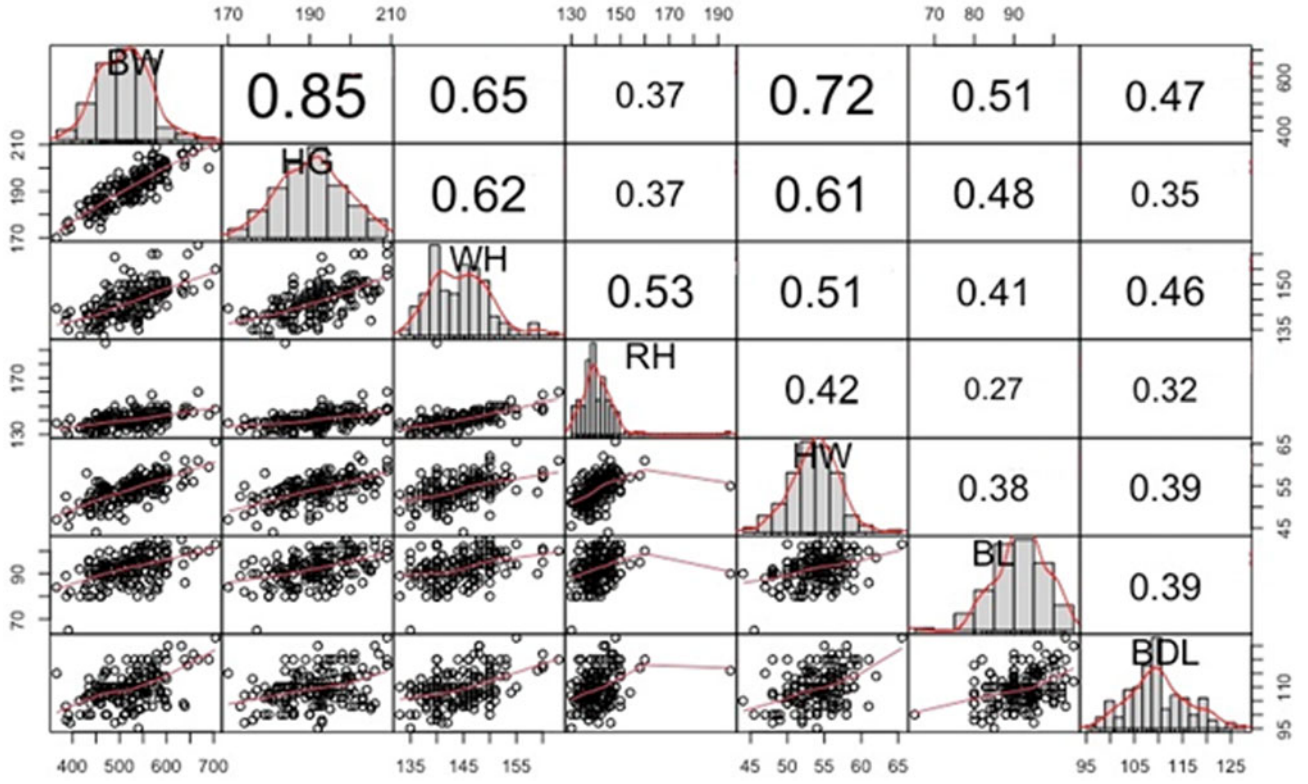


Figure 1. Pearson correlation coefficients between various biometric measures and body weight in cross-bred Holstein × Zebu cattle. All correlations were significant at $P < 0.001$. BW, body weight; HG, heart girth; WH, withers height; RH, rump height; HW, hip width; BL, body length; DBL, diagonal body length.

