

# Using Principal Component Analysis and Artificial Neural Networks for Fault Type Forecasting in an Automotive Company

Tülay Korkusuz Polat

\*Faculty of Engineering, Department of Industrial Engineering Sakarya University, Turkey

## Abstract

In this study, failures that occurred in the paint shop of an automotive company were discussed. The relationships between these failures and the probabilities of prospective occurrences were investigated. Any product produced in the company passes quality control at the end of production. Technical or operator-originated types of potential failures are examined during this control. Causes of failures in the paint shop and how they can be resolved pose a serious problem, just as in the other departments of the factory. This is because every failure encountered negatively influence the product quality and harm the company in terms of cost/productivity/image. The inability of the paint shop to predict the probability of failures in advance and its inability to establish a link between the types of failures also lead to its failure to pass quality control — which is the subsequent process — and cause its “production quality score” to fall, as well as other adversities. This study was carried out to determine which failures were usually caused by the activities in the paint shop and to develop a model that would predict the pass/fail state of the types of failures in question.

**Key words:** Principal Component Analysis, Artificial Neural Network, Fault Forecasting

## 1. Introduction

It is getting more and more difficult every passing day to meet increasing customer demands in today’s industry. The key factor in accomplishing a sustainable competitive edge in a challenging market environment is to increase productivity. The way to increase productivity is to create flexible production systems, to improve quality scores and to reduce production costs. Being able to predict failures that can occur in any production process not only improves quality scores but also reduces costs originating from faulty products.

In recent years, “quality” has been an essential criterion for survival in both local and global markets. There are a variety of methods to enhance quality and ensure sustainability. In this respect, it is important to be able to predict the types of failures in advance that negatively affect product quality and therefore customer satisfaction, and whether these types of failures can be corrected or rejected at the time of occurrence.

One of the sectors in which digitalization is to be experienced most heavily is the automotive sector (a technology-intensive sector bringing high added value), which is also one of the leading sectors of our country, Turkey. With the digitalization in the quality processes as well as in every stage of production, it has become easier to identify in advance the types of failures and whether these failures will be acceptable in the automotive sector where there is plenty of data.

\*Corresponding author: Address: Faculty of Engineering, Department of Industrial Engineering Sakarya University, 54187, Sakarya TURKEY. E-mail address: [korkusuz@sakarya.edu.tr](mailto:korkusuz@sakarya.edu.tr) Phone: +902642955687

In this study, an Artificial Neural Network (ANN) model was created to predict the types of failures occurring in the paint shop of an automotive company. In the second section of the article, a literature review is presented for the Principal Component Analysis (PCA) and ANNs, which were used in the study. Many computer programs are used for PCA. Since the data used in the application is kept in Excel, Analyse-It program is used with the Excel infrastructure. Analyse-It is a statistical data program that works with Excel infrastructure. Here, the correlation matrix of the data is generated. WEKA program is used to model artificial neural network. WEKA is a package program that performs data grouping, classification and forecasting and give test results. It has Java substructure. In the third section, the implementation that was carried out in the paint shop of the automotive company is addressed in detail. When the failure reports reaching the paint shop were examined, it was seen that ten types of failures appeared more often. In the initial part of the implementation, these ten types of failures were addressed. In the second part of the implementation, six types of failures that were identified by a Principal Component Analysis were used. In the fourth section, the statistical results of the implementation were discussed. In the conclusion section, suggestions are given for the firm to do in the future.

## **2. Materials and Method**

### **2.1. Principal Components Analysis**

As stated by Dunteman (1989), the idea of Principal Component Analysis was originally designed by Pearson in 1901 and was independently developed by the Hotelling in 1933. And also Dunteman (1989) refers to PCA is a statistical technique and can be used to convert the original set of variables to a smaller set of variables, but with the ability to represent most of the information in the original variable set. Smaller clusters of non-related variables are more preferable and easier to understand during analysis than a larger set of correlated variables.

In addition, Vidal and his friends (2016) mention that calculating a subspace from missing or corrupted data is an important problem. Since there is a common problem in practice, many extensions have been proposed for PCA over the years (although most of these methods are heuristic, greedy or ad hoc) in order to process these defective data in different application areas. And Vidal and his friends (2016) add that recent developments in large-volume statistics have brought with them the right and efficient methods to find the optimal subspace from data even if they are too much missing or distorted.

