
Poultry Science Journal
ISSN: 2345-6604 (Print), 2345-6566 (Online) http://psj.gau.ac.ir
DOI: 10.22069/psj.2024.21337.1934

GUASNR

# Investigation of Egg Quality Characteristics Affecting Egg Weight of Lohmann Brown Hen with Data Mining Methods

**Thobela Louis Tyasi[1] & Senol Celik[2]**

[1]School of Agricultural and Environmental Sciences, Department of Agricultural Economics and Animal Production, University of Limpopo, South Africa

[2]Department of Animal Science, Biometry Genetics Unit, Agricultural Faculty, Bingol University, Bingol, Turkey

Poultry Science Journal 2024, 12(1): 107-117

**Keywords**
Egg
MARS
Chicken
Bagging MARS
Random forest

**Corresponding author**
Thobela Louis Tyasi
louis.tyasi@ul.ac.za

**Article history**
Received: May 08, 2023
Revised: December 06, 2023
Accepted: January 09, 2024

## Abstract

Hen egg weight is one of the most important traits in the egg production industry; however, the egg traits influencing it are poorly understood. Random forest (RF), multivariate adaptive regression spline (MARS), classification and regression trees (CART), bagging MARS, chi-square automatic interaction detector (CHAID), and exhaustive CHAID were used in egg weight (EW) prediction from selected egg quality characteristics in chicken. A total of 400 egg weight (EW), egg length (EL), egg width (EWD), shell weight (SW), yolk weight (YW), and albumen weight (AW) predictors were turned into account. The goodness-of-fit criteria were used to select the best model to estimate Lohman Brown hen egg weight. The data was separated into train and test datasets for validation using a 10-fold cross-validation. The most significant EW predictors were albumen weight, egg width, and egg length. The correlation coefficient (r) value ranged from 0.957 (CHAID) to 0.99999 (MARS and Bagging MARS). The lowest RMSE (0.001) was found for MARS and bagging MARS algorithms and the highest (2.154) was obtained for CHAID. In general, the implemented algorithms excellently predicted the EW of hens. The ascertainment of the egg quality characteristics associated with EW using data mining algorithms can be considered an indirect selection criterion for further chicken breeding programs.

## Introduction

The production of eggs with high-quality egg shells and high-quality eggs within is essential to the global egg industry's economic sustainability. The business loses millions of dollars annually due to egg quality issues. Thus, it is crucial to comprehend the variables influencing the quality of the egg's interior and shell (Ahmadi and Rahimi, 2011).

Pires *et al*. (2021) also indicated that eggs are the cheapest source of animal protein, minerals, and vitamins. Egg quality is a determinant of the reproductive fitness of the parents and was defined by Stadelman (1977) as the characteristics of an egg that affects its admissibility to the consumers. Various factors such as hen's age and genotype, nutrition, type of rearing system, and laying time affect egg quality characteristics (Ahmadi and Rahimi, 2011; Yang *et al.*, 2014). The Lohmann Brown chicken breed is a cross between lines of Rhode Island Red and White Rock breeds with a high yield of eggs and friendly temperament. It is mainly bred for egg production but can be used for meat (Tutkun *et al.*, 2018).

Chi-square Automatic Interaction Detection (CHAID) and exhaustive CHAID were computed following the procedure of Biggs *et al.* (1991). Classification and Regression Tree (CART) algorithm was conducted using the procedure of Breiman *et al.* (1984). Briefly, CART is a tree method that divides a node into child nodes, starting with the root node that contains the whole learning sample. CART is one of the most common algorithms used in machine learning for regression and classification (Troncoso *et al.*, 2015). The Multivariate Adaptive Regression Spline (MARS) algorithm is a flexible data mining algorithm, which Friedman (1991) noticed, enabling highly delicate predictions.

Please cite this article as Thobela Louis Tyasi & Senol Celik. 2024. Investigation of Egg Quality Characteristics Affecting Egg Weight of Lohmann Brown Hen with Data Mining Methods. Poult. Sci. J. 12(1): 107-117.

This study aimed to evaluate egg length, egg width, shell weight, yolk weight and albumen weight properties that affect egg weight in chickens. In addition, CHAID, Exhaustive CHAID, CART, Random Forest, MARS and Bagging MARS algorithms were applied to determine the effects of factors affecting egg weight.

## Materials and Methods

### Study area, experimental birds, management and study design

The current study was conducted at Kitamu farm in Ntsima village in the Limpopo Province, South Africa. The temperature ranges, rainfall patterns and coordinates are the same as Kutu and Asiwe (2010) explained. A total of 100 Lohmann Brown chicken breeds were utilized for the experiment. The birds were bought at 18 weeks and raised under an intensive production system. The laying mash was purchased from Angels Feeds in Polokwane, South Africa. The chickens were raised following the ordinary husbandry practices of feeding systems, housing, vaccination and health care as Alabi et al. (2012) described. The chicken house was cleaned before the chickens arrived and disinfected with Virokill disinfectants to avoid transmission of pathogenic diseases to the chickens. The biosecurity protocols were followed in the area where the footbaths with disinfectant were placed at the door for disinfecting before entering the chicken house. The drinkers and feeders were bought at NTK in Polokwane, South Africa. The chickens were fed the egg-laying mash for 18 weeks, and water was provided ad libitum.

### Egg collection and egg quality characteristics measurements

A total of 400 eggs were collected and used to determine EW and egg quality characteristics. Five egg quality characteristics were measured, which include egg length (EL), egg width (EWD), shell weight (SW), yolk weight (YW) and albumen weight (AW). The egg quality characteristics were measured as Ukwu et al. (2017) explained.

