REGRESSION TREE ANALYSIS OF FACTORS AFFECTING FIRST LACTATION MILK YIELD OF DAIRY CATTLE

MIKAIL, N.[1*] – BAKIR, G.[2]

[1]*Department of Animal Science, Faculty of Agriculture, Siirt University, 56100 Siirt, Turkey*

[2]*Department of Animal Science, Faculty of Agriculture, Sütçü İmam University, 46100, Kahramanmaraş, Turkey*
(phone: +90-344-300-212-7432; fax: +90-344-280-2002)

**Corresponding author*
e-mail: naziremikail@siirt.edu.tr; phone: +90-484-212-1111/2818; fax: +90-484-212-1112

(Received 28th Jan 2019; accepted 6th Mar 2019)

Abstract. The objective of this study was to examine the effects of management and environmental effects on first lactation milk yield by means of regression tree method. Regression tree method is useful to determine the effects of several factors on specified depended variables. The independent variables used in the model were factors such as farm, breed, year of calving, season of calving, age at calving and days in milk. Data used in this study were obtained from three state farms. The resulting data set consisted of 754 records from 1st farm, 1120 records from 2nd farm and 324 records from 3rd farm. The average total milk yield and days in milk were 5413.11 ± 2033.18 kg and 324 ± 64 days, respectively. Year of calving and days in milk were important variables affecting first lactation milk yield of dairy cattle, followed by farm and breed ($p < 0.01$). The determination coefficient of the prediction was found as 78.9%. As a result of the study 14 distinct paths from the root node to the leaves were presented.
Keywords: *multiple linear regression, farm, breed, year of calving, days in milk, cow*

Introduction

First lactation milk yield is a very important characteristic in all dairy cattle. Factors affecting this trait can be divided into genetical (breed) and environmental, such as farm, year of calving, season of calving, age at calving and days in milk. Understanding the factors that change the environment of the dairy cattle can be used to take advantage of some improvement in milk yield that occur in normal lactation. During first lactation at an average age of 2.5 years cow produces approximately 76% of the milk produced by a mature cow (Nirish, 2010). Milk production is usually less during the summer because of the higher environmental temperatures and the prevalence of green-forage scarcity. Thus the season of calving has got a marked effect on the total production. Cows freshening shortly before winter months produce more total yield than those calving at other times of the year. The increase is probably due to more favorable temperature and more digestible feeds available during the winter. The farmer has no effect on the physiological factors of the cow, but has some effects on environmental factors affecting the cow. Factors that alter the environment of dairy cattle can be used to take advantage of some changes in milk composition and the yield during the lactation (Irshad, 2015).

An alternative approach to nonlinear regression is to sub-divide the space into smaller regions, where the interactions are more manageable. The general model consists of two parts: one is only a recursive section, and the other is a simple model for each cell of the section. The regression trees use the tree to represent the recursive part.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1702_52935303

Each terminal node or leaf of the tree represents a cell of the section and has just added a simple pattern applied to it in this cell. This is due to CART's desirable properties as, automatic handling of the variable selection, variable interaction modeling, local effect modeling, nonlinear relationship modeling. CART also is more robust in the presence of outliers and not affected by monotonic transformations of variables. Not all variables have to be in the same type; some of them can be continuous, some can be discrete, etc. Regression trees have several advantages: there are no complicated calculations and hence the estimation is fast: it is easy to understand what variables are important when estimating. In practice, it is possible to build CART models with dirty data (i.e. missing values, lots of variables, nonlinear relationships, outliers, and numerous local effects).

Many studies have been carried out in the application of regression trees in animal husbandry. Lots of them concerning determination of factors effecting milk yield (Mirtagioglu et al., 2008; Bakır et al., 2010; Topal et al., 2010; Cak et al., 2013; Eyduran et al., 2013) and body weight (Topal et al., 2010) of animals. But also there are some researches concerning the application of regression trees to the different data from animal science (Eyduran et al., 2008, 2016; Koç et al., 2017; Takma et al., 2017)

In the present study we are interested in examining the management and environmental effects affecting cows at first lactation by means of regression tree analysis.

Materials and methods

Materials

Data on total lactation milk production were obtained from Tahirova, Konuklar and Malya farms. These farms are the state enterprises belonged to General Directorate of Agricultural Enterprises (*Fig. 1*). Konuklar Farm is located in the 57 km from Konya. The climate is a typical continental climate of Central Anatolia; summers are hot and dry, winters are cold and rainy (GDA, 2016). The average annual rainfall is 322.4 mm, and the altitude is 1050 m. The average annual temperature is 11.6 °C (GDM, 2017). Malya Farm is located 27 km north east of Kırşehir. This farm is located in the Central Anatolia Region, at an altitude of 985 m. Again, the climate is the typical continental climate of Central Anatolia, summers are hot and dry and winters are cold and rainy (GDA, 2016). The average annual rainfall is 378.4 mm. The average annual temperature is 11.4 °C (GDM, 2017). Tahirova Farm is located in Balıkesir province; the farm is situated at an altitude of 166 m. Summers are hot, winters are rainy. The average annual rainfall is 583.7 mm. The average annual temperature is 14.6 °C (GDM, 2017).

