MJAVL
Manas Journal of Agriculture Veterinary and Life Sciences
ISSN 1694-7932 | e-ISSN 1694-7932
Volume 12 (Issue II) (2022) Pages 115-121 https://doi.org/10.53518/mjavl.1198225

Investigation of Egg External Quality Characteristics of Linda Geese with Data Mining Methods

Harun YONAR[1a], Emre ARSLAN[2b*], Kemal KIRIKÇI[2c]

[1]Department of Biostatistics, Faculty of Veterinary Medicine, Selcuk University, Konya, TÜRKIYE
[2]Department of Animal Science, Faculty of Veterinary Medicine, Selcuk University, Konya, TÜRKIYE

[a]*https://orcid.org/0000-0003-1574-3993* , [b]*https://orcid.org/0000-0002-4609-8395*
[c]*https://orcid.org/0000-0002-6649-1127*

*Corresponding author: emre.arslan@selcuk.edu.tr

ABSTRACT

In this study, the external quality characteristics of eggs belonging to Linda accidents, a poultry species, were investigated with a data mining approach. The 288 Linda goose eggs used in the study were 36 weeks old; Their width was 52.34 mm, egg length was 76.8 mm, egg weight was 120.43 g, and shape index value was 68.26. Eggs were clustered according to shape index and weight measurements using the K-means clustering algorithm, a data mining approach. Statistically significant differences were found between the clusters in width, height, shape index, and weight ($p<0.05$). The findings of our study showed that eggs with a low index had high- weight; eggs with a high index were low-weight eggs. According to the results of this study, it was concluded that the shape index value might be related to egg weight. More detailed inferences can be made using data mining algorithms for different poultry species.

ARTICLE INFO

Research article
Received: 02.11.2022
Accepted: 09.12.2022

Keywords:
Linda, goose, data mining method, egg, quality

INTRODUCTION

It is accepted that geese are one of the first domesticated poultry species after the chicken (Kozák 2019). Although it seems to be a developed sector in some regions of Turkey, it is widely produced and consumed in Eastern Anatolia, especially in the provinces of Kars and Ardahan and its surroundings (Wilson and Yilmaz 2013, Saatci et al. 2021). Geese are generally divided into three classes according to their size. To light goose breeds; Chinese and Roman geese, medium-sized goose breeds; Pilgrim, Heavy bodied goose breeds; Examples are the Embden and Linda geese. Linda geese were obtained by combining the domestic geese of Russia and the goose breeds such as Adler, Gorky and Solnechnogorks, especially the Chinese goose. It is reported that the registration of this goose breed took place in 1994, which was recent. In addition, the annual egg production of Linda geese is low compared to other laying geese, and it is reported that the egg weight varies between 119.6-192.3 g (Sari et al. 2019; Karabulut 2021a; Saatci et al. 2021). Linda geese have been reported to be one of the preferred hybrids in Turkey in the last 10 years (Kaya and Yurtseven 2021). There is a limited number of studies on Linda geese in geese raised in Turkey (Sarı et al. 2019, Karabulut 2021a). Egg weight depend on age (Eroğlu and Yetisir 2022; Kucharska-Gaca et al. 2022), breed (Karabulut 2021a, Akın and Çelen 2022a), management methods (Saatci et al. 2021) and laying period factors (Kucharska-Gaca et al. 2022) Since egg weight affects hatching performance and chick quality, it is one of the most important egg quality characteristics for the hatching sector (Uçar et al. 2022).

Cluster Analysis

Clustering methods are defined as unsupervised learning methods. They are methods that can make evaluations without needing a control variable to determine the densities of the distributions of similar observations in terms of the examined characteristics. While the observations within the clusters formed are similar, there are differences in observations between clusters (Altunkaynak 2022). Cluster analysis is divided into hierarchical and non-hierarchical

To Cite: Yonar H, Arslan E, Kırıkçı K. 2022. Investigation of Egg External Quality Characteristics of Linda Geese With Data Mining Methods, MJAVL Sciences. 12 (2) 115-121

methods. In hierarchical clustering analysis, the number of clusters is not certain, and it is the examination of the data by combining or separating according to clustering around the determined centers. Non-hierarchical clustering analysis is based on the principle that densities suitable for the specified number of clusters form clusters.