### Statistical analysis

The MARS model was used as explained by Zhang and Goh (2016).

$$f(x) = \beta_0 + \sum_{m=1}^{m} \beta_m \lambda_m(x)$$

where $f(x)$ is the anticipated response, $\beta_0$ and $\beta_m$ are factors computed to cause the best data fit, and m is the number of BFs in the model. General Cross-Validation (GCV) was used to eliminate duplicate BFs following Kornacki and Cwik (2005) procedure.

$$GCV = \frac{\frac{1}{N}\sum_{i=1}^{n}[y_i - f'(x_i)]^2}{\left[1 - \frac{C(B)}{N}\right]^2}$$

where N represents points in the data, while C(B) is a complexity penalty that improves with the number of BFs in the model, and it is determined as Goh *et al.* (2016) method. The 10-fold cross-validation was considered a resampling method in the MARS algorithm. It was used a 10-fold cross-validation to yield the optimal regularization parameter, minimizing the sum of least square plus shrinkage penalty by using the R glmnet package (James *et al*., 2013). The bagging (Bootstrap aggregating) MARS algorithm was computed as Kunn and Johnson (2013) explained. Bagging MARS is a helpful tool employed to improve the predictive precision of the MARS model. The Random Forest model algorithm was performed following the procedure of Breiman (2001). The grid search approach can be used to find the best settings for the Random Forest algorithm. Random Forests is a popular classification approach that generates a large number of classification trees to identify the class label of new objects. A majority vote is used on each tree to establish the class label of new objects (Ramadhan *et al*., 2017). RF is the most at premium algorithm for classification and regression in decision tree learning. The estimated values for unseen instances x is computed by averaging the estimation findings from all regression trees as following:

$$f(x) = \frac{1}{B}\sum_{b=1}^{B} t_b(x)$$

Here f(x) is egg weights. Other egg characteristics are independent variables affecting egg weights. By modelling different trees instead of a single tree, the random forest model presents preferable and more accurate predictions (Tso and Yau, 2007).

### The goodness of fit test

In order to comparatively the predictive performances of the CART, the RF, the MARS, and the Bagging MARS in the 10-fold cross-validation, the following model evaluation criteria were calculated (Willmott and Matsuura, 2005; Liddle, 2007; Takma *et al.*, 2012; Chen and Li, 2014; Chen *et al.*, 2016).

Coefficient of Determination

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

Adjusted Coefficient of Determination

$$Adj.R^2 = 1 - \frac{\frac{1}{n-k-1}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

Error mean square

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

Root-mean-square error (RMSE) is expressed by the following formula:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}$$

The Mean absolute deviation (MAD) is the average absolute prediction error. It is less sensitive to outliers. The formula is presented by

$$MAD = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right|$$

Mean absolute percentage error (MAPE);

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right|.100$$

Standard Deviation Ratio;

$$SD_{ratio} = \sqrt{\frac{\frac{1}{n-1}\sum_{i=1}^{n}(\varepsilon_i - \bar{\varepsilon})^2}{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

An R software was used for analyses, taking cross-validation as 10 (R Core Team, 2021). Results were obtained by using the RF algorithm random forest", MARS and Bagging MARS algorithms "earth" packages. Data mining algorithms' model evaluation performance criteria were evaluated using the "ehaGoF" package (Eyduran, 2020). Using the R software "corrplot" package, Pearson correlation coefficients between egg weight and other egg characteristics were calculated. Also, the multicollinearity problem between the independent variables was tried at the outset of the analysis, and it was discovered that there was none. One of the machine learning methods is to work with this type of data. CHAID, Exhaustive CHAID and CART algorithms were analyzed with the SPSS V. 26.0 package program (IBM Corp, 2019).

## Results

### Descriptive statistics

Descriptive statistics for the measurements of quality characteristics in hen's eggs are shown in Table 1.

### Correlation matrix

Correlation coefficients between egg quality characteristics are also indicated in Table 2. Notably, the correlation coefficients between egg weight and egg length, egg width, yolk weight and albumen weight were high, 0.74, 0.764, 0.601 and 0.933, respectively. A low correlation coefficient (0.23) was estimated for shell weight and egg weight, but it was statistically significant ($P < 0.001$). The lowest correlation coefficient was recorded as -0.043 between albumen weight and shell weight. Also, correlations between all other egg quality characteristics were found to be positive.

**Table 1.** Descriptive statistics of egg quality traits

| Parameters | EW | EL | EWD | SW | YW | AW |
|---|---|---|---|---|---|---|
| Number | 400 | 400 | 400 | 400 | 400 | 400 |
| Mean | 65.4 | 57.6 | 44.7 | 9.34 | 17.2 | 38.9 |
| Standard error of means | 0.264 | 0.126 | 0.063 | 0.0581 | 0.0687 | 0.224 |
| Standard deviation | 5.28 | 2.52 | 1.26 | 1.16 | 1.37 | 4.49 |
| Minimum | 55.8 | 50.6 | 42.2 | 7.69 | 12.9 | 32.3 |
| Maximum | 78.3 | 63.4 | 47.5 | 14.8 | 20.1 | 50.2 |

EW: Egg weight (g), EL: Egg length (mm), EWD: Egg width (mm), SW: Shell weight (g), YW: Yolk weight (g), AW: Albumen weight (g).

**Table 2.** Correlation coefficients between of egg quality traits

| Traits | EW | EL | EWD | SW | YW | AW |
|---|---|---|---|---|---|---|
| EW | — | | | | | |
| EL | 0.74*** | — | | | | |
| EWD | 0.764*** | 0.25*** | — | | | |
| SW | 0.23*** | 0.129** | 0.075 | — | | |
| YW | 0.601*** | 0.574*** | 0.453*** | 0.179*** | — | |
| AW | 0.933*** | 0.662*** | 0.74*** | -0.043 | 0.355*** | — |

$^{ns}$: Not significant, **: p < 0.01, ***: p < 0.001, EW: Egg weight, EL: Egg length, EWD: Egg width, SW: Shell weight, YW: Yolk weight, AW: Albumen weight.

### The goodness of fit criteria

Summary results of the CHAID, Exhaustive CHAID, CART, Random Forest (RF), MARS, and Bagging MARS algorithms in the predictive accuracy are presented in Table 3. The goodness of fit criteria was calculated based on the validation set. The findings showed that MARS and Bagging MARS methods had superior performance over regression tree and Random Forest methods. The estimated value of the coefficient of determination ($R^2$) was higher for the

MARS and Bagging MARS (0.9999) and RF (0.9974) compared with that for CHAID (0.916), Exhaustive CHAID (0.945) and CART (0.966). All other quality measures (RMSE, MSE, SD ratio, MAD, MAPE and AIC) were generally lowest for the MARS and Bagging MARS compared with other used models.

