

COMPARISON OF SELECTED DECISION TREE ALGORITHMS IN THE PREDICTION OF BODY WEIGHT IN AWASSI LAMBS

Adile Tatlıyer Tunaz

Department of Animal Science, Faculty of Agriculture, Kahramanmaraş Sütçü İmam University,
Kahramanmaraş/Turkey

Corresponding author's email: atatliyer@ksu.edu.tr

ABSTRACT

The present study was conducted to evaluate the comparative predictive performances of Classification and Regression trees (CART), Chi Squared Automatic Interaction Detector (CHAID) and Exhausted CHAID algorithms used to predict body weights of Awassi lambs at 60-d (W60) and 90-d (W90) of age. For this purpose, 730 Awassi lamb records were collected from 3 base flocks in Turkey in 2014-2016. The potential predictors included in this study were dam age, sex, birth type, flock, lambing season and birth year. In order to determine the best one among these decision tree algorithms, model evaluation criteria i.e. *RMSE*, *MAPE*, *RAE*, *SD_{ratio}*, *MAD*, Pearson coefficient, Coefficient of determination (R^2) and adjusted coefficient of variation (R^2_{Adj}) values were calculated. For the prediction of W60 and W90, the best decision tree algorithm was found to be the CART algorithm. R^2 for W60 and W90 were 0.614 and 0.978 and RMSE estimates for W60 and W90 were 0.94 and 0.321, respectively. The influential predictors affecting W90 were flock and W60. However, flock, birth weight (BW), birth type and birth year were found as significant factors for W60. In conclusion, CART algorithm may be a useful tool in describing breed standards of the Awassi for selection purposes in animal breeding. Also, it outperformed Exhausted CHAID and CHAID decision tree algorithms in predictive performance to predict W60 and W90 of Awassi lambs.

Key words: Awassi, Body Weight, Regression Tree.

<https://doi.org/10.36899/JAPS.2021.4.0288>

Published online December 15, 2020

INTRODUCTION

In animal breeding, selection of superior individuals to be parents of next generation provides genetic improvement for economic characteristics such as birth weight. (Olfaz *et al.*, 2019). It is imperative in any breeding program to develop selection strategies and to determine the factors that affect the quantitative characteristics of economic importance (Koç, 2016). In this context, analysis of the data on performance traits using robust statistical techniques along with proper recording system is inevitable to make an accurate decision for breeding purposes (Eyduran *et al.*, 2017; AYTEKİN *et al.*, 2018).

Meat yield characteristics are influenced by various environmental and genetic factors and reproductive conditions (Balta and Topal, 2018). In addition to these factors, several researchers reported some other factors i.e. lamb sex, dam age, birth year, breed, parity, season of birth, birth type, low birth weight, injury and mothering ability were able to influence lamb viability (Morris *et al.*, 2000; Susic *et al.*, 2005; Khan *et al.*, 2006; Mandal *et al.*, 2007; Sawalha *et al.*, 2007; Vatankhah and Talebi, 2009). Generally least squares method had been used to determine the relationship among these factors (Sevgenler, 2019). However, the presence of multicollinearity among the predictors affecting response

variables like body weights at different ages generally reduces the reliability of the least squares method and makes it difficult to interpret the results (Khan *et al.*, 2014). In order to remove the multicollinearity problem, which results from very strong correlations between predictors, it was reported that alternative statistical techniques were available; namely, ridge regression analysis (Orhan *et al.*, 2016), factor analysis scores (Tariq *et al.*, 2012) and principal component scores (Eyduran *et al.*, 2013a) in multiple linear regression analysis. Apart from these studies, path analysis and multiple linear regression analysis to reveal the causality relationships between dependent and predictor variables were used by some researchers (Çankaya and Abacı, 2012; Önder and Abacı, 2015). However, statistical interpretation was not satisfactory due to the complexity of the data and violation of linearity, constant-variance, normality and multicollinearity assumptions; so more sophisticated approaches such as data mining and artificial neural network algorithms were used to achieve high predictive accuracy with recent developments of the computer information management systems (Grzesiak and Zaborski, 2012; AYTEKİN *et al.*, 2018; Olfaz *et al.*, 2019).

In recent years, there were numerous studies on popular data mining algorithms (CHAID, CART, Exhausted CHAID and MARS) in sheep and goat breeding (Yakubu, 2012; Ali *et al.*, 2015; Eyduran, 2016; Eyduran

et al., 2016; Çelik *et al.*, 2017; Eyduran *et al.*, 2017; Karabacak *et al.*, 2017; Balta and Topal, 2018; Çelik *et al.*, 2018; Aksoy *et al.*, 2019; Olfaz *et al.*, 2019).

The data mining techniques are used for classification and regression type problems (Eyduran *et al.*, 2019). In the data mining technique, classification algorithms are used in Decision Trees and Neural Networks, on the other hand, prediction algorithms are used in Regression, Association Rules and Clustering (Palaniappan and Awang, 2008; Kovalchuk *et al.*, 2017, 2018). Decision tree building algorithms include Classification and Regression trees (CART), Chi Squared Automatic Interaction Detector (CHAID) and Exhausted CHAID. These techniques do not require normality assumptions unlike classical statistical techniques (Akin *et al.*, 2017, 2018, 2020). Furthermore, these powerful algorithms, visually-flexible modern analytical techniques, can analyze ordinal and nominal data as well as a continuous data set (Eyduran *et al.*, 2013b). In addition, in case the existence of multicollinearity problem among independent variables, outliers and missing values in data set, these algorithms can produce effective solutions with a different process (Ali *et al.*, 2015). Among these tree-based algorithms, CHAID and Exhausted CHAID that is derived from CHAID allow more than two splits while CART employs binary splits (Kusiak *et al.*, 2010). Although Exhausted CHAID includes the comprehensive merging with respect to CHAID, both algorithms have three essential steps: merging, splitting and stopping (Orhan *et al.*, 2016). Splitting and stopping steps in Exhaustive CHAID algorithm are the same with CHAID algorithm; However, the exhaustive CHAID at merging step operates an exhaustive procedure to merge any similar pair until only a single pair is obtained. In addition, CHAID and Exhausted CHAID algorithms produce adjusted p values in the decision tree structure by using Bonferroni adjustment (Çelik *et al.*, 2017).

