Avrasya Sosyal ve Ekonomi Araştırmaları Dergisi (ASEAD)
Eurasian Journal of Researches in Social and Economics (EJRSE)
ISSN:2148-9963 **www.asead.com**

FORECASTING TOURIST ARRIVALS FIVE-STAR HOTEL IN ANTALYA, ISTANBUL AND MUGLA

Dr. Mehmet Emin AKKAYA[1]

ABSTRACT

Unable to predict the number of tourists is an important problem. Estimation is required to make price optimization and planning. It is also important to identify variables that increase the number of tourists. Variables affecting the number of tourists occur with the estimation method. Using these variables we can increase the number of tourists.

This is a comparative study for the provinces of Antalya, Istanbul and Mugla which has the highest tourist overnight stay numbers of Turkey by time series analyses performed by Linear and MLP regression analyses methods. Annual data range has been used to estimate the tourism demand. Multivariate data have been used. WEKA 3.8 data mining software was used in this study where the estimation methods were applied. In some cases, estimation work has been done on the number of foreign tourists. These results were compared with 2 different regression methods. Different regression analysis methods gave the best results for different tourism destinations in forecasting studies. Determined that the regression analysis that gives the best result for the destination of the forecasting study is determined and the closest values can be reached by the regression analysis which gives the result suitable for the destination.

Keywords: Accommodation Estimates, Tourism Forecasting, Time Series Regression

[1] eminakkaya@windowslive.com

ANTALYA, İSTANBUL VE MUĞLA'DAKİ BEŞ YILDIZLI OTELLERE TURİST GELİŞLERİNİN TAHMİNİ

ÖZET

Turist sayısının tahmin edilememesi önemli bir sorundur. Fiyat optimizasyonu ve planlaması yapmak için tahmin gereklidir. Turist sayısını arttıran değişkenlerin belirlenmesi de önemlidir. Turist sayısını etkileyen değişkenler tahmin yöntemi ile ortaya çıkmaktadır. Bu değişkenleri kullanarak turist sayısını arttırabiliriz. Bu çalışma, gelen turist sayısının tahmin edilebilmesi için yapılmıştır. Turizm, son 60 yılda savaşlar, bölgesel salgınlar ve mali krizler gibi kısa vadede turist hareketlerini önemli ölçüde etkileyen bazı engellere rağmen sürekli bir büyüme ve çeşitlilik sağlamıştır. Doğru talep tahminleri, fiyatlandırma ve kurumsal stratejiler konusunda turizm ve otellerle ilgili kararların alınmasında temel oluşturur.

Lineer ve MLP regresyon analiz yöntemleriyle yapılan zaman serisi analizleri ile Türkiye'nin en yüksek turist geceleme sayısına sahip illeri Antalya, İstanbul ve Muğla illeri için karşılaştırmalı bir çalışmadır. Turizm talebini tahmin etmek için yıllık veri aralığı kullanılmıştır. Çok değişkenli veri kullanılmıştır. Tahmin yöntemlerinin uygulandığı bu çalışmada WEKA 3.8 veri madenciliği yazılımı kullanılmıştır. Bazı durumlarda, yabancı turist sayısı ile ilgili tahmin çalışmaları yapılmıştır. Bu sonuçlar 2 farklı regresyon yöntemi ile karşılaştırılmış, tahmin çalışmalarında farklı turizm destinasyonları için farklı regresyon analiz yöntemleri en iyi sonuçları vermiştir. Tahmin çalışmasının varış yeri için en iyi sonucu veren regresyon analizinin belirlendiği ve varış noktasına uygun sonucu veren regresyon analizi ile en yakın değerlere ulaşılabileceği belirlenmiştir.

Anahtar Kelimeler: Konaklama Tahminleri, Turizm Tahmini, Zaman Serisi Regresyon

INTRODUCTION

For example, suppose that one of the variables affecting the number of tourists is the number of travel agencies. Considering that the number of tourists will increase when we increase the number of agencies, we see the importance of estimation. Forecasting is necessary to eliminate future uncertainties and to make strategic planning. It is not possible to make strategic planning without having the predictive information about the sector. This work was done to eliminate this problem.

