jotaf Journal of Tekirdag Agricultural Faculty
Tekirdağ Ziraat Fakültesi Dergisi

Eylül/September 2024, 21(4)
Başvuru/Received: 14/11/23
Kabul/Accepted: 27/07/24
DOI: 10.33462/jotaf.1390307

http://dergipark.gov.tr/jotaf
http://jotaf.nku.edu.tr/

ARAŞTIRMA MAKALESİ **RESEARCH ARTICLE**

A New Perspective on Türkiye's Sheep Population: Classification with Decision Trees

Türkiye Koyun Varlığına Yeni Bir Bakış: Karar Ağacı ile Sınıflama

Çağrı KANDEMİR[1], Çiğdem TAKMA[2], Turgay TAŞKIN[3]

Abstract

Decision trees are data mining algorithms that make interpreting the cause-effect relationship or classification between features with visual diagrams easy and do not require parametric assumptions. The aim of this study is to classify different sheep breeds raised in Turkey according to their origin and tail structure characteristics using a decision tree algorithm. It has been seen that the CART (Classification and Regression Trees) algorithm is sufficient for the classification of sheep breeds by obtaining low-risk values. According to the results of the study, it is seen that domestic sheep breeds are distributed to Eastern Anatolia, Mediterranean, Black Sea, and Central Anatolia Regions and imported breeds are distributed to all regions of Türkiye. The total population of these breeds was determined as 43.889.918 heads. As it can be understood, the Akkaraman breed ranks first with a rate of 40.88% of the total sheep stock. Next comes the Morkaraman breed with a rate of 11.68%. Thirdly, the most cultivated breed is Kıvırcık, whose rate is 8.82%. Merino sheep comes in fourth place with 8.43%. Awassi sheep are among the most common breeds in the fifth place. Hereby the distribution of sheep breeds in Türkiye has also been revealed and the distribution of breeds according to environmental conditions has been determined. The decision tree model generated using the CART algorithm for the distribution of sheep breeds in Turkey based on tail structures has been found to have a value (0.47). The proportion of sheep with a thin tail structure is 52.2%, while those with a semi-fatty tail structure are 17.5%, and those with a fatty tail structure are 30.2%. It has been determined that sheep breeds with thin tails are more commonly raised in Turkey. In the study, a decision tree model was also created using the CART algorithm to analyze the distribution of sheep populations in regions based on their origin, and it was found to have a value (0.26) with an accuracy rate of 74%. According to the research, 73.9% of sheep are domestic, while 26.1% are imported. When examining the distribution of origin by region, the region with the least number of imported sheep is Eastern Anatolia, while the highest proportions are found in the Mediterranean and Black Sea regions. The region with the lowest proportion of domestic sheep is the Mediterranean and Black Sea (70.4%), whereas the highest proportion is in the Eastern Anatolia region (82.6%).

Keywords: Sheep number, Classification, Decision tree, Sheep breeds

[1]***Sorumlu Yazar/Corresponding Author** Çağrı Kandemir, Ege University, Faculty of Agriculture, Department of Animal Science, İzmir, Turkey. E-mail: cagri.kandemir@ege.edu.tr OrcID: 0000-0001-7378-6962

[2]Çiğdem Takma, Ege University, Faculty of Agriculture, Department of Animal Science, İzmir, Turkey. E-mail: cigdem.takma@ege.edu.tr OrcID: 0000-0001-8561-8333

[3]Turgay Taşkın, Ege University, Faculty of Agriculture, Department of Animal Science, İzmir, Turkey. E-mail: turgay.taskin@ege.edu.tr OrcID: 0000-0001-8528-9760

Atıf: Kandemir, Ç., Takma, Ç., Taşkın, T. (2024). Türkiye koyun varlığına yeni bir bakış: Karar ağacı ile sınıflama. *Tekirdağ Ziraat Fakültesi Dergisi,* 21(4): 966-979.

Citation: Kandemir, Ç., Takma, Ç., Taşkın, T. (2024). A new perspective on Türkiye's sheep population: Classification with decision trees. *Journal of Tekirdag Agricultural Faculty,* 21(4): 966-979.

Öz

Karar ağaçları özellikler arasındaki neden-sonuç ilişkisini ya da sınıflandırmayı görsel diyagramlarla kolay yorumlanmasını sağlayan, parametrik varsayımlara gerek duymayan veri madenciliği algoritmalarıdır. Bu çalışmanın amacı, Türkiye'de yetiştirilen farklı koyun ırklarının orijin ve kuyruk yapısı özelliklerine göre karar ağacı algoritması kullanarak sınıflandırmaktır. Düşük risk değerleri elde edilerek, koyun ırklarının sınıflandırılmasında CART algoritmasının yeterli olduğu görülmüştür. Çalışma sonuçlarına göre, yerli koyun ırklarının Doğu Anadolu, Akdeniz, Karadeniz ve İç Anadolu Bölgelerine ve ithal ırkların ise Türkiye'nin tüm bölgelerine dağıldığı görülmektedir. Bu koyun ırkların toplam popülasyonu 43.889.918 baş olarak belirlenmiştir. Toplam koyun varlığı içerisinde Akkaraman ırkı %40,88 oranı ile ilk sırada yer almaktadır. Bunu %11.68 oranla Morkaraman ırkı takip etmektedir. Üçüncü sırada ise en çok yetiştirilen tür %8.82 ile Kıvırcık koyunudur. Merinos koyunu ise %8.43 ile dördüncü sırada yer almaktadır. İvesi koyunları ise en yaygın ırklar arasında beşinci sırada yer almaktadır. Böylelikle koyun ırklarının Türkiye'deki dağılımı da ortaya konmuştur ve çevresel şartlara göre ırkların dağılımı belirlenmiştir. Türkiye'deki koyun ırklarının kuyruk yapılarına göre dağılımına ilişkin CART algoritması ile oluşturulan karar ağacı modelinin risk değeri düşük (0.47) bulunmuştur. İnce kuyruk yapısına sahip koyunların oranı %52.2 iken, yarı yağlı kuyruk yapısına sahip koyunların oranı %17,5, yağlı kuyruk yapısına sahip koyunların oranı ise %30.2'dir. Türkiye' de ince kuyruklu koyun ırkları daha yaygın olarak yetiştirildiği belirlenmiştir. Çalışmada koyun varlığının kökenine göre göre bölgelere dağılımına yönelik CART algoritması ile oluşturulan karar ağacı modelinin risk değeri düşük bulunmuştur (0.26) ve doğru sınıflandırma oranı %74' dir. Araştırmada koyunların yüzde 73.9'u yerli, yüzde 26.1'i ise ithal koyunlardan olduğu tespit edilmiştir. Bölgelere göre köken dağılımı incelendiğinde ithal koyun sayısının en az olduğu bölge Doğu Anadolu'da, en yüksek oranda ise Akdeniz ve Karadeniz bölgeleridir. Yerli koyun oranının en düşük olduğu bölge Akdeniz ve Karadeniz (%70.4) olurken, en yüksek oranın ise %82.6 ile Doğu Anadolu bölgesi olduğu belirlenmiştir.