K-means Clustering Method

The k-means clustering method is a non-hierarchical clustering method that can be applied to large and complex data structures. The number of clusters is determined by the experiments of the researcher and the approaches used to determine the optimal number of clusters, and each observation belongs to only one cluster.

Steps of the K-means clustering algorithm; determining the number of clusters, choosing k centers randomly, assigning each observation to the nearest k center by calculating its distance from the k center, calculating the new center for the formed clusters, choosing the cluster structure of the observations in which the sum of the within-cluster error squares is minimum after the determined iteration is selected as the final clustering (Altunkaynak 2022).

Metrics used in calculating distances in the K-means clustering algorithm; ıt can be listed as Euclidean distance, Manhattan distance, Pearson correlation distance, Spearman correlation distance, and Chebychev distance (Abdulhafedh 2021). The most commonly used measure among these is the Euclidean distance, and according to this measure, the i. observation and the j. The formula for the distance between the observations is given in the equation (Altunkaynak 2022, Lantz 2019).

$$d_{ij} = \sqrt{\sum_{m=1}^{p} (X_{im} - X_{jm})^2}$$

In K-means clustering analysis, the determination of the number of clusters and the determination of the starting point cause the results to vary. Therefore, different cluster structures should be handled comparatively, and business information to be segmented should be prioritized. The optimum number of clusters is determined by comparing the Total Error Sum of Squares values calculated for each clustering study created according to different cluster values. In the literature, different hierarchical and non-hierarchical algorithms can be alternatives to K-means clustering methods. The fact that these algorithms are handled with different scenarios according to business information and determined purposes affects the quality of the knowledge levels of the clusters.

In this study, the variation of egg weight according to shape index in Linda goose eggs was investigated with the k-means clustering method, which is a non-hierarchical clustering algorithm. In the comparisons made for the cluster structures, there were differences between the clusters in terms of external quality characteristics. In particular, it was concluded that the eggs in the cluster with a high shape index had low weight.

MATERIALS AND METHOD

Material

The material of this study consists of 288 fresh eggs obtained from a 36-week-old Linda goose flock of the same breeder reared with traditional methods in the Konya region. Laying goose feed (17% HP) and water were given ad libitum.

Methods

Each egg was numbered, and egg weights were determined with a balance with a precision of 0.01 g. The distance between the blunt and sharp end of the eggs is the length of the egg; The width of the equatorial region was evaluated as egg width and measured with a digital caliper with 0.01 mm precision (Karabulut 2021b). The shape index value, one of the egg's external quality characteristics, was determined by the formula "*Shape index = (Egg length / Egg width) x 100*".

Statistical analysis

The descriptive statistics of the quantitative variables used in the study were reported with mean, standard deviation, minimum, maximum, and interquartile range. Cluster structures were created with the K-means clustering algorithm, one of the clustering algorithms. The Kolmogorov-Smirnov test was used to examine normality. Welch's F test was used when the homogeneity of the variances was not met to determine the mean differences between the clusters and the differences between the clusters Games-Howell Post-Hoc test was performed the as the multiple comparisons. IBM SPSS 22 and R studio (version 4.1) were used for the study's statistical analysis. A p-value less than 0,05 was statistically significant.

RESULTS

A summary of the descriptive statistics of eggs for width, height, weight, and Shape Index measurements were given in Table 1.

Table 1. *Descriptive statistics of eggs for width, height, weight, and Shape Index measurements*

Parameter	Mean	Standard Deviation	Minimum	Maximum	Interquartile Range (IQR)
Egg Width	52,34	1,49	47,47	56,24	1,74
Egg Length	76,8	3,61	66,99	89,47	4,07
Egg Weight	120,43	10,13	92,86	147,67	12,36
Shape Index	68,26	3,02	58,21	75,3	3,61

No outlier observations were detected when examining the shape index and weights of the observations together. In Figure 1, the distribution of measurements can be seen in the histogram graphs of the frequency of the variable separately. It has been seen that the standard deviation of the weight variable was higher than the other variables, and standardized data were used to perform the clustering.

