The Journal of Animal & Plant Sciences, 30(3): 2020, Page: 560-567
ISSN (print): 1018-7081; ISSN (online): 2309-8694

DESCRIBING FACTORS AFFECTING BIRTH WEIGHT AND GROWTH TRAITS IN HEMSIN LAMBS USING DECISION TREE METHODS

B. Balta[1*] and M. Topal[2]

[1]Animal Science, East Anatolian Agricultural Research Institute, Republic of Turkey
[2]Department of Biostatistics, Faculty of Medicine, Kastamonu University, Kastamonu/Turkey
[*]Corresponding author e-mail: burcuhanb@gmail.com

ABSTRACT

The main goal of this study was to determine the best decision tree algorithm in order to determine the effects of the birth type, herd type, main age, pasture type, sex and lamb color variables in the Hemsin lambs. These data, taken during certain periods of the Hemsin lambs, were subjected to simulation. The obtained data were evaluated by Classification and Regression Tree (CART), Boosting, Bagging (Bag) and Random Forest (RF) Decision Tree Algorithms. The best model of lamb birth weight, with the lowest Error Squares Mean (MSE) = 0.469, with the lowest Mean Absolute Error (MAE) = 0,471 with Random Forest and the lowest Symmetric Percentage Mean Absolute Error (SMAPE) = 3,63 The best model according to daily live weight increases of lambs, the lowest MSE (3620.67), MAE (4878.71) and SMAPE(4.80) values formed by Random Forest algorithm and the best model according to the daily live weight gain constructed by Random Forest (RF) algorithm. It was determined that the Bagging algorithm with the lowest MSE (970.09), MAE (1362.65) and SMAPE (3.03) was formed. The achieved results showed that the best algorithm was Random Forest, followed by Bagging algorithm.

Keywords: Random Forest, Bagging, Classification and Regression Tree, Boosting, Simulation.
https://doi.org/10.36899/JAPS.2020.3.0066 *Published online March 25, 2020*

INTRODUCTION

The primary objectives in sheep businesses are milk yield, live weight, wool yield, number of lambs born per dam and lamb viability, since these provide important economic gains. The viability of lambs born in businesses leads to an increase in other yield characteristics. Therefore, it is necessary to accurately identify number of lambs born per dam and factors affecting lamb viability and rank these factors according to their importance in selection studies. Yield characteristics are affected by various factors such as genotype, environment, breeding conditions and diseases. Effects of these factors may be determined with linear and non-linear models and classified with decision tree algorithms.

The popularity and the number of scientific research increase during the last century. As a result, the amount of the scientific data soar which causes the use of some conventional method limited. Due to the fact that the analysis of large data sets and interpreting the obtained results are extremely difficult the importance and application of data mining, artificial neural networks, decision inference, logic programming, decision trees and genetic algorithms, are gaining importance. In recent years, data mining technology has been widely adopted to support institutions and individuals in every field (Dongsong and Lina 2004). One of the most used data mining techniques is Decision Trees.

Decision Trees were proposed by Breiman and Friedman. Decision Tree algorithms are a models used to make forward-looking estimation by classifying the data according to a certain feature. The decision-making process of tree-shaped decision structures from the data set is called decision trees. These trees are formed by dividing large volumes of data into smaller subgroups by decision-making rules. A Decision Tree model separates large heterogeneous groups into homogeneous subgroups according to the set target variable, ie, creating a rule from the root to the leaf (Berry and Linoff, 2004). The fact that decision trees can be easily applied to data sets that have very large and missing values, and that both continuous and categorical variables can be studied are very useful because the results aremore understandable and interpretable. In addition, the Decision Trees method is a nonparametric method which is an alternative to the least squares and logistic regression method and does not include the assumptions required for regression type problems. Different decision trees have been formed with different algorithms. However, due to the excessive adaptation of single decision tree algorithms to trees, severalstudies have been conducted to constructdifferent methods (Breiman 2003). With the development of computer programs, Decision Tree algorithms have been developed in community rather than single decision trees. These are Bagging, Boosting and Random Forest algorithms.