At all three farms the cows were grazed and winter supplement was also provided. In the current paper, 2198 first lactation records were assessed for 1874 Brown-Swiss dairy cattle reared at Konuklar and Malya State Farms, and 324 Holstein dairy cattle reared at Tahirova State Farm. Records covered the period of calving from 1987 to 2007.

Methods

Regression tree method

Multiple linear regression (*Eq. 1*) is a way of making quantitative predictions with multiple independent variables $X_i \equiv \{X_1, X_2, ... Xn\}$

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1702_52935303

$$Y = \beta_0 + \beta^{T} X + \varepsilon, \qquad \text{(Eq.1)}$$

Independent variables have an additive effect on Y, also an interaction (*Eq. 2*),

$$Y = \beta_0 + \beta^{T} X + \gamma X X^{T} + \varepsilon \qquad \text{(Eq.2)}$$

Prediction trees use the tree to represent the recursive partition. Each of the terminal nodes, or leaves, of the tree represents a cell of the partition, and has attached to it a simple model which applies in that cell only (*Fig. 2*).

Figure 1. *Locations of investigated farms (Google Earth, 2019)*

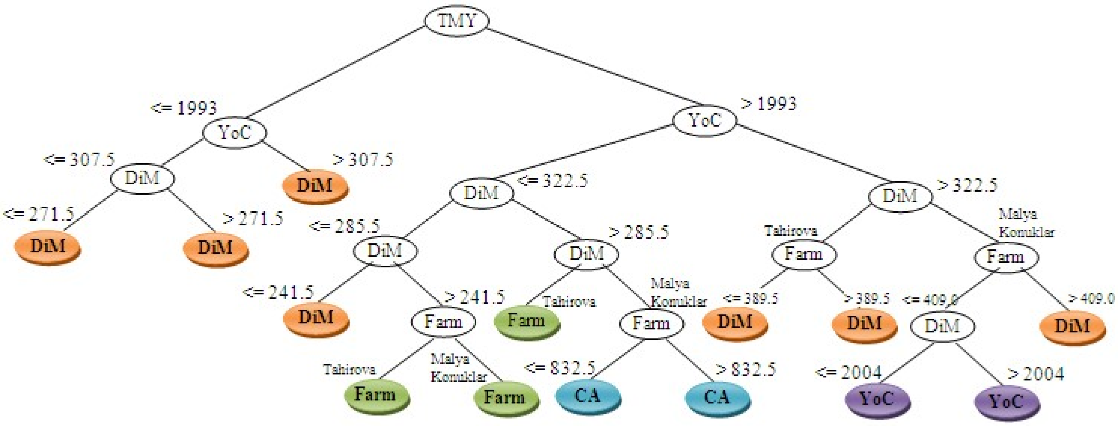

***Figure** 2. Scheme of CART of present work. TMY - total milk yield, YoC - year of calving, DiM - days in milk, CA - calving age*

Classification and Regression Tree splits the data into segments that are as homogeneous as possible with respect to the dependent variable (*Fig. 3*) (Lahmann and Kottner, 2011).

A regression tree can be seen as a kind of additive model (Hastie and Tibshirani, 1990) of the form:

$$m(x) = \sum_{i=1}^{n} k_i \times I(x \in D_i) \qquad \text{(Eq.3)}$$

where k_i are constants; I(.) is an indicator function returning 1 if its argument is true and 0 otherwise; D_i are disjoint partitions of the training data D such that $\bigcup_{i=1}^{n} D_i = D$ and $\bigcap_{i=1}^{n} D_i = \emptyset$.

A terminal node in which all cases have the same value for the dependent variable is a homogeneous, "pure" node. Regression trees (RT) is a method, where the target variable is continuous and tree is used to predict its value.

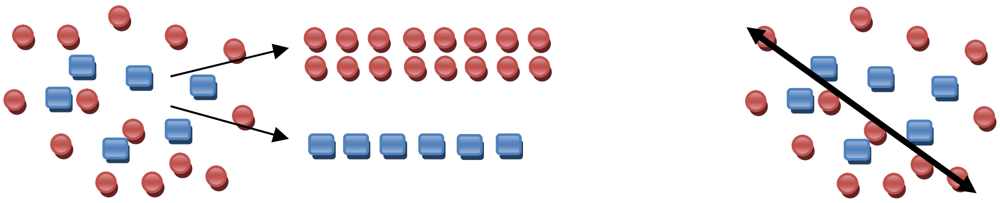

Figure 3. *Classification (left) and regression tree structure (right) for a model with 2 classes*

Each path from the root of the tree to a leaf corresponds to a region. Each inner node of the tree is a logical test on a predictor variable. In the particular case of binary trees, there are two possible outcomes of the test, true or false. This means that associated to each partition D_i we have a path p_i consisting of a conjunction of a logical tests on the predictor variables. This symbolic representation of the regression function is an important issue when one wants to have a better understanding of the regression surface (Soman et al., 2006).