Classification and Regression trees (CART) analysis, generates homogenous nodes until achieving the strongest correlation between predicted and observed values in the response variable. The difference between CHAID and CART (developed by Breiman (1984)) is that pruning option continuously is active for avoiding from unnecessary branches in CHAID algorithms apart from CART algorithms which is needed for activation by scientists in IBM SPSS software (Eyduran, 2016). Orhan *et al.* (2016) reported that continuous variables were converted to ordinal variables in the CHAID algorithms.

Regression tree (RT) data mining algorithms provide great convenience in the interpretation of data visually without any mathematical assumption (Eyduran *et al.*, 2019). For this purpose, live body weights of Awassi lambs at 60-d and 90-d (W60 and W90) of age were predicted with the most commonly used RT data mining algorithms such as CHAID, Exhausted CHAID and CART. Predictive performances of these algorithms were

evaluated and influential factors were determined in the tested algorithms

Among sheep breeds raised in Turkey, Awassi breed is the fat-tailed sheep having high adaptability to high hot temperatures and insufficient feeding conditions: it has a superior breeding material in milk and lamb production for breeding and adaptation studies, and has gained importance due to increasing demand in red meat (Üstüner and Oğan, 2013). To our knowledge; however, there are few studies on the application of these powerful algorithms on lambs of Awassi sheep. Weaning time and after weaning time for sheep breeding is important in order to get the expected efficiency in breeding and to reach the desired result in a shorter time. Therefore, the aim of the present study was to comparatively evaluate predictive performances of data mining algorithms in order to predict live weights of Awassi lambs at 60-d and 90-d of age using with birth type, birth year, dam age, birth season, sex and lambing season.

MATERIALS AND METHODS

Data Collection: In the present study, 730 Awassi lambs raised under breeding conditions from base flock in Turkey were used. Animal records of base flock were taken from three different farms ($n_A=193$, $n_B=209$, $n_C=308$) which are located in the same region of Turkey in the years of 2014, 2015 and 2016. Live body weights of Awassi lamb at 60-d and 90-d (W90) were considered as response variables. Body weight at 60-d was early weaning weight in the flock. The predictors included in the study were dam age (2, 3, 4, 5 and 6 years), sex (male and female), birth type (single and twin), lambing season (major season defined as in October, November and December and minor season defined as in January, February, and March) and birth year (2014, 2015 and 2016).

Statistical models: The CART, CHAID and Exhausted CHAID regression tree analyses were performed. A ten-fold cross validation was considered in handling decision tree algorithms. Pruning operations was activated in CART. In all the statistical analyses, to achieve the optimum decision tree diagram, parent node and child node were preferred at 40:20, respectively. When substitution cost was very close to cross-validation cost for each of the handled decision tree algorithms, elimination of overfitting problem was provided and then, it was achieved to the optimum tree structures.

To determine the ideal algorithms, numerous goodness of fit criteria was used as described by Ali *et al.* (2015):

- Coefficient of determination (R^2)

$$R^2 = \left[1 - \frac{\sum_{i=1}^n (y_i - y_{ip})^2}{\sum_{i=1}^n (y_i - \bar{Y})^2} \right]$$

- Adjusted coefficient of determination (R^2_{Adj})

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

- Pearson correlation coefficient (r) between actual and predicted values in each of dependent variables,

- Akaike information criterion (AIC):

$$AIC = n \cdot \ln \left[\frac{1}{n} \sum_{i=1}^n (y_i - y_{ip})^2 \right] + 2k$$

- Root-mean-square error ($RMSE$):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_{ip})^2}$$

- Mean error (ME):
- Mean absolute deviation (MAD):

$$MAD = \frac{1}{n} \sum_{i=1}^n |y_i - y_{ip}|$$

- Standard deviation ratio (SD_{ratio}):
- Global relative approximation error (RAE):

$$RAE = \sqrt{\frac{\sum_{i=1}^n (y_i - y_{ip})^2}{\sum_{i=1}^n y_i^2}}$$

- Mean absolute percentage error ($MAPE$):

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - y_{ip}}{y_i} \right| \cdot 100$$

Where; n : number of animals in data, k : number of model parameters, y_i : the real value of dependent variable, y_{ip} : the predicted value of dependent variable, s_m : standard deviation of residuals and s_d : standard deviation of dependent variable.

The most suitable algorithm was chosen according to the lowest $RMSE$, $MAPE$, RAE , SD_{ratio} and MAD values and the highest Pearson correlation coefficient, R^2_{Adj} and R^2 values (Mohammad *et al.*, 2012; Ali *et al.*, 2015).

Statistical evaluations of the decision tree algorithms were performed using SPSS 2006 (version

15.0). To estimate goodness of fit criteria, ehaGoF package written by Eyduran (2019) was employed in R studio software (R Core Team, 2018).

RESULTS

Descriptive statistics and Tukey multiple comparison results of live body weights at birth, 60-d and 90-d are shown in Table 1.

The decision tree diagrams generated by the CART algorithms for W60 and W90 are shown in Figures 2 and 3. The Pearson correlation coefficient (r) between observed and predictive live weights of Awassi lambs at 60-d was 0.783 for CART algorithm ($P < 0.01$).