The literature contains only a limited number of studies using regression tree analysis in animal production, particularly in birth weight and milk yield. In a previous study, carried out on Norduz and Karakaş sheep in Turkey (Eyduran *et al.*, 2008), The regression tree method was used to determine the effects of race, sex, birth type, year of birth and mother's age on birth weight.In other study was to evaluate predictive performances of CHAID, Exhaustive CHAID, and CART regression tree methods for different combinations of parent node: child node in the data set regarding animal science. To achieve the aim, 1884 Mengali lambs were provided for predicting weaning weight from sex, birth type, birth year, farm, birth weight, dam age, and dam weight (Koc *et. al*. 2017).

Waltner *et al.* (1993) used multiple linear and nonlinear regression models; Domecq *et al.* (1997) used multiple linear regression and principal component analysis; Roshe *et al.* (2006) used mixed models and Berry *et al.* (2007) used linear and nonlinear regression models to determine the relationship between body condition scores and milk yield in Holstein cattle. Turkyılmaz *et al.* (2005); Yaylak and Kumlu (2005); Akçay *et al.* (2007) and Petrovic *et al.* (2009) used analysis of variance; Ray *et al.* (1992) used analysis of variance and regression analysis to determine the lactation number and calving seasons in Holstein cattle.

Topal *et al*. (2017) used to Chaid and Logistic Regression for assessing the effects of non-genetic factors on lamb mortality. Piwczyński (2009) established factors responsible for the number of lambs reared from fertilized mother using classification tree. Classification trees and logistic regression were used to obtain mortality of Polish Merino lambs between their birth and weaning time (Piwczyński *et al*. 2012). Classification trees and logistic regression were used to obtain relation between PrP genotypes and litter size in Polish Merino, Black-headed, Ile de France and Berrichon du Cher breeds (Grochowska *et al*. 2014). Regression tree was used to detect relationship between body weight and morphometric traits of Uda sheep (Yakubu, 2012).

In this study, data from certain periods of Hemsin were taken and simulation was applied to these data. The data obtained by Classification and Regression Tree (CART), Bagging and Random Forest decision tree algorithms by determining the important factors affecting the dependent variables by determining the best decision tree model is intended to determine the algorithm.

Estimations can be made with less errors if tree community methods are applied in animal husbandry studies. Since random forests will determine which variables are most important by calculating the importance of the variables in the studies on animal husbandry, the variables to be used in the studies to be done accordingly can be determined. In addition, the Random Forest method is a good guide to be used in animal breeding and genetic studies, especially where a large number of variables are used.

MATERIALS AND METHODS

Materials: The datawere provided from 23278 Hemsin lambs rearedin the province of Artvin between the years 2007-2010. Herd type, pasture type, age of dam, lamb sex, lamb color and birth type (single, twin and triplets) were considered as independent variables lamb birth weight, starting time of pasture period and end of pasture period live weight gain were used as dependent variables.

Methods: The data were prepared in accordance with the RStudio Windows NT 6.1 statistical program, the necessary software for the analysis of the package programs to be installed, the algorithms to be used and the algorithms to be made with these algorithms are written and the results of the analysis results are obtained in tables. Mean squares error, mean absolute error and symmetric mean absolute percentage errors were calculated by comparing algorithms.

Classification and regression tree algorithm (Camdeviren *et. al.* 2005): This method is employed to capture the independent variables that are influential with the aim of revealing the relationships between dependent and independent variables, taking into account the high level interactions between the independent variables. In the classification and regression tree (CART), each variable in the tree is divided into branches according to the importance level and the variable that divides the tree is determined. Other variables to be included in the tree, such as the first variable, by dividing both nodes in a similar way the tree diagram is created. The data to be used in the analysis were repeated 5000 times with simulation.