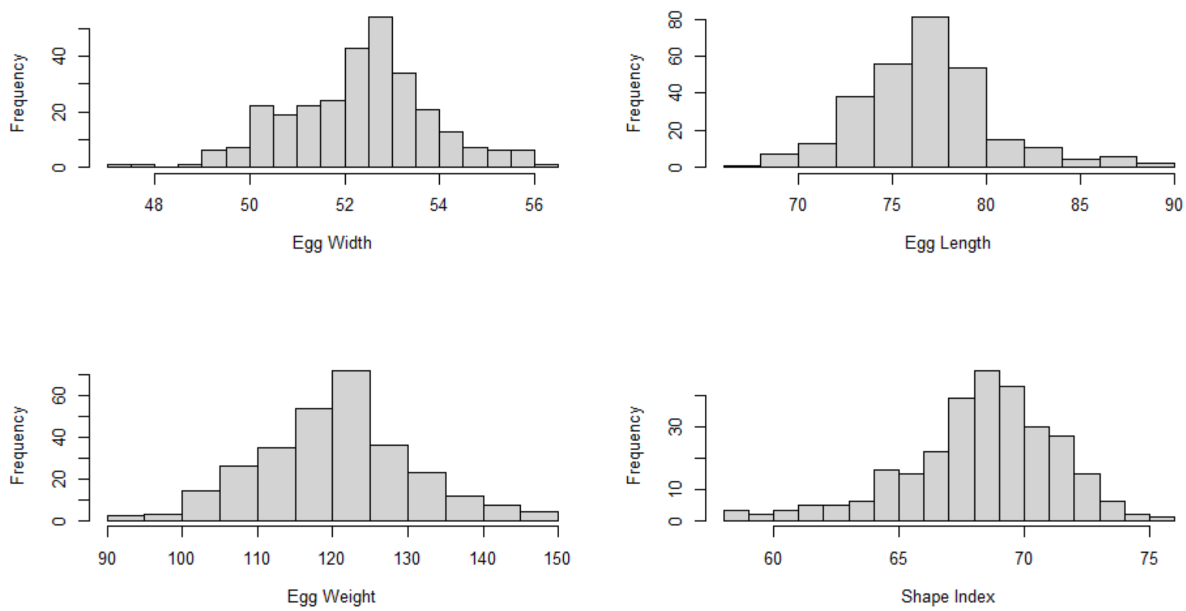

Figure 1: Histogram plots of some external quality characteristics of eggs

K-means cluster structures were formed using weight and shape index measurements of 288 Linda goose eggs. The optimal number of clusters was determined using mathematical metrics (Elbow and Silhouettes methods) (Kassambara 2017, Rençber 2019). The results of the two approaches used to determine the optimal cluster number were given in Figure 2. The optimal number of clusters has been determined as four according to the silhouette width in Figure 2(a). In Figure 2(b), the within-cluster sums of squares value gradually decreases after four, so it was the appropriate number of clusters.

