



Measuring Predictive Performance of Data Mining and Artificial Neural Network Algorithms for Predicting Lactation Milk Yield in Indigenous Akkaraman Sheep

Koksal Karadas¹, Muhammad Tariq², Mohammad Masood Tariq^{3*} and Ecevit Eydurhan⁴

¹Department of Agricultural Economics, Agricultural Faculty, Igdir University, Igdir, Turkey

²Department of Livestock Management, University of Agriculture, Faisalabad, Pakistan Sub-Campus Toba Tek Singh

³Center for Advanced Studies in Vaccinology and Biotechnology, University of Balochistan, Quetta, Balochistan, Pakistan

⁴Biometry Genetics Unit, Department of Animal Science, Agricultural Faculty, Igdir University, Igdir, Turkey

ABSTRACT

The aim of the investigation was to determine several factors affecting lactation milk yield amounts of indigenous Akkaraman ewes reared in Sanliurfa (137 farms) and Hakkari (113 farms) provinces of Turkey. For the determination process, the statistical predictive accuracy of GLM (General Linear Model), CART (Classification and Regression Tree), CHAID (Chi-square Automatic Interaction Detector), Exhaustive CHAID, and a ANN type MLP (Multilayer Perceptron) in the prediction of lactation milk yield in sheep was measured comparatively by model assessment criteria, such as R^2 , $R^2_{ADJUSTED}$, SD_{RATIO} , $CV(\%)$, RMSE, RAE, MAPE, MAD and Pearson correlation coefficient (r) between actual and predicted milk yield values, respectively. Minimum farm numbers for parent and child nodes were assigned at 10:5 for producing the best predictive accuracy in CART and both CHAID algorithms. According to the criteria, the significance order in the predictive accuracy for lactation milk yield was recorded as CHAID=Exhaustive CHAID > ANN > GLM > CART. The decision tree results of CHAID algorithm displayed that ewe age, number of milking and lactation length for lactation milk yield should be taken into account and lactation lengths for Akkaraman sheep milked more than 2 a day should be longer than 150 days at 3 year old age group, and longer than 160 days at 4 year old age group for sustaining high productivity in lactation milk yield. Consequently, it is recommendable that determinant factors such as age, number of milking and lactation length on lactation milk yield should be considered on the basis of the powerful algorithm giving high predictive power.

Article Information

Received 22 June 2016

Revised 17 July 2016

Accepted 23 July 2016

Available online 04 September 2016

Authors' Contributions

KK contributed in experiment design, acquisition of data, analysis and drafting of the manuscript.

EE statistically evaluated the data. MMT and MT interpreted the data, drafted and revised the manuscript. All of the authors read and approved the final manuscript.

Key words

CHAID, CART, Exhaustive CHAID, MLP, ANN

INTRODUCTION

Sheep husbandry is practiced throughout the world, and has been fundamental to many civilizations and is most often associated with pastoral, Arcadian imagery. Sheep is a multifaceted livestock species having a large number of breeds over the World. With respect to [FAO \(2014\)](#) database, the number of sheep was 1210 million heads in the world and 31 million heads in Turkey. The amount of sheep milk production in the World and Turkey were recorded as 10.1 and 1.1 in million tones, respectively. In the past years, sheep production was considered as one of the important animal husbandry activities due to climate, topographic structure and rich flora of Turkey ([Karaca et al., 2003](#)).

In Turkey, Akkaraman fat-tailed breed (44% of the total sheep breeds) is found mostly as a domestic breed, and therefore in the place of the conservation of the native gene sources, a great attention must be drawn on scientific studies concerning all economically important traits of Akkaraman breed for the sustainability of its yield traits, including meat, milk and wool ([Karaca et al., 2003](#)). Like the other fat-tailed sheep breeds in the country, it is very hardy and strives well under poor feeding and extreme climatic conditions ([FAO, 1986](#)). Among the traits, lactation milk yield can be affected by some environmental factors (age, birth type, gender at lambing, parity, season, grazing period, lambing year and month) and genetic factors (breed, breeding pattern, breeding systems etc). Investigation of these factors is very prominent for fostering breeding strategies, which is presumptive to support powerful statistical analysis approaches ([Kucuk and Eydurhan, 2010](#)). With the aim of

* Corresponding author: tariqkianiraja@hotmail.com
0030-9923/2017/0001-0001 \$ 9.00/0
Copyright 2017 Zoological Society of Pakistan

using lamb production, usability of indigenous breeds is of great importance in improvement of the crossbreeds with superior offspring and milk yield (Unal *et al.*, 2002).