Tourism has provided a continuous growth and diversity in the last 60 years despite some obstacles having significant effects on tourist movements in the short term such as wars, regional epidemic and financial crises. Accurate demand estimates form the basis on which decisions on tourism and hotels are made regarding pricing and corporate strategies. At the same time, medium and long term tourism and hotel demand estimates are required for investment decisions of the private actors and public infrastructure investments. Demand modelling and estimation is a significant area for tourism and accommodation researches.**(1)**

Demand for accommodation at hotels is measured by different variables from various perspectives. Number of rooms sold and occupancy rate,**(2)** some variables such as arrivals of visitors, number of overnight stays.**(3)** are related to the demand scale.

Arrival of tourists at a destination is the traditional and most commonly used indicator of the tourism demand. Other two popular indicators are tourism expenditures**(4-5)** and number of overnight stays **(6-7).**

Artificial Neural Network which is a nonparametric and data oriented technique has attracted great attention due to its ability of mapping the linear or nonlinear function without an assumption forced by the modelling process. Layers simulating the biological neural system, especially the human brain including input and output have one or more neurons. These neurons are related to each other in information and data processing process.**(8)**

In terms of data set, annual data have been used by many researchers for tourism and hotel demand modelling and estimation studies**.(9)** These studies normally focus on factors affected by tourism (or hotel) demand and/or long term relationships**(10)** between medium and long term trend estimates.**(11)**

Different ANN models have been implemented for tourism and hotel estimation applications such as multilayer perceptron(MLP),**(12)** radial basis function (RBF), general regression neural network (GRNN) and Elman neural network (Elman NN). **(13)** MLP is the most commonly used ANN (Artificial Neural Network) model **(14)** and has more three or more neural layers with non-linear activation function.**(15)**

In the study, annual time series data of the variables have been cited from the websites of the following institutions:Republic of Turkey Ministry of Tourism**(16),**Turkish State Institute of Statistics**(17),**Central Bank of Republic of Turkey**(18),** European Central Bank (ECB)**(19)**and TURSAB**(20)**.

1. DATA SET and MULTIVARIATE APPROACH for TOURISM DEMAND MODELLING

Annual time series analyses have been used to make future estimations by data set time series analyses and regression methods of past figures. The multiple variables related to estimation method,USD exchange rate, occupancy rates, arriving foreign tourist numbers, average accommodation duration of tourists, number of overnight stays and CPI, have been used and implemented in the study by analysis via time series methods.

The main factors affecting the tourism demand estimation reveal the relation between variables (USD exchange rate, occupancy rates, arriving foreign tourist numbers, average accommodation duration of tourists, number of overnight stays, CPI) and demand and determinants explaining the data mining methods.

2. IMPLEMENTATION

An analysis of the time series of foreign tourists in provinces of Antalya, Istanbul and Mugla, which are the three most frequented tourist destinations in Turkey, was carried out by estimating the number of nights spent on the number of years. Linear and nonlinear regression analysis methods were compared.Linear Regression and Multilayer Perceptron regression analyzes were compared with actual values.

In the implementation stage, WEKA 3.8 data mining software has been used. Weka was originally developed by Waikato University in New Zealand. It has widespread use for machine learning and data mining. The state-of-the-art machine written in Java contains a large collection of algorithms for learning and data mining.WEKA includes regression, classification, clustering, association rules, visualization and data preprocessing tools. WEKA is very popular for academic and industrial researchers. And is widely used for teaching purposes. WEKA is the most popular and open source data mining tool.**(21)**

Linear and nonlinear comparative studies are available in the literature.**(22)**

Implementation Phase: Total Number of Overnight Stays in the Province of Antalya (Foreign Tourist)