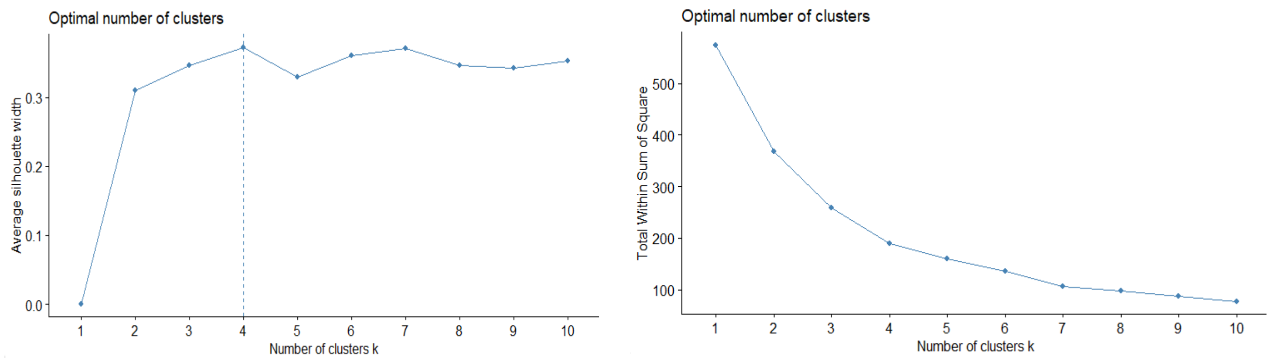

Figure 2. Determination of the optimal number of clusters using the Elbow and Silhouette methods

It was examined using the weight and shape index of 288 eggs to determine the optimal number of clusters. The number of observations in the clusters consists of 76, 55, 131, and 26 observations, respectively, and the cluster structures were given in Figure 3.