In sheep dairy production systems, milk yield, fertility and healthiness are some of the most momentous economic traits affecting the gainfulness (Kucuk and Eydurán, 2010). The attainment of success in sheep dairy farm operation is related to the accurate managerial and rearing decisions at different stages of milk production, strongly aided by powerful statistical methods (Grzesiak *et al.*, 2006). In order to organize managerial conditions and to foster sheep breeding strategies, lactation curve models and their biological parameters are substantial tools to model the non-linear relationship between lactation milk yield and certain time period of the lactation (Keskin and Dag, 2006; Kucuk and Eydurán, 2010).

General linear models without including interaction effects on the basis of least squares method have been used in widespread manner in the prediction of lactation milk yield by using environmental factors (Yilmaz *et al.*, 2004). The ANNs, CART and CHAID algorithms are specified for ordinal, nominal and continuous variables all together within the framework of non-parametric techniques with the purpose of minimizing residual variance. Apart from ANNs, CART and CHAID algorithms produces a decision tree structure that helps ones to identify non-linear and interaction relationships of the significant effects in the model built. The CART algorithm employs binary splitting criteria, repeatedly whereas CHAID algorithms use multisplitting criteria, recursively and Bonferroni adjustment. The ANNs make the best of analyzing most complex relationships between output and input variables compared with traditional statistical techniques as well as resemble to human brain in a network structure with input, hidden and output layers (Ali *et al.*, 2015).

In recent studies, an attention on powerful data mining algorithms (CART, CHAID, Exhaustive CHAID and ANNs) has been paid for modeling and predicting lactation milk yield in dairy goat (Eydurán *et al.*, 2013a) and cattle (Dogan, 2003; Grzesiak *et al.*, 2006; Bakir *et al.*, 2010; Topal *et al.*, 2010; Eydurán *et al.*, 2013b) in the definition of significant factors on lactation milk yield. Eydurán *et al.* (2013a) implemented the Exhaustive CHAID algorithm to predict first lactation milk yield by means of udder traits and somatic cell count in Turkish Saanen goat. Similarly, the documentation on applicability of the ANNs algorithms in dairy sheep is very limited (Ince and Sofu, 2013).

In dairy cattle production, some data mining algorithms like CHAID and ANNs have been applied efficaciously for 305-d milk yield modeling (Grzesiak *et al.*, 2003; Grzesiak *et al.*, 2006; Bakir *et al.*, 2010; Takma *et al.*, 2012; Eydurán *et al.*, 2013b; Ali *et al.*, 2015),

whereas no obtainable paper in the prediction of lactation milk yield in sheep was found in literature apart from Ince and Sofu (2013) using only ANNs for the prediction procedure in Awassi sheep. But, it is obvious that an accurate prediction of lactation milk yield per ewe with the contribution of a good algorithm is more important for farmers and milk processors for decision making process (Grzesiak and Zaborski, 2012), in place of regulating their managerial conditions and sustainability policies (Ince and Sofu, 2013). Hereafter, the main purpose of this study was to predict lactation milk yield per ewe by means of several factors for a total of 240 sheep farms in Sanliurfa and Hakkari provinces of Turkey and to find significant ones among the factors. With the exception of very few reports, particular data of model evaluation criteria have not yet been keynoted (Grzesiak and Zaborski, 2012; Ali *et al.*, 2015) and so, in the predictive accuracy, GLM (General Linear Model), CART (Classification and Tree), CHAID (Chi-square Automatic Interaction Detector), Exhaustive CHAID, and MLP (Multilayer Perceptron), one of ANNs types were compared for the prediction procedure with a great number of model assessment criteria submitted in materials and methods section.

MATERIALS AND METHODS

Data structure

The current study was conducted on Akkaraman sheep reared in a total of 240 farms in Sanliurfa (137 farms) and Hakkari (113 farms) provinces of Turkey in order to predict lactation milk yield per ewe by some factors (province, number of sheep, total amount of feed, veterinary costs, total labor costs, amount of salt, grazing period, lambing month, age of ewe, lactation length and number of milking). Average lactation milk yield per ewe was taken a basis for each farm.

Data sampling

In the study, for each of two provinces (Sanliurfa and Hakkari) of Turkey, the required sample sizes were determined by the following formula of simple random sampling method (Cicek and Erkan, 1996).

$$n = \frac{N S^2 t^2}{(N - 1) d^2 + S^2 t^2}$$

Where n is Sample size, N is Population size, S² is Variance of population, d² is Acceptable error proportion and t is critical value.

$$\text{Sample size of Hakkari province} = \frac{992 \times 70,8^2 \times 1,65^2}{(992 - 1) 10,3^2 + 70,8^2 \times 1,65^2} = 112,8$$

$$\text{Sample size of Sanliurfa province} = \frac{1128 \times 144,79^2 \times 1,65^2}{(1128 - 1) 20^2 + 144,79^2 \times 1,65^2} = 126,76$$

Total sample size for two provinces was taken as: 113+127=240.

