ISPEC Tarım Bilimleri Dergisi
6(3): 539-546, 2022

Araştırma Makalesi

E-ISSN:2717-7238
www.ispecjournal.com

ISPEC Journal of Agr. Sciences
6(3): 539-546, 2022

Research Article

Comparison of Cattle Breeds with Classification Tree According to Milk Components

Hasan ERGUN[1a*]
İsmail KESKİN[1b]

[1]Selcuk University, Faculty of Agriculture, Department of Animal Science, Konya

[1a]**ORCID**: 0000-0002-9797-2855
[1b]**ORCID**: 0000-0001-9358-7522

[*]Corresponding author (Sorumlu yazar):
hs4n@hotmail.com

DOI
https://doi.org/10.5281/zenodo.6993838

Abstract
This study was carried out to classify Jersey, Charolaise, Simmental and Holstein cattle breeds by using decision tree analysis according to components of milk sample (fat, protein, lactose, density, pH, conductivity). In the study, there are a total of 16 nodes in classification tree, formed for CART decision tree for 4 different dairy cattle. It is seen that root node (Node 0) is divided into two lower nodes as Node 1 (pH≤6.86) and Node 2 (pH>6.86) in terms of pH, and that pH of raw milk is effective in classifying cows. In the study, correct classification rate actualized as 63.0% in Charolaise cows; 83.0% in Holstein cows, 78.9% in Jersey cows and 86.7% in Simmental cows. As a conclusion, it was stated that the different cattle breeds could be correctly classified by CART algorithm in the rate that may be considered high.

Alınış (Received): 26/04/2022
Kabul Tarihi (Accepted): 29/05/2022

Keywords
CART algorithm, regression tree, holstein, charolaise, jersey, simmental, milk components

INTRODUCTION

In dairy cattle breeding, the contents of dry substance, protein and fat in the milk produced are highly important, and they are also criteria used in not only in drinking milk but also milk processing industries for classifying milk (Ergül et al., 2019). However, milk composition varies, depending on the factors such as the breed of animal, lactation order, lactation period, season, feeding applications and udder health (Özek, 2015). Knowing the factors modifying milk composition is quite important for milk processing plants. Because these plants form their production and marketing systems according to the properties of incoming raw milk (Yaylak et al., 2007). If total content of dry substance in raw milk is more than expected, it will be suitable to use it in the production of the products such as milk powder, concentrated milk, and yoghurt. In classification of raw milk in Turkey, the rescript (Tebliğ No: 2019/64), published with the number of 31019 in the Official Gazette on the date of January 25, 2020, has been used. According to this rescript, classification of raw cow milk is like in Table 1.

Table 1. Classification of raw cow milk

Classes	Protein Value (%)		Fat Value (%)	
	October-March	April-September	October-March	October-March
A	3.20 and over	3.10 and over	3.60 and over	3.50 and over
B	3.0 ≤ protein value < 3.2	3.0 ≤ protein value < 3.1	3.3 ≤ fat value < 3.6	3.2 ≤ fat value < 3.5
C	2.9 ≤ protein value < 3.0	2.9 ≤ protein value < 3.0	fat value < 3.3	fat value < 3.2