**Table 3.** The goodness of fit criteria

| Criteria of algorithms | CHAID | Exhaustive CHAID | CART | RF | MARS | Bagging MARS |
|---|---|---|---|---|---|---|
| r | 0.957 | 0.972 | 0.983 | 0.995 | 0.99999 | 0.99999 |
| $R^2$ | 0.916 | 0.945 | 0.966 | 0.9974 | 0.9999 | 0.9999 |
| $R^2$ adjusted | 0.915 | 0.944 | 0.965 | 0.9970 | 0.9999 | 0.9999 |
| SD ratio | 0.289 | 0.233 | 0.185 | 0.028 | 0.00001 | 0.00001 |
| MSE | 4.640 | 3.029 | 1.900 | 0.319 | 0.000001 | 0.000001 |
| RMSE | 2.154 | 1.740 | 1.378 | 0.102 | 0.001 | 0.001 |
| MAD | 1.085 | 0.850 | 0.730 | 0.037 | 0.000001 | 0.000001 |
| MAPE | 1.649 | 1.303 | 1.119 | 0.092 | 0.0003 | 0.0003 |
| AIC | 278.607 | 204.520 | 123.501 | -15.24 | -5196 | -469 |

CHAID, Exhaustive CHAID and CART algorithms

The diagram of the CHAID, Exhaustive CHAID and CART algorithms performed to determine egg quality characteristics was presented in Figure 1, Figure 2 and Figure 3, respectively. For the CHAID, Exhaustive CHAID and CART algorithms, the number of parent and child nodes was 30:15.

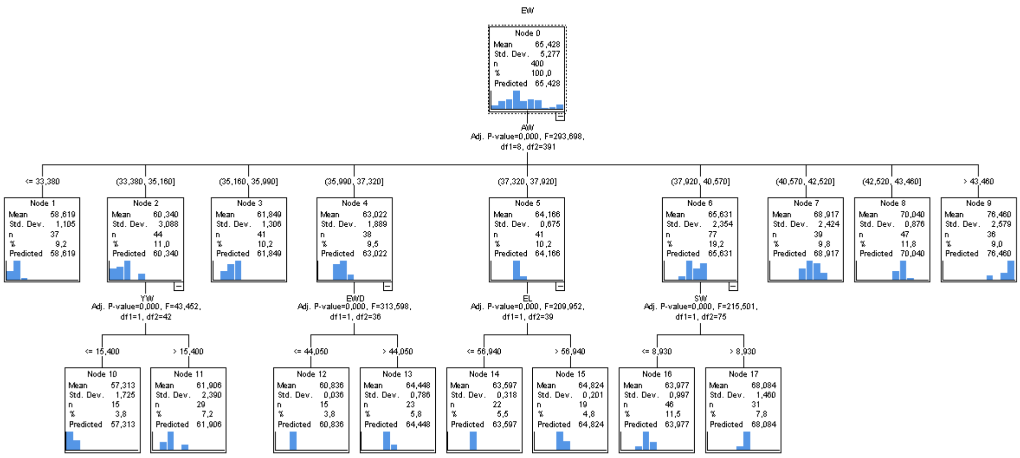

**Figure 1.** The regression tree diagram constructed by the CHAID algorithm. EW: Egg weight, AW: Albumen weight, EWD: Egg width, EL: Egg length, YW: Yolk weight, SW: Shell weight.

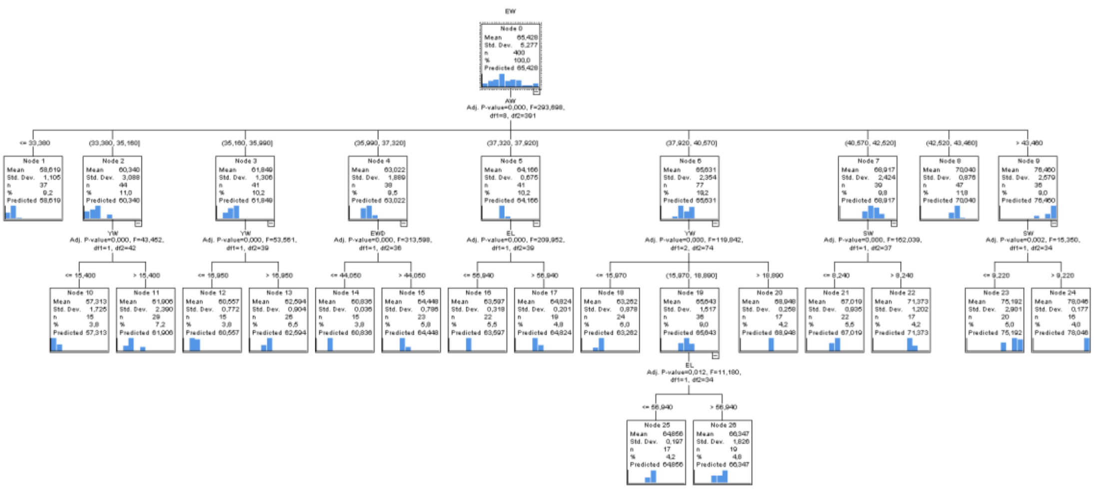

**Figure 2.** The regression tree diagram constructed by the Exhaustive CHAID algorithm. EW: Egg weight, AW: Albumen weight, EWD: Egg width, EL: Egg length, YW: Yolk weight, SW: Shell weight.