The predictive accuracy of GLM (General Linear Model), CART (Classification and Regression Tree), CHAID (Chi-square Automatic Interaction Detector), Exhaustive CHAID, and MLP (Multilayer Perceptron) as a ANNs type in the prediction of lactation milk yield in sheep were examined comparatively by using model assessment criteria such as R^2 , R^2_{ADJUSTED} , SD_{RATIO} , $CV(\%)$, RMSE, RAE, MAPE, MAD and Pearson correlation coefficient (r) between actual and predicted lactation milk yield values, respectively. More comprehensive knowledge on the examined algorithms in the study is accessible in the introduction section of the manuscript and especially in [Ali et al. \(2015\)](#). For CART, CHAID and Exhaustive CHAID, Cross-validation was used as 10 as also identified by [Ali et al. \(2015\)](#). Minimum farm numbers for parent and child nodes were identified at 10:5 for producing the best predictive accuracy in CART and both CHAID algorithms. Also, Multilayer Perceptron, a type of ANNs, was used in training (80%) and testing (20%) sets with one hidden layer, hyperbolic tangent as a activation function, and an identity activation function for output layer when the least difference in each of model assessment criteria between training and testing sets were found.

Formulas of the model assessment criteria measured for comparing their predictive accuracy each other are given as follows:

Coefficient of Determination	$R^2 = \left[1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \right]$
Adjusted Coefficient of Determination	$R^2_{\text{ADJUSTED}} = \left[1 - \frac{\frac{1}{n-k-1} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2} \right]$
Coefficient of Variation (%)	$CV(\%) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2} \cdot \frac{1}{\bar{Y}} \cdot 100$
Standard Deviation Ratio	$SD_{\text{RATIO}} = \sqrt{\frac{\frac{1}{n-1} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2}{\frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2}}$
Relative Approximation Error (RAE)	$RAE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n Y_i^2}}$
Root Mean Square Error	$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}}$
Mean Absolute Deviation	$MAD = \frac{\sum_{i=1}^n Y_i - \hat{Y}_i }{n}$

$$\text{Mean Absolute Percentage Error MAPE} = \frac{\sum_{i=1}^n \frac{|Y_i - \hat{Y}_i|}{Y_i}}{n} \quad Y \neq 0_i$$

Where Y_i is the actual lactation milk yield value of i^{th} sheep farm, \hat{Y}_i is the predicted lactation milk yield value of i^{th} sheep farm, \bar{Y} is average of the actual lactation milk yield values of the sheep farms, ε_i is the residual value of i^{th} sheep farm, $\bar{\varepsilon}$ is average of the residual values, k is number of significant independent variables in the model and n is total farm size.

The residual value of each sheep farm is equal to $\varepsilon_i = Y_i - \hat{Y}_i$

The Pearson correlation coefficients between the actual (observed) and predicted lactation milk yield values were estimated individually for each of the algorithms. The most suitable algorithm in the predictive performance produces the highest values in r , R^2 , R^2_{ADJUSTED} and the lowest values in $CV(\%)$, SD_{RATIO} , RAE, RMSE, MAD and MAPE, respectively ([Ali et al., 2015](#)). $SD_{\text{RATIO}} < 0.40$ or $SD_{\text{RATIO}} < 0.10$ shows a good fit or a very good fit ([Grzesiak et al., 2006](#); [Ali et al., 2015](#)). Statistical evaluation was conducted using statistical package IBM SPSS 23 software SPSS Inc., 2015.

RESULTS AND DISCUSSION

The documentation on the comparison of the approaches in sheep has been discussed here for the first time. [Table 1](#) presents the results of model assessment criteria measuring the predictive accuracy of GLM, CART, CHAID, Exhaustive CHAID, and MLP as an ANN type in the prediction of lactation milk yield in sheep.