For all analysis handled in this study IBM SPSS Statistics version 20 was used.

Goodness of fit

The coefficient of determination (*Eq. 4*) was used for determination of goodness of fit of the regression tree model.

$$R^2 = 1 - \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2 / \sum_{i=1}^{n} (y_i - \bar{y})^2 \qquad \text{(Eq.4)}$$

Results and discussion

There are 2 continuous (days in milk (DIM), calving age (CA) and 4 categorical (farm, breed, season of calving and year of calving) variables used to predict total milk yield (TMY). *Table 1* shows descriptive statistics of continuous variables. *Table 2* represents the frequency and percentage of categorical variables.

Table 1. *Descriptive statistics for the continuous variables*

Traits	N	Mean	SD	Min.	Max.	CV (%)
Total milk yield (kg)	2198	5413.11	2033.178	1098	12032	37.56%
Days in milk (days)	2198	324.19	63.899	137	675	19.7%
Calving age (days)	2198	858.46	90.160	702	1194	10.5%

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1702_52935303

Table 1 shows that the most variation is in the total milk yield. At the same time, the lowest number of days in milk is 137 days and the lowest calving age is 702 days.

Table 2. *Frequency and percentage values for several categorical variables*

Farm	Frequency	Rate	Season	Frequency	Rate	Breed	Frequency	Rate
Konuklar	754	34.3%	Winter	522	23.7%	BS	1874	85.3%
Malya	1120	51.0%	Spring	656	29.7%	HO	324	14.7%
Tahirova	324	14.7%	Summer	559	25.4%			
			Autumn	461	21.0%			
Total	**2198**	**100**	**Total**	**2198**	**100.0**	**Total**	**2198**	**100.0**

YoC	Frequency	Rate	YoC	Frequency	Rate	YoC	Frequency	Rate
1987	27	1.2%	1995	101	4.6%	2003	162	7.4%
1988	63	2.9%	1996	104	4.7%	2004	98	4.5%
1989	52	2.4%	1997	131	6.0%	2005	231	10.5%
1990	55	2.5%	1998	130	5.9%	2006	134	6.1%
1991	87	4.0%	1999	141	6.4%	2007	49	2.2%
1992	84	3.8%	2000	127	5.8%			
1993	72	3.3%	2001	134	6.1%			
1994	75	3.4%	2002	141	6.4%			
Total = 2198								

Figure 4 shows the average TMY of different farms according to the calving year.

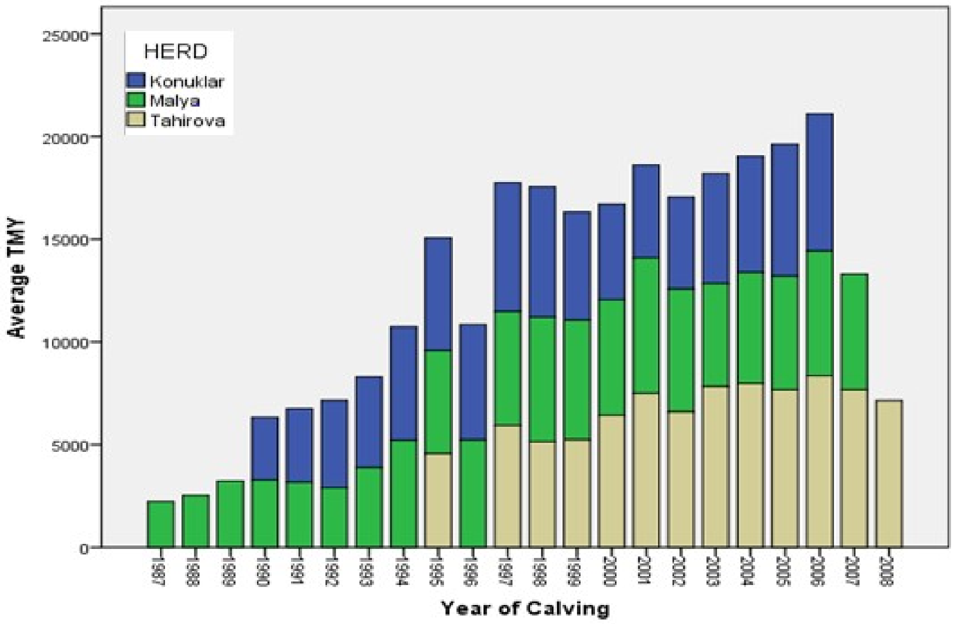

Figure 4. *The average TMY of different farms according to the calving year*

The CART algorithm is a process structured as a series of questions that determine what the next problem will be. The result of these questions is a tree-like structure with

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1702_52935303

endpoint nodes. On each intermediate node, a state goes to the lower left node if the condition is met (Loh, 2011).

The regression tree diagram is depicted in *Figure 5*.

