Animal Review
2017 Vol. 4, No. 3, pp. 45-52
ISSN(e): 2409-6490
ISSN(p): 2412-3382
DOI: 10.18488/journal.ar.2017.43.45.52

FORECASTING GOAT MILK PRODUCTION IN TURKEY USING ARTIFICIAL NEURAL NETWORKS AND BOX-JENKINS MODELS

Ferhan KAYGISIZ[1+]
Funda Hatice SEZGİN[2]

[1]*Istanbul University, Faculty of Veterinary Medicine, Department of Animal Breeding and Husbandry, Avcılar, Istanbul, Turkey*
Email: ferhan_64@yahoo.com Tel: +90 2124737070-17210
[2]*Istanbul University, Faculty of Engineering, Department of Industrial Engineering, Avcılar, Istanbul, Turkey*
Email: hfundasezgin@yahoo.com Tel: +90 2124737070-17472

(+ Corresponding author)

Article History
Received: 9 October 2017
Revised: 20 December 2017
Accepted: 26 December 2017
Published: 29 December 2017

Keywords
Artificial neural networks model
Box - Jenkins model
Forecasting
Goat milk.
Time series
Milk production

ABSTRACT

The demand for goat milk has gradually increased in Turkey in recent years and dairy goat breeding began to be seen as an alternative investment area. The aim of the study is to create the data that will contribute to policy formulation in the stockbreeding industry by making a 10-year forecast of output pertaining to the goat milk production in Turkey. In the study, the annual data of the goat milk production in Turkey during the time period 1961 and 2016 obtained from Turkish Statistical Institute and Food and Agriculture Organization was utilized. Box-Jenkins estimation models and artificial neural networks model were used to forecast the production of goat milk. It was identified that artificial neural networks model gave the best result and prospective estimations were made through this model. As a result of the study, the projected value of milk production for 2026 was found to be 495,536.1 tons. Following the forecasts, it was calculated that the average rate of increase in the goat milk production will be 0.12%.

Contribution/Originality: The study was conducted to estimate the amount of goat milk production in Turkey until 2026. Box-Jenkins and artificial neural networks models were used as prediction models. The results of this study contribute in the literature about the goat breeding production policies.

1. INTRODUCTION

Goat breeding is a traditional animal production branch usually practiced in underdeveloped and developing countries. Aforementioned production branch constitutes a significant source of living and nutrition for the families with low income living in rural and forested lands [1]. Goat breeding has made a significant progress in the world since the late 20th century in consequence of the development in agriculture, consumer demands related to social and economic indicators, the need for quality and healthy food products [2, 3]. Goat breeding is maintained as either in an agricultural enterprise or as village herds, transhumance or migratory herds. However, intensive businesses making cheese or providing milk for cheese making dairies have also been operating in Western Anatolia in recent years [4]. The amount of goat milk production which was 192,210 tons in Turkey in 2009 increased to 479,401 tons in 2016 by rising 2,5 times and the share of goat milk in the total milk production in Turkey was 2.59% [5].

In Turkey, goat milk and products have economically gained importance by means of their taste, aroma, and quality, and goat cheese has begun to be demanded more and more with the urban growth and the development of tourism [1]. Goat plays a significant part in the domestic and foreign trades as well as contributing to the nurturing of people, textile industry, and employment. Especially, the fact that EU (European Union) has not reached sufficiency in sheep-goat products creates a potential market for Turkey [1]. Therefore, it is important to promote goat breeding by implementing economic policies.

The significance of consistent estimations for creating realistic prospective policies regarding livestock sector is a fact. While forming consistent estimations, a sound data system and time series analysis methods modeling this data are necessary [6].

Researchers used Box-Jenkins [7-10] models successfully in analyses made with time series in agricultural output. It is observed that artificial intelligence methods have an increasing usage rate in recent years. Around the world and in Turkey, there is research that artificial neural networks method was used in the studies of estimating milk yield in livestock [11-13]. However, a limited number of studies [14] using artificial neural networks method in the estimation analyses of the amount of agricultural production were found in Turkey. This study was conducted to guide the policies which will be implemented in the field of livestock production by estimating the amount of goat milk production for the period of 2016-2026.