In identification of classification, it was also stated that providing the rates of fat and protein was difficult at the same time, in case that it cannot be provided it, that the lower value would be classified according to the class it takes places. In classifying the data obtained from animal breeding, a lot of statistical methods are used. In order to be able to analysis by these methods, it is necessary to perform some assumptions (that the data exhibit normal distribution, that variances are homogenous). However, any classification and regression trees are of the methods preferred due to some advantages such as not needing any assumption associated with distribution of independent variables of classification and regression tools, multicollinearity, outliers and not being affected from missing observations (Mendeş and Akkartal, 2009). Namely, classification and regression trees are reversed tree-shaped models serving to predict the class membership of non-continuous or continuous dependent variable without putting forward any condition (Akşahan and Keskin, 2015). The method of classification and regression tree is named as “regression tree” in case that dependent variable is continuous and as “classification tree” in case that dependent variable is categorical. Either regression or classification tree methods show the relationship between dependent and independent variables in the shape of tree. Using the combinations of one or more independent variables, categorically or continuously, models in the shape of reversed tree, which serves to reveal the variation in dependent variables with repetitive and dual homogenous divisions and predict the values of dependent variables, are referred to tree models. In classification and regression trees, the main aim is to divide the dataset into subgroups as homogenous as possible i.e., to obtain terminal nodes. The decision tree formed starts with root node and continues to form in the form of subgroups following each other repetitively (Çamdeviren et al., 2005). On the sub-nodes in tree, the best divided dependent variables are shown. Provided that enough homogeneity is provided in the structure of tree, division in subgroups (new subgroups) is not formed. In this stage, the subgroups, in which there is no division, are referred to terminal node. On the branches of these nodes, there are leaves indicating the value of distinctive dependent variables. To briefly summarize, they are the tree-shaped model, where the nodes take place by terminal nodes (leaves). In these trees, interclass distinction is maximized, and variation in each class is minimized (Özkan, 2012; Yücel, 2017). The classification and regression trees, since they do not need any assumption and that the results are visually presented, since it is difficult to understand and interpret them relatively easier compared to the other methods, are often used in the recent times. The classification and regression tree method has been used by some researchers in dairy cattle for different purposes: in determining of factors affecting milk yield in Holstein cows (Doğan, 2003), the relationship between mature age and cow age first mating age, lactation order, dry period, first calving age, calving season, birth type and sex of calf in Brown Swiss cows (Bakır et al., 2009), the effects of dry period, lactation parity, farm, calving season and age on 305-day milk yield in Holstein cows (Bakır et al., 2010), the effects of some factors affecting birth weight and actual milk yield in Swedish Red cattle (Topal et al., 2010), of environmental factors affecting 305-Day milk yield in Holstein Cows (Oruçoğlu, 2011), the factors affecting lactation milk yield in Brown Swiss cows (Çak et al., 2013), of some environmental factors (calving season, calving year, parity, calving interval and dry period) that are effective on 305-day milk yield in Brown Swiss cows (Eyduran et al., 2013), the relationship between mastitis and milk composition (pH, electrical conductivity, milk fat, milk freezing point, and milk color (L, a and b)) in Holstein and Brown Swiss cows (Aytekin et al., 2018). Apart from these, some researchers have also used Classification and Regression Tree methods in sheep (Eyduran et al., 2016; Karabacak et al., 2017; Altay et al., 2021). The factors affecting the continued consumption of foods of animal origin was also investigated by means of classification and regression tree analyses method (Mikail and Kaplan, 2021). In this study, it was aimed to classify Jersey, Charolaise, Simmental and Holstein cattle breeds, using decision tree analysis on their components of milk samples.

MATERIAL and METHOD

Material

The material of this study consists of 18 heads of Jersey, 53 heads of Holstein, 30 heads of Simmental and 27 heads of Charolaise in the first lactation, growing in the private enterprises in the different towns of Konya. The values of fat (%), protein (%), lactose (%), density (kg/m^3), pH and conductivity (µS/cm) for the milk samples, taken from 130 heads of cows into centrifugation tubes of 50 ml were identified by measuring once for each sample by means of MMC-30 milk analysis.

Method

The principal reason for commonly using regression trees in classification is that its analyzing is easier and more understandable. Basically, there are two stages for obtaining results from regression analysis. First is to install tree. In installing regression tree, for obtaining a result more rapidly and reliably, the questions to be able to divide the data into the biggest parts should be begun to be responded beginning from root node. After the structure of regression tree, it is passed to the second stage, and the available data are placed in the suitable branch of the tree (Akşahan and Keskin, 2015). In tree models, beginning from the first nodes, it is reached homogenous subgroups by means of dual repetitive divisions, and the state of dependent variable is defined in decision points. In this way, the observations taking place on the node points in regression trees are assigned to the suitable one of two child nodes. If the observations in the Node I are less than zero or equal to zero, they are assigned to Node II; if they are more than zero, to Node III. Similarly, if the observations in Node II are assigned to Node IV and Node V according to the value independent variable has, while the value in Node III are assigned to Node VI and Node VII (Temel, 2004). In this study, the numbers of parent and child node for CART algorithm was adjusted as 10:5.