researchers and producers alike. Only a few studies have focused on this issue, with most of the research conducted on Holstein or Zebu breeds. Nevertheless, some studies do exist, using some of the same biometric measurements as us (HG, WH, HW, BL: Reis *et al.*, 2008; Mota *et al.*, 2013; Oliveira *et al.*, 2013; Franco *et al.*, 2017). Our correlation coefficients among biometric measurements were estimated lower than the study of Putra *et al.*, (2020) on Pasundan cows, whilst in our hands the BW predictive power of WH was higher but BL and HG were lower than in Ongole cows (Putra, 2020). Bene *et al.* (2007) compared different beef breed cattle. Our correlation between BW and WH was higher than Angus, Hereford and Hungarian Simmental whereas in our hands RH was less useful as a predictor. We obtained similar correlations to those reported for BW and BL (Bene *et al.*, 2007). Our best correlation (HG) was lower than the results of Kashoma *et al.* (2011) for Tanzanian shorthorn Zebu cattle, and in another study using the same crossbred type of cattle as ourselves, Mota *et al.* (2013) found high correlation coefficients between BW and HG, hip height and rump height. On the other hand, our data may be more reliable since we used far more cattle (156 compared to 24).

Reis *et al.* (2008) report that the accuracy of estimating BW can be affected by breed, age, body size, body condition and physiological state. Franco *et al.* (2017) reported an R^2 of 0.83 between

Table 2. Goodness-of-fit criteria results of the different proportions of MARS models

	80:20		70:30		65:35	
Goodness-of-fit criteria	Train	Test	Train	Test	Train	Test
Root mean square error (RMSE)	26.695	27.697	26.233	24.719	23.884	32.01
Standard deviation ratio (SD_{ratio})	0.405	0.538	0.405	0.409	0.412	0.437
Coefficient of variation (CV)	5.15	5.44	5.06	4.78	4.6	6.22
Pearson's correlation coefficients (r)	0.914	0.844	0.914	0.924	0.911	0.900
Performance index (PI)	2.678	2.898	2.629	2.48	2.397	3.252
Relative approximation error (RAE)	0.003	0.003	0.002	0.002	0.002	0.004
Mean absolute percentage error (MAPE)	4.231	4.487	4.105	4.208	3.798	4.859
Coefficient of determination (R^2)	0.836	0.711	0.836	0.829	0.831	0.808
Akaike's information criterion (AIC)	844.252	202.636	739.276	302.68	672.026	372.469

https://doi.org/10.1017/S0022029924000578 Published online by Cambridge University Press

Table 3. Relative importance values for predicting body weight according to the MARS prediction model

	Proportion		
	65/35	70/30	80/20
HG	61	58	68
WH	6	8	0
HW	14	22	21
DBL	19	12	11

HG, heart girth; WH, withers height; HW, hip width; DBL, diagonal body length.

BW and HW in Holstein crossbred heifers. These authors conclude that although HW was highly correlated with BW, it showed a low R^2 with a high coefficient of variation when compared with other variables such as body length, hip height and rump height. Using HG and WH to estimate the BW of dairy cows in low-input systems in Senegal, Tebug *et al.* (2018) reported that R^2 varied from 0.77 to 0.94; they also reported that the RMSE of the developed models corresponded to 9.4 to 12.33% (29.27 to 39.24 kg) of the average BW of animals. Also, Bretschneider *et al.* (2014) determined that the RMSE of their model was 5.8% of the average BW (15.95 kg). Mota *et al.* (2013) concluded that the correlations between measurements and body development of heifers with different parentages are distinct, and that specific equations are necessary for predicting body weight. Additionally, Tedde *et al.* (2021) indicated that estimating BW through biometric measurements can be approached as a regression problem, where the input features are the body measurements, and the target value is the BW that the regression model predicts.

Ruchay *et al.* (2021) stated that the random forest regression algorithm, one of the machine learning methods, is the most effective algorithm in predicting the BW of Hereford cows and may be more effective than traditional models. Similarly, Dang *et al.* (2022) determined that the light generalized boosted regression tree-based model had the best performance, and will use these findings to develop a method for indirectly estimating the live weight of Hanwoo cows using machine vision technology that measures ten different body features. Celik (2019) compared MARS, Chi-square automatic interaction detection (CHAID), exhaustive-CHAID and classificastion and regression tree (CART) algorithms for predicting BW in Pakistani goats. Goodness of fit criteria were used to evaluate the model performances, and it was reported that the model obtained with the MARS procedure was most reliable among the models studied. Our evaluation criteria yielded similar results to theirs. Canga (2022) used MARS to predict hot carcass weight from several features, and once again the MARS algorithm gave similar results to the results of the current study within the scope of the model comparison criteria.