Anahtar Kelimeler: Koyun sayısı, Sınıflandırma, Karar ağacı, Koyun ırkları

1. Introduction

While the number of sheep and goats in Türkiye for the first period of 2022 is approximately 58.4 million, this number is 57.5 million in 2021, the number of sheep increased by 1.02% and the number of goats decreased by 0.99% (MFAL, 2023). The number of sheep milked is approximately 32 million heads, of which 19.8 million are sheep and 5.4 million are goats. According to the 2021 milk production data of the Turkish Statistical Institute, the total milk production is approximately 23.2 million tons, of which 2.1 million tons are obtained from sheep and goats. Of the total small ruminant milk obtained, 622.7 tons are sheep milk, 63.6 thousand tons are goat milk, 385.933 tons (19.77%) of the total red meat production (1.952.030 tons) are sheep meat, and 10,831 tons (5.54%) are goat meat (Sevinç et al., 2022; TÜİK, 2023).

The number of sheep and goats in Türkiye for the first period of 2022 was approximately 58.4 million, while the total number of sheep and goats in 2021 was 57.5 million heads, with a 1.02% increase in sheep and a 0.99% decrease in goats (MFAL, 2023). The number of sheep and goats milked is approximately 32 million heads, of which approximately 19.8 million are sheep and 5.4 million are goats. According to the milk production data of the Turkish Statistical Institute for 2021, total milk production was approximately 23.2 million tons, of which 2.1 million tons was obtained from sheep and goats. Of the total sheep and goats milk obtained, 622.7 tons of sheep milk and 63.6 thousand tons of goat milk, 385,933 tons (19.77%) of sheep meat and 10,831 tons (5.54%) of goat meat constitute the total red meat production (1.952.030 tons) (Sevinç et al., 2022; TÜİK, 2023).

Increases in ovine meat and milk production depend on the continuation of research on many issues such as animal breeding, management of the environment and natural resources, animal welfare, and social and cultural structure (Keskinkılıç, 2019; İzmir Commodity Excahnage, 2019; Yılmaz, 2019; Çiçek et al., 2022). It is important to have sufficient and accurate information about sheep breeding to meet the increasing need for meat and milk in the world in recent years and the sustainability of sheep breeding activities. Knowing how the number of ovine animals is dispersed across the geography of the country in terms of meat and milk production values affects the positive development of the factors affecting productivity (Cengiz et al., 2015; Özsayin and Everest, 2019; Kandemir et al., 2019; Cedden et al., 2020).

There are studies examining the effects of many factors on meat and milk production in sheep breeding. In the studies carried out, some yield characteristics and some structural characteristics were tried to be determined in sheep based on genotype or province (Özsayın and Everest, 2019; Kaymakçı and Taşkın, 2008; Sönmez et al., 2009; Kandemir et al., 2015; Semerci and Çelik, 2016; Tamer and Sarıözkan, 2017; Karakoç and Aygün, 2019; Kahraman et al., 2020; Demir and Aygün, 2021; Kırbaş et al., 2022; Hanoğlu et al., 2021). However, although there are studies that consider meat and milk production based on region and province, there are no studies that examine the genotype characteristics that directly affect these productions and the tail shape and distribution according to origins, which indirectly affect these productions, at the same time.

On the other hand, it is important to examine the effects of these factors with models using advanced software instead of classical approaches, in terms of making the results more reliable. Today, there are many new algorithms developed for processing data obtained from large information stacks. In this system known as data mining, artificial intelligence, and machine learning are especially valuable in the analysis of non-parametric data (Kononenko and Kukar, 2007). The decision tree method is a data mining algorithm that has been widely applied in various branches of science as a modern analytical technique (Abu-Hanna and De Keizer, 2003). This method has CHAID (Chi-square Automatic Interaction Detector), CART (Classification and Regression tree), QUEST (Fast Neutral Efficient Trees) and Exhaustive CHAID (Comprehensive Chi-square Automatic Interaction Detector) sub-algorithms for model analysis according to the research subject data structure. In the working principle of decision tree models, tree branches are formed like a tree structure with the answers taken from the basic simple questions of the independent variable. Tree branches formed in this way show which independent variable or variables affect the dependent variable (Temel, 2004; Sevgenler, 2019).

The classification and Regression tree method is a nonparametric analysis method that divides the data set into homogeneous subclasses and presents the relationship between dependent and independent variables visually (Kayri and Boysan, 2008; Akşahan and Keskin, 2015; Alev Çetin and Mikail, 2016). This method has some advantages such as not requiring any assumptions about the distribution of independent variables and not being negatively affected by multicollinearity, outliers, and missing observations (Mendeş and Akkartal, 2009).

This study aimed to classify the distribution of different sheep breeds in Türkiye in terms of origins and tail structures by considering the similarities and differences of the breeding areas, as well as to examine the current situation with the regression tree algorithm. It is thought that the findings obtained in the study will contribute to the creation of sheep breeding programs and sheep breeding maps to be made on a regional and provincial basis in Türkiye.

2. Material and Method

The material of the research consists of the data for the year 2021 obtained from the Livestock Information System (HAYBIS) belonging to the Republic of Turkish Ministry of Agriculture and Forestry (Anonymous, 2021a).

2.1. Study area

Seven geographical regions of Türkiye constitute the study area of the research. Four of these seven geographical regions are the name of the sea to which they are adjacent (Mediterranean Region, Black Sea Region, Aegean Region, Marmara Region), and the other three regions are named according to their positions in the whole of Anatolia (Central Anatolia Region, Eastern Anatolia Region, Southeastern Anatolia Region) (*Figure 1*). There are 81 provinces, 922 districts, 32175 neighborhoods, and 18292 villages in Türkiye. In this context, its population in 2022 is 84 million, and its surface area is 783,562 square kilometers (Yakar and Özgür, 2022).

Figure 1. Map of Türkiye by geographical regions

Sheep breeds, tail shapes (thin, fatted, semi-fatted), and their origins (imported or domestic) bred according to regions in Türkiye. The number of sheep in the provinces in each region determined in the Türkiye Domestic Genetic Resources Promotion Catalogue published by the Ministry of Agriculture and Forestry (Anonymous, 2021b) and the tail structure of these breeds (thin, semi-fatted and fatted) and origin (imported and domestic) classified according to their characteristics and descriptive statistics were given (Kalaycı, 2006; Alpar, 2013).