When Figure 2 was examined, significant independent variables providing the contribution to constructing CART diagram were found as farm, birth weight, birth type, and birth year. Node 0 ($N=730$) or root node was divided into two smaller groups which were Node 1 (farm = B) and Node 2 (farm =A; C) according to farm effect at the first tree depth. The mean live weight of lambs at 60-d in Node 1 and Node 2 were predicted as 16.326 ± 1.903 kg and 14.928 ± 1.097 kg, respectively. Node 1 ($N=209$) was branched into two smaller subgroups (Nodes 3 and 4) based on birth weight (BW). In the second tree depth of CART algorithm, Node 3 and Node 4 were not divided into new nodes at subsequent tree depth for live weight of lambs at 60-d hence, Node 3 and Node 4 were defined as terminal nodes. Mean W60 of lambs with $BW > 3.890$ kg raised in farm B in Node 4 were higher than that of lambs with $BW < 3.890$ kg raised in farm B (Node 3) (14.960 vs. 17.190 kg). Among the all nodes, Node 4 had the highest mean in W60. Similarly, Node 2 was splined into two homogenous groups (Node 5 and Node 6) based on BW as previously mentioned. Node 6 was observed to be a terminal node in CART diagram. Based on birth type, 412 lambs in Node 5 were partitioned into two new nodes which were Node 7 and Node 8. Node 7 was described as a group of single lambs with $BW < 4.375$ kg in Farms A and B with mean live weight of lambs at 60-d predicted as 14.860 ± 1.061 kg, and Node 8 (also called terminal node) was identified as a group of twin lambs with $BW < 4.375$ kg in Farms A and B with mean live weight of lambs at 60-d predicted as 13.795 ± 0.644 kg. The live weight of lambs at 60-d of twin-born lambs were usually weaker than those of singles when lambs with $BW < 4.375$ kg raised in Farms A and C were taken into consideration. Node 9 constituted by Node 7 was affected by the year of birth. Node 10 (a group of single-lambs with $BW < 4.375$ kg born in 2015 and 2016 in Farms A and C; mean live weight of lambs at 60-d 15.085 ± 1.044 kg, $n = 256$) was heavier than Node 9 (a group of single-lambs with $BW < 4.375$ kg born in 2014 in Farms A and C; mean live weight of lambs at 60-d = 14.356 ± 0.919 kg).

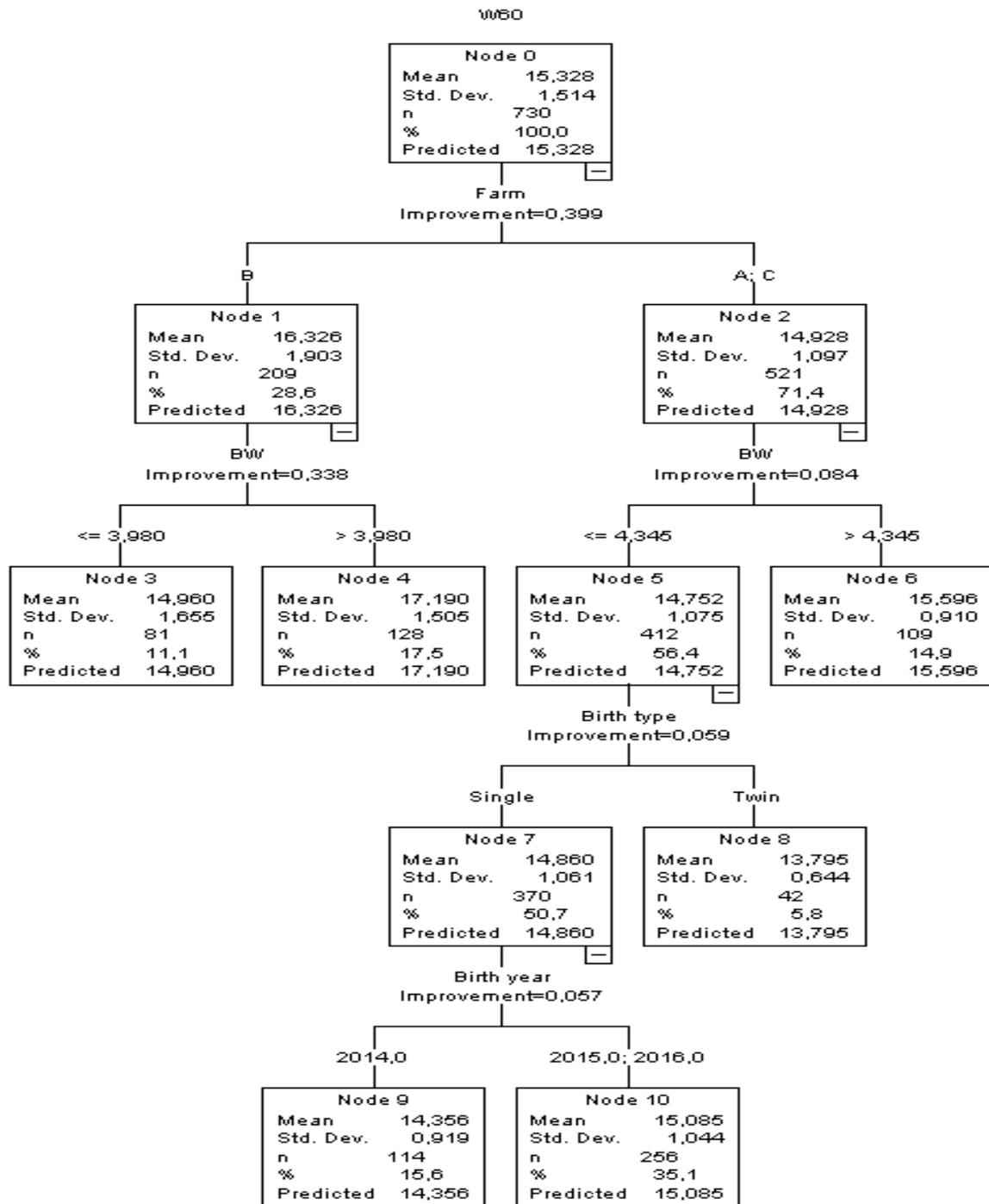


Figure 2. Decision tree structure constructed for CART algorithms for W60

The regression tree diagram obtained by CART algorithm for W90 had 10 child nodes by starting with root node (also known as Node 0 at the top of CART algorithm) with live weight of Awassi lambs at 90-d 21.304 ± 2.152 kg (n=730 Awassi lambs).

Regression tree diagram formed by CART algorithms for live weight of Awassi lambs at 90 days is

shown in Figure 3. Very strongly Pearson correlation coefficient (r) between observed and predicted W90 values for CART was estimated ($r=0.989$, $P<0.01$). Essentially when Figure 3 was examined, significant variables i.e. farm and W60 were entered into the constructed CART decision tree structure.