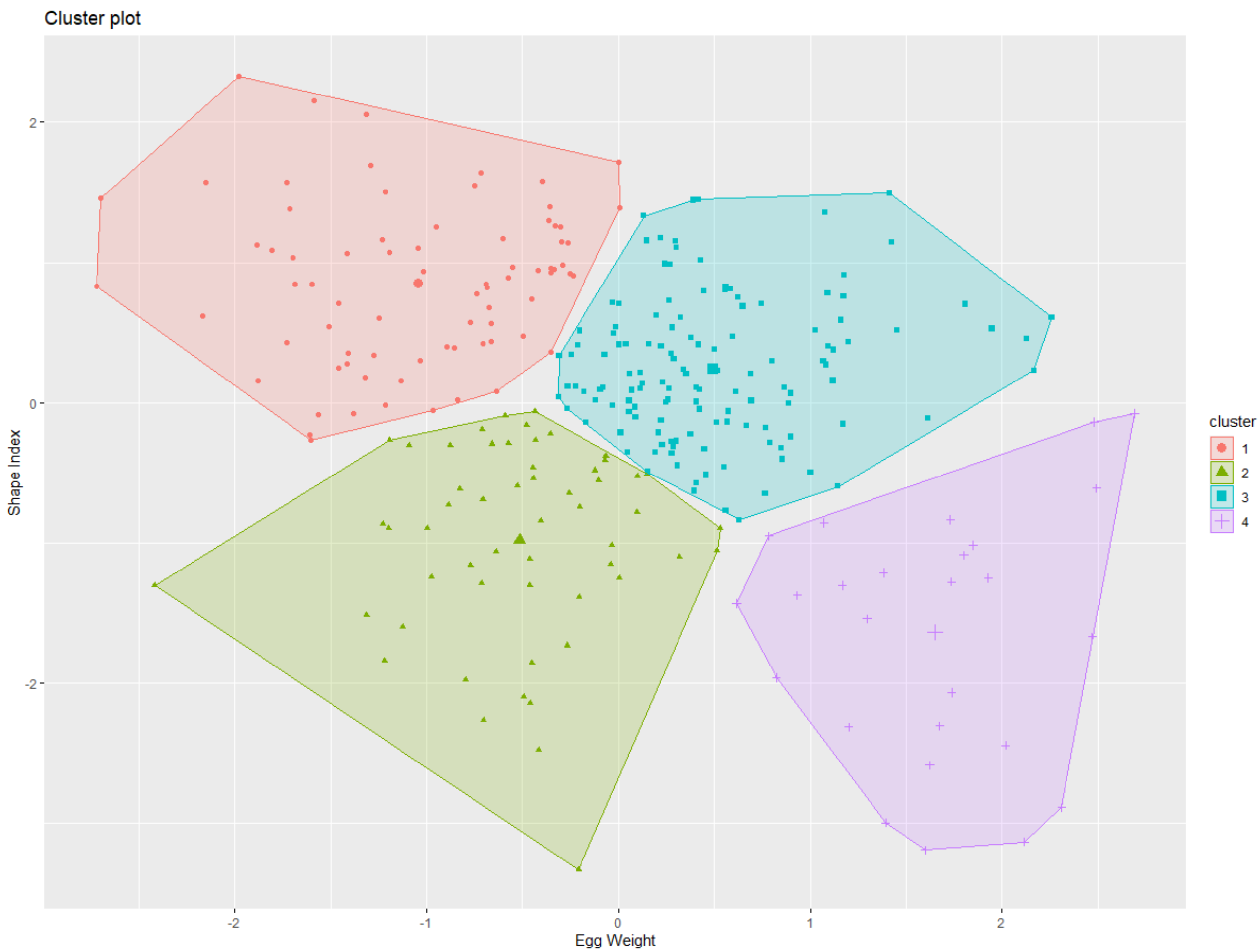

Figure 3. Clusters resulted by fitting K-means clustering based on the eggs' weight and Shape Index measures

When the cluster structures were examined, the weights of the observations in the third and fourth clusters were high, while those in the first and second clusters consisted of lower-weighted observations. It was seen that the eggs with high weight consist of observations with a higher shape index than those in the fourth cluster. When the first and second groups, which have low weight, were examined, it was observed that those in the second cluster had high shape index measurements. Regarding both dimensions, the observations were concentrated in a certain area for the second and fourth clusters. The examination of weight and shape index measurements was carried out through standardized observations; It was determined that the cluster with a high weight and shape index had the most observations (131); although they had a high weight, it was determined that the cluster consisting of a low shape index had the least number of (26) observations.

The cluster structures' width, height, weight, and shape index measurements were examined, and statistically significant differences were found between the clusters ($p<0.05$). In the mean comparisons made for each cluster, there is no difference between the third and fourth clusters for width measures; no statistically significant difference was detected for height measures in the second and third clusters ($p>0.05$). The statistics of the differences between the means were given in Table 2.

Table 2. *ANOVA(Welch's F) test results for clusters of egg external quality traits measures*

Parameter	Cluster	N	Mean	Std. Deviation	p	Post-Hoc; p
Egg Width	1	76	51,48	1,17	<0,001*[a]	1-2; 0,012 *[b] 1-3; <0,001*[b] 1-4;<0,001 *[b] 2-3; <0,001*[b] 2-4; <0,001*[b] 3-4 ;>0,05*[b]
	2	55	50,87	1,08		
	3	131	53,27	0,98		
	4	26	53,27	1,29		
Egg Length	1	76	72,71	1,85	<0,001*[a]	1-2; <0,001*[b] 1-3; <0,001*[b] 1-4; <0,001*[b] 2-4; <0,001*[b] 2-3;>0,05*[b] 3-4; <0,001*[b]
	2	55	77,93	1,97		
	3	131	77,24	1,51		
	4	26	84,22	2,73		
Egg Weight	1	76	109,87	6,28	<0,001*[a]	1-2;<0,001*[b] 1-3; <0,001*[b] 1-4; <0,001*[b] 2-3; <0,001*[b] 2-4; <0,001*[b] 3-4; <0,001*[b]
	2	55	115,21	5,16		
	3	131	125,43	5,34		
	4	26	137,14	5,79		
Shape Index	1	76	70,84	1,72	<0,001*[a]	1-2;<0,001*[b] 1-3; <0,001*[b] 1-4; <0,001*[b] 2-3; <0,001*[b] 2-4; 0,010*[b] 3-4; <0,001*[b]
	2	55	65,31	2,08		
	3	131	68,98	1,52		
	4	26	63,32	2,68		

*p<0,05; *[a] (Welch's F test); *[b] Games-Howell Post-Hoc test

DISCUSSION

It is an essential source of animal protein, with many amino acids and critical nutritional contents for the continuity of generations of egg poultry species (Sun et al. 2019). In various studies on goose eggs (Tilki and İnal 2004a; Mazanowski et al. 2005; Tilki et al. 2005; Chang et al. 2016; Kumbar et al. 2016; Alaşahan et al. 2019; Sarı et al. 2019; Karabulut 2021a); Akın and Çelen 2022b; Kucharska-Gaca et al. 2022) shape index value was reported to vary between 65-70.98 . In this study, the egg shape index value was determined as 68.26 and it is compatible with the literature.