2.2. Classification and regression tree

In this study, classification and regression tree algorithms were applied according to the tail structure and origin characteristics of sheep breeds. A classification and regression tree is an iterative partitioning tree procedure that is constructed by dividing data into binary subgroups using continuous or discrete arguments. When the dependent variable is categorical, it is called classification tree (CT), and when it is continuous, it is called regression tree (RT). The first node of the classification and regression trees is the root node where the dependent variable is located and has not yet been split. First, this root node of the regression tree splits into two parent branches and these two parts are called parent branches. Nodes formed because of separation and division are also called subsets (Kalaycı, 2006).

The basic principle in the construction of the regression tree is to iteratively split the response variable into two child nodes to ensure maximum homogeneity in the response variable. During the construction of the tree, the program tests all explanatory variables included in the model and determines the cut-off value (category if the explanatory variable is categorical) of the explanatory variable to ensure the highest homogeneity in the resulting node (Akşahan and Keskin, 2015). In this process, if homogeneity is ensured as much as possible in any progeny node that is recursively created in the response variables, the fragmentation process ends in these nodes and these nodes are called terminal nodes (Oruçoğlu, 2011).

Splitting points of nodes are usually determined based on splitting criteria such as Gini or Twoing. Gini is a classification method based on binary divisions; after each division, the smallest Gini value is selected. The tree is continuously divided using the Gini index and eventually the optimum tree is determined by evaluating it with independent test data. The minimum number of observations is important when determining tree depth, and 10% of the data set is generally preferred. Gini index is calculated with the following formula, C represents the number of classes and pi represents the proportion of data points belonging to class i:

$$Gini\ \text{İ}ndex = 1 - \sum_{i=1}^{C} P_i^2 \qquad \text{(Eq. 1)}$$

A splitting criterion called twoing is another splitting criterion used in the CART algorithm. Twoing takes into account the distribution of variable values when dividing the data set between two subgroups and tries to minimize this distribution. When splitting a node into two child nodes, Twoing calculates the probabilities for each of these child nodes separately. It then creates branches based on these possibilities. The Twoing method is especially preferred when there is imbalance in the data set and other criteria such as the Gini index are insufficient. The Twoing formula is as follows:

$$\text{Twoing } = \sum_{j=1}^{J} \frac{n_j}{N} * \frac{1}{4}\left(1 - \frac{n_j}{N}\right)\left(1 - \frac{\sum_{k\neq j}^{J} n_k}{N}\right) \qquad \text{(Eq. 2)}$$

N represents the total number of samples, and nj and nk represent the number of samples in j and k subgroups, respectively. J refers to the total number of subgroups (Breiman, 1984). In this study, the IBM SPSS v25 program was used to classify the tail structure and origin characteristics of sheep breeds with a regression tree.

3. Results

While the region with the highest number of sheep breeds in Türkiye was the Mediterranean with 48, the region with the least sheep breed was the Eastern Anatolia Region with 23. However, considering the number of sheep in Türkiye (43.889.918), the Eastern Anatolia Region ranks first among the regions with 11.803.377 heads (26.89%). This in order; It is followed by Central Anatolia with 23.07% and Southeast Anatolia with 14.74%. When the average number of sheep according to the regions is examined, the lowest 68,536 head belong to the Black Sea Region. The region with the highest average sheep presence is Eastern Anatolia with 513.190 heads. In the study, the presence of sheep according to the regions is given in *Table 1*.

Table 1. Some descriptive statistics of sheep presence by regions

Region	Number of races	Total number of sheep	Share in Türkiye (%)	Number of sheep in the provinces in the region	
				The province with the least sheep	The province with the most sheep
Aegean	43	5.298.59	12.07	1.933.18	5.298.59
Mediterrenian	48	3.599.19	8.20	1.267.39	3.599.19
Marmara	45	4.332.82	9.87	2.160.63	4.332.82
Black Sea	33	2.261.71	5.15	1.071.76	2.261.71
Southeastern Anatolia	30	6.468.77	14.74	3.118.12	6.468.77
Eastern Anatolia	23	11.803.38	26.89	6.382.28	11.803.38
Central Anatolia	46	10.125.45	23.07	6.499.14	10.125.45
Türkiye	**53***	**43.889.92**	100.00		

* It is the total number of races existing in Turkey, the races in each region vary.

When the sheep breeds bred in Türkiye are examined, it is determined that there are 53 different sheep breeds in total. Having so much diversity in the same species is a great wealth for Türkiye. Such a change in environmental conditions with global warming makes the future unknown. Being able to benefit from the different characteristics of each sheep breed will create an advantage for people living in future Türkiye. The number of sheep breeds reared in Türkiye and their proportions are given in *Table 2*. The total population of these breeds was determined as 43.889.918 heads. As it can be understood, the Akkaraman breed ranks first with a rate of 40.88% of the total sheep stock. Next comes the Morkaraman breed with a rate of 11.68%. Thirdly, the most cultivated breed is Kıvırcık, whose rate is 8.82%. Merino sheep comes in fourth place with 8.43%. Awassi sheep are among the most common breeds in the fifth place. In this way, other sheep breeds are bred in different regions of Türkiye, respectively.

The distribution of sheep breeds according to tail structures in Türkiye is given in *Figure 2*. Accordingly, the risk value of the decision tree model created with the CART algorithm for the distribution of sheep according to their tail structures was found to be low (0.47) and the correct classification rate was 53%. As seen in *Figure 1*, while the rate of sheep with thin tail structures is 52.2%, the rate of sheep with semi-fat tail structures is 17.5% and the rate of sheep with fatted tail structures is 30.2%. Thin-tailed sheep breeds are more commonly bred in a significant part or region of our country.

In the distribution of the tail structure according to the regions, a similar structure is observed in the Eastern, South-eastern, and Central Anatolian regions and the Black Sea. When the tail structures are examined, the total ratio of the East, Southeast, Central Anatolia, and Black Sea regions is 49.3%, while the total ratio of the sheep in the Mediterranean, Aegean, and Marmara regions is 50.7% (*Figure 2*). As shown in the section denoted by the 1st Node (Node 1), the ratio of thin, semi-fatted, and fat-tailed sheep in the East, Southeast, Central Anatolia, and Black Sea Regions, respectively; 47.0%, 17.4%, and 35.6%. In the 2nd Node, where the Mediterranean, Aegean, and Marmara Regions are located, when the distribution of sheep presence according to tail structures is examined; thin-tailed sheep have a rate of 57.4%, semi-fat 17.6%, and fat-tailed sheep have a rate of 25.0%. This situation is as expected. In other words, while thin and semi-fat-tailed sheep are dominant or more common in the Western Anatolia Region, the proportion of fat and semi-fat-tailed sheep is relatively high in the Eastern, Southeastern, and Central Anatolian Regions.