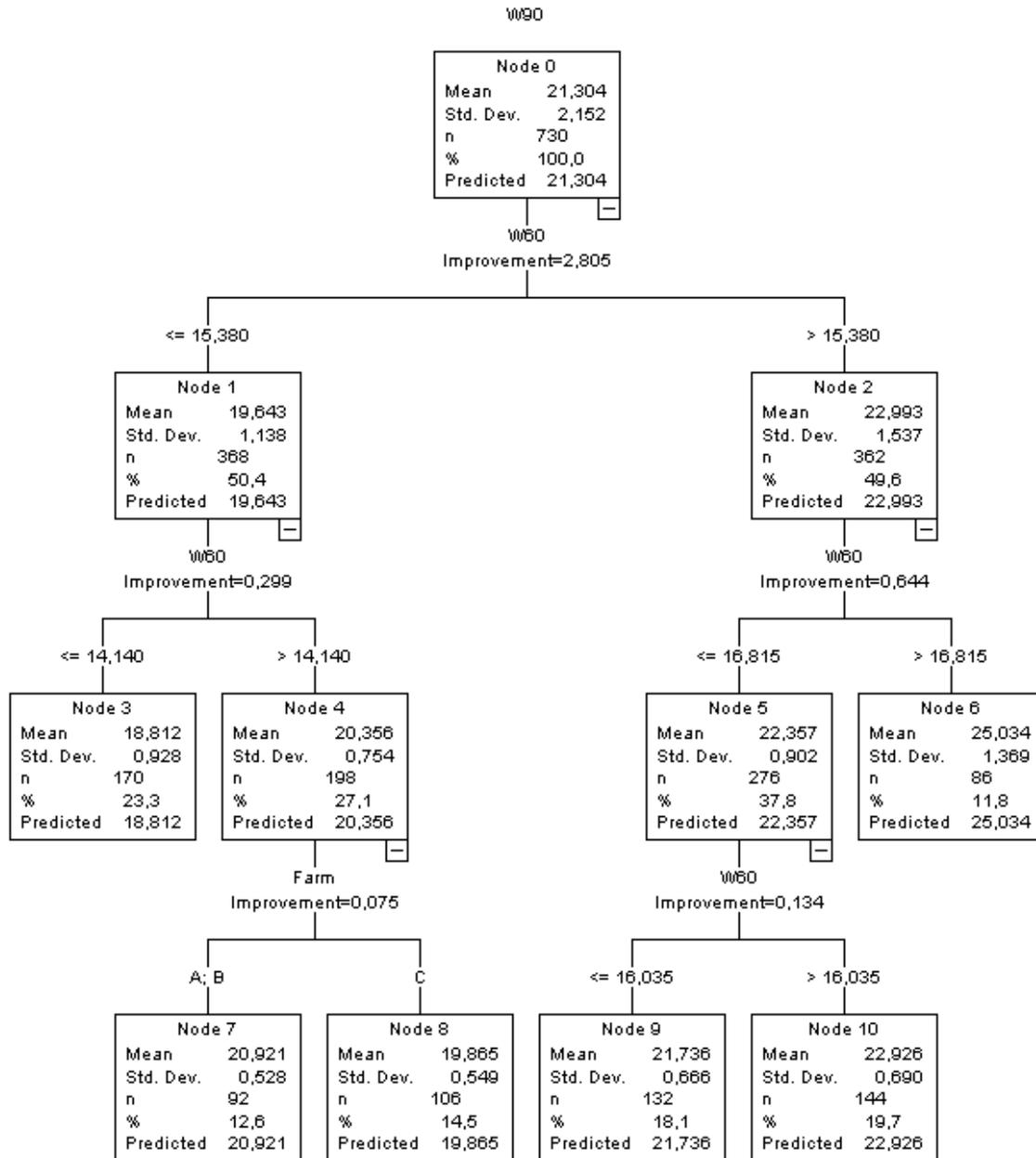


Figure 3. Regression tree diagram generated by CART algorithm for W9

Nodes 3 and 6-10 were the terminal nodes. The root node, defined as Node 0, was first divided into new

child nodes, Node 1 and Node 2, based on live weight of lambs at 60-d.

Table 1. Descriptive statistics of body weight, live weights of Awassi lambs at 60-d and 90-d.

Predictors		N	BW (Mean ± SD)	W60 (Mean ± SD)	W90 (Mean ± SD)
Dam Age	2	85	3.77 ± 0.60 ^b	15.59 ± 1.82 ^a	21.36 ± 2.57 ^{ab}
	3	195	3.90 ± 0.58 ^{ab}	15.28 ± 1.55 ^a	21.02 ± 2.24 ^b
	4	202	4.01 ± 0.56 ^a	15.22 ± 1.44 ^a	21.13 ± 2.05 ^b
	5	167	4.01 ± 0.65 ^a	15.38 ± 1.49 ^a	21.50 ± 2.02 ^{ab}
	6	81	3.94 ± 0.68 ^{ab}	15.35 ± 1.30 ^a	21.97 ± 1.81 ^a
Sex	Female	367	3.72 ± 0.56 ^b	15.16 ± 1.51 ^b	21.02 ± 2.11 ^b

Birth Type	Male	363	4.17 ± 0.57 ^a	15.50 ± 1.50 ^a	21.59 ± 2.16 ^a
	Single	682	3.99 ± 0.59 ^a	15.43 ± 1.50 ^a	21.40 ± 2.18 ^a
	Twin	48	3.30 ± 0.49 ^b	13.92 ± 0.77 ^b	19.99 ± 1.08 ^b
Lambing Season	Major	336	3.90 ± 0.58 ^b	15.14 ± 1.53 ^b	20.95 ± 2.20 ^b
	Minor	394	3.98 ± 0.63 ^a	15.49 ± 1.49 ^a	21.60 ± 2.06 ^a
Farm	A	193	3.95 ± 0.78 ^b	14.97 ± 1.16 ^b	21.68 ± 1.77 ^b
	B	209	4.08 ± 0.58 ^a	16.33 ± 1.90 ^a	22.66 ± 2.37 ^a
	C	328	3.85 ± 0.48 ^b	14.90 ± 1.06 ^b	20.22 ± 1.58 ^c
Birth Year	2014	185	3.92 ± 0.65 ^a	15.08 ± 1.56 ^b	21.03 ± 2.28 ^a
	2015	324	3.92 ± 0.57 ^a	15.51 ± 1.65 ^a	21.38 ± 2.28 ^a
	2016	221	4.00 ± 0.63 ^a	15.27 ± 1.21 ^{ab}	21.42 ± 1.80 ^a

^{a,b,...}: Means with different letter in each columns are statistically different (P<0.05).