We can conclude that the use of statistical procedures for predicting BW from biometric measurements is an important and useful tool that can be applied to crossbred cattle. They are relatively easy to use and require minimal effort. The MARS algorithm is particularly useful. It is a non-parametric approach, which allows for the incorporation of non-linear relationships between the independent and dependent variables. This makes it especially useful for predicting body weight from body measurements, as these can often be subject to non-linear relationships. The MARS algorithm also provides an efficient and accurate way to construct prediction models, without the need to perform many variable transformations. Furthermore, MARS allows for the incorporation of both continuous and categorical variables, thus making it an ideal method for predicting BW from biometric measurements.

References

Akin M, Eyduran SP, Eyduran E and Reed BM (2020) Analysis of macro nutrient related growth responses using multivariate adaptive regression splines. *Plant Cell, Tissue and Organ Culture* **140**, 661–670.

Altay Y and Delialioğlu RA (2022) Diagnosing lameness with the random forest classification algorithm using thermal cameras and digital colour parameters. *Mediterranean Agricultural Sciences* **35**, 47–54.

Arthur CK, Temeng VA and Ziggah YY (2020) Multivariate adaptive regression splines (MARS) approach to blast-induced ground vibration prediction. *International Journal of Mining, Reclamation and Environment* **34**, 198–222.

Bene S, Nagy B, Kiss B, Polgar JP and Szabo F (2007) Comparison of body measurements of beef cows of different breeds. *Archiv fur Tierzucht* **50**, 363–373.

Bretschneider G, Cuatrin A, Arias D and Vottero D (2014) Estimation of body weight by an indirect measurement method in developing replacement Holstein heifers raised on pasture. *Archivos de Medicina Veterinaria* **46**, 439–443.

Canga D (2022) Use of MARS data mining algorithm based on training and test sets in determining carcass weight of cattle in different breeds. *Journal of Agricultural Sciences* **28**, 259–268.

Canga D and Boğa M (2022) Detection of correct pregnancy status in lactating dairy cattle using MARS data mining algorithm. *Turkish Journal of Veterinary & Animal Sciences* **46**, 809–819.

Celik S (2019) Comparing predictive performances of tree-based data mining algorithms and MARS algorithm in the prediction of live body weight from body traits in Pakistani goats. *Pakistan Journal of Zoology* **51**, 1447–1456.

Coşkun G, Şahin Ö, Delialioğlu RA, Altay Y and Aytekin I (2023a) Diagnosis of lameness *via* data mining algorithm by using thermal camera and image processing method in Brown Swiss cows. *Tropical Animal Health and Production* **55**, 50.

Coşkun G, Şahin Ö, Altay Y and Aytekin I (2023b) Final fattening live weight prediction in Anatolian merinos lambs from somebody characteristics at the initial of fattening by using some data mining algorithms. *Black Sea Journal of Agriculture* **6**, 47–53.

Cruz-Tamayo AA, Ramírez-Bautista MA, Mota-Rojas D, Escobar-España JC, García-Herrera R, Gurgel ALC, Dias-Silva TP, de Araújo MJ, Santana JCS, Aguiar IOM, Ítavo LCV and Chay-Canul AJ (2024) Relationship between body weight and hip width in dairy buffaloes (*Bubalus bubalis*). *Journal of Dairy Research* **91**, 180–183.

Dang CG, Choi TJ, Lee SS, Lee SH, Alam M, Park MN, Han S, Lee JG and Hoang DT (2022) Machine learning-based live weight estimation for Hanwoo cow. *Sustainability* **14**, 12661.

Dingwell RT, Wallace MM, McLaren CJ, Leslie CF and Leslie KE (2006) An evaluation of two indirect methods of estimating body weight in Holstein calves and heifers. *Journal of Dairy Science* **89**, 3992–3998.

Eyduran E (2020) ehaGoF: calculates goodness of fit statistics. *R package version* 0.1.1.