Eastern Anatolia region differs from Southeast and Central Anatolia Region and Black Sea Region (*Figure 1*). While the rate of fat tail sheep in the Eastern Anatolia region is 43.5% (3rd Node), this rate is 33.9% in the aforementioned regions (4th Node). The rates are close to each other in the Mediterranean (5th Node) and Aegean and Marmara regions (6th Node) where thin-tailed sheep are concentrated. While the rate of thin-tailed sheep is 54.2% in the Mediterranean Region, this rate is 59.1% in the classification that includes the Aegean and Marmara regions (6th Node). However, it is noteworthy that the ratio of fat-tailed sheep (27.1% and 23.9%, respectively) is higher than that of semi-fat-tailed sheep in both the Mediterranean, Aegean, and Marmara regions.

Table 2. Sheep breeds raised in Türkiye.

Breed	Presence	Rate	Breed	Presence	Rate
Acıpayam	9.579	0.02	Kıvırcık (3)	3.870.648	8.82
Akkaraman (1)	17.940.185	40.88	Koçeri	54.608	0.12
Alman Siyah Bas	8.184	0.02	Lacaune	8.405	0.02
Anadolu Merinosu	331.402	0.76	Langhe	279	0.0005
Assaf	868	0.00	Malya	18.623	0.04
Bafra	28.931	0.07	Menemen	23.622	0.05
Bandırma	732	0.00	Merinos (4)	3.698.157	8.43
Bergamasca	19	0.00	Mor Karaman (2)	5.124.472	11.68
Çine Çapari	73.427	0.17	Norduz Koyunu	33.451	0.08
Dağlıç	128.831	0.29	Orta Ana Merinosu	120.394	0.27
Dogu Friz (Mars)	3.967	0.01	Ödemiş	1.250	0.003
Dorper	7.226	0.02	Pırlak	2.722.164	6.20
Esme	137.395	0.31	Plevne	2.393	0.01
Gökçeada	112.666	0.26	Polatlı	7.133	0.02
Hamdani	1.677.950	3.82	Ramlıç	53.241	0.12
Hasak	1.023	0.00	Romanov	98.890	.0.23
Hasmer	1.199	0.00	Sakız	1.506.754	3.43
Hemsin	237.350	0.54	Sarole	6.045	0.01
Herik	72.082	0.16	Sönmez	10.292	0.02
Ile de Frans	26.578	0.06	Suffolk	18.550	0.04
İvesi (5)	3.002.326	6.84	Tahirova	155.309	0.35
Kangal Akkaraman	568.288	1.29	Teksel	2.253	0.01
Karacabey Merinosu	368.275	0.84	Tuj	32.156	0.07
Karagül	8.771	0.02	Turcana	4.288	0.01
Karakas	7.076	0.02	Türkgeldi	281	0.0005
Karayaka	1.153.923	2.63	Zom	308.768	0.70
Karya	99.239	0.23	**TÜRKİYE**	**43.889.918**	**100.00**

In the study, the distribution of sheep existence by origin is given in *Figure 2*. Accordingly, the risk value of the decision tree model created with the CART algorithm for the distribution of sheep to regions according to their structures of origin was found to be low (0.26) and the correct classification rate was 74%. In the study, 73.9% of the sheep are domestic, while 26.1% are imported. When the distribution of origin by region is examined, Eastern Anatolia differs from other regions. The ratios of domestic and imported sheep in this region (Eastern Anatolia, 1st Node) are respectively; 82.6% and 17.4%. Whereas, in the group that includes the Mediterranean, Aegean, Southeast and Central Anatolia, and the Black Sea and Marmara Regions (2nd Node), respectively; 73.1% versus 26.9%. When the Mediterranean, Aegean, Southeast, and Central Anatolia and the Black Sea and Marmara regions (2nd Node) are examined within themselves, the ratio of domestic and imported sheep in the 3rd Node, where the Mediterranean and Black Sea is a group, is respectively; 70.4% and 29.6%. In the 4th Node group, which includes Aegean, Southeast, Central Anatolia and Marmara regions, the rates are respectively; 74.4% domestic and 25.6% imported. To summarize, the region with the lowest number of imported sheep in Eastern Anatolia, while the highest rate is in the Mediterranean and Black Sea regions. The region with the lowest rate of domestic sheep was the Mediterranean and Black Sea regions (70.4%), while the highest rate was the Eastern Anatolia region with 82.6%.