The CHAID algorithms were found similar to each other together with having the highest performance ($R^2_{\text{ADJUSTED}} > 0.80$ and $R^2 > 0.80$) in their predictive accuracy of the model fit criteria. This figure was supported by [Grzesiak et al. \(2003\)](#) who reported that a model with $R^2 > 0.70$ gave a very good fit. [Takma et al. \(2012\)](#) informed that R^2_{ADJUSTED} estimates of ANN models ranged between 0.62 (5th lactation) and 0.85 (1st lactation) for 305-d milk yield of Holstein Friesians in the first five lactations. [Ince and Sofu \(2013\)](#) found marvelous predictive performance of ANN with multilayer perceptron in the lactation milk yield prediction of Awassi sheep, which was better than our performance of ANN with MLP applied in the study. This result might be attributed to variability in breed, managerial arrangement, sample size, variables used in the model.

In the study conducted with many factors, it is obvious that the significance order of the predictive performance was CHAID = Exhaustive CHAID > ANN > GLM > CART. In agreement with findings of the current study, some authors emphasized that the CHAID algorithms can occur

Table I.- Results of model assessment criteria.

Algorithm	r	R ²	R ² ADJUSTED	RMSE	RAE	CV (%)	SD RATIO	MAD	MAPE
CHAID	0.910	0.829	0.825	18.41	0.098	10.66	0.414	12.490	7.016
EX. CHAID	0.910	0.829	0.825	18.41	0.098	10.66	0.414	12.490	7.016
CART	0.855	0.731	0.727	23.08	0.122	12.62	0.519	16.103	9.046
ANN	Training	0.910	0.828	0.816	18.65	1.912	10.03	0.416	13.860
	Testing	0.893	0.791	0.725	18.78	0.754	10.85	0.452	14.360
GLM		0.902	0.813	0.798	19.21	0.102	10.51	0.423	14.114

more suitable tree structures in prediction procedures of the live body weights (Khan *et al.*, 2014; Ali *et al.*, 2015) and 305-d milk yield (Eyduvan *et al.*, 2013b) in the comparison with CART data mining algorithms. Also, SD_{RATIO} values of 0.414 for CHAID algorithms were estimated as nearly a good fit value proposed by Grzesiak *et al.* (2003) and Ali *et al.* (2015). Use of SD_{RATIO} as a goodness of fit criterion is rare in similar studies.

In the prediction of lactation milk yield per ewe, the visual tree analysis of the factors assigned to be significant among several factors (province, number of sheep, total amount of feed, veterinary costs, total labor costs, amount of salt, grazing period, lambing month, age of ewe, lactation length and number of milking) in CHAID algorithm is depicted in Figure 1. Overall average lactation milk yield per ewe was 63.771 kg in Node 0. Within the appreciated factors, age of ewe was observed to be the most significant factor in CHAID analysis, and at first step it divided Node 0 into five child nodes (Nodes 1-5), respectively. Node 1 was the ewe group with age ≤ 2 in the CHAID diagram. In accordance with rules of the CHAID algorithm, the ewe groups at 3, 4 and 5 years of age, were assigned to Node 2 (71.97 kg), Node 3 (77.637 kg) and Node 4 (52.629 kg), respectively. The ewe group older than 5 years of age was illustrated as Node 5 (42.479 kg). The lactation milk yield from Node 1 through Node 3 was increasing and then the yield reduced from Node 4 to Node 5.

Within the significant factors, amount of salt significantly influenced lactation milk yield of the ewe group in Node 1, which was split into three sub-ewe groups (Node 6 (49.139kg), Node 7 (57.674 kg) and Node 8 (53.492 kg)), respectively. Node 6 was the ewe group fed with 1.920 kg or less salt within the ewe group with age ≤ 2 . Node 7 was the ewe group fed with $1.920 \text{ kg} < \text{salt} \leq 2.450 \text{ kg}$ within the ewe group with age ≤ 2 . Node 8 was determined as the ewe group fed with salt $> 2.45 \text{ kg}$ in the same ewe group.

In lactation milk yield, the ewe groups in Nodes 2 and 3 were divided by number of milking into Node 9 (69.060kg) and Node 10 (81.595kg), respectively. Province split only Node 9, which was the 3 year-ewe group milked 2 or less, into Node 13 (64.999 kg) and

Node 14 (70.249 kg) but, with CHAID algorithm, Node 10 was defined as the 3 year-ewe group milked more than 2 and divided by lactation length into new two Nodes (15 and 16), respectively. Node 15 was the ewe group with lactation length 150 days or shorter at the 3 year-ewe group milked more than 2 per day, but Node 16 was identified as the ewe group with lactation length longer than 150 days at the 3 year-ewe group milked more than 2 per day. According to Node 16, it is obvious that the lactation milk yield increased when lactation length was longer than 150 at the 3 year-ewe group milked more than 2 per day.