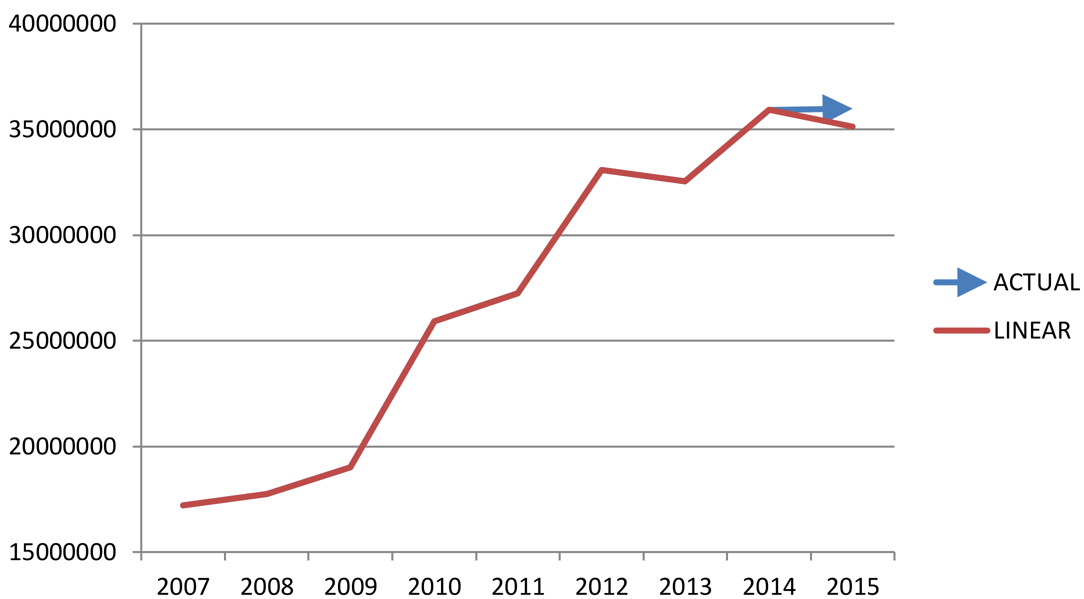

Figure 1: Values Estimated via Linear Regression Analysis Forecasting Model for the Province of Antalya

Linear Regression Analysis Model Performance for the Province of Antalya (Number of overnight stays in 2015) In this graph, Linear is analyzed for Antalya province. The values estimated by the Linear regression method for the province of Antalya are compared with the actual values. Scheme LinearRegression -S 0 -R 1.0E-8 -num-decimal-digit 4, Training Time 500 gave the closest result to the true value.

Table 1: Forecasting Performance of The Linear Regression Analysis Model Models for Antalya

Scheme	Training Time	Actual Values	Forecasting Values
LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4	500	35.978.369	35.113.625

Table 2: Forecasting Performance of The MLP-Based Models for Antalya

MLP Model ID	Scheme	Learning Rate	Momentum	Training Time	Actual Values	Forecasting Values
1	MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a	0.3	0.2	500	35.978.369	33.901.785
2	MultilayerPerceptron -L 0.2 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a	0.2	0.2	500	35.978.369	34.303.619
3	MultilayerPerceptron -L 0.1 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a	0.1	0.2	500	35.978.369	35.287.552
4	MultilayerPerceptron -L 0.1 -M 0.3 -N 500 -V 0 -S 0 -E 20 -H a	0.1	0.3	500	35.978.369	35.027.604
5	MultilayerPerceptron -L 0.1 -M 0.1 -N 500 -V 0 -S 0 -E 20 -H a	0.1	0.1	500	35.978.369	35.572.385
6	MultilayerPerceptron -L 0.1 -M 0.0 -N 500 -V 0 -S 0 -E 20 -H a	0.1	0.0	500	35.978.369	35.846.865

Values Estimated via Multilayer Perception Analysis Model for the Province of Antalya