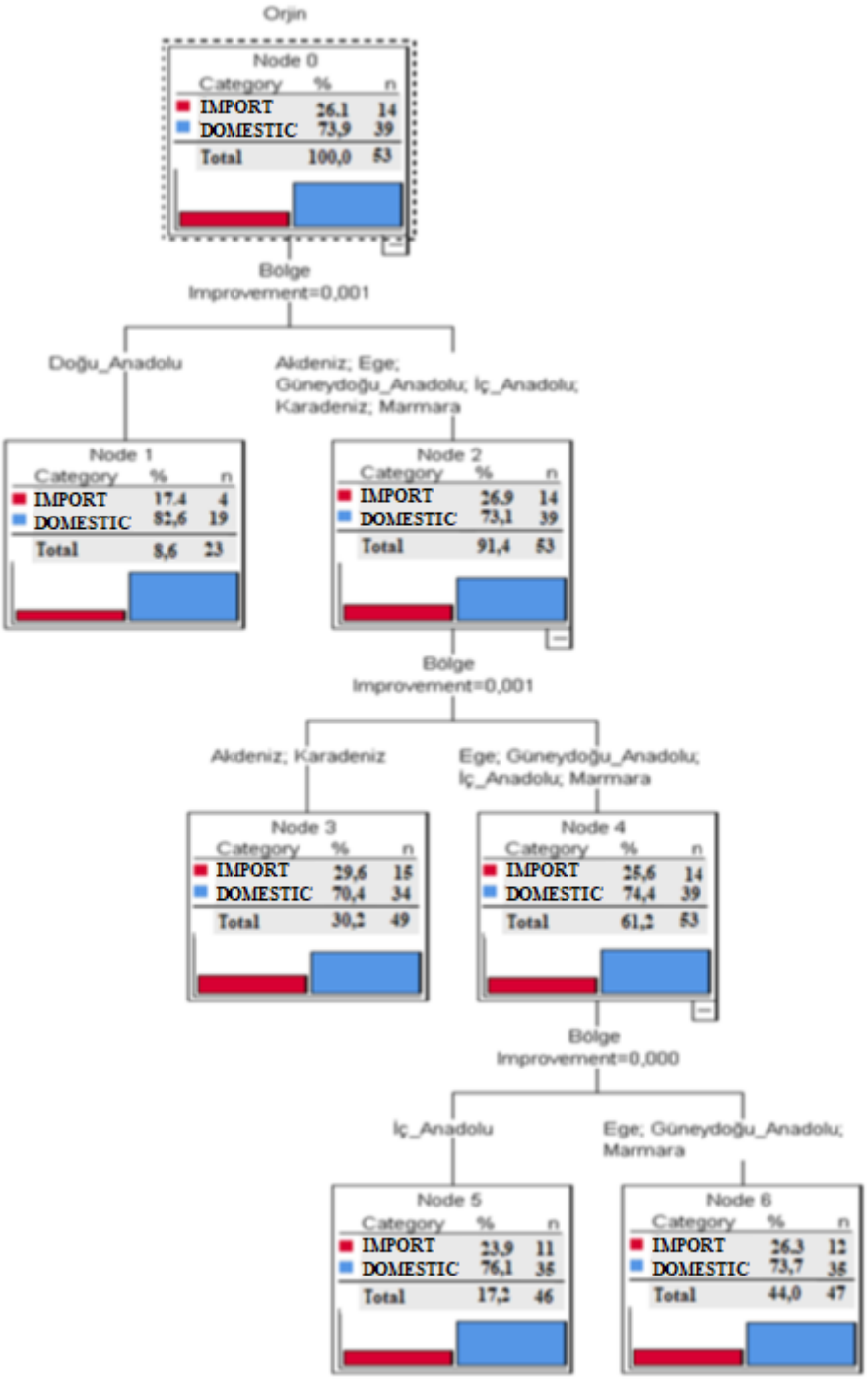

Figure 2. Distribution of sheep according to origine according to the decision tree method

4. Discussion

Regression trees are used to pre-select the features that affect the continuous dependent variable. Regression trees are algorithms that easily interpret the cause-effect relationship or classification between measurements in different structures of research with visual diagrams and do not require any parametric assumptions. On the other hand, although there are few studies on the use of decision trees in small cattle raised in Türkiye, different decision tree algorithms have been used in studies examining the effects of various factors on weaning weight, fleece weight, and birth weight. Olfaz et al. (2019) investigated the effects of gender, birth type, holding type, birth weight, and weight measurement time on weaning weight in Karayaka sheep with CART and CHAID decision trees.

In conclusion, according to the CHAID algorithm, the effect of measurement time, gender, and farm type on weaning weight was found to be significant. The results of the CART algorithm, on the other hand, determined

that the type of birth was effective on the weaning weight. (Eyduran et al., 2017). In addition, the CART method was applied to evaluate the effects of genotype, sex, birth type, year, and maternal age on birth weight in Eyduran et al. (2008), Karakaş, and Norduz lambs. While the effect of gender was insignificant in single-born lambs, the effect of genotype was significant in twin lambs. Balta and Topal (2018) examined the effects of year, flock type, maternal age, lamb gender, birth type, and lamb color on birth weight in Hemşin lambs by the CART method. The most important variables affecting birth weight are respectively, year, birth type, sex, herd type, and maternal age. In this study, the distribution and classification of different sheep breeds bred in Türkiye according to regions were examined according to the regression tree algorithm. Origin and tail structures were taken into consideration as classification criteria in sheep breeds. The low-risk values obtained according to the results of the study showed that the CART algorithm was sufficient in explaining the distribution of sheep breeds according to their tail and origin characteristics. This result was reported by Olfaz et al. (2019), Eyduran et al. (2008), and Balta and Topal (2018).

As a result, according to the CHAID algorithm, the effect of measurement time, sex, and holding type on weaning weight was found to be significant. The results of the CART algorithm showed that the type of birth was effective on weaning weight. Eyduran et al. (2017) estimated the fleece weight of Akkaraman and İvesi ewes from some physical properties of fleece with the help of CHAID analysis, which can be a useful method for determining high-yielding ewes. In addition, Eyduran et al. (2008) applied the CART method to evaluate the effects of genotype, sex, type of birth, year, and age of dam on birth weight in Karakaş and Norduz lambs. While the effect of sex was insignificant in single-born lambs, the effect of genotype was found to be significant in twin-born lambs. Balta and Topal (2018) examined the effects of year, flock type, dam age, lamb sex, birth type, and lamb color on the birth weight of Hemşin lambs using the CART method, and the most important variables affecting birth weight were found to be the year, birth type, sex, flock type, and dam age, respectively. In this study, the distribution and classification of different sheep breeds bred in Türkiye according to regions were analyzed by regression tree algorithm. Origin and tail structures were taken into consideration as classification criteria for sheep breeds. The low-risk values obtained according to the results of the study showed that the CART algorithm is sufficient to explain the distribution of sheep breeds according to tail and origin characteristics. This result is consistent with Olfaz et al. (2019), Eyduran et al. (2008), and Balta and Topal (2018) who applied the CART decision tree algorithm in sheep and goat breeds.

Feeding in sheep breeding in Türkiye is largely based on village common pastures. There are a total of 14.6 million meadow pasture areas in Türkiye, of which 1.45 million are meadows and 13.1 million are pastures. The Eastern Anatolia Region has the largest meadow and pasture area, with a rate of 37.54%. The Central Anatolia Region (29.68%) takes second place. Other regions, respectively; Black Sea Region (11.97%), Southeastern Anatolia Region (6.90%), Aegean Region (5.49%), Mediterranean Region (4.51%), and Marmara Region (3.90%) come from (Kuşvuran et al. 2011; Çaçan and Yüksel, 2016). In the research, the surface measurements of the meadow and pasture areas owned by the regions in Türkiye and the presence of sheep belonging to the regions are directly proportional and in harmony. The difference between the animal existence in Eastern Anatolia and Central Anatolia Regions and other regions is statistically significant. It is understood that these two regions are the locomotive of sheep breeding in Türkiye.

Although the decrease in agricultural production values in recent years has shown that the sector has contracted (Sezenler, 2013; Karaman, 2018), the high input costs that have occurred have had a significant impact on the sustainability of animal husbandry (Güler and Saner, 2021). Accordingly, Türkiye's imports of live animals and carcasses have also increased. In Türkiye, after 2010, especially to reduce meat prices, the import of livestock, mostly for fattening and slaughtering, has accelerated. Between 2010 and 2019, approximately 8 million heads of live sheep/cattle animals with a total value of 7.08 billion dollars were imported. Between 2010 and 2019, 296.487 tons of boned/boneless/carcass meat with a total value of 1.406 billion dollars were imported (Anonymous, 2020). Grants, loans, and support given to the livestock sector in the last fifteen years in Türkiye; it is seen that some provinces and most large-scale enterprises in these provinces benefit. It is aimed to increase the meat and milk productivity of large enterprises and to supply the animal products that Türkiye needs domestically thus preventing the increase in product prices (Oğul, 2022).