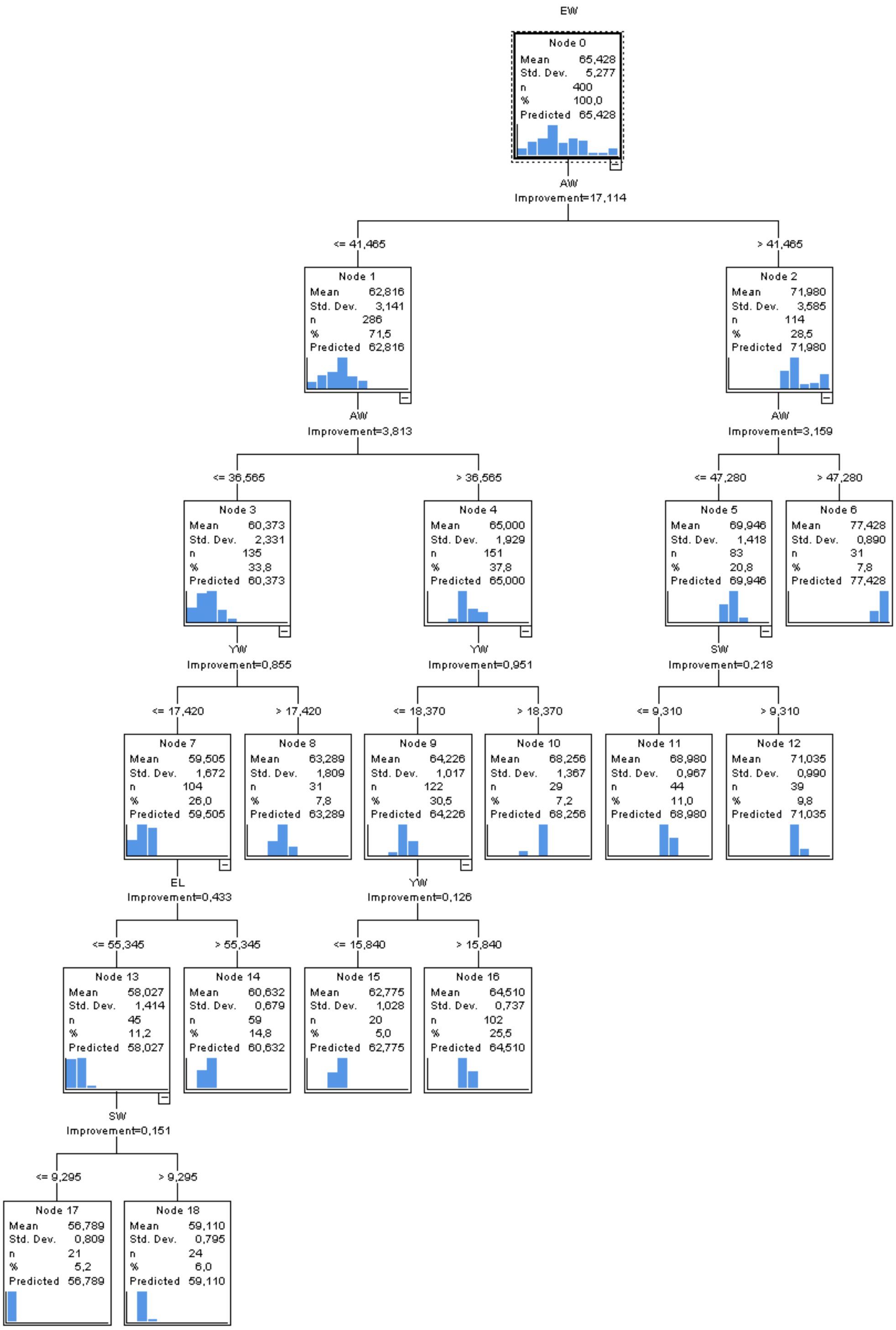


**Figure 3.** The regression tree diagram constructed by the CART algorithm. EW: Egg weight, AW: Albumen weight, EL: Egg length, YW: Yolk weight, SW: Shell weight.

**Random forest algorithm**

In a random forest application, the dataset is divided into 75% training and 25% test data. Test data are used to verify the outcomes, and training data are utilized to train machine learning algorithms. A total of 75% of the data in this study were utilized for training, while 25% were used for testing. Put another way, out of the 400 total data, 300 were utilized for training and 100 were used for testing the algorithm's output. The data set for the 10-fold validation method was randomly split into ten sections. Nine pieces were used for training, while the remaining one was set aside for testing. This process was carried out ten times, setting aside a separate tenth for testing each time. EL, EWD, SW, YW and AW variables were given as inputs and egg weight (EW) was predicted. When five independent variables in the data set were given as input, the rate of explaining the variance with the random forest algorithm in estimating the egg weight was found to be 99.74%. RF Algorithm error rate graph obtained according to the number of trees in the RF model is displayed in Figure 4. The significance level of the independent variables in the RF model is given in Figure 5.