The PCA method has several characteristic features (Yıldız, et al., 2010);

- It tries to find the strongest pattern in the data. So it can be used as a pattern-finding technique,
- Often, the diversity of data can be captured by a small set of sizes selected from the entire size set. Thus, size reduction operations using PCA result in smaller sized datasets, so that techniques that are not suitable for high-dimensional data can work comfortably on this new dataset.
- Noise from data is more than patterns these noise can be cleared as a result of size reduction. This feature is particularly useful in both data mining and other data analysis algorithms.

In his study Naik (2018) mentioned the sparse PCA and requirements of sparse PCA. In general, there are physical descriptions of the dimensions of the data (For example, in financial or biological applications, each dimension may correspond to a particular physical entity or gene). For all that, as the loads we obtain with PCA are dense, the main component obtained can be a mixture of all dimensions. This makes it difficult to interpret. Therefore, if most of the inputs in Naik (2018) loading are zero (sparse), it mentions that each principal component will be a linear combination of several non-zero inputs and will facilitate the understanding of the main components as well as the physical meaning of the loads. He also pointed out that this is a rare PCA motivation, saying that if different uploads have different non-zero entries that correspond to different dimensions, the physical interpretation will be more explicit.

Feng et al. (2018) suggest that PCA is mostly used for data analysis and dimension reduction in multiple variance analyzes. And they also stated that they are a very powerful tool in image and signal processing.

Gupta and Barbu (2018) tested PPCA (Parameterized Principal Component Analysis) on artificial data based on two real data sets with different parameters and grouped observations according to the parameter values and compared PPCA with PCA, sparse PCA and IPCA (Independent Principal Component Analysis). As a result, they found that the PPCA has recovered the known functions with fewer errors and has made less restructuring errors in the data sets than the IPCA does.

Ren et al. (2018) mentioned the new model (Supervised Dirichlet Process Mixtures of Principal Component Analysis (SDPM-PCA)) they developed by incorporating PPCA (Probabilistic Principal Component Analysis) and DPGLM (Dirichlet Process Mixtures of Generalized Linear Models) in their study.

Jackson (2003) has given many examples to understand the general methodology of PCA. Tezbaşaran ve Gelbal (2018) worked with a group of 400 pre-service teachers, gathered data from these candidates in order to decide which of the scale structures obtained from different methods, and compared the results of confirmatory factor analysis based on PCA and ANN according to model fit and error indices.

## **2.2. Artificial Neural Networks (ANN)**

Artificial neural networks are computer programs developed based on the human working principles. There are many differences between traditional computer programs and human brain. The nonlinear complexity of the human brain shows superiority over traditional computers in many areas.

According to Rodriguez et al. (2018), Artificial Intelligence (AI) can increase the capabilities of computers by replicating human biological processing system. For example, ANN, which is one of the artificial intelligence technologies, can multiply the biological neural networks in the human brain, recognize the existing patterns and predict the future values.

Although the first article about the ANN was written a long time ago, in the beginning of the 90s, the subject was re-discussed and a lot of work is being done today. Silva et al. (2017), the smart applications worked with ANN are listed as follows:

- analyzing images obtained artificial satellites,
- speech and writing pattern classification,
- face recognition with computer vision,
- control of high-speed trains,
- stocks forecasting on financial market,
- anomaly identification on medical images,
- automatic identification of credit profiles for customers of bank, etc.

Liu et al. (2018) describe ANN as an artificial intelligence technique based on a series of processors or neurons. ANN is a widely used algorithm, in the most common form there are three components: input layer, hidden layer and output layer.

The elements of an ANN are artificial neurons, in figure 1, you can see the neuron can be represented by k (Rodriguez, et al., 2018).