Egg weights in geese Alaşahan et al. (2019) 124.11g; In white Chinese geese, Sun et al. (2019) 139.37 g; In Linda goose, Sarı et al. (2019) calculated the average egg weight value was 122.09, and it was found to be similar to the value in this study. On the other hand, Karabulut (2021a) reported the average egg weight value of the same breed as 192.3 g in his research. The average egg weight determined in this study was lower than what Karabulut (2021a) results. Since the egg weight increases with the increase in age in geese (Tilki and İnal 2004b, Biesiada-Drzazga 2016), it can be said that this difference occurs due to the use of eggs from animals of different ages.

Karabulut (2021a) measured Linda geese's egg length and width values as 5.32-6.13 cm and 7.89-9.73 cm, respectively. Chang et al. (2016) reported the same variables as 56.9-57.3 mm and 86.6-87.7 mm in white roman geese, respectively. When compared with these values, egg width was determined as 52.34 mm and egg length as 76.80 mm in the present study. According to this, in terms of egg width, Karabulut (2021a) and Alaşahan et al (2019) were lower than that Chang et al. (2016). The fact that egg width and length values were not similar may be due to genotype and age differences.

CONCLUSIONS

In this study, cluster analysis, a data mining method, was used to segment Linda's geese's eggs according to external quality characteristics. Cluster structures were created based on the use of shape indexes and weights of eggs. The cluster with heavy eggs' shape index values was low; low-weight eggs were those with a high shape index value. In addition to the low shape index of heavy eggs, it was also determined that the eggs were longer than those in other clusters.

Eggs with a low shape index are pointed shape; It was determined that eggs with a higher index were more oval-shaped eggs. It was predicted that the shape index value could be related to the egg weight. In addition, it was concluded that it could be used for egg weight estimation with data mining methods for different poultry species. In this sense, modeling studies can be done using different data mining algorithms and predictive variables.

ETHICAL STATEMENT

Selcuk University Experimental Research and Application Center, Animal Experiments Ethics Committee 2022/112 Number Ethics Committee Decision

CONFLICT OF INTERESTS

The authors did not report any conflict of interest. In addition, they have done this work without any financial support.

AUTHORS CONTRIBUTION

Motivation / Concept: Harun YONAR/ Emre ARSLAN
Design: Harun YONAR/ Emre ARSLAN
Control/Supervision: Kemal KIRIKÇI / Harun YONAR
Data Collection and / or Processing: Emre ARSLAN
Analysis and / or Interpretation: Harun YONAR
Literature Review: Emre ARSLAN/Harun YONAR/ Kemal KIRIKÇI
Writing the Article: Emre ARSLAN/Harun YONAR/ Kemal KIRIKÇI
Critical Review: Kemal KIRIKÇI/Harun YONAR

REFERENCES

Abdulhafedh A 2021. Incorporating k-means, hierarchical clustering and pca in customer segmentation. J City Dev. 3(1): 12-30.

Akın Y, Çelen MF 2022a. Comparisons of egg quality traits, egg weight loss and hatching results between Chinese and Mamut geese eggs reared under the countryside conditions. Iran J Appl Anim Sci. 12(1): 175-182.

Akin Y, Çelen MF 2022b. Slaughter, carcass and egg traits of domestic geese raised in the Aegean region of Turkey. S Afr J Anim Sci. 52(2): 214-222.

Alaşahan S, Garip M, Çağlayan T, Ateş CT 2019. Halk elinde yetiştirilen kaz, ördek ve hindi yumurtalarının bazı dış kalite özelliklerinin incelenmesi. Harran Üniv Vet Fak Derg. 8(1): 21-25.

Altunkaynak B 2022. Veri Madenciliği Yöntemleri ve R Uygulamaları, Genişletilmiş 3. Baskı. Seçkin Yayıncılık, Ankara.