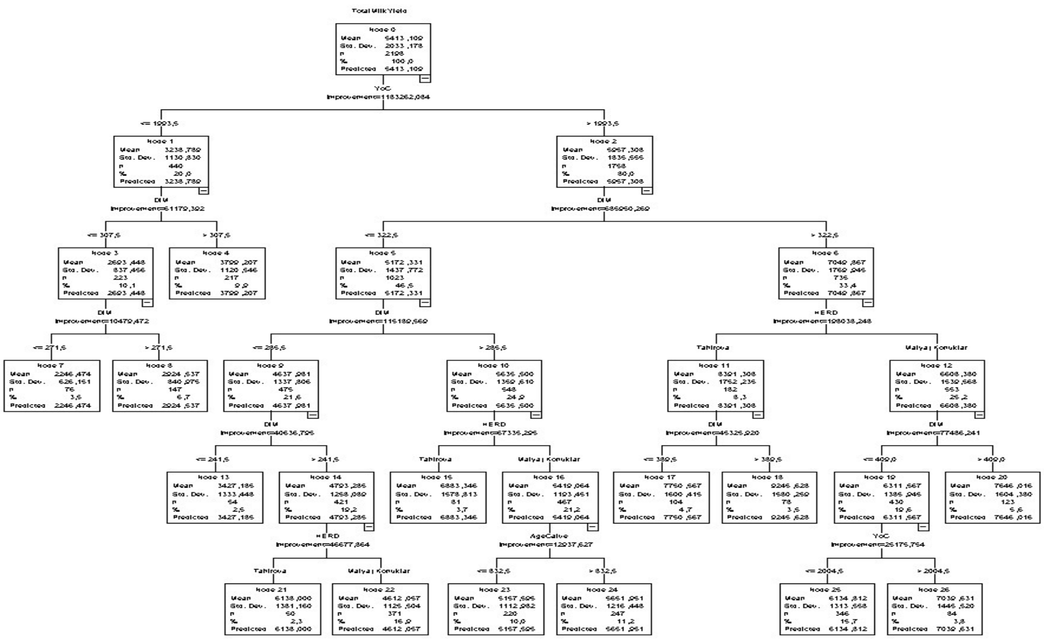

Figure 5. *Regression tree diagram for total milk yield*

The node definitions and splitting values for TMY were given in *Table 3*.

Table 3. *Node definitions and splitting values for TMY predictions*

Node	Predicted mean	SD	N	Percent	Parent node	Independent variable	F value	df1	df2	Split value
0	5413.11	2033.178	2198	100.0						
1	3238.79	1130.830	440	20.0	0	YoC	55.548**	21	2176	<= 1994
2	5957.31	1835.555	1758	80.0	0	YoC	55.548**	21	2176	> 1994
3	2693.45	837.456	223	10.1	1	DiM	3.897**	323	1874	<= 307.5
4	3799.21	1120.546	217	9.9	1	DiM	3.897**	323	1874	> 307.5
5	5172.33	1437.772	1023	46.5	2	DiM	3.897**	323	1874	<= 322.5
6	7049.87	1769.945	735	33.4	2	DiM	3.897**	323	1874	> 322.5
7	2246.47	626.151	76	3.5	3	DiM	3.897**	323	1874	<= 271.5
8	2924.54	840.975	147	6.7	3	DiM	3.897**	323	1874	> 271.5
9	4637.98	1337.806	475	21.6	5	DiM	3.897**	323	1874	<= 285.5
10	5635.50	1359.610	548	24.9	5	DiM	3.897**	323	1874	> 285.5
11	8391.31	1752.235	182	8.3	6	Farm	269.741**	2	2195	Tahirova
12	6608.38	1539.568	553	25.2	6	Farm	269.741**	2	2195	Malya; Konuklar
13	3427.19	1333.448	54	2.5	9	DiM	3.897**	323	1874	<= 241.5
14	4793.29	1258.089	421	19.2	9	DiM	3.897**	323	1874	> 241.5
15	6883.35	1578.813	81	3.7	10	Farm	269.741**	2	2195	Tahirova
16	5419.06	1193.451	467	21.2	10	Farm	269.741**	2	2195	Malya; Konuklar

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1702_52935303

17	7750.57	1600.415	104	4.7	11	DiM	3.897**	323	1874	<= 389.5
18	9245.63	1580.259	78	3.5	11	DiM	3.897**	323	1874	> 389.5
19	6311.57	1385.945	430	19.6	12	DiM	3.897**	323	1874	<= 409.0
20	7646.02	1604.380	123	5.6	12	DiM	3.897**	323	1874	> 409.0
21	6138.00	1381.160	50	2.3	14	Farm	269.741**	2	2195	Tahirova
22	4612.06	1125.504	371	16.9	14	Farm	269.741**	2	2195	Malya; Konuklar
23	5157.60	1112.982	220	10.0	16	CA	0.907	392	1805	<= 832.5
24	5651.95	1216.448	247	11.2	16	CA	0.907	392	1805	> 832.5
25	6134.81	1313.558	346	15.7	19	YoC	55.548**	21	2176	<= 2005
26	7039.63	1445.520	84	3.8	19	YoC	55.548**	21	2176	> 2005