2. MATERIAL AND METHODS

2.1. Material

The material of the study consists of the goat milk production data between 1961 and 2016. The statistical data used in the study were obtained from TSI (Turkish Statistical Institute) and FAO (Food and Agriculture Organization) [5, 15].

2.2. Statistical Analysis

In this study, future forecasts for goat milk production data were made and their compatibility was assessed by ARMA (Autoregressive Moving Average) which is one of the Box-Jenkins models frequently used in time series analyses and applied to level data, and ARIMA (Autoregressive Integrated Moving Average) methods applied to differentiated data along with ANN (Artificial neural networks models).

Box-Jenkins which is an analysis and estimation method in time series is based on discrete, linear stochastic processes. AR (Autoregressive), MA (Moving Average), ARMA and ARIMA are Box-Jenkins estimation models. While AR (p), MA (q) and their combination ARMA (p, q) models are applied to stationary processes, ARIMA (p, d, q) models are applied to non-stationary processes [16]. ARMA (p, q) models are the most general stationary stochastic process models and are a linear function of past observations and past error terms [17]. ARMA (p, q) models are usually as demonstrated below:

$$Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \cdots + \Phi_p Y_{t-p} + \delta + \alpha_t + \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} \quad (1)$$

In the equation above;

Y_{t-p} : is the past observation values,

$\Phi_1, \Phi_2, \ldots, \Phi_p$: is the coefficient for past observation values,

δ : is the constant value,

a_t, a_{t-1}, a_{t-p} : is the error term,

$\theta_1, \theta_2, \ldots, \theta_q$: is the coefficients regarding error terms [16].

In situations where time series is stationary, namely in situations where the average, variance, and covariance of the process is not changing depending on time and therefore series does not fluctuate, ARMA (p, q) or AR (p) or MA (q) models which are particular states of ARMA (p, q) can be used. However, the fluctuation of the average and variance of time series depending on time is frequently observed in reality. This situation is referred to as the non-stationary situation. This kind of time series can only be appropriate for the use of ARMA (p, q) models when they are made stationary.

The process of stabilizing can be done with differencing processes and differencing the time series which has a linear trend at first degree makes the time series stationary. However, the time series which demonstrate curvilinear trend can be stationary with the process of second differencing [16-18]. The models applied to the new series resulting from the process of stabilizing the non-stationary series by differencing is called "non-stationary linear stochastic models" or "integrated models". In this situation, the model is denoted as ARIMA (p, d, q). Where "d" is the stabilizing parameter of the series [19]. Namely, it states in which degree the series was differenced.

ANN was developed as a result of the studies of modeling the biological neural system mathematically. ANN (Artificial Neural Networks) produced by being inspired by the functions of the human brain is a method which provides learning and generalization by testing. The aforesaid method is used to make estimations as a result of its constant output production [20]. ANNs produce quite successful results in complicated estimations for the fact that it is a technique which has a high power of calculation and the ability to generalize.

ANN consists of stratified, feed-forward and interrelated artificial neural networks or nodes. Hence the network is feed-forward, unidirectional flow is provided and loop or return is not allowed. ANN model formed by the connection of numerous cells in various forms has a parallel distributed structure and the information which the network obtains is scattered on all the connections in the network. Therefore, the fact that some of the connections and cells of a trained ANN model become neutralized does not affect the accuracy of the information that the network produces significantly [21].

3. RESULTS AND DISCUSSION

Various models were tested to determine the most appropriate model for a ten-year prospective estimation using goat milk production data between 1961 and 2016 and an estimation method was developed by comparing them. First, correlogram test was applied to the data and it was observed that the result showed a slow decline indicating the unit root. ADF (Augmented Dickey-Fuller) test validated that the level data includes unit root, that it's non-stationary. While estimating the goat milk data which was observed to become stationary by first differencing, difference data was used firstly.