S branching criterion in any t node,

$\Psi(s/t) = 2P_L P_R \sum_{j=1}^{M} |P(C_J|t_L) - P(C_J|t_R)|$

t: node of branching, *c*: criter, *L*: Left side of tree, *R*: Right side of tree, P_L, P_R: Probability of a record in the learning set to the right or left, $P(C_J|t_L|)$ ve $P(C_J|t_R|)$: Probability that a C class record is on the right or left.

Boosting algorithm (Schapire 1999): In the selection of the data to be analyzed by the Boosting algorithm, the examples that were made in the previous analyzes were given priority. In other words, this method uses data that the previous classifier of the data cannot accurately determine. It is tried to make a more accurate estimation by adding the data found incorrectly in the training kit to be used later. The data was repeated 5000 times in the simulation and the data set to be used in the analysis was updated in each iteration.

The formation phase of the Boosting Algorithm (James *et al.* 2013),

1. For all *i* in training set $\hat{f}(x) = 0$ ve $r_i = y_i$
2. *b* = 1,2,……, Repeat until B
(a) Data (X, r) is divided by d (d + 1 inner node), a suitable $\hat{f}^b$ tree is formed.

(b) It is updated by adding the minimized $\hat{f}$ version of the new tree:

$$\hat{f}(x) \leftarrow \hat{f}(x) + \lambda \hat{f}^b(x)$$

(c) Structuring the residuals,

$r_i \leftarrow r_i - \lambda \hat{f}^b(x_i)$

3. Output of boosting model,

$$\hat{f}(x) = \sum_{b=1}^{B} \lambda \hat{f}^b(x)$$

Bagging algorithm: In the construction of the Bagging algorithm, samples were drawed randomly by using the bootstrap method by replacing it from the data set. The samples were put in place and the selection of the predictor was repeated. With the simulation study, 30% (5902) of the 22402 data in the actual data set and 70% (16500) of the data were used as training data. The training data to be used in the analysis is derived from the existing data by random selection. Simulation method was repeated 5000 times and the results were averaged.

Random Forest algorithm: While creating each decision tree, sample data is generated with the bootstrap method to the same extent as the number of variables (n) in the original data set. Two-thirds of this data was into of training data sets used to construct the tree and the remaining 1/3 was used as a test data set to test the internal error rate of the installed model, and these variables were chosen randomly in each analysis to repeated. In this analysis, 30% (5902) of the 22402 data in the real data set and 70% (16500) of the actual data set were used as training data. In the training data set, the multiple decision tree was analyzed by using the tree structure and the simulation method was repeated 5000 times and results were averaged.

The significance degrees of the variables were determined by using the shape MSE% increase criterion as a result of the analysis of the bagging algorithm and Random Forest algorithm. An internal estimation was madefor a generalization error with these observations which have never been used before. To obtain the error rate of unused data, each tree estimates a class value for the unused data set, and all these estimates are recorded, and at each point, the average of the error rate estimates in trees with unused data for each observation is taken as a percentage of MSE generalition error. The importance of variables increases with increasing MSE% increase in Bagging and Random Forest algorithms (Cutler *et al.* 2013).

RESULTS AND DISCUSSION

Today, with the increase in the amount of data, interest in data mining has gained great importance and has been widely used in many disciplines. Data mining enables the discovery of the desired and applicable information from a very large amount of information, as well as the storage and spreading of this information over many years and the ability to make forward-looking forecasts with the results obtained.

The basic information about the decision trees used in the study is given and the algorithms used in the analysis under the heading of decision tree algorithms are mentioned. In practice, lamb birth weight, daily live weight gain per pasture and end of pasture daily weight gain were taken as dependent variable, and pasture type, herd type, birth type, maternal age, gender and lamb color were taken as independent variables.

In order to determine the effects of the independent variables on dependent variables, Classification and Regression Tree (CART), Random Forest, Bagging and Acceleration methods, which are developed as an alternative to general statistical methods, were applied. Mean of error squares (MSE), Mean absolute error (MAE) and symmetric percentage mean absolute error (SMAPE) were compared. According to the results, the most important variables affecting the lamb birth weight, daily live weight gain per pasture and end of pasture daily live weight gain were found. According to these results, the algorithms constituting the best model were determined.