RESULTS and DISCUSSION

In terms of the milk components (fat, density, lactose, protein, pH and conductivity) in Jersey, Simmental, Holstein and Charolaise cows, in order to identify whether or not there is statistically significant difference, one way variance analysis was made. In comparison of means (breeds) turning out different, the results obtained by using Duncan analysis are given in Table 2 together with means and standard errors.

Table 2. The means and standard deviations of milk components according to the breeds

Milk Components	Cattle Breeds			
	Jersey	Simmental	Holstein	Charolaise
Fat	4.78±0.31^{ab}	4.34±0.32^{ab}	4.14±0.33^{b}	5.21±0.43^{a}
Density	42.60±1.13^{A}	38.85±0.81^{B}	35.43±0.54^{C}	36.12±1.12^{C}
Lactose	6.74±0.19^{A}	5.89±0.13^{B}	5.62±0.10^{B}	5.84±0.16^{B}
Protein	4.50±0.13^{A}	3.93±0.09^{B}	3.76±0.07^{B}	3.90±0.11^{B}
pH	7.12±0.02^{A}	6.74±0.03^{C}	6.94±0.02^{B}	6.98±0.04^{B}
Conductivity	4.59±0.11^{ab}	4.72±0.09^{a}	4.50±0.04^{b}	4.39±0.09^{b}

a, b: $p<0.05$; A, B, C: $p<0.01$

In terms of fat among milk components, while Charolaise breed has the highest value, Holstein breed had the highest value, and the difference between them was found statistically significant ($p<0.05$). Jersey and Simmental took place between Charolaise and Holstein, and the difference between it and Charolaise was found statistically insignificant. In almost all properties (density, lactose, protein and pH) that are important in turning raw milk into products, Jersey had higher values and the differences between the other breeds and it were found statistically significant ($p<0.01$). Classification tree diagram, formed by means of CART decision tree for 4 different cattle breeds according to the milk components, is given in Figure 1. On the top of CART based classification tree, all cattle breeds (Jersey, Simmental, Holstein and Charolaise) examined take place in Node 0, called root node. Root node (Node 0) was divided into two small subsets as Node 1 (pH≤6.86) and Node 2 (pH>6.86) in terms of pH. In Node 1, where pH is less than 6.86 or equals to 6.86, while there was not any Jersey, 26 out of 30 heads of Simmental cow were found in this node. In Node 2, where pH is more than 6.86, there were only 4 heads of Jersey. In addition, 22 of 27 heads of Charolaise cows took place in Node 2. Node 1 was reserved to Node 3 in terms of those, where fat rate was less 7.45 or equals to that, is divided into Node 3; in terms of those, where they were more than 7.45, to Node 4. In Node 4, any Jersey and Simmental did not take place. Node 2 was divided into Node 5 (density≤39.145) and Node 6 (density>39.145) in terms of milk density. In Node 5, where milk density is lower than 39.145 and equals to that, 41 out of 53 Holstein cows took place. Node 5 is divided into Node 7 and Node 8 in terms of electrical conductivity. Node 6 is divided into Node 9 and Node 10 in terms of milk density. Node 8, which consists of the cows, where the electrical density of the milk is higher than 4.085, was divided into Node 11 (pH≤7.015) and Node 12 (pH>7.015) in terms of pH of the milk. Node 9 is divided into Node 13 (density≤40.635) and Node 14 (density>40.635) in terms of milk density, while Node 12 is divided into Node 15 (fat≤4.145) and Node 16 (fat>4.145) in terms of fat content of the milk.