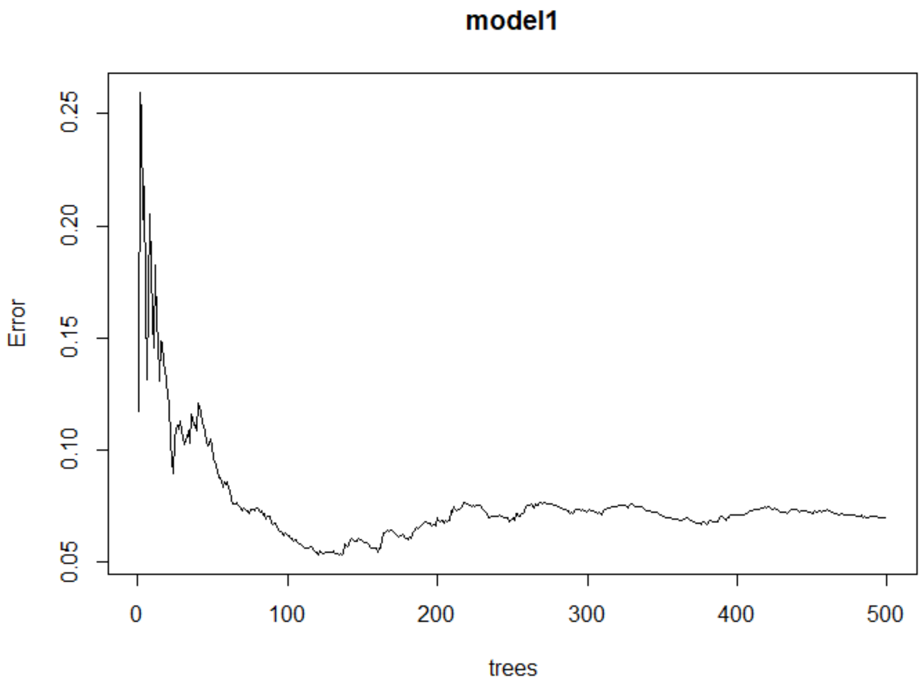


**Figure 4.** RF algorithm error rate of the model.

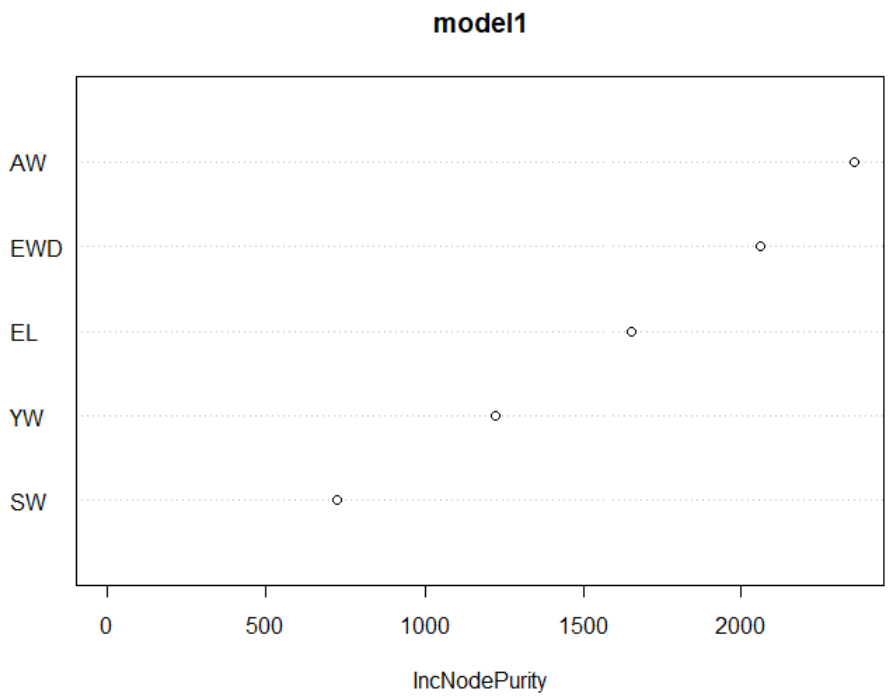


**Figure 5.** Significance graph of the variables of the RF Model. AW: Albumen weight, EWD: Egg width, EL: Egg length, YW: Yolk weight, SW: Shell weight.

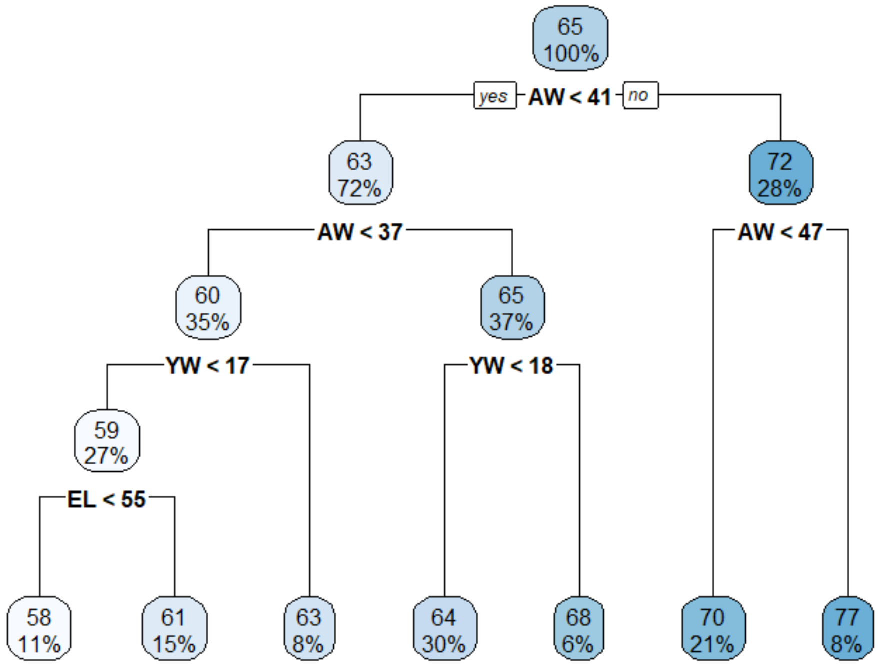


**Figure 6.** The regression tree diagram constructed by the RF algorithm. EW: Egg weight, AW: Albumen weight, YW: Yolk weight, EL: Egg length.

Figure 5 indicates the importance of 5 predictors (EL, EWD, SW, YW and AW) identified by the RF method for describing the egg weight of chickens. The most important variable was the albumen weight (AW) of chickens in the egg weight of the animals. The important score of the albumen weight (AW) variable is 2354.694. The important scores of other variables namely EWD, EL, YW and SW are 2058.814, 1653.559, 1224.283 and 723.886 respectively. The second important variable was egg width (EWD), and the lowest important variable was shell weight (SW). The significance level of the variables is listed as AW > EWD > EL > YW > SW. The diagram of the Random Forest algorithm to determine egg weight or quality characteristics was presented in Figure 6.

The RF algorithm results are summarized as follows:

N = 320 (The following explanations can be derived from the RF algorithm (n=320 nodes).

split, n, deviance, value, * denotes terminal node

1) root 320 9014.85600 65.32606

2) AW < 41.465 229 2165.32500 62.66878

4) AW < 36.565 112 604.77590 60.35973

8) YW < 17.42 85 237.56850 59.46424

16) EL < 55.345 36 76.61596 57.96694 *

17) EL >= 55.345 49 20.94940 60.56429 *

9) YW >= 17.42 27 84.45827 63.17889 *

5) AW >= 36.565 117 391.77010 64.87915

10) YW < 18.37 97 94.06907 64.19691 *

11) YW >= 18.37 20 33.58212 68.18800 *

3) AW >= 41.465 91 1163.35400 72.01308

6) AW < 47.28 66 122.90620 69.95030 *

7) AW >= 47.28 25 18.21746 77.45880 *

The RF algorithm results and the regression tree diagram shown in Figure 6 are summarized as follows.

- If AW < 41 g, then EW = 63 g.
- If AW ≥ 41 g, then EW = 72 g.
- If AW ≥ 41 g and AW <4 7 g, then EW = 70 g.
- If AW ≥ 41 g and AW ≥ 47 g, then EW = 77 g.
- If AW < 37 g, then EW = 60 g.
- If AW ≥ 37 g, then EW = 65 g.
- If AW < 37 g and YW ≥ 17 g, then EW = 63 g.
- If AW < 37 g and YW < 17 g, then EW = 59 g.
- If YW < 17 g and EL < 55 mm, then EW = 58 g.
- If YW < 17 g and EL ≥ 55 mm, then EW = 61 g.
- If AW < 37 g and YW < 18 g, then EW = 64 g.
- If AW ≥ 37 g and YW ≥ 18 g, then EW = 68 g.

The results of the prediction equation produced by MARS algorithm are presented in Table 4.

Selected 7 of 7 terms, and 3 of 5 predictors (nprune = 50). Termination condition: Reached maximum RSq 0.9990 at 7 terms. Importance: AW, YW, SW, EL-unused, EWD-unused

Number of terms at each degree of interaction: 16 (additive model): GCV = 2.2e-06;

RSS=0.000881, GRSq = 1, RSq = 1, CVRSq = 1

All the coefficients regarding MARS predictive model were statistically significant ($P$ < 0.001). The prediction equation produced by MARS algorithm is given below.

EW = 61.2 - 1 × max (0, 8.48 - SW) + 1 × max (0, SW - 8.48) - 1 × max (0, 17.8 - YW) + 1 × max (0, YW - 17.8) - 1 × max (0, 34.9 - AW) + 1 × max (0, AW - 34.9)

The concurrence between predicted and observed EW values in the MARS algorithm were examined as shown in Figure 7.