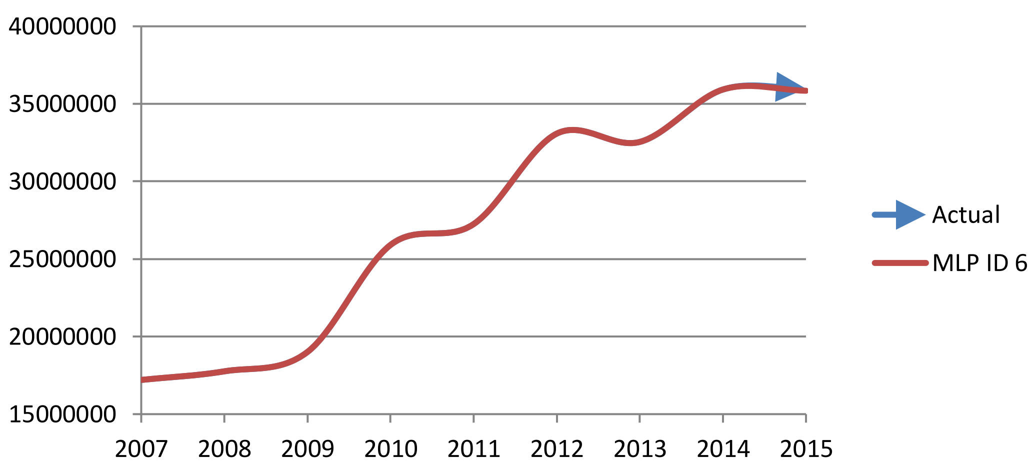

Figure 2: Comparison of Actual Values With The Results Obtained From MLP Model ID 6.

In this graph, the MLP method for Antalya province is analyzed. The values estimated by MLP regression method for Antalya province are compared with the actual values. Scheme Multilayer Perceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H, Learning rate 0.3, Momentum 0.2, Training Time 500 gave the closest result to the real value.

Table 3: Overall Performances of All The Methods For Antalya (Number Of Overnight Stays In 2015)

Model	Rank*	Actual Values	Forecasting Values
Linear	2	35.978.369	35.113.625
MLP	1	35.978.369	35.846.865

***Ranking values have been determined by proximity of estimated values to actual values.**

Total Number of Overnight Stays in the Province of Istanbul (Foregn Tourist)

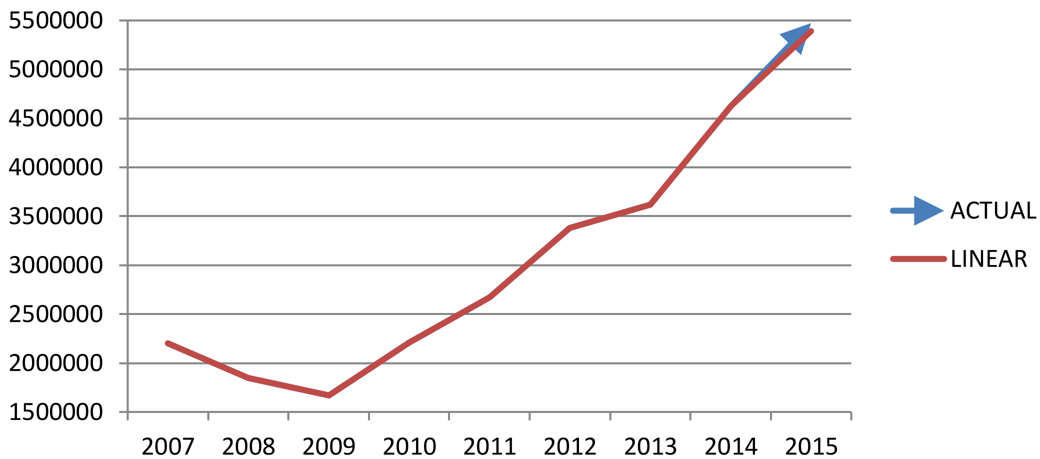

Figure 3: Values Estimated via Linear Regression Analysis Forecasting Model for the Province of Istanbul

Linear Regression Analysis Model Performance for the Province of Istanbul **(Number of overnight stays in 2015)** In this graph, the Linear method for the province of Istanbul is analyzed. The values estimated by the Linear regression method for Istanbul province are compared with the actual values. LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4, Training Time 500, worked best.