At the first CART tree depth, Node 1 was a subgroup of Awassi lambs with W60 < 15.380 kg, but Node 2 was determined to be a subgroup of Awassi lambs with W60 > 15.380 kg. Awassi lambs assigned to Node 1 were averagely lighter in W90 compared with those assigned to Node 2 (19.643 kg vs. 22.993 kg in W90). Nodes 1 and 2 were split into new smaller subgroups (Nodes 3-6) in W60, respectively. Node 3 was a subgroup of Awassi lambs with W60 < 14.140 kg, whereas Node 4 was understood to be a subgroup of Awassi lambs with

14.140 < W60 < 15.380 kg (Figure 2). Mean of Awassi lambs in Node 3 was estimated lighter in W90 compared with those in Nodes 4-6 at the second tree depth. Especially, Awassi lambs whose W60 was heavier than 16.815 kg had the averagely heaviest W90 value of 25.034 kg (Node 6). Node 4 was divided into two smaller subgroups i.e. Node 7 and Node 8 according farm factor. However, Node 5 was branched into two smaller subgroups i.e. Node 9 and Node 10 according to W60.

Table 2. Goodness of fit criteria for predicting live weights of Awassi lambs at 60-d and 90-d.

	W60			W90		
	CHAID	ExhaustedCHAID	CART	CHAID	ExhaustedCHAID	CART
* <i>pearcorr</i>	0.734	0.731	0.783	0.974	0.974	0.989
* <i>CV %</i>	6.709	6.739	6.138	2.288	2.298	1.507
* <i>SDratio</i>	0.679	0.682	0.622	0.227	0.227	0.149
* <i>RMSE</i>	1.028	1.032	0.94	0.487	0.489	0.321
* <i>ME</i>	0	-0.001	-0.001	0.001	0.001	-0.001
* <i>RAE</i>	0.067	0.067	0.061	0.023	0.023	0.015
* <i>MAPE</i>	5.141	5.189	4.765	1.491	1.508	1.217
* <i>MAD</i>	0.778	0.785	0.722	0.314	0.318	0.254
* <i>R²</i>	0.538	0.534	0.614	0.949	0.948	0.978
* <i>R²_{Adj}</i>	0.535	0.531	0.611	0.949	0.948	0.978
* <i>AIC</i>	49.842	56.373	-79.981	-1039.884	-1034.003	-1650.136
* <i>AICc</i>	49.925	56.456	-79.898	-1039.801	-1033.92	-1650.053

**CHAID* (Chi Squared Automatic Interaction Detector), *Exhausted CHAID* (Exhausted Chi Squared Automatic Interaction Detector), *CART* (Classification and Regression trees), *pearcorr* (Pearson correlation coefficient), *CV %* (coefficient of variation), *SDratio* (Standard deviation ratio), *RMSE* (Root-mean-square error), *ME* (Mean error), *RAE* (Global relative approximation error), *MAPE* (Mean absolute percentage error), *MAD* (Mean absolute deviation), *R²* (Coefficient of Determination), *R²_{Adj}* (Adjusted Coefficient of Determination), *AIC* (Akaike information criterion), *AICc* (Adjusted AIC)

The results of several goodness of fit criteria are summarized in Table 2. According to Table 2, coefficient of variation (*CV%*) values for CHAID, Exhausted CHAID and CART decision tree algorithms were estimated for W60 as 6.709, 6.739 and 6.138, respectively. *CV* values for W90 were estimated as 2.288, 2.298 and 1.507 for CHAID, Exhausted CHAID and CART, respectively. Likewise, *R²* estimates for W60 were 0.538, 0.534 and 0.614, respectively. Unlike W60, much higher *R²* values for W90 were estimated as 0.949, 0.948 and 0.978,

compared with those estimated for W60. Similarly, *R²_{Adj}* values for CART were estimated higher than CHAID and Exhausted CHAID values for both W60 and W90. Pearson correlation coefficients (*r*) for CART (0.783) were significantly higher than *r* for CHAID (0.734) and *r* for Exhausted CHAID (0.731) for W90 (P<0.05). Similarly, *r* value of W90 for CART (0.989) was higher than *r* value of W90 for CHAID and Exhausted CHAID. The relationship between actual values and predicted values for W60 and W90 is depicted in Figure 1.