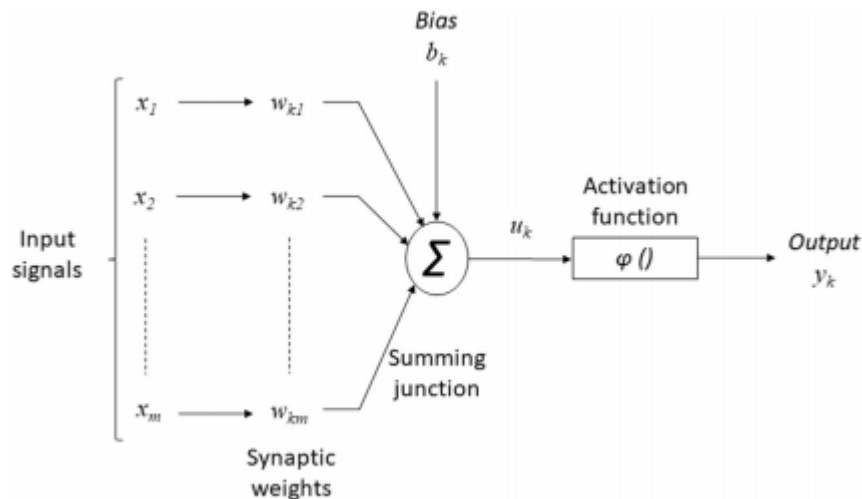


Figure 1: Artificial Neural Model

Hassoun (2003) indicates that ANN is a parallel calculation model consisting of interconnecting adaptive processing units and emphasizes two important features of ANN. The first is to take “learning by example” instead of “programming” in problem solutions with ANN. The second important feature is that when the ANN is applied in parallel digital computers or customized hardware, the solutions have parallel architecture that allows for fast computation.

Silva et al. (2017), the important features related to artificial nerve applications are as follows:

- Adapting from experience
- Learning capability,

- Generalization capability,
- Data organization,
- Fault tolerance,
- Distributed storage,
- Facilitated prototyping, etc.

We can reproduce samples studied with artificial neural networks. For example; KorkusuzPolat and Arslankaya (2010) have created an ANN model which estimates the amount of production by using the elements that constitute the cost in the factory as input. Çuhadar and Kayacan (2005) studied that using different variables with artificial neural networks, the occupancy rates in the accommodation establishments operating within a certain region or province, the expenditure on accommodation establishments of domestic and foreign guests, the number of overnight stays of tourists and the average duration of stay. Kutlu and Badur (2009) have done a study to model the Istanbul Stock Exchange index value with feed-forward neural networks. Arslankaya and Öz (2018) have estimated demand forecast using artificial neural networks in an automotive company.

### 3. Results

The implementation was carried out in the paint shop of an automotive company offering services in Sakarya. When the monthly failure records in the paint shop were examined, it was observed that the failures that had occurred had an adverse impact on the product quality and harmed the company in terms of cost. Failure reports including pass/fail states for paint shop operations of a total of 1500 products were investigated, and ten most common error types were identified. The inability of the company to predict the probability of failures in advance and its inability to establish a link between the types of failures lead to its failure to pass paint quality control and cause the production quality score of the paint shop to fall. This poses a major problem for both the department and the company.

When the failures in the paint shop were examined, it was seen that there were generally ten types of failures. These failures were determined as follows:

1. Operator-originated failure,
2. Misapplication of an operation,
3. Not observing the order of operations,
4. The technical structural failure of a part,
5. A part's being damaged during delivery,
6. Variations in robotic paint thickness,
7. Improper painting in the touch-up area,
8. Incorrect entry of the paint command for the part code,
9. Incomplete or faulty taping,
10. A malfunctioning robot.

These failures had been recorded on a daily basis. The pass and fail states of a total of 1500 final products were examined to be used in the study. For example, if the failure originating from the

operator gave an error that could not be repaired, 0 (in other words, fail) was assigned, and if it was a failure that could be fixed in the revision, 1 (in other words, pass) was assigned. The relationship between the failure types was examined using the Principal Component Analysis in the Analyse-It program, and a correlation matrix was created. Table 1 shows the correlation matrix.