Table 4: Forecasting Performance of The Linear Regression Analysis Model Models for Istanbul

Scheme	Training Time	Actual Values	Forecasting Values
LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4	500	5.473.498	5.391.374

Table 5: Forecasting Performance of The MLP-Based Models For Istanbul.

MLP Model ID	Scheme	Learning Rate	Momentum	Training Time	Actual Values	Forecasting Values
1	MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a	0.3	0.2	500	5.473.498	4.420.980
2	MultilayerPerceptron -L 0.4 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a	0.4	0.2	500	5.473.498	4.343.556
3	MultilayerPerceptron -L 0.4 -M 0.3 -N 500 -V 0 -S 0 -E 20 -H a	0.4	0.3	500	5.473.498	4.138.235
4	MultilayerPerceptron -L 0.2 -M 0.1 -N 500 -V 0 -S 0 -E 20 -H a	0.2	0.1	500	5.473.498	4.668.080
5	MultilayerPerceptron -L 0.1 -M 0.1 -N 500 -V 0 -S 0 -E 20 -H a	0.1	0.1	500	5.473.498	5.029.409
6	MultilayerPerceptron -L 0.1 -M 0.0 -N 500 -V 0 -S 0 -E 20 -H a	0.1	0.0	500	5.473.498	5.029.193

Values Estimated via Multilayer Perception Analysis Model for the Province of Istanbul.

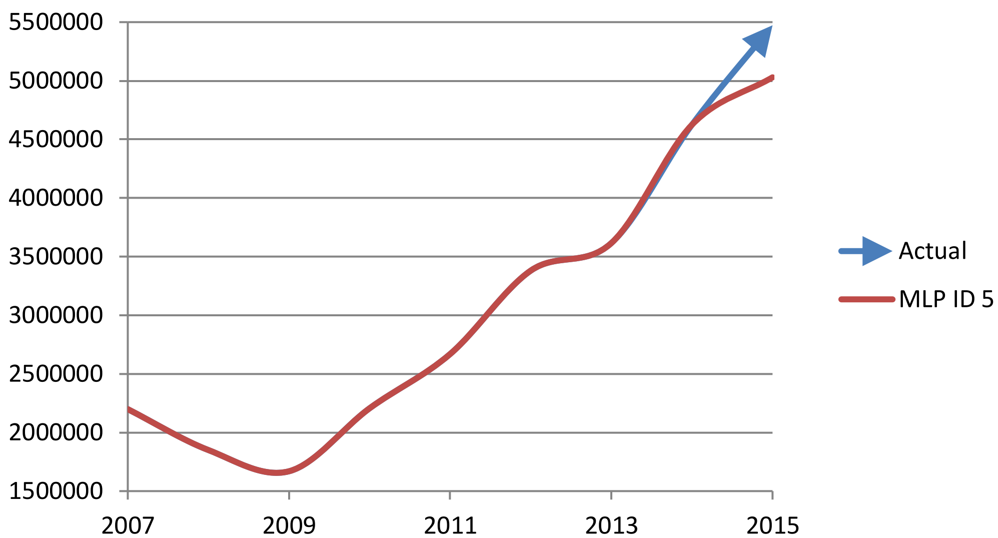

Figure 4: Comparison of Actual Values With The Results Obtained From MLP Model ID 5.

In this graph, the MLP method for Istanbul province is analyzed. The values estimated by MLP regression method for Antalya province are compared with the actual values. Scheme MultilayerPerceptron -L 0.1 -M 0.0 -N 500 -V 0 -S 0 -E 20 -H a, Learning rate 0.1, Momentum 0.0, Training Time 500 gave the closest result to the real value.

Table 6: Overall Performances of All The Methods For Istanbul (Number of Overnight Stays In 2015)

Model	Rank*	Actual Values	Forecasting Values
Linear	1	5.473.498	5.391.374
MLP	2	5.473.498	5.029.409

*Ranking values have been determined by proximity of estimated values to actual values.