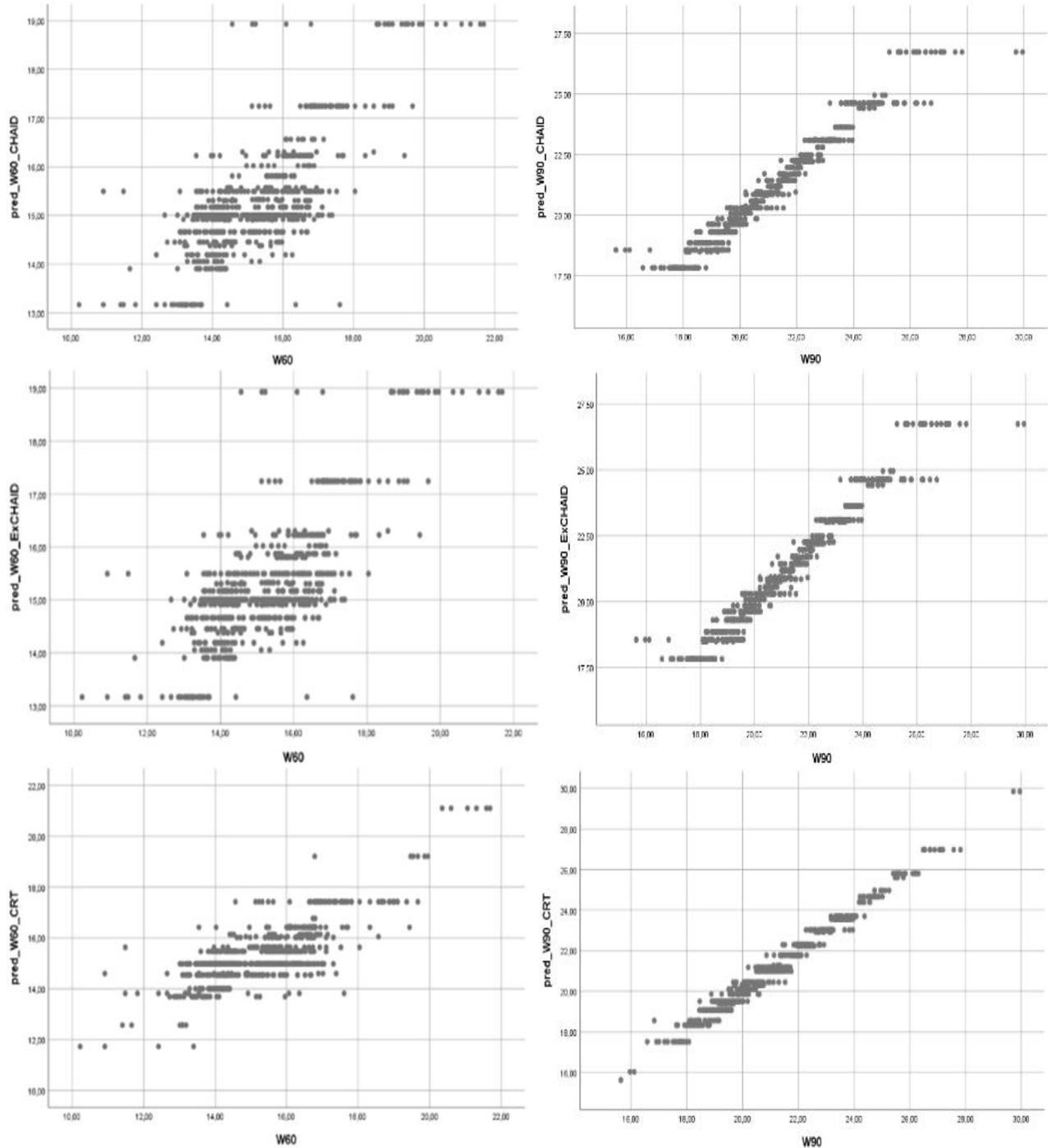


Figure 1. The relationship between observed and predicted values for W60 and W90 of Awassi lambs for RT data mining algorithms

The lowest RMSE, MAPE, RAE, SD_{ratio} and MAD values for W60 and W90, as well as the highest Pearson coefficient, R^2_{Adj} and R^2 values were obtained by CART algorithms. It is clear that the sequence of these tree-based data mining algorithms was found as CART > CHAID > Exhausted CHAID for W60 and W90. In the light of these results, CART was determined to be the most

suitable algorithm for especially W90 according to goodness of fit criteria in the present study. This means that there was still a need of including more influential factors for W60. SD ratio estimates obtained for all algorithms in prediction of W60 were able to be found within acceptable limits in having a good fit for regression type problems.

DISCUSSION

The ultimate goal in animal breeding is to select the superior breeding individuals and to use them efficiently in breeding to improve the population with high genetic merit. For this purpose, researchers desire to know the factors influencing selection strategies. In this regard, the present study was planned to evaluate comparative predictive performances of different data mining algorithms on the prediction of live body weights of Awassi lambs at 60 and 90 days. Eydurán (2016) emphasized that CART, CHAID and Exhausted CHAID data mining algorithms may be alternative methods when encountering multicollinearity problem among variables on live weight prediction studies of small ruminants.

The current results showed that the predictive performance of CART ($R^2=0.614$ for W60 and $R^2=0.978$ for W90) was much higher compared with those of other RT data mining algorithms (CHAID and Exhausted CHAID). The present R^2 estimate for W60 was in agreement with that reported by Yakubu (2012), who found 0.62 R^2 for body weight prediction of Uda sheep. It was determined that there was still a need of including more influential factors for especially W60. SD ratio estimates obtained for all algorithms in prediction of W60 were able to be found within acceptable limits in having a good fit for regression type problems. An earlier study by Ali *et al.* (2015) reported better predictive accuracy of decision tree algorithms for post-weaning period. However, the predictive performance results of the handled algorithms for W90 were found much better compared with Ali *et al.* (2015) for indigenous Harnai sheep of Pakistan.

In agreement with the present results regarding the highest R^2 estimates, Celik *et al.* (2017) reported that the predictive performance order was found as CART > CHAID \approx Exhausted CHAID > MARS_2 > MARS_1 > RBF > MLP in predicting body weight of Mengali rams in Pakistan. The present R^2 estimates were higher for CART decision tree algorithm than those reported by Mohammad *et al.* (2012) and Ali *et al.* (2015).

On the other hand, Olfaz *et al.* (2019) reported higher/lower R^2 estimates of 0.88 for CART and CHAID decision tree algorithms in the prediction of weaning weight of Karayaka lambs compared with corresponding estimates for W60/W90 in the present study. Koç *et al.* (2017) obtained lower predictive performances for CART, CHAID and Exhausted CHAID with 0.744, 0.743 and 0.716 R^2 in comparison with those estimated for W90 here. Balta and Topal (2018) studied on Hemşin lambs using regression tree method which explained 86.2 % of the variation in birth weight. They reported that the differences between years depended on environmental conditions, especially in pasture areas which had a significant impact on variations among lambs which is similar with this work.

When similar studies are examined, the differences in predictive performance in the literature may be attributed to variability in breed, type of birth, number of animals, management systems, year/season, dam age, body size of dam, algorithms used, structure of algorithms (minimum parent to child nodes, tree depth) and overfitting level of the used algorithms. The most effective predictions on body weights of Awassi lambs on 60-d and 90-d in this study were farm and birth weight. This result was similar to Topal (2018).