Table 1: Correlation matrix

Descriptives											
N	1500										
Correlation		Operator Kaynaklı Hata	Operasyonun Yanlış Uygulanması	Operasyon Sırasına Uyulmaması	Parçanın Teknik Yapı Hatası	Parçanın Sevkiyat Sırasında Zarar Görmesi	Robotun Boya Ağız Kalınlığının Farklı Olması	Touch-Up Bölgesinde Yanlış Boyama Yapılması	Parça Kodunun Boyama Komutunun Hatalı Girişi	Yapılan Tapalamanın Eksik veya Hatalı Olması	Robot Arızası
Operator Kaynaklı Hata	-	0,800	0,680	-0,141	-0,374	-0,461	-0,123	0,173	-0,159	0,671	
Operasyonun Yanlış Uygulanması	0,800	-	0,734	-0,120	-0,195	-0,348	-0,126	-0,193	-0,415	0,213	
Operasyon Sırasına Uyulmaması	0,680	0,734	-	-0,122	-0,230	-0,233	0,165	-0,162	-0,137	-0,088	
Parçanın Teknik Yapı Hatası	-0,141	-0,120	-0,122	-	-0,080	-0,279	-0,166	0,321	-0,123	-0,132	
Parçanın Sevkiyat Sırasında Zarar Görmesi	-0,374	-0,195	-0,230	-0,080	-	-0,045	0,156	-0,098	-0,096	-0,043	
Robotun Boya Ağız Kalınlığının Farklı Olması	-0,461	-0,348	-0,233	-0,279	-0,045	-	0,037	0,499	0,837	0,746	
Touch-Up Bölgesinde Yanlış Boyama Yapılması	-0,223	-0,226	0,165	-0,166	0,156	0,037	-	0,264	0,037	0,230	
Parça Kodunun Boyama Komutunun Hatalı Girişi	0,173	-0,193	-0,162	0,321	-0,098	0,499	0,164	-	0,142	0,073	
Yapılan Tapalamanın Eksik veya Hatalı Olması	-0,159	-0,415	-0,137	-0,123	-0,096	0,837	0,037	0,142	-	-0,675	
Robot Arızası	0,671	0,213	-0,088	-0,132	-0,043	0,746	-0,130	0,073	0,675	-	

The matrix shown in Table 1 is interpreted based on color tones. Light tones show weak and negative interactions between failure types, whereas dark tones show strong interactions. For example, the failure type with which the “operator-originated” failure establishes the strongest relationship is the “misapplication of an operation.” The failure type with which the “operator-originated” failure has the weakest interaction is the “variations in robotic paint thickness.” “Variances” were used to interpret which of these failure types was more frequent and which of them were less frequent. Table 2 shows the resulting variances for ten types of failures.

Table 2: variances for ten types of failures

Principal Components			
Variances			
Component	Variance	Proportion	Cumulative proportion
1	1,801	0,180	0,180
2	1,281	0,128	0,308
3	1,116	0,112	0,420
4	1,043	0,104	0,524
5	1,032	0,103	0,627
6	0,962	0,096	0,724
7	0,938	0,094	0,817
8	0,930	0,093	0,910
9	0,846	0,085	0,995
10	0,051	0,005	1,000

According to the variance calculation carried out through the Principal Component Analysis, the variance of the “operator-originated failure” — analyzed as the first failure — was 1.801, which was the most common failure. The “Proportion” section shown in Table 2 is interpreted as the rate of operator-originated failure in the data of 1500 products. The “Cumulative Proportion” section

shows the cumulative rate of failures. It is worth noting that the cumulative sum must always equal to 1.

After all of the ten failures were interpreted in this way, six most common failures which had strong interactions among themselves were determined.

Models were created in two different ways during the ANN application which was to be carried out to estimate the types of failures. The first of them was the ANN model including all ten failure types. The second of them was the ANN model which was formed using the six types of failures obtained after the correlation and variance studies through the principal component analysis.

Figure 2 shows the Multi-layer Artificial Neural Network Model prepared for the data sets of the ten failure types.

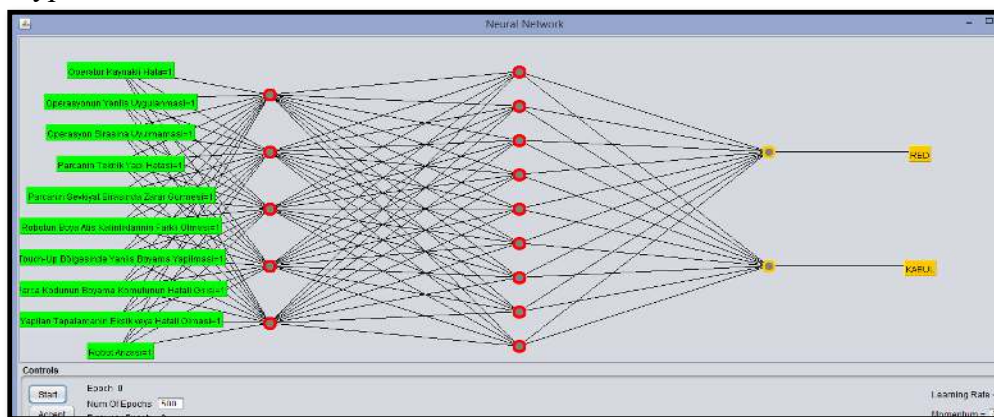


Figure 2: ANN model created for ten fault types

A total of 66% of the data of the ten failure types from 1500 final products were used for testing, and 33% were used for training. In the model, the **input layer** consisted of 10 process elements to represent the ten failure types. Since it was aimed that the model learned the pass or fail of the failure, the **output layer** consisted of 2 process elements which would represent the pass/fail state. As an intermediate layer, 5 process elements in the first layer and 9 process elements in the second layer were randomly identified after many attempts to ensure high success.