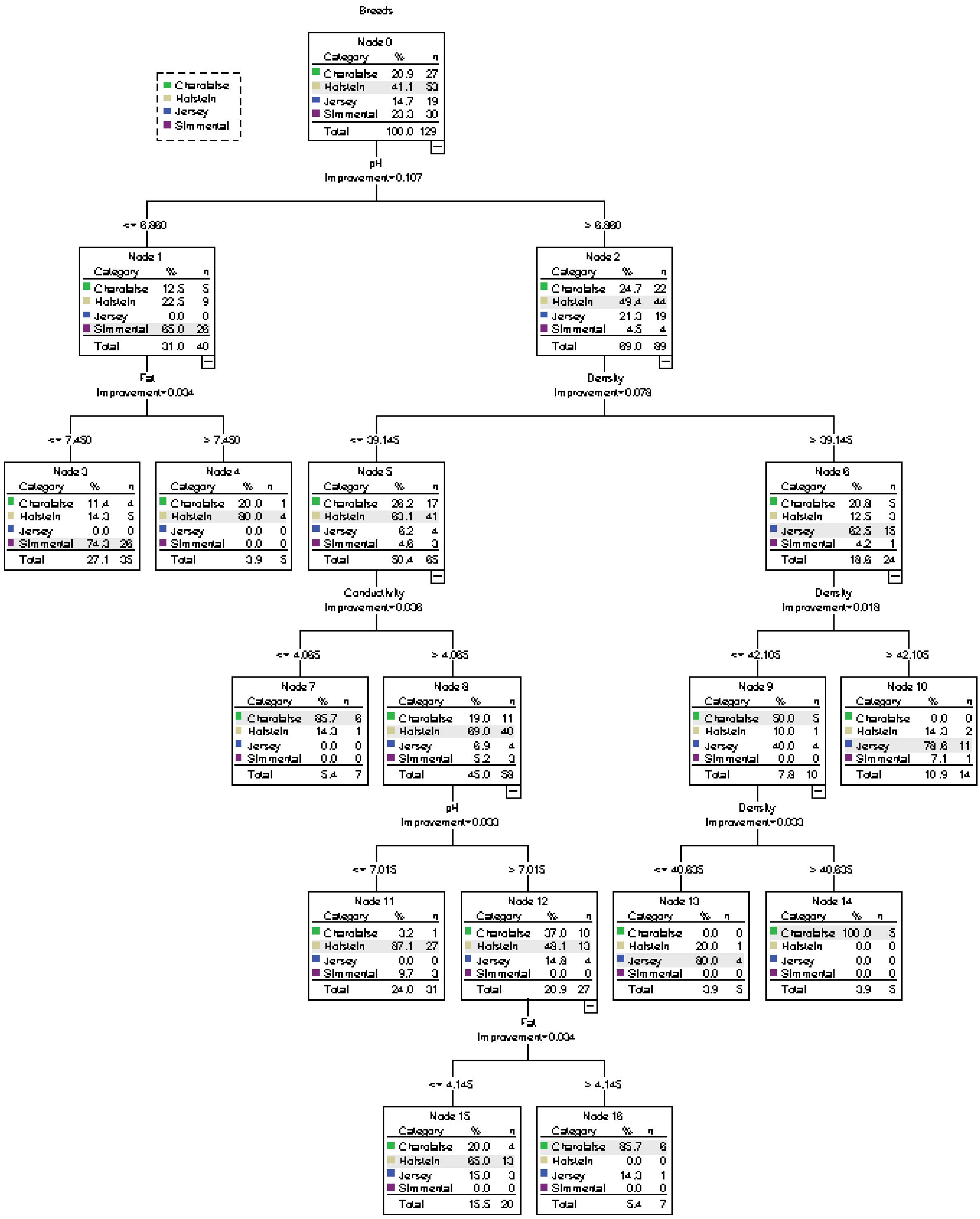

Figure 1. Classification tree diagram formed for four different dairy cattle according to milk components