Node 3 was divided by number of milking into Node 11 (70.922 kg) and Node 12 (86.270 kg), respectively. Node 11 was detected as the 4 year ewe age group milked 2 or less per day, but Node 12 was the ewe group milked more than 2 per day in same age group and re-branched by lactation length into three new ewe subgroups; Node 17 (77.825 kg), Node 18 (85.561 kg) and Node 19 (103.053 kg), respectively. Node 17 was found as the 4 year-ewe group milked more than 2 per day with the lactation length shorter than 150 days. Node 18 was identified as the 4 year-ewe group milked more than 2 per day with $150 < \text{lactation length} \leq 160$. Also, Node 19 was found as the 4 year-ewe group milked more than 2 per day with the lactation length longer than 160 days. The significant differences among Nodes 17, 18 and 19 were statistically obtained by CHAID algorithm. In GLM, the effect of number of milking, amount of salt consumed, and ewe age, lactation length on lactation milk yield was found statistically significant at ($P < 0.01$).

Visual decision tree diagram of CART algorithm with a recursively binary splitting in the prediction of the lactation milk yield in indigenous Akkaraman ewes is depicted in Figure 2. Number of milking, province and ewe age entered into the tree diagram structured by CART algorithm as significant factors. Node 2, a terminal node, was Akkaraman ewe group milked more than 2 (84.721 kg) per day. Node 1 was the Akkaraman ewe group milked 2 or less per day, and divided into two ewe subgroups (Nodes 3 and 4) according to province factor, respectively (52.227 vs. 70.148 kg). Node 3 was Akkaraman ewe group milked 2 or less per day in Sanliurfa province,