Notation for ARIMA (p, d, q) model is as follows: p: standard autoregressive level of the model; d: the degree of the difference that will make the series stationary; q: standard moving average degree. In this study, d=1 was used because it was differentiated from the first degree in order to make the series stationary. While choosing the most appropriate model among the calculated potential estimation models, the model having a lower SC (Schwarz criterion) value (SC = n Log (SEE) + k Log (n)), a lower SEE (Sum squared of regression) value and a higher R^2 (R-squared) value was determined to be more suitable and meaningful compared to other models.

Table-1. Estimation criteria of ARIMA models

ARIMA MODEL	SC	R^2	SEE
(1, 1, 0)	22.939	0.178	21933.62
(2, 1, 0)	23.054	0.090	23214.14
(1, 1, 1)	23.011	0.179	22131.10
(1, 1, 2)	22.991	0.196	21911.41
(0, 1, 1)	23.258	0.164	25749.31
(0, 1, 2)	23.331	0.101	26707.91
(2, 1, 1)	22.913	0.205	21908.42
(2, 1, 2)	23.124	0.094	23395.55

Source: Table was constitute from analyzed data of current study

With reference to Table 1, considering SC, R^2, SEE values; the model of (2, 1, 1) which was differentiated from first degree and with AR (2) and MA (1) added was decided to be the most appropriate ARIMA model. Estimation was tried to be made with ARIMA models on goat milk production data which was made stationary by differentiating from the first degree. Estimation results on the data which was not differentiated (level) were added to the study to make a comparison. ARMA models and AR and MA models which are specific versions of ARMA models were used on level data and their tests were carried out.

Table-2. Examination criteria of ARMA models

ARMA MODEL	SC	R^2	SEE
(1, 0)	23.340	0.935	26824.74
(2, 0)	24.227	0.840	41774.48
(1, 1)	23.221	0.945	24600.48
(1, 2)	23.294	0.941	25516.06
(0, 1)	24.893	0.722	58352.38
(0, 2)	25.122	0.651	65419.26
(2, 1)	23.145	0.948	23663.21
(2, 2)	24.088	0.868	37916.80

Source: Table was constitute from analyzed data of current study

The subjects to be considered while choosing the most appropriate model among the potential estimation models calculated for ARMA models are similar to the ones with ARIMA models. ARMA (2, 1) model which has the lowest SC and SEE values and the highest R^2 value among the given models was decided to be the most appropriate model (Table 2).

The ANN model set up in this study serve as a non-linear regression model. Minimum-maximum (min-max) normalization which is frequently used in literature for data normalization was made and input vector was the years between 1961 and 2016 and the output vector was annual goat milk production. LM (Levenberg Marquardt) model was chosen as the training method for the ANN model which was set as feed-forward, back-propagated and 70% was used in training, 15% used in validation and 15% in the testing stage. The interlayer number was taken as 10 and delay number was taken 2. These values were kept stabilized for the performance of the model was at the desired level.

Goat milk estimation performance of the network trained with LM method can be seen in Figure 1. Performance graph, input, and target data consist of three lines as they are divided into three sets. As seen in the figure, it declined to nearly zero error in 12th iteration. MSE (mean squared error) represents the proportional value of the difference between the values in the artificial neural network and estimated values. In this study, the training stops when the MSE value drops below 0.01. This value represents that the difference between the actual values and the estimated values will be 1% maximum. 0.05 and 0.01 values are frequently used in the literature [22]. For the reason that the error dropped below 0.01 in the training set, it can be said that the results are at an acceptable level. The training stopped when the verification error started to increase, at 100th iteration. A remarkable sign of memorization is not observed until the 12th iteration where the best verification performance is present because the error rate in the verification and test set does not increase as of this iteration. As verification error and test set error show similar characteristics and a significant memorization has not occurred, the performance of the network is at an acceptable level. In addition to this, MSE result of the goat milk production of the model is 0.0075064. As a result, error rate acquired and predetermined and acceptable performance target has been achieved. R correlation coefficient indicates how good the variations in outputs are explained by targets.