**Table 4.** Results of MARS model in the prediction of egg weight

| Coefficients | Estimate | Std. Error | t value | Pr (>\|t\|) |
|---|---|---|---|---|
| Intercept | 61.2 | 1.984e-04 | 308407 | < 2e-16 *** |
| bx[, -1]h(AW-34.93) | 1 | 2.097e-05 | 47678 | < 2e-16 *** |
| bx[, -1]h(34.93-AW) | -1 | 1.288e-04 | -7762 | < 2e-16 *** |
| bx[, -1]h(YW-17.79) | 1 | 1.502e-04 | 6660 | < 2e-16 *** |
| bx[, -1]h(17.79-YW) | -1 | 8.325e-05 | -12012 | < 2e-16 *** |
| bx[, -1]h(SW-8.48) | 1 | 8.052e-05 | 12419 | <2e-16 *** |
| bx[, -1]h(8.48-SW) | -1 | 4.532e-04 | -2209 | <2e-16 *** |

***: p < 0.001

As a result of the MARS algorithm, when SW >8.48 g, EW increases by 1 g (the contribution of this function to the model is 1). When YW > 17.8 g, EW increases by 1 g. Likewise, when AW > 34.9 g, EW increases by 1 g. The estimated egg weight (EW) value can be calculated by giving various estimated values to the independent variables. For example, the dataset could estimate the weight of chicken egg with egg quality characteristics: EL = 60 mm, EWD = 45 mm, SW= 10.25 g, YW = 19.5 g, and AW = 39.5 g. As follows by the MARS estimation equation:

EW = 61.2 - 1 × max (0, 8.48 - 10.25) + 1 × max (0, 10.25 - 8.48) - 1 × max (0, 17.8 - 19.5) + 1 × max (0, 19.5 - 17.8) - 1 × max (0, 34.9 - 39.5) + 1 × max (0, 39.5 - 34.9) = 69.25 g

In the same way, the weight of chicken egg with egg quality characteristics, which was EL = 58.75 mm, EWD = 46 mm, SW= 11 g, YW = 20.5 g, and AW = 41 g in the dataset, could be estimated using the MARS estimation equation:

EW = 61.2 - 1 × max (0, 8.48 - 11) + 1 × max (0, 11 - 8.48) - 1 × max (0, 17.8 - 20.5) + 1 × max (0, 20.5 - 17.8) - 1 × max (0, 34.9 - 41) + 1 × max (0, 41 - 34.9) = 72.7g

The plot between the predicted and observed EW value was displayed in Figure 8 for MARS algorithm. A graph of relative importance is given in Figure 9.

The predictive performances of the models built by Bagging MARS and MARS algorithms were found to be very close to each other. As a result of the bagging MARS algorithm, in the first bootstrap, an increase in egg weight can be expected for those with YW>19.42 g and AW>34.93 g. In the second bootstrap, an increase in egg weight can be expected for those with SW>8.26 g, YW>17.5 g and AW>41.06 g. In the third bootstrap, SW>8.48 g, YW>16.71 g and AW>34.93 g were predicted to increase egg weight. The fourth and fifth bootstraps projected that YW>17.79 g and AW>34.93 g would increase egg weight. Generally, chicken egg weight displayed an increasing tendency for SW>8.26 g, YW>16.71 g and AW>34.93 g. The plot between the predicted and observed EW value was displayed in Figure 9 for Bagging MARS algorithm.