**p < 0.01, $R^2 = 78.9\%$

At the top of regression tree diagram, Node 0, which gave general descriptive statistics of TMY, was divided into new two child nodes, with respect to YoC factor. The TMY averages of the cows were born in 1994 and early on (Node 1) was estimated as 3238.79 kg and the TMY average of the other cows (Node 2) was estimated as 5957.31 kg. Right side nodes always show high level mean. With an average of 5957.31 kg, Node 2 gave the highest TMY. Numbers (proportions) of lactation records were 440 (20%) for Node 1 and 1758 (80%) for Node 2. The YoC factor yielded a significant influence on TMY of cows whose lactation records were included in Node 1 (F = 55.548, df1 = 21, df2 = 2176, P < 0.01). This shows the increase in total milk yield over the years. Afterwards Node 1 was divided into two child nodes (Nodes 3 and 4), depending on DiM factor. Corresponding average values for these two new child nodes were 2693.45 kg and 3799.21 kg, respectively. Numbers of lactation records were established as 223 (10.1%) and 217 (9.9%). While TMY values of cattle with DiM of 307.5 days and less were 2693.45 kg and TMY values of cattle with DiM more than 307.5 days, were found as 3799.21 kg. On the other hand, Node 3 is divided into 2 branches according to DiM factor. While TMY value of 76 cows with DiM 271.5 days and less was found to be 2246.47 kg, TMY value of 147 cows with DiM more than 271.5 days was found as 2924.54 kg. DiM is the first criterion in the TMY classification of records from 1993 and after (Node 2). DiM factor remarkably influenced the TMY of cows whose lactation records were consisted of DiM less than 322.5 days in the Node 2. The TMY averages for these two nodes (Node 5 and Node 6) were found as: 5172.333 kg and 7049.87 kg, respectively. Numbers (proportions) of lactation records were established as 1023 (46.5%) and 735 (33.4%), respectively. According to DiM factor these records are also divided into animal classes which are less than 285.5 days (Node 9) and more than 285.5 days (Node 10). Node 9 is divided into classes with DiM less than 241.5 days (Node 13) and more than 241.5 days (Node 14). The Node 13 consists of 54 records and the average TMY value is 3427.19 kg. The Node 14 was divided into two child nodes: the Tahirova farm with an average TMY of 6138.00 kg and the Malya and Konuklar farms with an average TMY of 4612.06 kg. The records with DiM over 285.5 (Node 10) was divided into two groups. 81 animals with an average TMY value of 6883.35 kg were belonged to Tahirova (Node 15) farm and 467 animals with an average TMY value of 5419.06 kg belonged to the Konuklar and Malya farms (Node 16). The Node 16 was then further classified (Node 23 and Node 24) according to the CA factor that was not statistically significant. Animals with DiM of 322.5 days and more were divided into two child nodes on the basis of farm (Node 11

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1702_52935303

and Node 12). 182 animals with an average TMY value of 8391.31 kg were belonged to the Tahirova (Node 11) and 553 animals with an average TMY of 6608.38 kg were belonged to the Malya and Konuklar farms (Node 12). 182 animals (Node 11) in the Tahirova farm were divided into 2 homogeneous classes according to the DiM criteria: 104 animals with an average of 7750.57 kg milk and less than 389.5 DiM (Node 17) and 78 animals with an average of 9245.63 kg of milk and 389.5 and higher DiM (Node 18). The animals in the Malya and Konuklar farms (Node 12) were classified according to the DiM factor. Accordingly, the average TMY value of the 430 animals (Node 19) with a DiM of less than 409.5 days was 6311.57 kg, and the 123 animals (Node 20) with a DiM of more than 409.5 days were 7646.02 kg. Then, Node 19 was divided into two child nodes according to the YoC factor as before 2005 (Node 25) and after 2005 (Node 26).

On account of the fact that Nodes 4, 7, 8, 13, 15, 17, 18, 20, 21, 22, 23, 24, 25 and 26 are terminal nodes, there is not any separation for providing homogeneity in those nodes.

The following algorithms were used for the estimation of the TMY in the regression tree method we applied:

$P_1 = (\boldsymbol{YoC \le 1994}) \cap (\boldsymbol{DiM > 307.5})$	$k_1 = 3799.2$ kg
$P_2 = (YoC \le 1994) \cap (DiM \le 307.5) \cap (DiM \le 271.5) = (\boldsymbol{YoC \le 1994}) \cap (\boldsymbol{DiM \le 271.5})$	$k_2 = 2246.5$ kg
$P_3 = (YoC \le 1994) \cap (DiM \le 307.5) \cap (DiM > 271.5) = (\boldsymbol{YoC \le 1994}) \cap (\boldsymbol{271.5} < DiM \le 307.5)$	$k_3 = 2924.5$ kg
$P_4 = (YoC > 1994) \cap (DiM \le 322.5) \cap (DiM \le 285.5) \cap (DiM \le 241.5) = (\boldsymbol{YoC} > 1994) \cap (\boldsymbol{DiM \le 241.5})$	$k_4 = 3427.2$ kg
$P_5 = (YoC > 1994) \cap (DiM \le 322.5) \cap (DiM \le 285.5) \cap (DiM > 241.5) \cap (Farm = Tahirova) = (\boldsymbol{YoC} > 1994) \cap (\boldsymbol{241.5} < DiM \le 285.5) \cap (\boldsymbol{Farm = Tahirova})$	$k_5 = 6138.0$ kg
$P_6 = (YoC > 1994) \cap (DiM < 322.5) \cap (DiM < 285.5) \cap (DiM > 241.5) \cap (Farm = Malya \cup Konuklar) = (\boldsymbol{YoC} > 1994) \cap (\boldsymbol{241.5} < DiM \le 285.5) \cap (\boldsymbol{Farm = Malya \cup Konuklar})$	$k_6 = 4612.1$ kg
$P_7 = (YoC > 1994) \cap (DiM \le 322.5) \cap (DiM > 285.5) \cap (Farm = Tahirova) = (\boldsymbol{YoC} > 1994) \cap (\boldsymbol{285.5} < DiM \le 322.5) \cap (\boldsymbol{Farm = Tahirova})$	$k_7 = 6883.4$ kg
$P_8 = (YoC > 1994) \cap (DiM \le 322.5) \cap (DiM > 285.5) \cap (Farm = Malya \cup Konuklar) \cap (CA \le 832.5) = (\boldsymbol{YoC} > 1994) \cap (\boldsymbol{285.5} < DiM \le 322.5) \cap (\boldsymbol{Farm = Malya \cup Konuklar}) \cap (\boldsymbol{CA \le 832.5})$	$k_8 = 5157.6$ kg
$P_9 = (YoC > 1994) \cap (DiM \le 322.5) \cap (DiM > 285.5) \cap (Farm = Malya \cup Konuklar) \cap (CA > 832.5) = (\boldsymbol{YoC} > 1994) \cap (\boldsymbol{285.5} < DiM \le \boldsymbol{322.5}) \cap (\boldsymbol{Farm = Malya \cup Konuklar}) \cap (\boldsymbol{CA} > 832.5)$	$k_9 = 5652.0$ kg
$P_{10} = (YoC > 1994) \cap (DiM > 322.5) \cap (Farm = Tahirova) \cap (DiM \le 389.5) = (\boldsymbol{YoC} > 1994) \cap (\boldsymbol{322.5} < DiM \le 389.5) \cap (\boldsymbol{Farm = Tahirova})$	$k_{10} = 7750.6$ kg
$P_{11} = (YoC > 1994) \cap (DiM > 322.5) \cap (Farm = Tahirova) \cap (DiM > 389.5) = (\boldsymbol{YoC} > 1994) \cap (\boldsymbol{DiM} > 389.5) \cap (\boldsymbol{Farm = Tahirova})$	$k_{11} = 9245.6$ kg
$P_{12} = (YoC > 1994) \cap (DiM > 322.5) \cap (Farm = Malya \cup Konuklar) \cap (DiM > 409.5) = (\boldsymbol{YoC} > 1994) \cap (\boldsymbol{DiM} > 409.5) \cap (\boldsymbol{Farm = Malya \cup Konuklar})$	$k_{12} = 7646.0$ kg
$P_{13} = (YoC > 1994) \cap (DiM > 322.5) \cap (Farm = Malya \cup Konuklar) \cap (DiM \le 409.5) \cap (YoC \le 2005) = (\boldsymbol{1994} < YoC \le 2005) \cap (\boldsymbol{322.5} < DiM \le 409.5) \cap (\boldsymbol{Farm = Malya \cup Konuklar})$	$k_{13} = 6134.8$ kg
$P_{14} = (YoC > 1994) \cap (DiM > 322.5) \cap (Farm = Malya \cup Konuklar) \cap (DiM \le 409.5) \cap (YoC > 2005) = (\boldsymbol{YoC} > 2005) \cap (\boldsymbol{322.5} < DiM \le 409.5) \cap (\boldsymbol{Farm = Malya \cup Konuklar})$	$k_{14} = 7039.6$ kg

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1702_52935303

where, P_i shows the algorithm to the terminal Node. TMY obtained from each algorithm is shown with the variable k_i.

There are 14 distinct paths from the root node to the leaves. This three divides the input data in 14 different regions. Using the (*Eq. 3*), we obtain:

$$m(x) = \sum_{i=1}^{14} k_i \times I(P_i)$$

Figure 6 shows importance level of the independent variables. According to *Figure 6,* YoC was significantly identified as the most important factor influencing TMY.