Eyduran E, Akin M and Eyduran SP (2019) *Application of Multivariate Adaptive Regression Splines Through R Software*. Ankara. Türkiye: Nobel Academic Publishing, p. 104.

Faraz A, Tirink C, Eyduran E, Waheed A, Taukir NA, Nabeel MS and Tariq MM (2021) Prediction of live body weight based on body measurements in Thalli sheep under tropical conditions of Pakistan using CART and MARS. *Tropical Animal Health and Production* **53**, 301.

Fox J and Weisberg S (2019) *An {R} Companion to Applied Regression*, 3rd Edn. Thousand Oaks, CA: Sage.

Franco MDO, Marcondes MI, Campos JMDS, Freitas DRD, Detmann E and Valadares SDC (2017) Evaluation of body weight prediction equations in growing heifers. *Acta Scientiarum. Animal Sciences* **39**, 201–206.

https://doi.org/10.1017/S0022029924000578 Published online by Cambridge University Press

Friedman J (1991) Multivariate adaptive regression splines. *Annals of Statistics* **19**, 1–67.

Grzesiak W and Zaborski D (2012) Examples of the use of data mining methods in animal breeding. In Karahoca A (ed.), *Data Mining Applications in Engineering and Medicine*. London: InTech, pp. 303–324.

Heinrichs AJ, Rogers GW and Cooper JB (1992) Predicting body weight and wither height in Holstein heifers using body measurements. *Journal of Dairy Science* **75**, 3576–3581.

Herrera-López S, García-Herrera R, Chay-Canul AJ, González-Ronquillo M, Macías-Cruz U, Díaz-Echeverría VF, Casanova-Lugo F and Piñeiro-Vázquez A (2018) Desarrollo y evaluación de una ecuación para predecir el peso vivo en novillas cruzadas usando el ancho de cadera. *ITEA-Información Técnica Económica Agraria* **114**, 368–377.

Kashoma IPB, Luziga C, Werema CW, Shirima GA and Ndossi D (2011) Predicting body weight of Tanzania shorthorn zebu cattle using heart girth measurements. *Livestock Research for Rural Development* **23**, 94.

Kuhn M (2022) caret: Classification and regression training. *R package version* 6.0–93.

Lesosky M, Dumas S, Conradie I, Handel IG, Jennings A, Thumbi S, Toye F and de Clare Bronsvoort BM (2012) A live weight–heart girth relationship for accurate dosing of East African shorthorn zebu cattle. *Tropical Animal Health and Production* **45**, 311–316.

Lukuyu MN, Gibson JP, Savage DB, Duncan AJ, Mujibi FDN and Okeyo AM (2016) Use of body linear measurements to estimate live weight of crossbred dairy cattle in smallholder farms in Kenya. *SpringerPlus* **5**, 1–14.

Magaña MJG, Ríos AG and Martínez GJC (2006) Los sistemas de doble propósito y los desafíos en los climas tropicales de México. *Archivos Latinoamericanos de Producción Animal* **14**, 105–114.

Mota DA, Berchielli TT, Canesin RC, Rosa BL, Ribeiro AF and Brandt HV (2013) Nutrient intake, productive performance and body measurements of dairy heifers fed with different sources of protein. *Acta Scientiarum. Animal Sciences* **35**, 273–279.

Olfaz M, Tirink C and Onder H (2019) Use of CART and CHAID algorithms in Karayaka sheep breeding. *Kafkas Üniversitesi Veteriner Fakültesi Dergisi* **25**, 105–110.

Oliveira AS, Abreu DC, Fonseca MA and Antoniassi PMB (2013) Development and evaluation of predictive models of body weight for crossbred Holstein-Zebu dairy heifers. *Journal of Dairy Science* **96**, 6697–6702.

Peterson BG and Carl P (2020) Performance analytics: econometric tools for performance and risk analysis. *R package version* 2.0.4.

Putra WPB (2020) The assessment of body weight of Sumba Ongole cattle (*Bos indicus*) by body measurements. *Manas Journal of Agriculture Veterinary and Life Sciences* **10**, 52–57.