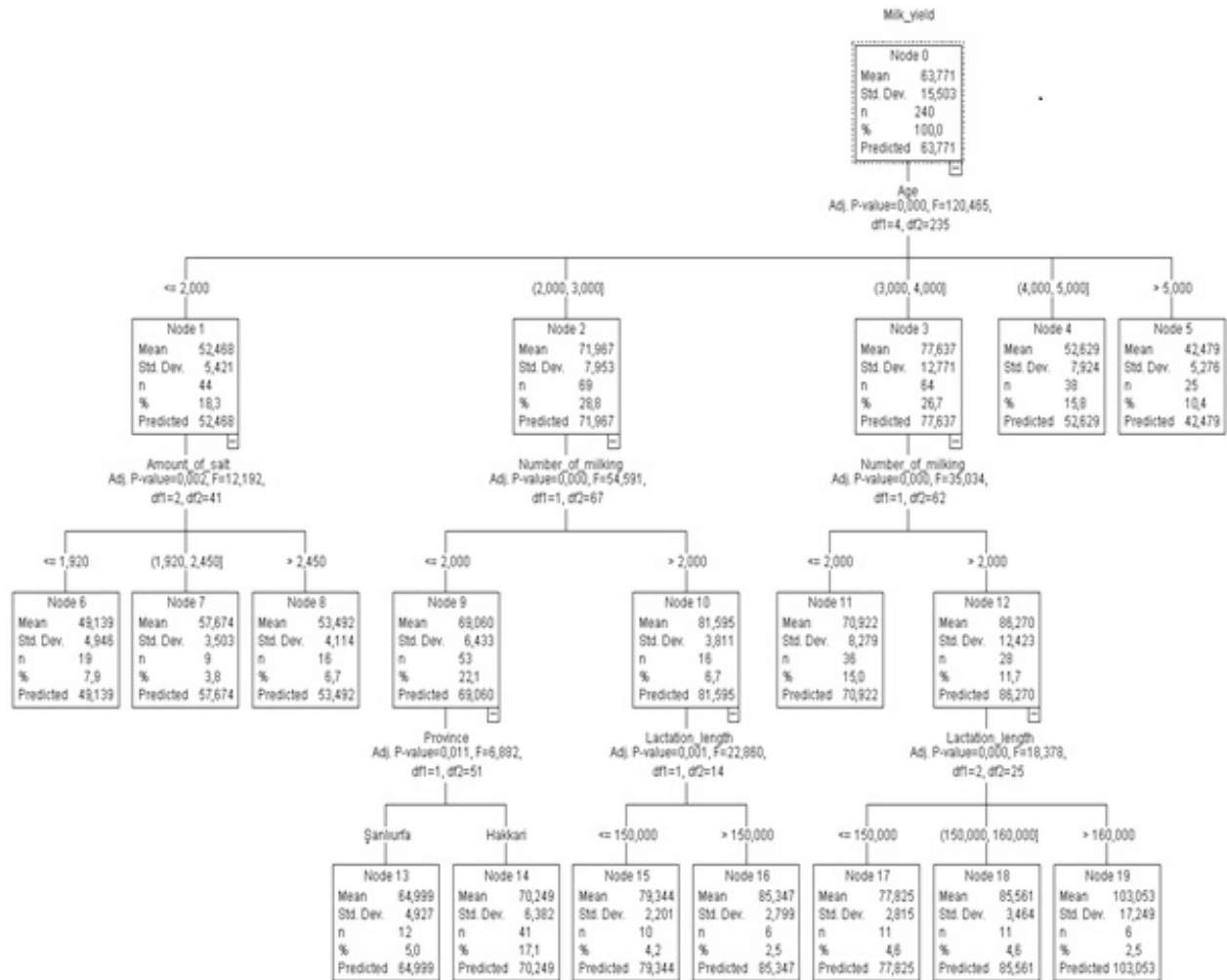


Fig. 1. The decision tree diagram constructed by CHAID algorithm.

but Node 4 is a terminal node that had the ewe group reared in Hakkari province and milked 2 or less per day. Node 3 was split by ewe age factor into Nodes 5 and 6, respectively (54.740 vs. 42.479 kg).

Yildiz and Denk (2006) observed and reported that breed, birth type, live weight, age, feeding, and body form of sheep can affect lactation milk yield. In addition, Sahin and Akmaz (2004) reported that lambing season, number of lambs sucking, rearing system, feed quality and lactation length, and interval between two milking also affected lactation milk yield. Lactation length illustrated a considerable impact on the milk yield of East Friesian ewes (Kuchtik *et al.*, 2008) and the yield of Norduz sheep (Yilmaz *et al.*, 2004), whereas in the current study, it which was interacted with ewe age and number of milking factors had an important influence on lactation milk yield of the ewe groups in Nodes 10 and 12. Yilmaz *et al.* (2004)

reported that in GLM factors, age, weight and lactation length of ewe significantly affected lactation milk yield for Norduz sheep, a variety of Akkaraman breed, and mentioned that regression coefficient (0.389) of lactation milk yield on lactation length was significant. In lactation milk yield, similar conclusions were noted by Yilmaz *et al.* (2004), who found remarkable differences at increasing amounts among 2, 3, 4 and 5 years of age in Norduz sheep. Sezenler *et al.* (2016) recorded significant effects of year ($P<0.01$), parity ($P<0.05$) except for birth type on lactation milk yield of Bandırma (German Black Head Mutton x Kivircik) sheep.

It was recorded by Sahin and Akmaz (2004) that number of milking in sheep enhanced the productivity in lactation milk yield, which was in agreement with the visual output of CART algorithm in Figure 2, but common effect of number of milking with lactation length, age and

province was increasing lactation milk yield in CHAID algorithm. The authors affirmed that ewes milked three times a day had more milk yield by 5-23% compared to those milked twice a day.

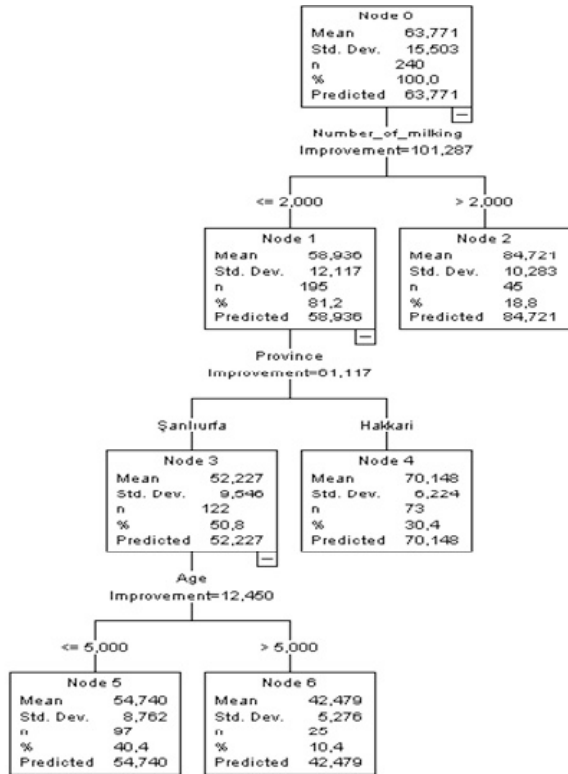


Fig. 2. The decision tree diagram constructed by CART algorithm.