The fact that the R-value approaches 1 signifies that the relationship strengthens and that it approaches zero signifies that the relation weakens [22]. Figure 2 shows the correlation between outputs and targets. It represents a good compatibility that R-value is quite close to 1. Training data shows a quite good harmony. While R correlation coefficient is 0.99 for training data, R correlation coefficient values which are above 0.97 are acquired for test and verification data. Scatter graph is important in that the particular data points show a weak harmony. When training, verification and test graphs are examined; it can be observed that there is no significant inconsistency between the actual value and network inputs. 20-years of the output of ANN, ARIMA and ARMA models used for estimation and the actual output of last 10 years are as given in Table 3.

Table-3. Values of 2007-2016 and estimation results for 2007-2026

DATE	ACTUAL	YSA	ARIMA	ARMA
2007	237487	220867.1	223829.0	348809.0
2008	209570	216894.4	202292.4	348265.9
2009	192210	235414.8	188506.3	347764.5
2010	272811	282890.3	270884.3	347301.6
2011	320588	335389.6	320005.4	346874.2
2012	369429	377973.5	371207.2	346479.7
2013	415743	420999.0	419396.1	346115.4
2014	463270	460443.2	468105.2	345779.2
2015	481174	484143.5	487482.0	345468.7
2016	479401	494410.3	486374.4	345182.1
2017		443125.0	485491.5	344917.5
2018		446462.5	485484.6	344673.2
2019		472664.7	485493.9	344447.6
2020		466844.8	485481.3	344239.4
2021		455216.3	485498.3	344047.2
2022		474135.0	485475.5	343869.7
2023		485646.4	485506.1	343705.9
2024		495665.7	485464.9	343554.6
2025		490479.1	485520.3	343415.0
2026		495536.1	485445.9	343286.0

Source: Table was constitute from analyzed data of current study

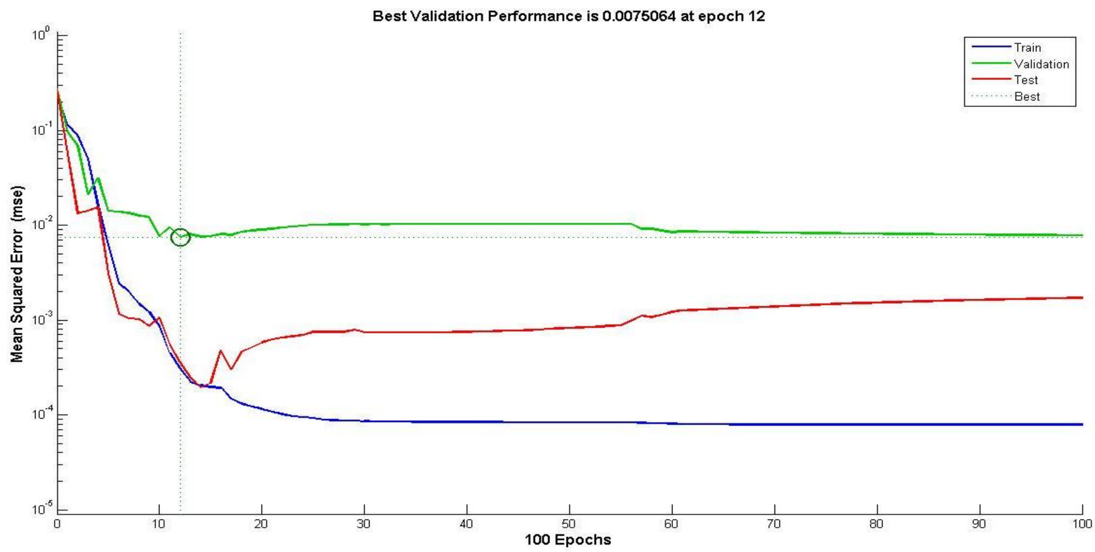

Figure-1. Performance graph of the network

Source: Figure was constitute from analyzed data of current study

MSE and MAPE (mean absolute percentage error) methods were applied to choose the estimation method which gives the best result among the models and the results were given in Table 4.