When the first ANN model was run, the learning success appeared to be 48%. This ratio was not sufficient to successfully estimate the pass/fail state of a failure to occur in the relevant failure type. For this reason, the number of failure types was reduced, and a new study was conducted.

Figure 3 shows the second ANN model that was established (only for the six failure types identified based on the Principal Component Analysis).

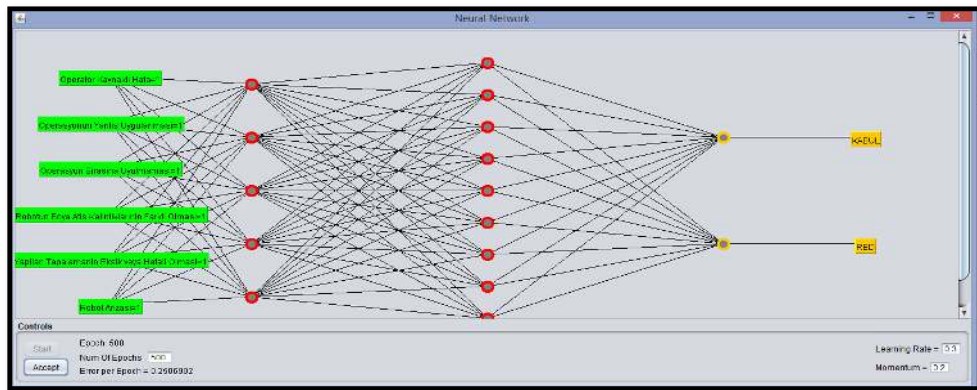


Figure 3: ANN model created for six fault types

As shown in Figure 3, the second ANN model was again constructed as a multi-layered model. A total of 66% of the data of the six failure types from 1500 final products that were identified through the Principal Component Analysis were used for testing, and 33% were used for training. In the model, the **input layer** consisted of 6 process elements to represent six failure types. Since it was aimed that the model learned the pass or fail of the failure, the **output layer** consisted of 2 process elements which would represent the pass/fail state. As an intermediate layer, 5 process elements in the first layer and 9 process elements in the second layer were randomly identified after many attempts to ensure high success. The model was set to run 500 iterations.

When the second ANN model was run, it was seen that the network learned at the rate of 89.41%. This rate is a rate that can help to estimate the pass or fail state of a failure type that can occur in a manufacturing company and to take measures accordingly.

#### 4. Discussion

It was seen that the artificial neural network model which was created by considering all of the ten failure types mentioned above regardless of the variance and correlations obtained through the Principal Component Analysis was not able to meet the necessary learning rate and had a low “pass” rate. It was observed that the artificial neural network model created using the failure types reduced to six after the Principal Component Analysis learned relatively successfully compared to the first model and had an increased “pass” rate.

The Analyse-It software program, which is a statistical data program that runs on Excel’s infrastructure, was used to identify the relationships between the types of failures in the data that were obtained. The Principal Component Analysis was carried out through the Analyse-It program. Which of the ten investigated failure types were related to each other was determined through the Principal Component Analysis. And, the degree of proximity of these relationships was determined. The types of failures were estimated by using the Artificial Neural Network algorithm in the WEKA software program in order to determine the probability that the types of failures — between which close relationships were observed — were likely to occur in products to be produced in the future.



## Conclusions

When the results of the failure test were examined, it was seen that the constructed artificial neural network model successfully learned the algorithm at a rate of 78.66%. With the help of this model, the probability of prospective failures was calculated in percentages. It was seen that there would be “pass” for 1501 vehicles with 89% probability in the prospective failure prediction.

The company can take many distinctive precautions to prevent failures that will occur in the future. For example, if the probability of failures is very high in the next month, the company can adjust the number of productions accordingly, or if the possibility of failures is low, the number of vehicles produced can be increased by increasing the capacity.

Based on the implementation in this study, the company has had a model that can predict the probability of failures in order to prevent the probable types of prospective failures. With the help of the “principal component analysis,” the company has also had chance to see the relationships between failure types that cannot be distinguished through the classical methods. The main purpose of the implementation was not to find one-to-one solutions to these failures. So, solutions were not sought in this study. However, if the company examines this study, it can create a new production plan for itself.