Hence, in complex data sets, including effects mentioned above, tree-based data mining diagrams that do not require any assumptions, can be used as a good-conspicuous method to explain the dependent variable. In addition, these visual diagrams enable to interpret the response variable easily with maximum variance among child nodes. With this approach, CART, CHAID and Exhausted CHAID algorithms can be practical applications for biological studies.

Conclusion: Predictive capabilities of decision tree algorithms in order to predict live weights of Awassi lamb at 60-d and 90-d of age were evaluated in the current study. Among those, CART had better predictive performances for especially W90 as compared to W60. The achieved CART results may help breeders to produce superior Awassi lambs. To generalize the obtained results, providing many Awassi lambs raised under different conditions is required. In conclusion, it could be suggested that use of decision tree algorithms may be useful information about the selection of superior lambs in ensuring higher lamb production

Acknowledgments: Thanks to Osmaniye Sheep and Goat Breeders' Association for providing help in data collection.

REFERENCES

- Akin, M., E. Eydurán and, B.M. Reed (2017). Use of RSM and CHAID data mining algorithm for predicting mineral nutrition of hazelnut. *Plant Cell Tiss. Org. (PCTOC)*, 128(2):303-316.
- Akin, M., C. Hand, E. Eydurán and, B.M. Reed (2018). Predicting minor nutrient requirements of hazelnut shoot cultures using regression trees. *Plant Cell Tiss. Org. (PCTOC)*, 132(3):545-559. DOI:10.1007/s11240-017-1353-x
- Akin, M., S.P. Eydurán, E. Eydurán and, B.M. Reed (2020). Analysis of macro nutrient related growth responses using multivariate adaptive regression splines. *Plant Cell Tiss. Org. (PCTOC)*, 140(3):661-670.
- Aksoy, A., Y.E. Ertürk, E. Eydurán and, M.M. Tariq (2019). Utility of MARS algorithm for describing non-genetic factors affecting pasture revenue of Morkaraman Breed and Romanov \times Morkaraman

- F1 Crossbred sheep under semi intensive conditions. *Pakistan J. Zool.* 51(1): 235–240.
- Ali, M., E. Eydurhan, M.M. Tariq, C. Tirink, F. Abbas, M.A. Bajwa, M.H. Baloch, A.H. Nizamani, A. Waheed, M.A. Awan and, S.H. Shah (2015). Comparison of artificial neural network and decision tree algorithms used for predicting live weight at post weaning period from some biometrical characteristics in Harnai sheep. *Pakistan J. Zool.* 47(6): 1579–1585.
- Aytekin, I., E. Eydurhan, K. Karadas, R. Akşahan and, I. Keskin (2018). Prediction of fattening final live weight from some body measurements and fattening period in young bulls of crossbred and exotic breeds using MARS data mining algorithm. *Pakistan J. Zool.* 50(1): 189–195.
- Balta, B., and, M. Topal (2018). Regression tree approach for assessing the effects of non-genetic factors on birth weight of Hemşin lamb. *Alinteri Zirai Bilimler Dergisi*, 33(1): 65–73,.
- Breiman, L., J.H. Friedman, C.J. Stone and, R.A. Olshen (1984). *Stone Classification and regression trees*. Chapman and Hall, Wadsworth Inc., New York, NY, USA.
- Celik, S., E. Eydurhan, A. Tatliyer, K. Karadas, M.K. Kara and A. Waheed (2018). Comparing predictive performances of some nonlinear functions and multivariate adaptive regression splines (MARS) for describing the growth of Daera Dın Panah (DDP) goat in Pakistan. *Pakistan J. Zool.* 50(3), 1187–1190.
- Celik, S., E. Eydurhan, K. Karadas and, M.M. Tariq (2017). Comparison of predictive performance of data mining algorithms in predicting body weight in Mengali rams of Pakistan. *Rev. Bras. Zootecn.* 46(11): 863–872. DOI:10.1590/s1806-92902017001100005
- Çankaya, S., and, S.H. Abacı (2012). Path analysis for determination of relationships between some body measurements and live weight of German Fawn x Hair Crossbred kids. *Kafkas Univ. Vet. Fak. Derg.* 18(5):796–773.
- Eydurhan, E., A. Waheed, M.M. Tariq, F. Iqbal and, S. Ahmad (2013a). Prediction of live weight from morphological characteristics of commercial goat in Pakistan using factor and principal component scores in multiple linear regression. *J Anim. Plant Sci.* 23(6): 1532–1540.
- Eydurhan, E., I. Yilmaz, M.M. Tariq and, A. Kaygisiz (2013b). Estimation of 305-d milk yield using regression tree method in brown Swiss cattle. *J Anim. Plant Sci.*, 23(3): 731–735.
- Eydurhan, E., İ. Keskin, Y.E. Erturk, B. Dag, A. Tatliyer, C. Tirink, R. Aksahan and, M.M. Tariq (2016). Prediction of fleece weight from wool characteristics of sheep using regression tree method (CHAID Algorithm). *Pakistan J. Zool.* 48(4): 957–960.
- Eydurhan, E. (2016). On the possibility of using data mining algorithms in prediction of live body weights of small ruminants. *Adv Cal Anal.* 1: 18–21, DOI: 10.21065
- Eydurhan, E., D. Zaborski, A. Waheed, S. Celik, K. Karadas and, W. Grzesiak (2017). Comparison of the predictive capabilities of several data mining algorithms and multiple linear regression in the prediction of body weight by means of body measurements in the indigenous Beetal goat of Pakistan. *Pakistan J. Zool.* 49(1): 257–265. DOI: 10.17582/journal.pjz/2017.49.1.257.265
- Eydurhan, E., M. Akin and, S.P. Eydurhan (2019). Application of Multivariate Adaptive Regression Splines in Agricultural Sciences through R Software. Nobel Bilimsel Eserler Sertifika No: 20779, Ankara. ISBN: 978-605-2149-81-2.
- Eydurhan, E. (2019). ehaGoF: Calculates Goodness of Fit Statistics. R package version 0.1.0. <https://CRAN.R-project.org/package=ehaGoF>
- Grzesiak, W. and, D. Zaborski (2012). Examples of the Use of Data Mining Methods in Animal Breeding, *Data Mining Applications in Engineering and Medicine*, pp 303-324. IntechOpen, DOI: 10.5772/50893. Available from: <https://www.intechopen.com/books/data-mining-applications-in-engineering-and-medicine/examples-of-the-use-of-data-mining-methods-in-animal-breeding>
- Karabacak, A., S. Celik, A. Tatliyer, I. Keskin, Y.E. Erturk, E. Eydurhan, Y. Javed and, M.M. Tariq (2017). Estimation of cold carcass weight and body weight from several body measurements in sheep through various data mining algorithms. *Pakistan J. Zool.* 49 (5): 1731–1738. DOI: <http://dx.doi.org/10.17582/journal.pjz/2017.49.5.1731.1738>
- Khan, A., M.A. Sultan, M.A. Jalvi and, I. Hussain (2006). Risk factors of lamb mortality in Pakistan Anim. Res., 55(4): 301–311. DOI: 10.1051/animres:2006017
- Khan, M.A., M.M. Tariq, E. Eydurhan, A. Tatliyer, M. Rafeeq, F. Abbas, N. Rashid, M.A. Awan and, K. Javed (2014). Estimating body weight from several body measurements in Harnai sheep without multicollinearity problem. *J Anim. Plant Sci.* 24(1):120–126
- Koç, Y., E. Eydurhan and, O. Akbulut (2017). Application of regression tree method for different data from animal science. *Pakistan J. Zool.* 49(2): 599–607.
- Koç, Y. Application of regression tree method for different data from animal science. MSc, Iğdır Universty, Iğdır, Turkey, 2016.
- Kovalchuk, I.Y., Z. Mukhitdinova, T. Turdiyev, G.