The research obtained it by crossing 24 domestic breeds, 16 imported breeds from 53 sheep breeds in Türkiye, and domestic breeds with imported breeds from the past to the present. It was determined that 13 native breeds

registered by the Ministry of Agriculture and Forestry were created (Anonymous, 2021a). Sheep breeding is a production branch that can be done in almost all regions of Türkiye. This branch of production is mostly made with local breeds and the feeding of these breeds is done by adhering to pasture (Semerci and Çelik, 2016; Acıbuca and Bostan Budak, 2018). In other words, the production system of sheep is predominantly based on the extensive production model. In Türkiye, state-supported small cattle enterprises have been established in many provinces and these enterprises continue to produce. Since these enterprises think that sheep breeding with domestic breeds is not very profitable, they approach the issue with the idea that it would be more appropriate to start with cultural breeds or crosses. As the breeding was done, there were animal sales from newly established enterprises from imported breeds to surrounding enterprises, and in this way, the dispersal of imported breeds to that region was realized. However, due to different climatic difficulties, it is analyzed that imported breeds could not become widespread in some regions due to climatic difficulties and remained limited to locally imported farms.

In the research, the Eastern Anatolia Region is the first to differentiate due to the high rate of domestic sheep breeds. The proportion of imported breeds in this region is quite low. The result of the study was determined by the regional study Kandemir and Taşkın (2022a). It has been determined that there are 23 breeds in total in the Eastern Anatolia Region, of which 19 breeds are domestic and 4 breeds are imported. This is because; It is thought that because of the very cold and long winters in the Eastern Anatolia Region and the very limited vegetation, it is thought that the races coming from outside cannot adapt to the difficult conditions. Imported breeds grown; It is thought that sheep breeds such as Charolais, Ile de France, and Romanov can be raised by intensive feeding in cold winter periods. The Black Sea Region is the second region to differentiate. Since the livestock enterprises in the region know that sheep breeding with local breeds is not very profitable, they act with the thought that it would be more appropriate to start with culture breeds or crosses. Karayaka sheep constitute half of the sheep presence in the region due to their adaptation to high rainfall and humidity (Kandemir and Taşkın, 2022b). The fact that imported breeds are not preferred over domestic breeds in intensive production in the Black Sea Region, which receives high rainfall, causes a differentiation. About 5 per thousand of the 10 million sheep assets in the Central Anatolia Region consist of imported breeds (Taşkın and Kandemir, 2022a). This is thought to be one of the reasons for the variation in sheep breeds bred in the region.

In sheep, the tail structure is a morphological feature and is a transmitted breed character that is transferred from generation to generation. It is accepted as a constant measure in the classification of sheep breeds (Taşkın et al., 2015). Tail structures in Türkiye are registered as fat tail, the half fat tail, and thin tail, and the distinctions are made accordingly (Anonymous, 2021b). In the research, the fat-tailed sheep breeds, which started from the Eastern Anatolia Region, which had the harshest winter, left their place to the thin-tailed sheep breeds in the milder regions. Eastern, South-eastern, Central Anatolian, and Black Sea Regions are in harmony with the regional studies that are predominantly oil-tailed (Kandemir and Taşkın, 2022a, Kandemir and Taşkın, 2022c, Kandemir and Taşkın, 2022d). Thin-tailed sheep breeds are mainly grown in the Marmara, Mediterranean, and Aegean regions. Similarly, this finding is in harmony with the regional sheep breeding studies (Kandemir and Taşkın, 2022c; Kandemir and Taşkın, 2022e, Taşkın and Kandemir, 2022a; Taşkın and Kandemir, 2022b).

On the other hand, decision trees are considered a powerful tool for data analysis and creating predictive models in terms of their advantages. As a matter of fact, since decision trees have a simple structure, it is quite easy to interpret and explain the results. The tree structure increases understandability as it helps visualize the decision process (Quinlan, 1985). Since decision trees can work with categorical and numerical data, they are used effectively in data sets where different types of data coexist (Breiman et al., 1984). Decision trees have fast training and prediction times even on large data sets. This feature provides a significant advantage in real-time applications and big data analysis (Genuer, 2010). In addition, decision trees can be used directly in data sets, which reduces data preprocessing steps (Hastie et al., 2009).

Besides the mentioned advantages, decision trees also have disadvantages. For this reason, some issues should be carefully evaluated when using it in small cattle breeding. Because decision trees tend to overfit the training data set, this leads to the drawback of reducing generalization ability (Domingos, 1997). Additionally, decision trees can be quite sensitive to small changes in the input data (Hastie, 2009). In unbalanced data, decision trees can focus on these classes and ignore others, considering that larger classes have more samples. Thus, while a high accuracy is achieved in the majority class, the small class may misclassify. Due to these drawbacks, when using decision trees, data quality should be increased with data pre-processing steps, tree depth should be limited and care should be taken to select the correct node division criteria for the nodes in the trees.

5. Conclusion

In this study, which shows the distribution of sheep breeds according to geographical regions in Türkiye, the distribution of sheep according to their origin and tail structure is shown with a statistical method, a decision tree. It has been determined that there is a wide diversity in the geographical regions of Türkiye depending on different environmental conditions. While the presence of fat-tailed breeds is high in regions with cold climatic environments, it is seen that tail structures become thinner in milder temperate regions. Sheep breeds that are not in different continents and Türkiye have been brought to Türkiye due to the importation that has been going on for many years. Although the Eastern Anatolia, Mediterranean, Black Sea, and Central Anatolia Regions differ by evaluating the sheep according to their origins in geographical regions, imported sheep breeds are bred in all regions of Türkiye, although predominantly local breeds.

In the accompaniment of this study, some suggestions that we think may be useful for the future of sheep breeding are given. These recommendations are as follows; In Türkiye, there are studies carried out to protect the diversity of local gene resources and to ensure their sustainability. However, further studies at the molecular level will contribute to the determination of genetic diversity as well as the level of inbreeding, the creation of gene resource protection programs and the development of breeding strategies, especially for the development of sheep breeding. To produce high-quality roughage, which is also an important problem in Türkiye's sheep breeding, it should be ensured that the forage crops culture in field agriculture is developed as well as the existing pasture areas. Finally, it should be ensured that shepherd or animal care should change the perspective of young people, especially in rural areas, towards sheep breeders, who do not have social security or a profession, to be respected in society. Instead of temporary solutions, necessary legal arrangements including the technical and economic dimension of the problem should be made as soon as possible. Otherwise, the number of people who leave or leave this production branch in many regions of our country where sheep breeding is carried out will increase gradually.

Acknowledgement

We would like to thank the Izmir Provincial Director and Staff of the T.C. Ministry of Agriculture and Forestry, who helped carry out the study and share the current figures.

Ethical Statement

There is no need to obtain permission from the ethics committee for this study.

Conflicts of Interest

We declare that there is no conflict of interest between us as the article authors.

Authorship Contribution Statement

Concept: ÇK, ÇT, TT; Design: ÇK; Data Collection or Processing: ÇK, TT; Statistical Analyses: ÇT; Literature Search: TT; Writing, Review and Editing: ÇK, ÇT, TT.