## REFERENCES

- [1] G.H. Dunteman “*Principal Components Analysis – Series: Quantitative Applications in the Social Sciences*”, SAGE University Paper, 1<sup>st</sup> Edition, printed in USA, ISBN-10: 0803931042 ISBN-13: 978-0803931046, 1989
- [2] R. Vidal, Y. Ma, S.S. Sastry “*Generalized Principal Component Analysis*”, Springer, Interdisciplinary Applied Mathematics Volume 40, New York, 2016
- [3] G.R. Naik (Editor) “*Advances in Principal Component Analysis – Research and Development*” Springer Nature Singapore Pte Ltd, ISBN-13: 978-9811067037 ISBN-10: 9811067031, 2018
- [4] K. Yıldız, Y. Çamurcu, B. Doğan, “*Veri Madenciliğinde Temel Bileşenler Analizi ve Negatıfsız Matris Çarpanlarına Ayırma Tekniklerinin Karşılaştırmalı Analizi*”, Akademik Bilişim’10 – XII. Akademik Bilişim Konferansı Bildirileri, Muğla Üniversitesi, 10-12 Şubat 2010
- [5] Y.T. Feng, T. Zhao, M. Wang D.R.J. Owen “*Characterising particle packings by principal component analysis*”, Computer Methods in Applied Mechanics and Engineering, 340, pp 70-89, 2018
- [6] A. Gupta, A. Barbu, “*Parameterized Principal Component Analysis*”, Pattern Recognition 78, pp 215-227, 2018
- [7] J. Ren, K. Li, C. Ghen, “*Supervised Dirichlet Process Mixtures of Principal Component Analysis*”, Neurocomputing 305, pp 15-26, 2018
- [8] J.E. Jackson “*A User’s Guide to Principal Components*”, John Wiley & Sons, Wiley Series in Probability and Statistics, New Jersey, ISBN-10: 0471471348 ISBN-13: 978-0471471349, 2003

- [9] E. Tezbaşaran, S. Gelbal “*Temel Bileşenler Analizi ve Yapay Sinir Ağı Modellerinin Ölçek Geliştirme Sürecinde Kullanılabilirliğinin İncelenmesi*”, Mersin Üniversitesi Eğitim Fakültesi Dergisi 14(1), ss 225-252, 2018
- [10] F. Rodriguez, A. Fleetwood, A. Galarza, L. Fontan “*Predicting solar energy generation through artificial neural networks using weather forecasts for microgrid control*”, Renewable Energy 126, pp 855-864, 2018
- [11] I. N. Silva, D. H. Spatti, R. A. Flauzino, L. H. B. Liboni, S. F. R. Alves “*Artificial Neural Networks – A Practice Course*”, Springer International Publishing, Switzerland, ISBN-13: 978-3319431611 ISBN-10: 3319431617, 2017
- [12] R. Liu, B. Yang E. Zio, X. Chen “*Artificial intelligence for fault diagnosis of rotating machinery: A review*”, Mechanical Systems and Signal Processing 108, pp 33-47, 2018
- [13] M. H. Hassoun “*Fundamentals of Artificial Neural Networks*”, MIT Press, ISBN-10: 0262514672 ISBN-13: 978-0262514675, USA, 2003
- [14] T. KorkusuzPolat, S. Arslankaya “*The cost forecasting application in an enterprise with artificial neural networks*”, Proceedings of the World Congress on Engineering Vol III, London, UK, June 30 – July 2 2010
- [15] M. Çuhadar, C. Kayacan “*Yapay sinir Ağları Kullanılarak Konaklama İşletmelerinde Doluluk Oranı Tahmini: Türkiye’deki Konaklama İşletmeleri Üzerine Bir Deneme*”, Anatolia: Turizm araştırmaları Dergisi, Cilt 16, Sayı 1, Bahar: 24-30, 2005
- [16] B. Kutlu, B. Badur “*Yapay Sinir Ağları ile Borsa Endeksi Tahmini*”, Yönetim, Yıl:20, Sayı: 63, Haziran 2009
- [17] S. Arslankaya, V. Öz “*Time Series Analysis of Sales Quantity In An Automotive Company and Estimation By Artificial Neural Networks*”, Sakarya University Journal of Science, 22(5), pp 1482-1492, 2018