Table-4. Error results for the estimation models

	MSE	MAPE
ARMA	6028086102	0,21251266
ARIMA	274473715	0,028086212
YSA	209308290	0,028060791

Source: Table was constitute from analyzed data of current study

While MAPE values of which estimation values below 10% were regarded to be at high accuracy degree, models between 10% and 20% were classified as accurate estimation model [23, 24]. As a result of the error tests demonstrated in Table 4, MAPE value of the ARMA model was found as 0.212, MAPE value of the ARIMA model and ANN model which were quite close to each other was found as 0.028. When MSE values, another comparison measure, were examined, the model with the least error rate was found to be the model established by ANN. It is indicated that it is necessary to work with over 70 data for Box-Jenkins methods which include ARIMA and ARMA models subject to this study [19]. ANN is a method used frequently in time series estimations and it produces quite successful results. It allows working with less data by nature. However, despite the number of the data is not sufficient in this study, it is observed that ARIMA model resulted well with a close degree to ANN. Estimation model made with ANN was the model which gave the most accurate result among examined time series estimation methods. When regression outputs of the model were examined, it was observed that it showed good compatibility and MAPE value was found at an accurate estimation level.

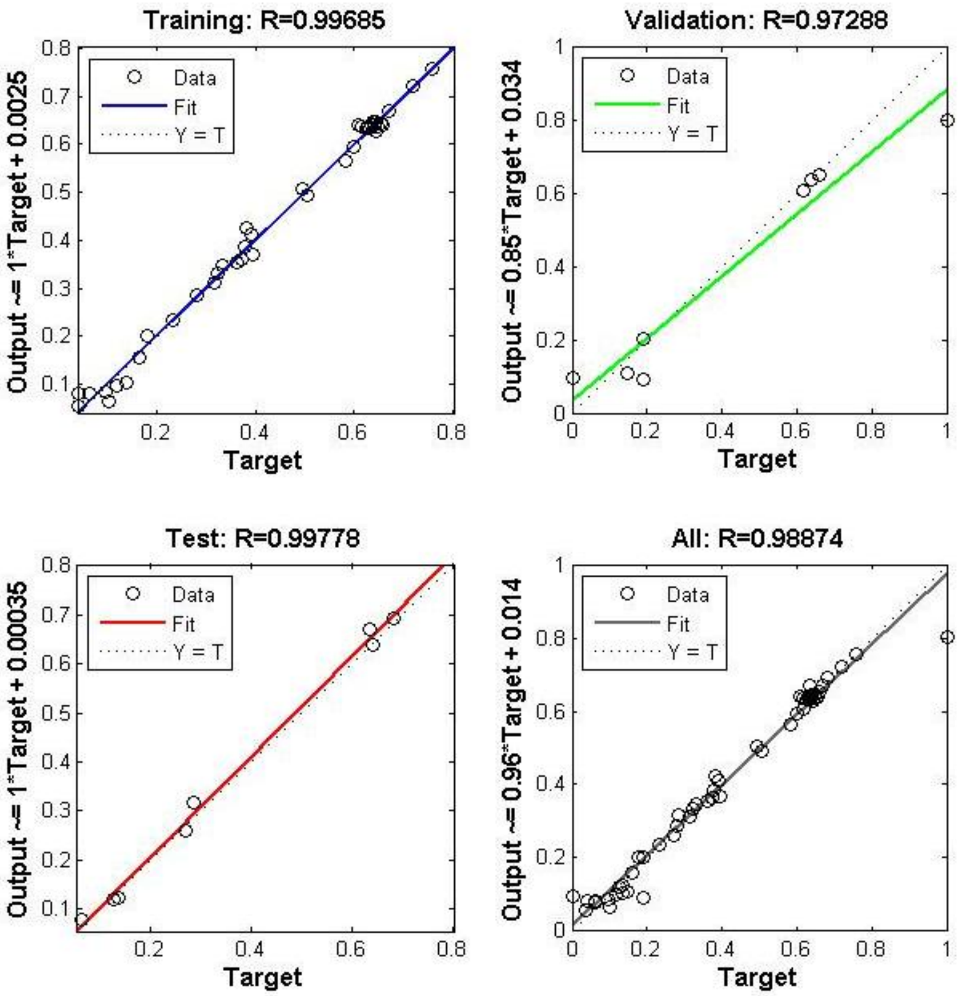

Figure-2. Regression output of the network

Source: Figure was constitute from analyzed data of current study

In this study, various estimation models with Box-Jenkins and ANN were tested by using the data of the goat milk produced between the years 1961-2016 and the models built were compared to each other. According to ANN results, forecasted value of the milk production belonging to 2026 is 495,536.1 tons. According to the forecast values in Table 4, it was forecasted that a fair amount of milk decrease will be observed in goat milk amount starting from 2017 when compared to the production amount of 2016; however, it was forecasted that the production will increase as of 2023. Following the forecasts, it was calculated that annual average rate of increase in the goat milk production will be 0.12%.

4. CONCLUSION

The aim of the study is to obtain the data that will contribute to forming a policy in the field of animal production by making a 10-year forecast of output belonging to the goat milk production in Turkey. As a result of the estimation, it was concluded that ANN models are more successful and goat milk production will follow an upward trend in 10 years.