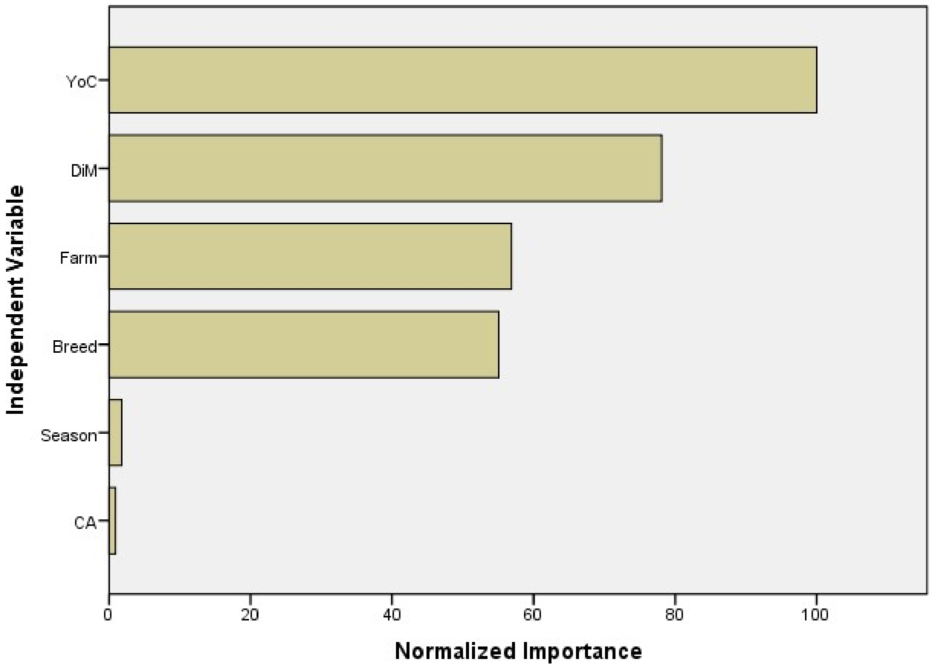

Figure 6. *Importance level of independent variables affecting TMY*

Their importance rate was given in *Table 4*.

Table 4. *Importance level of independent variables affecting TMY*

Independent Variable	Importance	Normalized importance (%)
Year of Calving	1299236.915	100.0
Days in Milk	1052872.027	81.0
Farm	765055.898	58.9
Breed	741247.200	57.1
Season	29218.096	2.2
Calving Age	24202.305	1.9

The R^2 of the tree is 78.9%, which is significantly higher than that of a multiple linear regression fit to the same data (R^2 = 59.1%). In current YoC was found most

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1702_52935303

important factor affecting total milk yield followed by days in milk (81%). Teke and Akdağ (2010); Bayril and Yılmaz (2010) in their study indicated that the effect of calving year on lactation milk yield were significant in Holstein cows. The categorical variable farm, as we expected, is grouped separately Malya and Konuklar (close to each other geographically) and Tahirova. The other common characteristics of Konuklar and Malya farms are that they have the same breed animal (Holstein). As a result of the analysis, it is expected that the importance rate of both farm and breed characteristics are close (58.9% and 57.1%, respectively). Gürses and Bayraktar (2012) reported that effects of enterprise, calving year were found significant on 305 days milk yield. Season and age at calving was not found so important in the total milk yield prediction (2.2% and 1.9%, respectively). In the study of Topal et al. (2010) season of calving of Swedish Red cattle also was found the least important factor (5.9%) affecting actual milk yield. But there are some other works were claiming the opposite (Bakır and Çetin, 2003; Özçakır and Bakır, 2003; Sehar and Özbeyaz, 2005; Erdem et al., 2007; Akçay et al., 2007; Çilek, 2009).

Conclusion

Due to the fact that CART is not a parametric method, the data used here is not required be belonged to a particular type of distribution. Also, this method can easy determine effects of both continuous and categorical variables on dependent continuous variable.

In regression tree diagram, the year of calving trait was determined to be the most affective factor for total milk yield prediction, followed by days in milk, farm and breed ($P < 0.01$). Thus, while the average total milk yield in the first lactation was 3238.79 ± 1130.83 until 1993, this yield increased in years and reached 6134.81 ± 1313.56 kg in 2005. This yield has increased to 7039.63 ± 1445.52 kg in Malya and Konuklar farms after 2005. In the Tahirova farm, we can say from the P_{11} path that the total milk yield increased after 1994 and reached the average of 9245.63 ± 1580.26 kg. Insignificant input variables such as season and age at calving were excluded from the regression tree.

As a result, the routing of the algorithms leading to the terminal node will help to determine the criteria for the classification in the regression tree and the application of the test data to this algorithm.

REFERENCES

[1] Akcay, H., Ilaslan, M., Koc, A. (2007): Effect of calving season on milk yield of Holstein cows raised at Dalaman State Farm Turkey. – Ziraat Fak Itesi Dergisi 4: 59-61.

[2] Bakır, G., Çetin, M. (2003): Breeding characteristics and milk yield traits of Holstein cattle in Reyhanlı Agricultural Farm. – Turk. J. Vet. Anim. Sci. 27: 173-180.

[3] Bakır, G., Mirtagioğlu, H., Keskin, S. (2010): Determination of the effective factors for 305 days milk yield by regression tree method. – J. Anim. Vet. Adv. 9: 55-59.