Table 3. Correct classification rates of four different cattle breed according to milk components

Observed	Predicted				
	Charolaise	Holstein	Jersey	Simmental	Percent Correct
Charolaise	17	6	0	4	63.0%
Holstein	1	44	3	5	83.0%
Jersey	1	3	15	0	78.9%
Simmental	0	3	1	26	86.7%
Overall Percentage	14.7%	43.4%	14.7%	27.1%	79.1%

Correct classification rates of four different cattle breed are given in Table 3. As seen from the Table 3, 17 of 27 Charolaise cows were correctly classified (they took place in Charolaise group), while 10 heads of Charolaise cows took place in the wrong groups (16 in Holstein, 4 in Simmental). In Charolaise cows, correct classification rate has been 63.0%. 44 out of 53 heads of Holstein cows were correctly classified and correct classification rate has been 83.0%. 15 out of 19 heads of Jersey cows were correctly classified (took place in Jersey group), while 4 heads of Jersey cows took place in the wrong group (1 Charolaise, 3 Holstein). Any Jersey cow did not take place in Simmental group according to milk components. Correct classification rate in Simmental cows according to milk components actualized as 86.7%. Only 4 of 30 heads of Simmental cows took place in the wrong place (3 Holstein, 1 Jersey).

CONCLUSION and SUGGESTIONS

Although there are many statistical methods in classification of the data, since regression tree does not require any assumption on the variables taking place in the model, it is commonly used. In addition, in case that the available data in more number and complex structure, it is a strong statistical method that is easily applicable in classification analysis. When the diagram, formed by means of regression analysis, is examined, it can be easily observed which independent variables affect dependent variables. According to milk components, classification tree diagram, formed by means of CART decision tree for the discrimination of four different dairy cattle, is examined, it is seen that root node (Node 0) is divided into two sub-nodes as Node 1 (pH≤6.86) and Node 2 (pH>6.86) in terms of pH. In Node 1, where pH of raw milk is less than 6.86 or equals to that, any Jersey cow did not take place. However, it is seen that 26 of 30 heads of Simmental cows are in this node (Node 2). In Node 2, where pH is more than 6.86, that there are only 4 heads of Jersey, and 22 heads of Charolaise cows demonstrates that pH of raw milk is effective in classifying cows. In this study, when correct classification rates of four different cattle breeds are examined, it is seen that the lowest rate (63.0%) actualizes in Charolaise cows (86.7%) and the highest rate (86.7%) in Simmental cows. It is also seen that general correct classification rate (regardless breeds) is 79.1%. As a conclusion, using milk components, it can be said that different cattle breeds can be correctly classified in the rate that may be considered high.

REFERENCES

Akşahan, R., Keskin, İ. 2015. Determination of the some body measurements effecting fattening final live weight of cattle by the regression tree analysis. Selcuk Journal of Agriculture and Food Sciences, 2(1): 53-59.

Altay, Y., Boztepe, S., Eyduran, E., Keskin, İ., Tariq, M.M., Bukhari, F.A., Ali, I. 2021. Description of factors affecting wool fineness in karacabey merino sheep using chaid and mars algorithms. Pakistan J. Zool., 53: (2): 691-697.

Aytekin, I., Eyduran, E., Keskin, I. 2018. Detecting the relationship of california mastitis test (CMT) with electirical conductivity, composition and quality of the milk in Holstein-Friesian and Brown Swiss cattle breeds using CART analysis. Fresenius Environmental Bulletin, 27(6): 4559-4565.

Bakır, G., Keskin, S., Mirtagioğlu, H. 2009. Evaluating the relationship between mature age milk yield and several traits using CHAID analysis in Brown Swiss Cows. Journal of Animal and Veterinary Advances, 8 (3): 587-589.