Putra WPB, Said S and Arifin J (2020) Principal component analysis (PCA) of body measurements and body indices in the Pasundan cows. *Black Sea Journal of Agriculture* **3**, 49–55.

R Core Team (2018) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ramos-Zapata R, Dominguez-Madrigal C, García-Herrera R-A, Camacho-Perez E, Lugo-Quintal JM, Tyasi TL, Gurgel ALC, Ítavo LCV and Chay-Canul AJ (2023) Predicting live weight using body volume formula in lactating water buffalo. *Journal of Dairy Research* **90**, 138–141.

Reis GL, Albuquerque FHMAR, Valente BD, Martins GA, Teodoro RL, Ferreira MBD, Monteiro JBN, de Almeida M and Madalena FE (2008) Predição do peso vivo a partir de medidas corporais em animais mestiços Holandês/Gir. *Ciência Rural* **38**, 778–783.

Revelle WR (2017) Psych: procedures for personality and psychological research.

Rojo-Rubio R, Vázquez-Armijo JF, Pérez-Hernández P, Mendoza-Martínez GD, Salem AZM, Albarrán-Portillo B, González-Reyna A, Hernández-Martínez J, Rebollar-Rebollar S, Cardoso-Jiménez D, Dotantes-Coronado EJ and Gutierrez-Cedillo JG (2009) Dual purpose cattle production in Mexico. *Tropical Animal Health and Production* **41**, 715–721.

Román-Ponce SI, Ruiz-López FDJ, Montaldo HH, Rizzi R and Román-Ponce H (2013) Efectos de cruzamiento para producción de leche y características de crecimiento en bovinos de doble propósito en el trópico húmedo. *Revista Mexicana de Ciencias Pecuarias* **4**, 405–416.

Ruchay AN, Kolpakov VI, Kalschikov VV, Dzhulamanov KM and Dorofeev KA (2021) Predicting the body weight of Hereford cows using machine learning. *IOP Conference Series: Earth and Environmental Science* **624**, 1–5.

Stajnko D, Brus M and Hočevar M (2008) Estimation of bull live weight through thermographically measured body dimensions. *Computers and Electronics in Agriculture* **61**, 233–240.

Tatliyer A (2020) The effects of raising type on performances of some data mining algorithms in lambs. *Journal of Agriculture and Nature* **23**, 772–780.

Tebug SF, Missohou A, Sourokou Sabi S, Juga J, Poole EJ, Tapio M and Marshall K (2018) Using body measurements to estimate live weight of dairy cattle in low-input systems in Senegal. *Journal of Applied Animal Research* **46**, 87–93.

Tedde A, Grelet C, Ho PN, Pryce JE, Hailemariam D, Wang Z, Plastow G, Gengler N, Brostaux Y, Froidmont E, Dehareng F, Bertozzi C, Crowe MA, Dufrasne I, GplusE Consortium Group and Soyeurt H (2021) Validation of dairy cow bodyweight prediction using traits easily recorded by dairy herd improvement organizations and its potential improvement using feature selection algorithms. Animals 11, 1288.

Tırınk C, Önder H, Francois D, Marcon D, Şen U, Shaikenova K, Omarova K and Tyasi TL (2023a) Comparison of the data mining and machine learning algorithms for predicting the final body weight for Romane sheep breed. *PLoS ONE* **18**, e0289348.

Tırınk C, Piwczyński D, Kolenda M and Önder H (2023b) Estimation of body weight based on biometric measurements by using random forest regression, support vector regression and CART algorithms. *Animals* **13**, 798.

Wangchuk K, Wangdi J and Mindu M (2018) Comparison and reliability of techniques to estimate live cattle body weight. *Journal of Applied Animal Research* **46**, 349–352.

Zaborski D, Ali M, Eyduran E, Grzesiak W, Tariq MM, Abbas F, Waheed A and Tirink C (2019) Prediction of selected re-productive traits of indigenous Harnai sheep under the farm management system *via* various data mining algorithms. *Pakistan Journal of Zoology* **51**, 421–431.

https://doi.org/10.1017/S0022029924000578 Published online by Cambridge University Press