In recent years both around the world and in Turkey, the demand for goat products have increased gradually with the rising significance of the goat milk with regards to nutrition. It is necessary to take a number of measures in terms of increasing the production of goat products and its sustainability in line with this development. Direct and indirect government interventions with technical measures will play an important role in production policies in improving goat breeding. Subventions and regulations regarding the price formation of meat, milk, and mohair are compulsory within production policies. On the other hand, legal regulations which will promote breeder to organize under cooperatives must be implemented and measures protecting goat health must be taken for Turkey to compete with other countries in exportation [4]. When the measures are taken regarding the technical and economic problems that producers encounter, goat breeding will contribute greatly to the national economy.

Funding: This study received no specific financial support.
Competing Interests: The authors declare that they have no competing interests.
Contributors/Acknowledgement: Both authors contributed equally to the conception and design of the study.

REFERENCES

[1] M. Kaymakçı and S. Engindeniz, "Goat breeding in Turkey: Problems and solutions." Retrieved: http://www.akademik.adu.edu.tr/myo/cine/webfolders/File/Yeni%20klasor/Kav%C5%9Fit%20K%C3%B6y%C3%BC%20ve%20Y%C3%B6resi%20Ke%C3%A7icilik%20Projesi%20Geli%C5%9Fmeler.pdf. [Accested 01.06.2017], 2017.

[2] A. Günlü and S. Alaşahan, "Evaluations on the future of goat breeding in Turkey," *Journal Turkish Veterinary Medicine Association*, vol. 81, pp. 15-20, 2010. *View at Google Scholar*

[3] K. Ş. Uzun and Ş. Ö. Altınçekiç, "Situation of products obtained from goats in Turkey and in the world." Retrieved: http://www.akademik.adu.edu.tr/myo/cine/webfolders/File/Yeni%20klasor/Kav%C5%9Fit%20K%C3%B6y%C3%BC%20ve%20Y%C3%B6resi%20Ke%C3%A7icilik%20Projesi%20Geli%C5%9Fmeler.pdf. [Accested 01.06.2017], 2017.

[4] S. Engindeniz and K. Uçar, "Alternative investment opportunities for rural area: Dairy goat farming." Retrieved: http://www.tarekoder.org/wp-content/uploads/2014/12/samsuncilt_I.pdf. [Accested 03.06.2017], 2017.

[5] TSI Turkish Statistical Institute, Retrieved from http://tuikapp.tuik.gov.tr/hayvancilikapp/hayvancilik.zul. [Accested 14.06.2017], 2017.