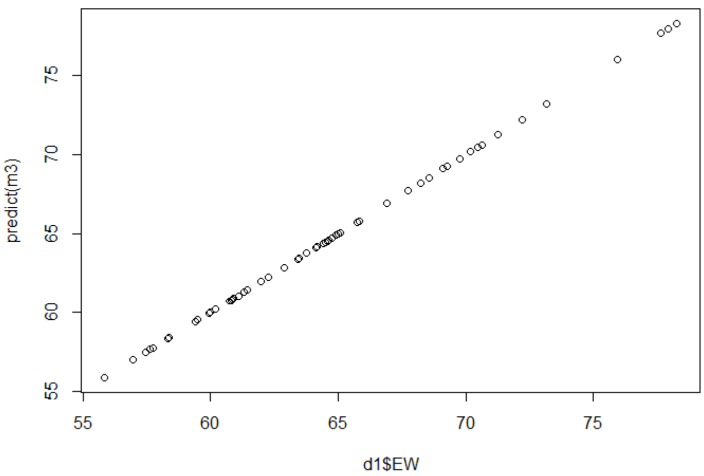


**Figure 7.** The concurrence between predicted and observed EW values in the MARS algorithm.

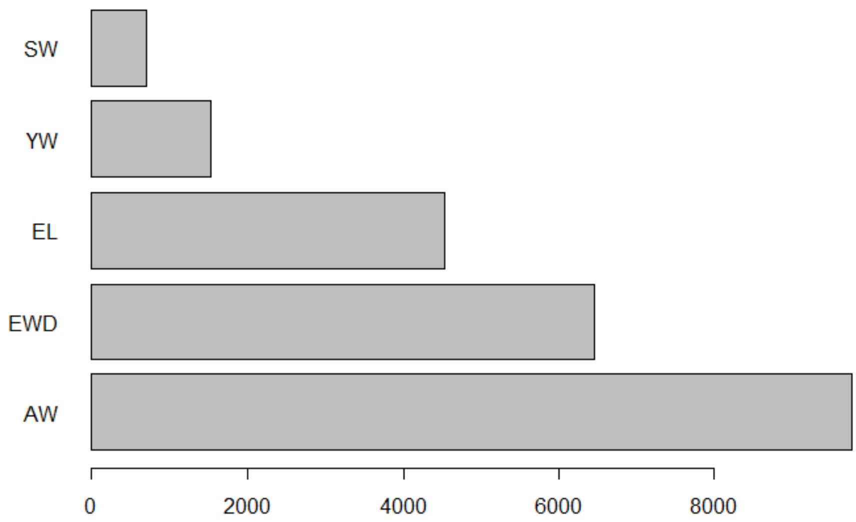


**Figure 8.** Graph of relative importance.

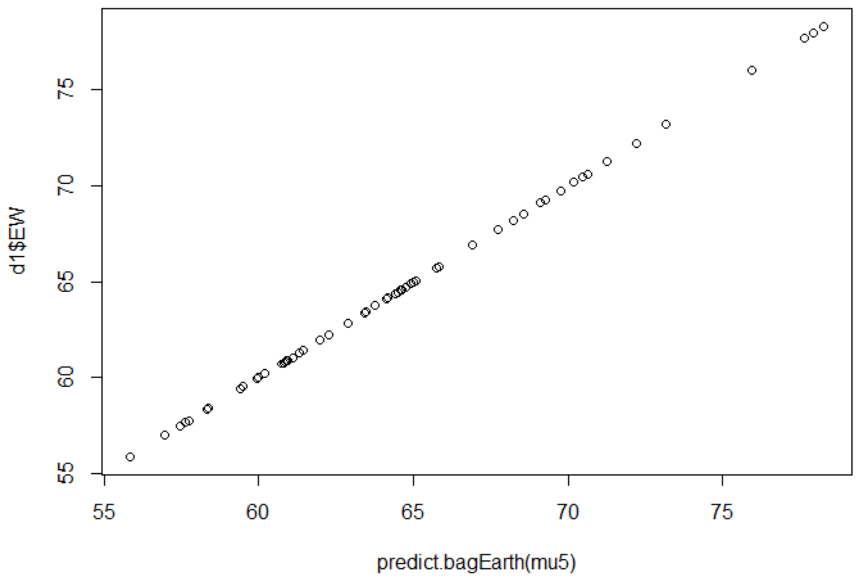


**Figure 9.** The concurrence between predicted and observed EW values in the Bagging MARS algorithm.

**Discussion**

In a study, the egg weight of chickens in Brovan Nera and Isa Brown genotypes was 56.38 and 55.43 g, respectively; egg length 5.54 and 5.61 cm; egg width 4.26 and 4.22 cm; shell weight was obtained as 6.20 and 5.04 g. The values of these properties in this study were found to be higher. This difference can be thought to be due to the genotype (Olaswumi and Ogunlade, 2009). In a different study, the mean egg weight of hens raised in Belgrade, Serbia, between 53 and 57 days was found to be 64.62, 55.10, and 56.60 g for the Tetra SL, Banat Naked Neck, and Svrljig hen genotypes, respectively. Banat Naked Neck and Svrljing chickens have significantly lower egg weight than the commercial hybrid Tetra SL, according to Škrbić *et al*. (2020). Different genotypes and ages generate considerable variances in egg weight and quality attributes. In the study of Olaswumi and Ogunlade (2008), the correlation coefficients between some internal and external quality characteristics of Exotic Brown Layer Breeders chicken eggs were investigated. The correlation coefficient between egg length-egg width was 0.34, the correlation between egg weight-egg length was 0.61, and the correlation coefficient between egg weight-shell weight was 0.66. Correlations between egg-egg width, egg length-shell weight, egg width-shell weight, egg width-yolk weight, egg weight-yolk weight, eight weight-albumen weight, egg length-albumen weight, shell weight-albumen weight 0.88, 0.27, 0.61, 0.48, 0.55, 0.91, 0.64 and 0.35, respectively. The difference in correlation coefficients obtained in this study may be due to race. According to Mathapo *et al.* (2022), the correlation coefficient does not indicate the effect of independent variables on dependent variables. Hence, data mining methods were used to indicate the effect of egg characteristics on the egg weight of the Lohmann Brown chicken breed. The results indicated that all the data mining algorithms used produced good results for the estimation of egg weight using egg characteristics of Lohmann Brown chicken breed. The CART data mining technique results revealed that albumen weight was the greatest egg quality feature for predicting egg weight. These findings contrast with the CART findings of Alapatt *et al*. (2022), who discovered that shell weight is the best predictor of egg weight in the White Leghorn chicken breed. While Liswaniso *et al*.'s (2021) CART algorithm revealed that egg length is the key predictor of egg weight in Zambian indigenous free-range chicken. RF, MARS and Bagging MARS algorithms findings indicated that albumen weight, egg width and egg length were the most critical predictors influencing egg weight of Lohmann Brown layers. Using the MARS algorithm, Canga et al. (2021) found that the egg shape index is the best predictor of egg weight in Lohmann LSL Classic White hybrid hens. Canga and Boga (2019) Bagging MARS and MARS algorithms found that feed consumption positively influenced egg weight of Lohman breed chickens. Based on the goodness of fit criteria findings of this study MARS and Bagging MARS algorithms are the best algorithm followed by RF, CART, Exhaustive CHAID and CHAID being the last algorithm. In conclusion, albumen weight was the best estimator of egg weight in all the data mining methods. In the CART, albumen weight was determined as the most effective EW predictor and albumen weight greater than 47.28 g had heavier EW, In the RF, MARS and Bagging MARS methods, albumen weight, egg width and egg length were the most important traits. The supremacy order in the predictive accuracy of the algorithms was found as MARS ≈ Bagging MARS > RF > CART > Exhaustive CHAID > CHAID. This study suggests that data mining methods could be used to estimate egg weight from egg quality traits and might be useful for chicken breeding programs.

## Declarations

**Ethical approval**

The experimental procedures were conducted following the University of Limpopo Animal Research Ethics Committee (AREC/14/2021: PG).

**Acknowledgments**

The authors thank the Kitamu farm manager in Ntsima village, Polokwane, Limpopo Province of South Africa for allowing data collection.

**Conflicts of Interest**

The authors declare no conflict of interest.

## References

Ahmadi F & Rahimi F. 2011. Factors affecting quality and quantity of egg production in layers: a review. World Applied Sciences Journal, 12(3): 372–384.

Alabi OJ, Ng’ambi JW, Norris D & Egena SSA. 2012. Comparative study of three indigenous chicken breeds of South Africa: body weight and linear body measurements. Agricultural Journal, 7(3): 220–225. DOI: 10.3923/aj.2012.220.225

Alapatt A, Chaudhary JK, Shyamsana N, Tolenkhomba TC, Kalita G & Jagan Mohanarao G. 2022. Prediction of egg weight from egg quality characteristics by using regression analysis methods in White Leghorn chicken. International Journal of Livestock Research, 12(2): 40–48. DOI: 10.5455/ijlr.20210929030220

Biggs D, De Ville B & Suen B. 1991. A method of choosing multiway partitions for classification and decision trees. Journal of Applied Statistics, 18: 49–62. DOI: 10.1080/02664769100000005

Breiman L, Friedman JH, Olshen R & Stone CJ. 1984. Classification and regression tree, Wadsworth Brooks/Cole Advanced Books and Software, Pacific California.

Breiman L. 2001. Random forest, Mach. Learn., 45: 5–32.