Biesiada-Drzazga B, Banaszewska D, Charuta A, Koncerewicz A 2016. Influence of age on egg characteristics and reproduction features of Koluda® White geese. Europ Poult Sci. 80.

Chang SC, Lin MJ, Fan YK, Lee TT 2016. Effects of lighting intensity on growth and reproductive performance of breeder geese. J Appl Poult Res. 25(3): 315-321.

Eroglu M, Erisir Z 2022. The effects of breeder age and laying period on hatchability and some external egg quality characteristics in breeder geese. Anim Sci Pap Rep. 40(1): 89-98.

Karabulut O 2021a. Estimation of the external quality characteristics of goose eggs of known breadth and length. Vet Med. 66(10): 440-447.

Karabulut O 2021b. A Study on the Effect of Altitude on Shell Thickness in Goose Eggs. Indian Journal of Animal Research. 55(9): 1116-1120.

Kassambara, A. Practical guide to cluster analysis in R: unsupervised machine learning 1st edn (CreateSpace, North Charleston, 2017).

Kaya Z, Yurtseven S 2021. Effects of feeding systems on yield and performance of Linda geese (Anserinae sp.) in hot climatic conditions. S Afr J Anim Sci, 51(5): 628-634.

Kozák J 2019. Variations of geese under domestication. World's Poult Sci J. 75(2): 247-260.

Kucharska-Gaca J, Adamski M, Biesek J 2022. The age of the geese from the parent flock and the laying period affect the features of the eggs. Poult Sci, 101(10): 1-9.

Kumbar V, Nedomova S, Trnka J, Buchar J, Pytel R 2016. Effect of storage duration on the rheological properties of goose liquid egg products and eggshell membranes. Poult Sci, 95(7): 1693-1701.

Lantz B 2019. Machine Learning with R: Expert Techniques for Predictive Modeling. Packt Publishing Ltd.

Mazanowski A, Kisiel T, Adamski M 2005. Evaluation of some regional varieties of geese for reproductive traits, egg structure and egg chemical composition. Ann Anim Sci. 5: 67-83.

Rençber ÖF 2019. Veri madenciliğinde kullanılan kümeleme algoritmalarının karşılaştırılması üzerine bir inceleme. Düzce Üniversitesi Bilim ve Teknoloji Dergisi. 7(3): 1671-1685.

Sari M, Buğdaycı KE, Akbaş AA, Saatci M, Oğuz MN 2019. The effect of laying period on egg quality traits and chemical composition of Lindovskaya (Linda) geese reared under breeder conditions. Turk J Vet Anim Sci. 43(5): 662-669.

Saatcı M, Tilki M, Sarı M, Şahan Yapıcer Ö 2021. Her yönüyle kaz yetiştiriciliği, 2.baskı, Kutlu&Avcı Ofset Yayınevi, Antalya.

Sun C, Liu J, Yang N, Xu G 2019. Egg quality and egg albumen property of domestic chicken, duck, goose, turkey, quail, and pigeon. Poult Sci. 98(10): 4516-4521.

Tilki M, İnal Ş 2004a. Yield traits of geese of different origins reared in turkey I. Hatching traits. Turk J Vet Anim Sci. 28: 149-155.

Tilki M, İnal Ş 2004b. Quality traits of goose eggs: 1. Effects of goose age and storage time of eggs. Arch für Geflügelk. 68(4): 182-186.

Tilki M, Saatci M, Kirmizibayrak T, Aksoy AR 2005. Effect of age on growth and carcass composition of Native Turkish Geese. Arch für Geflügelk. 69: 77-83.

Uçar A, Boz MA, Erensoy K, Sarıca M 2022. The Effect of Hatching System and Egg Weight on Hatching Traits in Turkish Geese: Hatch time, Hatchability and Gosling Quality Traits. TURJAF. 10(4): 686-692.

Wilson RT, Yilmaz O 2013. The Domestic Livestock Resources of Turkey: Populations, production and pathology of ducks and geese. Int J Poult Sci. 12(9): 553-560.