[6] N. Cenan, "The research of the amount of the livestock (Cow, Buffalo, Sheep, Goat) and the meat production from these animals according to the methods of time series analysis," Ph.D. Thesis Ankara University, Ankara, Turkey, 2011.

[7] F. Yavuz, A. Bilgiç, M. Terin, and I. O. Güler, "Policy implications of trends in Turkey's meat sector with respect to 2023 vision," *Meat Science*, vol. 95, pp. 798-804, 2013. *View at Google Scholar* | *View at Publisher*

[8] Ş. Çelik, "The modeling of annual red meat production of Turkey by using box-Jenkins method and projection of production," *Journal of Animal Production*, vol. 53, pp. 32-39, 2012.

[9] N. Cenan and İ. S. Gürcan, "Forward projection of the number of farm animals of Turkey: ARIMA modeling," *Veteriner Hekimler Derneği Dergisi*, vol. 82, pp. 35-42, 2011. *View at Google Scholar*

[10] N. Iqbal, K. Bakhsh, A. Maqbool, and A. S. Ahmad, "Use of the ARIMA model for forecasting wheat area and production in Pakistan," *Journal of Agricultural and Social Sciences*, vol. 1, pp. 120-122, 2005. *View at Google Scholar*

[11] Ç. Takma, H. Atıl, and V. Aksakal, "Comparison of multiple linear regression and artificial neural network models goodness of fit to lactation milk yields," *Journal of the Faculty of Veterinary Medicine*, vol. 18, pp. 941-944, 2012. *View at Google Scholar* | *View at Publisher*

[12] P. Hosseinia, M. Edrisi, M. A. Edriss, and M. A. Nilforooshan, "Prediction of second parity milk yield and fat percentage of dairy cows based on first parity in formation using neural networks system," *Journal of Applied Sciences*, vol. 7, pp. 3274-3279, 2007. View at Google Scholar | View at Publisher

[13] A. K. Sharma, R. K. Sharma, and H. S. Kasana, "Prediction of first lactation 305-day milk yield in Karan Fries dairy cattle using ANN modeling," *Applied Soft Computing*, vol. 7, pp. 1112-1120, 2007. View at Google Scholar | View at Publisher

[14] M. Karahan, "Statistical demand forecasting methods: An application of product demand forecast with artificial neural networks method," Ph.D. Thesis, Selçuk University, Konya, Turkey, 2011.

[15] FAO Food and Agriculture Organization of the United Nations, Retrieved from http://www.fao.org/faostat/en/#data/QL. [Accested 14.06.2017], 2017.

[16] C. Hamzaçebi and F. Kutay, "Electric consumption forecasting of Turkey using artificial neural networks up to year, 2010," *Journal of the Faculty of Engineering and Architecture of Gazi University*, vol. 19, pp. 227-233, 2004. View at Google Scholar

[17] H. Kayım, *Statistical forecasting methods*. Ankara: Hacettepe University Faculty of Economics and Administrative Sciences Publications, 1985.

[18] M. Hibon and S. Makrıdakis, "Arma models and the box–Jenkins methodology," *Journal of Forecasting*, vol. 16, pp. 147-163, 1997. View at Google Scholar | View at Publisher

[19] G. E. P. Box and G. M. Jenkins, *Time series of analysis, forecasting and control*. California: Holden Day, 1976.

[20] D. T. Larose, *Discovering knowledge in data: An ıntroductionto data mining*. New Jersey: John Wiley and Sons Inc, 2005.

[21] H. Ergezer, M. Dikmen, and E. Özdemir, "Artificial neural networks and recognition systems," *Pivolka*, vol. 2, pp. 14-17, 2003. View at Google Scholar

[22] S. Kalaycı, *SPSS applied multivariate statistical techniques*. Ankara: Asil Publications, 2010.

[23] S. F. Wıtt and C. Wıtt, *Modeling and forecasting demand in tourism*. London: Academic Press, 1992.

[24] C. D. Lewıs, *Industrial and business forecasting methods*. London: Butterworhts Publishing, 1982.