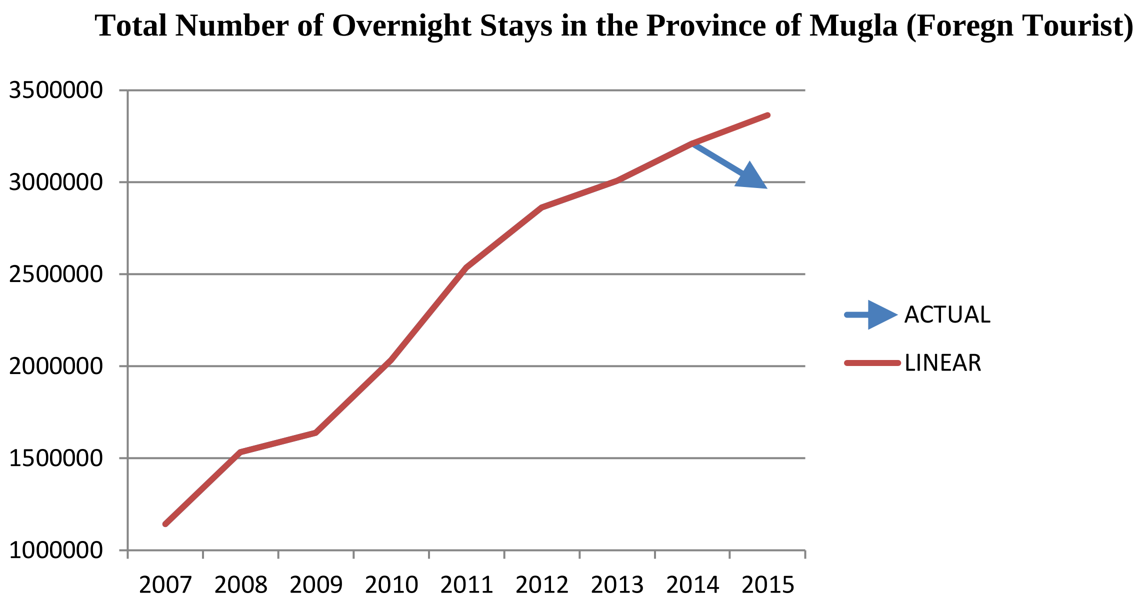

Figure 5: Values Estimated via Linear Regression Analysis Forecasting Model for The Province of Mugla

Linear Regression Analysis Model Performance for the Province of Mugla **(Number of overnight stays in 2015)** In this graph, the Linear method for the province of Mugla is analyzed. The values estimated by the Linear regression method for Mugla province are compared with the actual values. LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4, Training Time 500, worked best.

Table 7: Forecasting performance of the Linear Regression Analysis Model models for Mugla.

Scheme	Training Time	Actual Values	Forecasting Values
LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4	500	2.966.021	3.364.982

Table 8: Forecasting Performance of The MLP-Based Models for Mugla.

MLP Model ID	Scheme	Learning Rate	Momentum	Training Time	Actual Values	Forecasting Values
1	MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a	0.3	0.2	500	2.966.021	3.286.003
2	MultilayerPerceptron -L 0.4 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a	0.4	0.2	500	2.966.021	3.247.641
3	MultilayerPerceptron -L 0.5 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a	0.5	0.2	500	2.966.021	3.244.282
4	MultilayerPerceptron -L 0.6 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a	0.6	0.2	500	2.966.021	3.222.892
5	MultilayerPerceptron -L 0.7 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a	0.7	0.2	500	2.966.021	3.216.926
6	MultilayerPerceptron -L 0.7 -M 0.3 -N 500 -V 0 -S 0 -E 20 -H a	0.7	0.3	500	2.966.021	3.242.392
7	MultilayerPerceptron -L 0.7 -M 0.4 -N 500 -V 0 -S 0 -E 20 -H a	0.7	0.4	500	2.966.021	3.326.003
8	MultilayerPerceptron -L 0.7 -M 0.1 -N 500 -V 0 -S 0 -E 20 -H a	0.7	0.1	500	2.966.021	3.202.422
9	MultilayerPerceptron -L 0.7 -M 0.0 -N 500 -V 0 -S 0 -E 20 -H a	0.7	0.0	500	2.966.021	3.386.742
10	MultilayerPerceptron -L 0.8 -M 0.1 -N 500 -V 0 -S 0 -E 20 -H a	0.8	0.1	500	2.966.021	3.239.016

Values Estimated via Multilayer Perception Analysis Model for the Province of Mugla.