Ince and Sofu (2013) suggested that lactation milk yield can be predicted by using priori information on ewe age, number of lactation, milk yield per month and lactation duration in ANNs model. In dairy cattle, prediction effort of 305-d milk yield by some factors through CHAID, Exhaustive CHAID and ANNs was quite good in predictive success. But, very great effort on the prediction of lactation milk yield in sheep with same statistical tools must be made.

Our results are exactly different from the previous ones obtained with various animal materials and breeding systems, and the existence of using different statistical models, factors and interaction effects. In future surveys, with same data mining tools, firstly, significant factors affecting lactation curve parameters could be determined after biological parameters of a lactation curve model for each ewe are calculated. Secondly, lactation milk yield could be predicted by using lactation curve parameters and environmental factors. Thus, we can observe the interaction effect of lactation curve parameters, the interaction effects

of lactation curve parameters by other environmental factors and thirdly, after individual inbreeding coefficient of each ewe in the pedigree was calculated, the interaction effect of inbreeding coefficient with curve parameters and environmental factors could be identified.

As a result, the obtainment of higher predictive performance in lactation milk yield can be important for selecting effective factors and a powerful algorithm. It could be suggested that all the statistical approaches under the survey were sufficient in this regard.

CONCLUSIONS

The major interest of the work was to examine the influence of some factors on lactation milk yield of the ewe with the aid of computational different data mining algorithms. To know effective factors in this regard is important to make future economic decisions in the arrangement of managerial conditions on increasing overall lactation milk yield of the farm. With this study, the following results have been enumerated:

- The performance efficiency of the algorithms was sequentially found as follows: CHAID>ANN>GLM>CART respectively.
- Important degrees of factors handled in both CHAID algorithms were ewe age > number of milking = amount salt > lactation length = province. In the groups of 3 and 4 year-old age Akkaraman sheep (Nodes 2 and 3), the positive effect of number of milking and lactation length on the lactation milk yield was observed. It could be advised that Akkaraman sheep with 3 and 4 year old age should be milked more than 2 per day. But, lactation lengths for Akkaraman sheep milked more than 2 per day should be longer than 150 days at 3 year old age group, and longer than 160 days at 4 year old age group.
- In joint group comprising the first and second year old age Akkaraman sheep, amount of salt given to the sheep could be suggested between $1.920 < X \leq 2.450$ in lactation milk yield.
- As also seen from Nodes 1 and 2 of CART algorithm, number of milking should be more than 2 per day (58.936 vs. 84.721 kg) in the productivity of the lactation milk yield. Among the group of Akkaraman sheep milked once and twice per day, Hakkari Akkaraman sheep gave higher lactation milk yield amount compared to Sanliurfa sheep (52.227 vs. 70.148kg). In Sanliurfa sheep (Node 3), ewe age should not be more than 5 (54.740 vs. 42.479) as also understood from Nodes 5 and 6.

In conclusion, it could be advised that age, number of milking and lactation length for lactation milk yield should be considered depending upon the selection of ideal algorithm (CHAID tree based algorithm). Thus it

could ultimately help in accurate prediction of milk yield per ewe with the contribution of a good algorithm that is more important for farmers and milk processors of Turkey for decision making process, in place of regulating their managerial conditions and heading toward policies leading to sustainability.

Conflict of interest statement

We declare that we have no conflict of interest.