Table 1. Simulation results of algorithms of lamb birth weight.

Algorithm / Variable	Birth Weight			
	Bagging (% MSE Inc)	Random Forest (% MSE Inc)	Boosting (% Significance)	CART (%Significance)
Herd type	207.57	339.04	42.60	30
Birth type	153.85	422.25	30.11	24
Age of dam	128.83	177.76	11.01	19
Pasture type	126.54	197.67	7.47	16
Sex of lamb	44.21	215.32	7.75	7
Color of lamb	27.82	88.07	1.05	4

Table 2. Comparison results of algorithms according to lamb birth weight.

Algorithms	Birth Weight		
	MSE	MAPE	SMAPE
Bagging	0.4694	0.4741	3.67
Random Forest	0.5343	0,4711	3,63
Boosting	0.5388	0.5202	3.68
CART	0.5425	0.5420	3.74

The most important variables according to the methods used to determine the variables effecting birth weight in lambs,

In the Bagging algorithm, herd type, birth type, age of dam, pasture type, sex of lamb and lamb color were respectively 207.57, 153.85, 128.83, 126.54, 44.21 and 27.82,

In the Random Forest algorithm, birth type, herd type, sex of lamb, pasture type, age of dam and lamb color were respectively 422.25, 339.04, 215.32, 197.67, 177.76 and 88.07,

In the Boosting algorithm, herd type, birth type, age of dam, pasture type, sex of lamb and lamb color were respectively 42.60, 30.11, 11.01, 7.75, 7.47 ve 1.05,

In the CART algorithm, herd type, birth type, age of dam, pasture type, sex of lamb and lamb color were respectively calculated as 30, 24, 19, 16, 7 ve 4.

The best result for lamb birth weight, was obtained by Bagging with MSE (0.46), Random Forest algorithm with MAPE (0.471) and SMAPE (3.63), followed by Boosting and CART algorithm respectively.

Table 3.Simulation results of algorithms of daily live weight ıncrease of lambs in starting time of pasture period.

Algorithm / Variable	Daily Live Weight Increase of Lambs in Starting Time of Pasture Period			
	Bagging (% MSEInc)	Random Forest(%MSE Inc)	Boosting (%Significance)	CART (%Significance)
Herd type	211.99	186.31	48.84	47
Brith type	70.33	100.85	1.15	1
Age of dam	208.84	205.77	41.31	48
Pasture type	132.20	107.58	4.96	2
Sex of lamb	11.79	45.65	1.44	1
Color of lamb	43.20	121.27	2.30	1

Table 4. Comparison results of algorithms according to daily live weight ıncrease of lambs in starting time of pasture period.

Algorithms	Daily Live Weight Increase of Lambs in Starting Time of Pasture Period		
	MSE	MAPE	SMAPE
Bagging	3880.624	5137.46	4.82
Random Forest	3620.674	4878.71	4.80
Boosting	3758.37	5265.144	4.85
CART	3953.928	5194.82	5.02

The most important variables according to the methods used to determine the variables effecting daily live weight increase of lambs in per pasture,

In the Bagging algorithm, herd type, age of dam, pasture type, birth type, lamb color and sex of lamb were 211.99, 208.84, 132.20, 70.33, 43.20 ve 11.79, respectively. In the Random Forest algorithm, age of dam, herd type, lamb color, pasture type, birth type and sex of lamb were 205.77, 186.31, 121.27, 107.58, 100.85 ve 45.65, respectively.

In the Boosting algorithm, herd type, age of dam, pasture type, lamb color, sex of lamb and birth type were 48.84, 41.31, 4.96, 2.30, 1.44 ve 1.15, respectively.

In the CART algorithm, age of dam, herd type, pasture type, birth type, lamb color and sex of lamb were calculated as 48, 47, 2, 1, 1, 1, respectively. The best result for daily live weight gain of lambs in the Bagging of grazing was possible, with Random Forest as MSE (3620.67), MAPE (4878.71) ve SMAPE (4.80), followed by Bagging, Boosting and CART algorithm, respectively.