Bakır, G., Keskin, S., Mirtagioğlu, H. 2010. Determination of the Effective Factors for 305 Days Milk Yield by Regression Tree (RT) Method, Journal of Animal and Veterinary Advances, 9(1): 55-59.

Çak, 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(4): 677-682.

Çamdeviren, H., Mendeş, M., Özkan, M.M., Toros, F., Şaşmaz, T., Öner, S. 2005. Determination of depression risk factors in children and adolescents by regression tree methodology. Acta Medica Okayama, 59(1): 19-26.

Doğan, İ. 2003. Investigation of the factors which are affecting the milk yield in holstein by CHAID analysis. Veterinary Journal of Ankara University, 50: 65-70.

Ergül, Ş., Ergül, A., Göncü, S. 2019. The effects of feeding strategies on milk yield and milk composition in dairy cattle. International Journal of Eastern Mediterranean Agricultural Research, 2(2): 145-165.

Eyduran E, Yilmaz I, Tariq M.M., Kaygisiz A. 2013. Estimation of 305-D milk yield using regression tree method in brown swiss cattle. The Journal of Animal & Plant Sciences, 23(3): 731-735.

Eyduran, E., Keskin, I., Erturk, Y.E., Dag, B., Tatliyer, A., Tirink, C., Aksahan, R., Tariq, M.M. 2016. Prediction of fleece weight from wool characteristics of sheep using regression tree method (Chaid Algorithm). Pakistan J. Zool., 48: (4): 957-960.

Karabacak, A., Senol Celik, S., Tatliyer, A., Keskin, I., Erturk, Y.E., Eyduran, E., Javed, Y., Tariq, M.M. 2017. Estimation of cold carcass weight and body weight from several body measurements in sheep through various data mining algorithms. Pakistan J. Zool., 49(5): 1731-1738.

Mendeş, M., Akkartal, E. 2009. Regression tree model in predicting slaughter weight of broiler. Italian Journal of Animal Science, 8: 615-624.

Mikail, N., Kaplan, M.Z. 2021. Effect of COVID-19 pandemic on animal-source food consumption in Turkey. ISPEC Journal of Agricultural Sciences, 5(3): 616–626.

Oruçoğlu, O. 2011. Holstein irkı ineklerin 305 günlük süt verimini etkileyen çevre faktörlerinin regresyon ağacı ile belirlenmesi. Master Thesis, Süleyman Demirel Üniversitesi, Fen Bilimleri Enstitüsü, Isparta.

Özek, K. 2015. Factors affecting composition of milk in dairy cattle and relation between nutrition and milk composition. Journal of Bahri Dagdas Animal Research, 4(2):37-45.

Özkan, K. 2012. Modelling ecological data using classification and regression tree technique (CART). SDU Faculty of Forestry Journal, 13: 1-4.

Tebliğ No: 2019/64. Çiğ inek sütünün sınıflandırılmasına ilişkin tebliğ. Resmi Gazete No: 31019. https://www.resmigazete.gov.tr/eskiler/2020/01/20200125-41.htm (Date of Access: 25.01.2020)

Temel, Ö.G. 2004. Sınıflama ve regresyon ağaçları. Master Thesis, Mersin Üniversitesi, Sağlık Bilimleri Enstitüsü, Mersin.

Topal, M., Aksakal, V., Bayram, B., Yaganoglu, M. 2010. An analysis of the factor affecting birth weight and actual milk yield in swedish red cattle using regression tree analysis. The Journal of Animal & Plant Sciences, 20: 63-69.

Yaylak, E., Alçiçek, A., Konca, Y., Uysal, H. 2007. Determination of variations on some nutrients and physical characteristics of milk collected in winter months by dairy in Izmir’s county. Journal of Animal Production, 48(1): 26-32.

Yücel, Y.B. 2017. The determination of factors that affect life satisfaction with classification and regression trees. İstanbul Commerce University, Graduate School of Applied and Natural Sciences, Master Thesis.