References

Abu-Hanna, A. and De Keizer, N. (2003). Integrating classification trees with local logistic regression in Intensive Care prognosis. *Artificial Intelligence in Medicine*, 29(1-2): 5-23.

Acıbuca, V. and Bostan Budak, D. (2018). Place and Importance of Medicinal and Aromatic Plants in the World and Türkiye. *Çukurova Tarım ve Gıda Bilimleri Dergisi*. 33(1): 37- 44 (In Turkish).

Akşahan, R. and Keskin, İ. (2015). Determination of the somebody measurements effecting fattening final live weight of cattle by the regression tree analysis. *Selçuk Tarım Bilimleri Dergisi*, 2(1): 53-59.

Alev Çetin, F. and Mikail N. (2016). Data Mining Aplications in Livestock. *Turkish Journal of Agricultural Research,* 3(1): 79-88 (In Turkish).

Alpar, R. (2013). Applied Multivariate Statistical Methods. (4th Edition) Detay Publising, Ankara (In Turkish).

Anonymous (2020). Meat and Milk Institution 2019 Sector Evaluation Report. https://www.esk.gov.tr/upload/Node/10255/files/2019_Yili_Sektor_Degerlendirme_Raporu.pdf (Accessed Date: 18.06.2020)

Anonymous (2021a). T.C Ministry of Agriculture and Forestry Animal Information System, (https://haybis.tarimorman.gov.tr/IsletmeGiris.aspx (Accessed Date: 02.01.2021)

Anonymous, (2021b). Türkiye Pet Genetic Resources Promotional Catalog. T.C. Ministry of Agriculture and Forestry https://www.tarimorman.gov.tr/TAGEM/Belgeler/yayin/Katalog%20T%C3%BCrk%C3%A7e.pdf (Accessed Date: 02.03.2021)

Balta, B. and Topal, M. (2018). Regression tree approach for assessing the effects of non-genetic factors on birth weight of Hemşin lamb. *Alinteri Journal of Agriculture Science,* 33(1): 65–7.

Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J. (1984). Classification and Regression Trees. Wadsworth International Group.

Çaçan, E. and Yüksel, A. (2016). Influence of Meadows and Pastures on Regional Development. *UNIDAP International Regional Development Conference*, Muş, Türkiye. (In Turkish).

Cedden, F., Cemal, İ., Daşkıran, İ., Esenbuğa, N., Gül, S., Kandemi,r Ç., Karaca, O., Keskin, M., Koluman, N., Koyuncu, M., Savaş, T., Taşkın, T., Tölü, C., Ulutaş, Z., Yılmaz, O. and Yurtman, İ.Y. (2020). Current Situation and Future in Türkiye's Sheep Breeding. *Turkish Agricultural Engineering IX. Technical Congress,* 13-17 Ocak 2020, Ankara, s. 133-152 (In Turkish).

Cengiz F., Karaca S., Kor A., Ertuğrul M., Arık İ.Z. and Gökdal Ö. (2015). Changes and New Quests in Small Livestock Breeding. *Turkish Agricultural Engineering VIII. Technical Congress*, 12-16 Ocak 2015, TMMOB Chamber of Agricultural Engineering, C:2, 809-837, Ankara, 2015 (In Turkish).

Çiçek, A., Ayyıldız, M., Erdal, G. and Erdal, H. (2022). Importance and economic analysis of sheep breeding in Türkiye. *MAS Journal of Applied Sciences (Special İssue)*: 1303–1322 (In Turkish).

Demir, E. and Aygün, T. (2021). Characteristics of reproduction, milk and fleece yield of Hırik (Hamdani x Akkaraman crossbred) sheep raised in rural conditions. *Journal of Animal Production*, 62 (1), 35-44 (In Turkish).

Domingos, P. (1997). Why Does Bagging Work? A Bayesian Account and Its Implications. *Proceedings of The Second International Conference On Knowledge Discovery and Data Mining*, 155-165.

Eyduran, E., Karakus, K., Keskin, S. and Cengiz, F. (2008). Determination of factors influencing birth weight using regression tree method, *Journal of Applied Animal Research* 34: 109-112.

Eyduran, E., Zaborski, D., Waheed, A., Celik, S., Karadas, K., Grzesiak, W. (2017). Comparison of the predictive capabilities of several data mining algorithms and multiple linear regression in the prediction of body weight by means of body measurements in the indigenous Beetal goat of Pakistan. *Pakistan Journal of. Zoology* 49: 257-265.

Genuer, R., Poggi, J. M. and Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern recognition letters*, 31(14): 2225-2236.

Güler, D. and Saner, G. (2021). The measurement of efficiency of dairy farms: The cases of Izmir and Manisa. *Yuzuncu Yıl University Journal of Agricultural Sciences,* 30(2): 386-397. (In Turkish).

Hanoğlu, Oral.H., Kuz, H.İ., Dayanıklı, C., Önaldı, A.T., Alarslan, E. and Duman, E. (2021). Structural characteristics of extensive small ruminant farms and transition opportunities to organic animal husbandry: The Case of Balıkesir province, Türkiye. *Turkish Journal of Agricultural Research,* 8(3): 320-330. (In Turkish).

Hastie, T., Tibshirani, R. and Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Science & Business Media.

İzmir Commodity Exchange. (2019). Future of Sheep Farming in Türkiye and Sustainable Production Project. Workshop Final Report, 31 January 2019, İzmir (In Turkish).

Kahraman, M. and Yüceer Özkul. B. (2020). Investigation of milk fatty acid profiles in some domestic and hybrid sheep genotypes. *Harran University Journal of the Faculty of Veterinary Medicine,* 9(2): 126-132. (In Turkish).

Kalaycı, Ş. (2006). SPSS Applied Multivariate Statistics Techniques. (2th Edition), Asil Publishing Distribution, Ankara. (In Turkish).

Kandemir, Ç. and Taşkın, T. (2022a). Current situation of sheep breeds raised in the Eastern Anatolia Region *Journal of Animal Production*, 63 (1): 57-65 (In Turkish).

Kandemir, Ç. and Taşkın, T. (2022b). Current situation and future of sheep breeds: Mediterranean Region. *Turkish Journal of Agricultural Research* 9(1): 97-106 (In Turkish).

Kandemir, Ç. and Taşkın, T. (2022c). Current state and future of sheep breeds in Türkiye: Blacksea Region. *Van Yuzuncu Yil University Institute of Natural and Applied Sciences.* 27(1):101-112 (In Turkish).

Kandemir, Ç. and Taşkın, T. (2022d). Status of sheep breeds by regions in Türkiye: Central Anatolia Region. *International Journal of Anatolia Agricultural Engineering Sciences,* 4(4): 97-105 (In Turkish).

Kandemir, Ç. and Taşkın, T. (2022e). Current situation of native and cultured sheep breeds in the Marmara Region. *Voice of Nature,* 5(9): 17-33 (In Turkish).