Table 5. Comparison results of algorithms according to daily live weight ıncrease of lambs in the end of pasture period.

Algorithm / Variable	Daily Live Weight Increase of Lambs in End of Pasture Period			
	Bagging (% MSE Inc)	Random Forest (% MSE Inc)	Boosting (%Significance)	CART (%Significance)
Herd type	199.67	118.85	12.17	31
Birth type	113.31	108.23	1.18	22
Age of dam	361.79	217.12	48.37	21
Pasture type	307.39	204.93	31.82	14
Sex of lamb	73.73	146.32	5.28	7
Color of lamb	60.11	100.86	1.17	5

Table 6. Comparison results of algorithms according to daily live weight ıncrease of lambs in end pasture period.

Algorithms	Daily Live Weight Increase ofLambs in the End of Pasture Period		
	MSE	MAPE	SMAPE
Bagging	970.088	1362.65	3.03
Random Forest	1050.857	1404.448	3.06
Boosting	1014.468	1410.241	3.11
CART	1129.754	1411.121	3.32

The most important variables according to the methods used to determine the variables affecting daily live weight gain of lambs in the end of pasture period, In the Bagging algorithm, relative importance values onage of dam, pasture type, herd type, birth type, sex of lamb and lamb color were 361.79, 307.39, 199.67, 113.31, 73.73 ve 60.12, respectively.

In the Random Forest algorithm, age of dam, pasture type, sex of lamb, herd type, birth type and lamb color were 217.12, 204.93, 146.32, 118.85, 108.23 ve 100.86, respectively. In the Boosting algorithm, age of dam, pasture type, herd type, sex of lamb, lamb color and birth type were 48.37, 31.82, 12.17, 5.28, 1.18 ve 1.18, respectively. In the CART algorithm, herd type, birth type, age of dam, pasture type, sex of lamband lamb color were calculated as 31, 22, 21, 14, 7 ve 5, respectively. Daily live weight gain were calculated at the end of grazing period and the best accurate results MSE (970.09), MAPE (1362.65) and SMAPE (3.03) were reached by Bagging Algorithm, is followed by Random Forest, Boosting and CART algorithm respectively.

Shahinfar *et al.* (2014) showed similarity with our results. They analyzed the data from Holstein breed cows by using Bagging, Bayesian networks, Bayes classifier and Random Forest algorithms and they concluded that Random Forest algorithm wasa better model compared with other algorithms. The present results were in agreement with the results ofthe Regression tree method by Eyduran *et al*. (2008) in order to determine the factors affecting birth weight in Karakas and Norduz lambs of Turkey. The previous results illustrated that the most influential predictors which affected birth weight of the lambs were the birth type, sex, age of dam and breed. Topal *et al.* (2010) investigated the influential factors affecting birth weight and actual milk yield on Swedish Red Cattle through the Regression Tree method and birth weight were found to had a significant affect on the type of birth. The affects of birth type and gender on birth weight were in disagreement with our present results and previous authors stated that genotype and race had significant effects (Thieme *et al*., 1999; Matika *et al*., 2003; Hassen *et al*., 2004; Cemal *et al*., 2005).

Ghafouri *et al.* (2016), investigated predictive performances of the Random Forest, Support Vector Machines and Boosting Algorithms. They found that the best performance was reached with Boosting Algorithm, which was different from our results. Akman *et al* (2011) obtained similar results for the Random Forest, CART and Bagging algorithm. The Random Forest algorithm, with the lowest error rate in the present study, produced the best results (error rates; Random Forest 3,33%, Bagging 5,54% and CART 8,75%). Akcetin and Celik (2014) 's best alternative decision tree, the best decision tree, logical analysis of data decision tree, C4.5, Naive Bayes, CART, Random Forest decision tree methods such as the results of their work to compare the results of the best accuracy rate and the success of the success of the random forest decision tree that shows the success of the large data sets, indicating that the analysis was similar to this study.