REFERENCES

- Ali, M., Eydurán, E., Tariq, M.M., Tirink, C., Abbas, F., Bajwa, M.A., Baloch, M.H., Nizamani, A.H., Waheed, A., Awan, M.A., Shah, S.H., Ahmad, Z. and Jan, S., 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**:1579-1585.
- Bakir, G., Keskin, S. and Mirtagioglu, H., 2010. Determination of the effective factors for 305 days milk yield by regression tree (RT) method. *J. Anim. Vet. Adv.*, **9**:55-59. <http://dx.doi.org/10.3923/javaa.2010.55.59>
- Cicek, A. and Erkan, O., 1996. *Research and sampling methods in agricultural economics*. Gaziosmanpasa University Agricultural Faculty Publications No: 12, Course book series, No: 6, Tokat.
- Dogan, I., 2003. Investigation of the factors which are affecting the milk yield in Holstein by CHAID analysis. *Ankara Univ. Vet. Fac. J.*, **50**:65-70.
- Eyduran, E., Yilmaz, I., Kaygisiz, A. and Aktas, Z. M., 2013a. An investigation on relationship between lactation milk yield, somatic cell count and udder traits in first lactation Turkish Saanen goat using different statistical techniques. *J. Anim. Pl. Sci.*, **23**:956-963.
- Eyduran, E., Yilmaz, I., Tariq, M.M. and Kaygisiz, A., 2013b. Estimation of 305-d milk yield using regression tree method in brown swiss cattle. *J. Anim. Pl. Sci.*, **23**:731-735.
- FAO, 1986. Sheep and goats in Turkey. Accessed on June 5, 2016. <http://www.fao.org/docrep/009/ah224e/AH224E00.htm>
- FAO, 2014. FAOSTAT. <http://faostat3.fao.org/browse/Q/QA/E> (08.05.2016).
- Grzesiak, W. and Zaborski, D., 2012. *Examples of the use of data mining methods in animal breeding*. ISBN 978-953-51-0720-0 <http://dx.doi.org/10.5772/50893>
- Grzesiak, W., Blaszczyk, P. and Lacroix, R., 2006. Methods of predicting milk yield in dairy cows- Predictive capabilities of Wood's lactation curve and artificial neural networks (ANNs). *Comput. Electron. Agric.*, **54**:69-83. <http://dx.doi.org/10.1016/j.compag.2006.08.004>
- Grzesiak, W., Lacroix, R., Wójcik, J. and Bxlaszczyk, P., 2003. A comparison of neural network and multiple regression predictions for 305-day lactation yield using partial lactation records. *Can. J. Anim. Sci.*, **83**:307-311.
- Ince, D. and Sofu, A., 2013. Estimation of lactation milk yield of Awassi Sheep with Artificial Neural Network modeling. *Small Rumin. Res.*, **113**:15-19. <http://dx.doi.org/10.1016/j.smallrumres.2013.01.013>
- Karaca, O., Akyuz, N., Andic, S. and Altin, T., 2003. Milk yield characteristics of Karakafi Sheep. *Turk. J. Vet. Anim. Sci.*, **27**:589-594.
- Keskin, I. and Dag, B., 2006. Comparison of different mathematical models for describing the complete lactation of Akkaraman Ewes in Turkey. *Asian-Aust. J. Anim. Sci.*, **19**:1551-1555.
- Khan, M.A., Tariq, M.M., Eydurán, E., Tatliyer, A., Rafeeq, M., Abbas, F., Rashid, N., Awan, M.A. and Javed, K., 2014. Estimating body weight from several body measurements in Harnai sheep without multicollinearity problem. *J. Anim. Pl. Sci.*, **24**:120-126.
- Kuchtik, J., Sustova, K., Urban, T. and Zapletal, D., 2008. Effect of the stage of lactation on milk composition, its properties and the quality of rennet curdling in East Friesian Ewes. *Czech J. Anim. Sci.*, **53**:55-63.
- Kucuk, M. and Eydurán, E., 2010. Determination of the best nonlinear model for describing complete lactation of Akkaraman and German Blackheaded Mutton x Akkaraman crossbreed (F1) cheep. *Bulg. J. agric. Sci.*, **16**:247-251.
- Sezenler, T., Ceyhan, A., Yuksel, M.A., Koncagul, S., Soysal, D. and Yildirim, M., 2016. Influence of year, parity and birth type on milk yield and milk components of Bandırma Sheep (German Black Head Mutton x Kivircik). *J. agric. Sci.*, **22**:89-98.
- Sahin, E.S. and Akmaz, A., 2004. Milk yield characteristic and control in sheep. *Vet. Bil. Derg.*, **20**:5-11.
- Takma, C., Atil, H. and Aksakal, V., 2012. Comparison of multiple linear regression and artificial neural network models goodness of fit to lactation milk yields. *Kafkas Univ. Vet. Fac. J.*, **18**:941-944.
- Topal, M., Aksakal, V., Bayram, B. and Yaganoglu, M., 2010. An analysis of the factor affecting birth weight and actual milk yield in swedish red cattle using regression tree analysis. *J. Anim. Pl. Sci.*, **20**:63-69.
- Unal, N., Atasoy, F., Aytac, M. and Akcapinar, H., 2002. Akkaraman, Sakız x Akkaraman F1, Kivircik x

- Akkaraman F1 ve Sakız x Karayaka G1 koyunlarda ilk laktasyon sut verim ozellikleri. *Turk. J. Vet. Anim. Sci.*, **26**:617-622.
- Yildiz, N. and Denk, H., 2006. Van bolgesinde halk elinde yetistirilen Akkaraman koyunlarda cesitli verim ozellikleri 1. dol ve sut verimi ozellikleri. *F.U. Saglik Bil. Dergisi.*, **20**:21-27.
- Yilmaz, O., Denk, H. and Nursoy, H., 2004. Milk yield characteristics of Norduz Sheep. *YYU. Vet. Fak. Derg.*, **15**:27-31.