Kandemir, Ç., Adanacıoğlu, H., Taşkın, T. and Koşum, N. (2019). Sheep and sheep meat prices of comparison analysis of scaling. *Journal of Tekirdag Agricultural Faculty,* 16(3), 315-327 (In Turkish).

Kandemir, Ç., Alkan, İ., Yılmaz, H.İ., Ünal, H.B., Taşkın, T., Koşum, N. and Alçiçek, A. (2015). General situation and development opportunities to the geographical locations of small ruminant farms in İzmir Region. *Journal of Animal Production*, 56(1):1-17 (In Turkish).

Karakoç, T. and Aygün, T. (2019). The live weight after shearing and the greasy wool yield of Zom ewes at different raising conditions in Türkiye. *Journal of Advanced Agricultural Technologies*, 6(4): 267-271.

Karaman, S. (2018). Comparative analysis of turkey's crop and livestock productıon values at regional level using a panel data index. *Yuzuncu Yıl University Journal of Agricultural Sciences.* 28(2):168-174 (In Turkish).

Kaymakçı, M. and Taşkın, T. (2008). Sheep crossbreeding studies in Türkiye. *Journal of Animal Production,* 49 (2), 43-51 (In Turkish).

Kayri, M. and Boysan M. (2008). Assesment of relation between cognitive vulnerability and depression's level by using classification and regression tree analysis *Harran University Journal of Education*, 34: 168-177 (In Turkish).

Keskinkılıç, K. (2019). Sustainability of Sheep Farming Activity. Izmir Commodity Exchange Publications No:99, İzmir. p:195 (In Turkish).

Kırbaş, M., Bülbül, B. and Kal, Y. (2022). Some reproductive traits in Central Anatolian Merino sheep under breeder conditions. *Journal of Advances in Veterinary Bio Science and Techniques*, 7(1), 14-18 (In Turkish).

Kononenko, I. and Kukar, M. (2007). Machine Learning and Data Mining: Introduction to Principles and Algorithms. Horwood Publishing.

Kuşvuran, A., Nazlı, R.İ. and Tansı, V. (2011). Current situation of meadow-rangelands, animal existence and cultivation for forage crops in Türkiye and East Black Sea Region. *Journal of Agricultural Faculty of Gaziosmanpaşa University,* 28(2), 21-32 (In Turkish).

Mendeş, M and Akkartal E. (2009). Regression tree analysis for predicting slaughter weight in broilers. *Italian Journal of Animal Science*, 8(4):615-624.

MFAL. (2023). Minestery of Food and Agricultutal and Livestock. (Accessed Date:02.03.2021)

Oğul, B. (2022). The relationship of agricultural subsidies and agricultural production: empirical findings on the Turkish economy. *The Journal of Agricultural Economics Researches*, 8(1): 44-56 (In Turkish).

Olfaz, M., Tırınk, C. and Önder, H. (2019). Use of CART and CHAID algorithms in Karayaka sheep breeding. *Kafkas Universitesi Veteriner Fakültesi Dergisi*, 25(1): 105–110.

Oruçoğlu, O. (2011). *Determination of environmental factors affecting 305-day milk yield of Holstein cows by regression tree method*. (MSc Thesis) Süleyman Demirel University, Institute of Science and Technology, Isparta, Türkiye. (In Turkish)

Özsayın, D. and Everest, B. (2019). Socio-economic structure of farmers that make sheep breeding and practices related to their sheep breeding activities. *Journal of Agriculture and Nature*, 22 (Supplementary Issue 2): 440-448. (In Turkish)

Quinlan, J. R. (1985). Decision Trees and Multi-Valued Attributes. In J. E. Hayes and D. Michie (Eds.), Machine intelligence 11. Oxford University Press.

Semerci, A. and Çelik, A.D. (2016). General overview of ovine breeding in Türkiye. *Journal of Agricultural Faculty of Mustafa Kemal University,* 21(2): 182-196. (In Turkish)

Sevgenler H. (2019). *Comparison of data mining algorithms (Cart, Chaid and Mars) used to determine the effects of some characteristics of goats on live weight*. (MSc Thesis) Iğdır University, Institute of Science, Department of Animal Science, Iğdır, Türkiye. (In Turkish).

Sevinç G., Şahin Z. and Aydoğdu M.H. (2022). The trend analysis of the developments of ovine presence and its milk production in Türkiye. *Journal of Academic Social Resources*, 7(35): 377-38. (In Turkish)

Sezenler T., Soysal, D., Yildirir, M., Yüksel, M. A., Ceyhan, A., Yaman Y., Erdoğan İ. and Karadağ O. (2013). Influence of some environmental factors on litter size and lamb growth performance in Karacabey Merino sheep. *Journal of Tekirdağ Agricultural Faculty,*10(1): 40-47. (In Turkish)

Sönmez, R., Kaymakçı, M., Eliçin, A., Tuncel, E., Wassmuth, R. and Taşkın, T. (2009). Improvement Studies in Türkiye Sheep Husbandry. *Türkiye Sheep Breeding Congress,* 12-13 Şubat 2009, İzmir, Türkiye. (In Turkish).

Tamer, B. and Sarıözkan, S. (2017). Socio-economic structure and production costs of sheep farms in central district of Yozgat province. *Journal of The Faculty of Veterinary Medicine Erciyes University*, 14(1): 39-47. (In Turkish).

Taşkın, T. and Kandemir, Ç. (2022a). Current situation and future of sheep breeds: Aegean Region. *Journal of Agriculture Faculty of Ege University*, 59 (3):485-498. (In Turkish).

Taşkın, T. and Kandemir, Ç. (2022b). Current state and future of sheep breeds in Türkiye: Southeast Anatolia Region. *Journal of Kadirli Faculty of Applied Sciences,* 2(1): 93-105. (In Turkish).

Taşkın, T., Ünal, H. B. and Canbolat, Ö. (2015). Basic Principles of Sheep Breeding. Hasad Publishing. Ltd. Şti, Ümraniye-İstanbul (In Turkish).

Temel, Ö. G. (2004). *Classification and regression trees*. (MSc Thesis) Mersin University, Health Information Institute, Department of Biostatistics, Mersin, Türkiye. (In Turkish).

Türkiye İstatistik Kurumu (2023). Livestock data, https://biruni.tuik.gov.tr/medas/?kn=101&locale=tr (Accessed Date: 09.10.2023)

Yakar, M. and Özgür, E. M. (2022). The aging population in Türkiye is ringing alarm bells at the local level. *Journal of Geography,* (44): 231-250 (In Turkish).

Yılmaz, Ş. G. (2019). *Socio-economic analysis and efficiency of small livestock enterprises: the example of the Western Mediterranean Region*. (PhD Thesis) Isparta University of Applied Sciences Institute of Science, Isparta, Türkiye. (In Turkish).