Conclusion: Tree-based community methods are methods that are easy to use and interpret compared to other methods and give very successful results. Studies have shown that the result of more than one tree coming together gives more successful results than the results of

a single tree. The Random Forest algorithm, which is included in the tree-based community methods, adds randomness to the model unlike other tree-based community methods. It can also be used easily if the data set is large.

In the present study, predictive performances of Bagging, Boosting, CART and Random Forest algorithms were comparatively examinedand the best algorithm was found to be Random Forest. In the Random Forest algorithm, the error rate is expected to be less than the other algorithms that we have used because of the random selection of the variables to be used in the separation of each node, as well as the individual power of all the trees and the connection between these trees. Studies on animal husbandry can be done with less error if tree community methods are applied. By calculating the importance of the variables in the studies on animal husbandry by random forest method, the variables which can be used in the studies to be done accordingly can be determined. In addition, a good guideline has been developed for use in animal breeding and genetic studies using a large number of variables with the Random Forest method.

Acknowledgements: The study was carried out under The Project "The Breeding of Hemsin Sheep raised in local conditions" (TAGEM/06/08/01/01). This research was supported by East Anatolian Agricultural Research Institute. This study is summarized for the first author'sPhD thesis.

REFERENCES

Akcay, H., M. İlaslan and A. Koc (2007). Effect of calving season on milk yield of Holstein cows raised at Dalaman state farm in Turkey. ADU J. Agriculture, 4: 59-61.

Akcetin, E. and U. Celik (2014). Performance comparison of decision tree algorithms in detecting spam. J. Internet Applications and Management, 5 (2).

Akman, M., Y. Genc, and H. Ankara (2011). Random forest method and an application in the field of health. Turkey Clinics, Biostat, 3 (1): 36-48.

Berry, M. J. A. and G. S. Linoff (2004). Data mining techniques, marketing, sales, and customer releationship management, Wiley Publishing, Inc., Indiapolis, Indiana.

Berry, D. P., F. Buckley and P. Dillon (2007). Body condition score and live-weight effects on milk production in Irish Holstein-Frisian dairy cows. Animal, 1(9): 1351-1359.

Breiman, L., J. Friedman, R. Olshen and C. Stone (2003). Classification and RegressionTrees. Boca Raton. Florida: Chapman and Hall.

Cemal, I., O. Karaca, T. Altin and M. Kaymakci (2005). Live Weights of Kivircik ewes andlambs in some periods under extensive management conditions, Turk J. Vet. Anim Sci., 29: 1329-1335.

Cutler, A., D. R. Cutler and J. R. Stevens (2013). Tree-based methods, p.21

Camdeviren, H., M. Mendeş, M. Ozkan, F. Toros, T. Sasmaz and S. Oner (2005). Determination of depression risk faktors in children and adolescents by regression tree methodology. Acta Med. Okayama, 59(1): 19-26.

Dongsong, Z. and Z. Lina (2004). Discovering golden nuggets, data mining in financial application, IEEE Transactions on Systems, Man and Cybernetics-Part C: Applications and Reviews, Vol:34, No:4, 513-515.

Domecq, J. J., A. L. Skidmore, J. W. Lloyd and J. B. Kaneene (1997). Relationship between body condition scores and milk yield in a large dairy herd of high yielding Holstein cows. J. Dairy Sci., 80: 101-112.

Eyduran, E., K. Karakus, S. Keskin and F. Cengiz (2008). Determination of factors influencing birth weight using regression tree (RT) method. J. Appl. Anim.Res., 34(2), 109-112.

Ghafouri-Kesbi, F., G. Rahimi-Mianji, M. Honarvarand and A. Nejati-Javaremi (2016). Predictive ability of random forest, boosting, support vector machines and genomic best linear unbiased prediction in different scenarios of genomic evaluation. Animal Pro. Sci. 57 (2) 229-236.

Grochowska, E., D. Piwczyński, B. Portolano and S. Mroczkowski (2014). Analysis of the influence of the PrP genotype on the litter size in Polish sheep using classification trees and logistic regression. Livest. Sci., 159: 11–17.