[4] Bayrıl, T., Yılmaz, O. (2010): Milk yield traits of Holstein cows raised in Kazova Vasfi Diren Agriculture Farm. – J. Fac. Vet. Med. Uni. Yüzüncü Yıl 21: 115-119.

[5] Cak, B., Keskin, S., Yılmaz, O. (2013): Regression tree analysis for determining of affecting factors to lactation milk yield in Brown Swiss cattle. – Asian Journal of Animal and Veterinary Advances 8: 677-682.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1702_52935303

[6] Çilek, S. (2009): Milk yield traits of Holstein cows raised at Polatlı State Farm in Turkey. – J. Vet. Anim. Adv. 8: 6-10.
[7] Erdem, H., Atasever, S., Kul, E.(2007): Milk yield and fertility traits of Holstein cows raised at Gökhöyük State Farm fertility traits. – J. Fac. Agric. 22: 47-54.
[8] Eyduran, E., Karakuş, K., Keskin, S., Cengiz, F. (2008): Determination of factors influencing birth weight using regression tree method. – J. App. Anim. Res. 34: 109-112.
[9] Eyduran, E., Yılmaz, I., Tarıq, M., Kaygısız, A. (2013): Estimation of 305-d milk yield using regression tree method in Brown Swiss cattle. – J. Anim. Pl. Sci. 23: 731-735.
[10] Eyduran, E., Keskin, İ., Ertürk, Y. E., Dağ, B., Tatlıyer, A., Tirink, C., Aksahan, R., Tarıq, M. M. (2016): Prediction of fleece weight from wool characteristics of sheep using regression tree method (Chaid Algorithm). – Pakistan J. Zool. 48(4).
[11] GDA (2016): General Directorate of Agriculture. – https://www.tigem.gov.tr-/Isletmeler.aspx (date of access: 14.07.2016.
[12] GDM (2017): Official statistics. General Directorate of Meteorology. – https://www.mgm.gov.tr/veridegerlendirme/il-ve-ilceler-istatistik.aspx (date of access: 04.05.2017).
[13] Gürses, M., Bayraktar, M. (2012): Some milk production and reproductive traits of Holstein cattle raised in different regions of Turkey. – J. Fac. Vet. Med. Uni. Kafkas 18: 273-280.
[14] Hastie, T. J., Tibshirani, R. J. (1990): Generalized Additive Models. – Chapman & Hall/CRC, London.
[15] Irshad, A. (2015): Factors Affecting Quality and Quantity of Milk. – Department of MSC and Tech VCRI, Namakkal.
[16] Koç, Y., Eyduran, E., Akbulut, Ö. (2017): Application of regression tree method for different data from animal science. – Pakistan J. Zool. 49(2): 599-607.
[17] Lahmann, N. A., Kottner, J. (2011): Relation between pressure, friction and pressure ulcer categories: a secondary data analysis of hospital patients using CHAID methods. – Int. J. Nurs. Stud. 48(12): 1487-1494.
[18] Loh, W. Y. (2011): Classification and regression trees. – Data Mining and Knowledge Discovery 1(1): 14-23.
[19] Mirtagioğlu, H., Keskin, S., Bakır, G. (2008): Regression tree analysis for 305 day milk yield in Holstein cows. – Indian Vet. J. 85: 943-945.
[20] Nirish, C. S. (2010): Prospects of Commercial Dairy Farming in Haryana. – Division of Dairy Extension National Dairy Research Institute (I.C.A.R.), Karnal, India.
[21] Özçakır, A., Bakır, G. (2003): Milk yield and reproduction characteristics of Holstein Friesian cattle raised in Tahirova Agricultural Facility: 1. Milk yield traits. – J. Fac. Agric. AU. 4: 145-149.
[22] Sehar, O., Özbeyaz, C. (2005): Some production traits of Holstein cows at a state farm in middle Anatolian conditions. – J. Lalahan Livestock Central Res. Inst. 45: 9-19.
[23] Soman, K. P., Diwakar, S., Ajay, V. (2006): Insight into Data Mining: Theory and Practice. – Prentice Hall India Learning Private Limited, Delhi, pp. 83-84.
[24] SPSS (2011): SPSS for Windows, Version 20. – SPSS Inc., Chicago, USA.
[25] Takma, Ç., Gevrekçi, Y., Karahan, A., Atıl, H., Çevik, M. (2017): Determining some traits effect on egg production of layers using regression tree analysis. – The Journal of Faculty of Agriculture of Ege University 54(4): 459-463 (original in Turkish).
[26] Teke, B., Akdağ, F. (2010): Effects of some environmental factors and the length of dry period on milk yield of Jersey cows raised at Karakoy State Farm. – J. Lalahan Livest. Res. Inst. 50: 65-72.
[27] Topal, M., Aksakal, V., Bayram, B., Yağanoğlu, A. M. (2010): An analysis of the factors affecting birth weight and actual milk yield in Swedish Red cattle using regression tree analysis. – J. Anim. Plants Sci. 20: 63-69.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1702_52935303