Canga D & Boga M. 2019. Hayvancılıkta Mars Kullanımı Ve Bır Uygulama. III. International Scientific and Vocational Studies Congress – Science and Health 27–30 June 2019, Ürgüp, Nevşehir / Turkiye.

Canga D, Yavuz E & Efe E. 2021. Prediction of egg weight using MARS data mining algorithm through R. KSÜ Tarım ve Doğa Derg, 24(1): 242–251.

Chen JL & Li GS. 2014. Evaluation of support vector machine for estimation of solar radiation from measured meteorological variables. Theoretical and Applied Climatology, 115: 627–638. DOI: 10.1007/s00704-013-0924-y

Chen T, Guestrin C & Boost XG. 2016. A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. ACM; 785–794.

Eyduran E. 2020. ehaGoF: Calculates Goodness of Fit Statistics. R package version 0.1.0. URL: https://CRAN.Rproject.org/package=ehaGoF. (access date: September 27, 2022).

Friedman JH. 1991. Multivariate adaptive regression splines. The Annals of Statistics, 19: 1-67.

Goh ATC, Zhang W, Zhang Y, Xiao Y & Xiang Y. 2016. Determination of earth396pressure balance tunnel-related maximum surface settlement: a multivariate adaptive regression splines approach. Bull. Bulletin of Engineering Geology and the Environment, 77: 489–500.

IBM Corp. 2019. Released. IBM SPSS Statistics for Windows, Version 26.0. Armonk, NY: IBM Corp. DOI: 10.3923/aj.2012.220.225

James G, Witten D, Hastie T, Tibshirani R. 2013. An introduction to statistical learning: with applications in R. New York: Springer-Verlag.

Kornacki J & Cwik J. 2005. Statistical learning systems (in Polish). WNT, Warsaw.

Kunn M & Johnson K. 2013. Applied predictive modeling. NY. DOI: 10.1007/978-1-4614-6849-3.

Kutu FR & Asiwe JAN. 2010. Assessment of maize and dry bean productivity under different intercrop systems and fertilization regimes. African Journal of Agricultural Research, 5: 1627–1631.

Liddle AR. 2007. Information criteria for astrophysical model selection. Monthly Notices of the Royal Astronomical Society: Letters, 377: L74-L78.

Liswaniso S, Qin N, Tyasi TL & Chimbaka IM. 2021. Use of data mining algorithms CHAID and CART in predicting egg weight from egg quality traits of indigenous free-range chickens in Zambia. Advanced Animal and Veterinary Science, 9(2): 215–220. DOI: 10.17582/journal.aavs/2021/9.2.215.220

Mathapo MC, Mugwabana TJ & Tyasi TL. 2022. Prediction of body weight from morphological traits of South African non-descript indigenous goats of Lepelle Nkumbi Local Municipality using different data mining algorithm. Tropical Animal Health and Production, 54: 102. DOI: 10.1007/s11250-022-03096-9.

Olaswumi SO & Ogunlade JT. 2008. Phenotypic correlation between some external and internal egg quality traits in the Exotic Isa Layer Breeds. Asian Journal of Poultry Science, 2(1): 30–35.

Olaswumi SO & Ogunlade JT. 2009. The effect of genotype and age of layer breeders on egg quality traits. Nigerian Journal of Animal Production, 36(2): 228–236. DOI: 10.51791/njap.v36i2.1339

Pires PGS, Baveresco C, Prato BS, Wirth ML & Moraes PO. 2021. The relationship between egg quality and hen housing systems - A systematic review. Livestock Science, 250: 104597. DOI: 10.1016/j.livsci.2021.104597

R Core Team. 2021. R: A Language and environment for statistical computing. (Version 4.1) [Computer software]. Retrieved from https://cran.r-project.org. (R packages retrieved from MRAN snapshot 2022-01-01).

Ramadhan MM, Sitanggang IS, Nasution FR & Ghifari. 2017. Parameter Tuning in Random Forest Based on Grid Search Method for Gender Classification Based on Voice Frequency. International Conference on Computer, Electronics and Communication Engineering (CECE 2017) ISBN: 978-1-60595-476-9

Škrbić Z, Lukić M, Petričević V, Bogosavljević-Bošković S, Rakonjac S, Dosković V & Tolimir N. 2020. Quality of eggs from pasture rearing layers of different genotypes. Biotechnology in Animal Husbandry, 36(2): 181–190. DOI: 10.2298/BAH2202125S

Stadelman WJ. 1977. Quality identification of shell eggs in egg science and technology. Ed. W.J. Stadelman, D.J. Cotterill, AVI Publishing Company Inc. Wesport, Connecticut 2nd Edition, pg. 33.

Takma C, Atil H & Aksakal V. 2012. Comparison of multiple linear regression and artificial neural network models goodness of fit to lactation milk yields, Kafkas Univ Vet Fak Derg. 18: 941–944.

Troncoso A, Salcedo-Sanz S, Casanova-Mateo C, Riquelme JC & Prieto L. 2015. Local models-based regression trees for very short-term wind speed prediction. Renewable, Energy, 81: 589–598. DOI: 10.1016/j.renene.2015.03.071

Tso GKF & Yau KKW. 2007. Predicting electricity energy consumption: a comparison of regression analysis, decision tree and neural networks. Energy, 32(9): 1761–1768. DOI: 10.1016/j.energy.2006.11.010

Tutkun M, Denli M & Demirel R. 2018. Productivity and egg quality of two-layer hybrids kept in free-range system. Turkish Journal of Agriculture - Food Science and Technology, 6(10): 1444–1447. DOI: 10.24925/turjaf.v6i10.1444-1447.2070

Ukwu HO, Ezihe CO, Asaa SK & Anyogo ME. 2017. Effect of egg weight on external and internal egg quality traits of Isa Brown egg layer chickens in Nigeria. Veterinary and Animal Science, 2: 126–132. https://doi. org/10.31248/JASVM2017.051

Willmott A & Matsuura K. 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research, 30: 79–82. DOI: 10.3354/cr030079

Yang HM, Yang Z, Wang W, Wang ZY, Sun HN, Ju XJ & Qi XM. 2014. Effects of different housing systems on visceral organs, serum biochemical proportions, immune performance and egg quality of laying hens. European Poultry Science, 78: 1–9. DOI: 10.1399/eps.2014.48

Zhang W & Goh ATC. 2016. Multivariate adaptive regression splines and neural network models for prediction of pile drivability. Geoscience Frontiers, 7: 45–52. DOI: 10.1016/ j.gsf.2014.10.003