Hassen, Y., J. Solkner and B. Fuerst-Waltl (2004). Body weight of Awassi and Indigenous Ethiopian sheep and their crosses. Small Rumin. Res., 55: 51-56.

James, G., D. Witten, T. Hastie, and R. Tibshirani (2013). An ıntroduction to statistical learning with application in R. Springer New York Heidelberg Dordrecht London, ISBN 978-14614-7138-7 (ebook), p. 322.

Koc, Y., E. Eyduran and O. Akbulut (2017). Application of regression tree method for different data from animal science. Pakistan J. Zool., 49(2), pp 599-607.

Matika, O., J. B. Van Wyk, G. J. Erasmus and R. L. Baker (2003). A description of growth, carcass and reproductive traiths of Sabi sheep in Zinbabwe. Small Rumin. Res., 48:119-126.

Petrovic, M. D., Z. Skalicki, M. M. Petrovic and V. Bogdanovic (2009). The effect of systematic

factors on milk yield in Simmental cows over complete lactations. Biotechnology in Animal Husbandry, 25: 61-71.

Piwczynski, D., B. Sitkowska and E. Wisniewska (2012). Application of classification trees and lojistic regression to determine factors responsible for lamb mortality**.** Small Rumin. Res. 103: 225-231.

Piwczynski, D. (2009). Using classification trees in statistical analysis of discert sheep reproduction traits. *J Central Eurupean Agri 10, 303-310.*

Ray, D. E., T. J. Halbach and D. V. Armstrong (1992). Season and lactation number effects on milk production and reproduction of dairy cattle in Arizona. J. Dairy Sci., 75: 2976-2983.

Roshe, J. R., J. M. Lee and K. A. Macdonald (2006). Relationships among body condition score, body weight, and milk production variables in pasture-based dairy cows. J. Dairy Sci., 90: 3802-3815.

Schapire, R. E. (1999). Theoretical views of boosting and applications, Proceedings of the Tenth International Conference on Algorithmic Learning Theory, 13-25.

Shahinfar, S., D. Page, J. Guenther, V. Cabrera, P. Fricke and K. Weigel (2014). Prediction of insemination outcomes in Holstein dairy cattle using alternative machine learning algoritms. J. Dairy Science, 97(2): 731-742.

Thime, O., M. Karazeybek, M. A. Azman and A. Ugurlu (1999). Performance of willage sheep flocks in Central Anatolia. I . Growth of Lambs. Turk. J. Vet. Anim. Sci., 23:467-474.

Topal, M., V. Aksakal, B. Bayram, and M. Yaganoğlu (2010). An analysis of the factors affecting birth weight and actual milk yield in Swedish Red Cattle using regression tree analysis. The J. Anim.Plant Sci., 20(2):63- 69, ISSN: 1018-7081.

Topal, M., E. Emsen and A. M. Yağanoğlu (2017). CHAID and logistic regression approaches for assessing the effects of non-genetic factors on lamb mortality. The J. Anim. Plant Sci., 27(1): 40-47.

Turkyilmaz, M. K., H. E. Bardakcıoğlu and A. Nazligul (2005). Effect of some factors on milk yield in Holstein cows. J. Faculty of Vet. Med. Univ. Kafkas, 11: 69-72.

Yakubu, A. (2012). Application of regression tree methodolojy in predicting the body weight of Uda sheep. Anim. Sci. Biotech. 45: 484-490.

Yaylak, E. and S. Kumlu (2005). The effects of body condition score and some environmental factors on 305-day milk yield of Holstein cows. J. Agri. Faculty Ege Univ., 42(3):55-66.

Waltner, S. S., J. P. McNamara and J. K. Hillers (1993). Relationship of body condition score to production variables in high production Holstein dairy cattle. J. Dairy Sci., 76: 3410-3419.