- Madiyeva, M. Akin, E. Eyduran and, B.M. Reed (2017). Modeling some mineral nutrient requirements for micropropagated wild apricot shoot cultures. *Plant Cell Tiss. Org. (PCTOC)*, 129(2):325-335.
- Kovalchuk, I.Y., Z. Mukhitdinova, T. Turdiyev, G. Madiyeva, M. Akin, E. Eyduran and, B.M. Reed (2018). Nitrogen ions and nitrogen ion proportions impact the growth of apricot (*Prunus armeniaca*) shoot cultures. *Plant Cell Tiss. Org. (PCTOC)*, 133(2):263-273.
- Kusiak, A., M. Li and, Z. Zhang (2010). A data-driven approach for steam load prediction in buildings. *Appl. Energy* 87(3): 925–933. DOI:10.1016/j.apenergy.2009.09.004
- Mandal, A., H. Prasad, A. Kumar, R. Roy and, N. Sharma (2007). Factors associated with lamb mortalities in Muzaffarnagari sheep. *Small Ruminant Res.* 71(1-3): 273–279.
- Mohammad, M.T., M. Rafeeq, M.A. Bajwa, M.A. Awan, F. Abbas, A. Waheed, F.A. Bukhari and, P. Akhtar (2012). Prediction of body weight from body measurements using regression tree (RT) method for indigenous sheep breeds in Balochistan, Pakistan *J Anim. Plant Sci.* 22(1):20–24.
- Morris, C.A., S.M. Hickey and, J.N. Clarke (2000). Genetic and environmental factors affecting lamb survival at birth and through to weaning. *New Zeal. J. Agr. Res.*, 43(4): 515–524. DOI:10.1080/00288233.2000.9513448
- Olfaz, M., C. Tırınk and, H. Önder (2019). Use of CART and CHAID algorithms in Karayaka sheep breeding. *Kafkas Univ. Vet. Fak. Derg.*, 25(1): 105–110. DOI: 10.9775/kvfd.2018.20388
- Orhan, H., E. Eyduran, A. Tatliyer and, H. Saygici (2016). Prediction of egg weight from egg quality characteristics via ridge regression and regression tree methods. *Rev. Bras. Zootecn.* 45(7):380–385. DOI: 10.1590/S1806-92902016000700004
- Önder, H. and, S.H. Abaci (2015). Saanen oğlaklarının canlı ağırlıkları üzerine vücut ölçümleri için path analizi. *Kafkas Univ. Vet. Fak. Derg.* 21(3): 351–354. DOI:10.9775/kvfd.2014.12500.
- Palaniappan, S. and, R. Awang (2008). Intelligent heart disease prediction system using data mining techniques. In: 2008 IEEE/ACS international conference on computer systems and applications. March 31 2008-April 4 2008. 8(8):108–115 R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Sawalha, R.M., J. Conington, S. Brotherstone and, B. Villanueva (2007). Analyses of lamb survival of Scottish Blackface sheep. *Animal* 1(1):151–157. DOI:10.1017/S1751731107340056
- Sevgenler, H. (2019). Comparison of data mining algorithms (CART, CHAID and MARS) used to determine the effects of some characteristics on body weight in goats. MSc, Iğdır University, Iğdır, Turkey.
- SPSS, 2006. SPSS for windows release. 15.0 SPSS Inc., Chicago.
- Susic, V., V. Pavic, B. Mioc, I. Stokovic and, A.E. Kabalin (2005). Seasonal variations in lamb birth weight and mortality. *Vet. Arhiv.* 75(5): 375–381.
- Tariq, M.M., E. Eyduran, M.A. Bajwa, A. Waheed, F. Iqbal and, Y. Javed (2012). Prediction of body weight from testicular and morphological characteristics in indigenous mengali sheep of Pakistan using factor analysis scores in multiple linear regression analysis. *Int. J.Agric. Biol.* 14(4): 590–594.
- Üstüner, H. and, M.M. Oğan (2013). Main productive performance of Awassi sheep in the Central Anatolian Region of Turkey. *Turkish J. Vet. and Anim. Sciences*, 37(3), 271-276. Vatankhah, M. and Talebi, M.A. (2009). Genetic and non-genetic factors affecting mortality in Lori-Bakhtiari lambs. *Asian Austral. J. Anim.* 22(4): 459–464. DOI: 10.5713/ajas.2009.80318
- Yakubu, A. (2012). Application of regression tree methodology in predicting the body weight of Uda sheep. *J. Anim. Sci. Biotechnol.* 45(2): 484–490.