Attachments

1. Bagging algorithm of lamb birth weight and algorithm code written in r program to obtain comparison criteria

```
# ***************************************************************************************************************************************
## bagging simulation:
set.seed(1)
bag.kuzdogag = randomForest(kuzdogag~., data=trainDF , mtry=6, importance =TRUE); bag.kuzdogag
yhat.bag     = predict (bag.kuzdogag , newdata =testDF)
yhat.bag2    = cbind(yhat.bag, newdata)
plot(yhat.bag, newdata)
abline(0,1)
mbag1 <- mean((yhat.bag-newdata)^2)
ibag1 <- importance(bag.kuzdogag)

bag.kuzdogag = randomForest(kuzdogag~., data=trainDF , mtry=6, ntree=5000, importance =TRUE); bag.kuzdogag
yhat.bag     = predict (bag.kuzdogag , newdata =testDF)
yhat.bag2    = cbind(yhat.bag, newdata)
plot(yhat.bag, newdata)
abline(0,1)
mbag2 <- mean((yhat.bag-newdata)^2)
ibag2 <- importance(bag.kuzdogag)
cbind(mbag1, mbag2); cbind(ibag1, ibag2)
#**************************************************************************************
```

2. Random forest algorithm of lamb birth weight and algorithm code written in r program to obtain comparison criteria

```
#*******************************************************************************************
## random forest simulation:
## target variable is "kuzdogag":
ind    <- sample.split(Y=Dataframe$kuzdogag, SplitRatio = 0.70)
trainDF<- Dataframe[ind,]
testDF <- Dataframe[!ind,]
modelRandom  <- randomForest(kuzdogag~., data=trainDF , mtry=2, ntree=5000, importance=TRUE); print(modelRandom)
yhat.bag     = predict (modelRandom , newdata =testDF)
yhat.bag2    = cbind(yhat.bag, newdata)
plot(yhat.bag, newdata)
abline(0,1)
mbag3 <- mean((yhat.bag-newdata)^2)
ibag3 <- importance(modelRandom)
varImpPlot(modelRandom)

cbind(mbag3); cbind(ibag3)
#*******************************************************************************************
```

3. Boosting algorithm of lamb birth weight and algorithm code written in r program to obtain comparison criteria

```
#*******************************************************************************
# boosted regression tree for kuzdogag:
gbmModel1 = gbm(kuzdogag~.,
                distribution   = "gaussian",
                data           = trainDF,
                n.trees        = 2500,
                cv.folds       = 5,
                shrinkage      = .01,
                n.minobsinnode = 20); summary(gbmModel1)
gbmTrainPredictions1 = predict(object  = gbmModel1,
                               newdata = testDF,
                               n.trees = 1500,
                               cv.folds= 5,
                               type    = "response")
yhat.boost   <- gbmTrainPredictions1
yhat.boost2  <- cbind(gbmTrainPredictions1, newdata)
plot(yhat.boost, newdata)
abline(0,1)
mbag4 <- mean((yhat.boost-newdata)^2)
cmse  <- cbind(mbag4);cmse                                    # compare mean squared errors for models

best.iter = gbm.perf(gbmModel1, method="cv")
plot.gbm(gbmModel1, 1, best.iter); plot.gbm(gbmModel1, 2, best.iter); plot.gbm(gbmModel1, 3, best.iter);
plot.gbm(gbmModel1, 4, best.iter); plot.gbm(gbmModel1, 5, best.iter);

#*****************************************************************
```

4. CART algorithm of lamb birth weight and algorithm code written in r program to obtain comparison criteria

```
#*********************************************************************
#CART (Classification and Regression Tree)
tree.fit = rpart (kuzdogag~., data=trainDF,
                  control = list(minsplit = 10, minbucket = 5, cp = 0.0001), method="anova" )

yhat.tree  = predict (tree.fit, newdata =testDF)
yhat.tree2 = cbind(yhat.tree, newdata)
plot(yhat.tree, newdata)
abline(0,1)

mtree <- mean((yhat.tree-newdata)^2); mtree

tree.fit$variable.importance

plot(tree.fit, uniform = TRUE)
text(tree.fit, digits = 3, use.n =TRUE)
#*********************************************************************
```