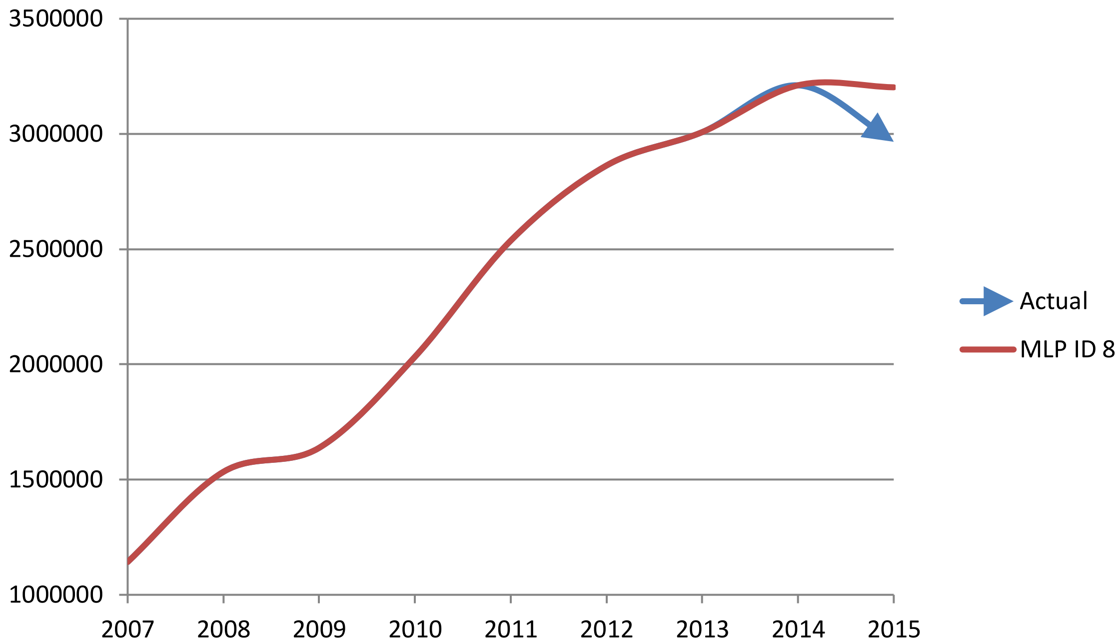

Figure 6: Comparison of Actual Values With The Results Obtained From MLP Model ID 8.

In this graph, the MLP method for Mugla province is analyzed. The values estimated by MLP regression method for Mugla province are compared with the actual values. Scheme MultilayerPerceptron -L 0.1 -M 0.0 -N 500 -V 0 -S 0 -E 20 -H a, Learning rate 0.1, Momentum 0.0, Training Time 500 gave the closest result to the real value.

Table 9: Overall Performances of All The Methods For Mugla (Number of Overnight Stays In 2015)

Model	Rank*	Actual Values	Forecasting Values
Linear	2	2.966.021	3.364.982
MLP	1	2.966.021	5.029.409

***Ranking values have been determined by proximity of estimated values to actual values.**

CONSIDERATION AND CONCLUSION

Different tourism destinations of Turkey have been dealt in this study. MLP analysis method have yielded the results closest to the actual values compared to Linear Regression methods in the estimation study performed on number of overnight stays for the province of Antalya. Linear Regression analysis has yielded the closest result in the in the estimation study performed on number of overnight stays for the province of Istanbul and Multilayer Perception have followed it respectively.

MLP analysis has yielded the closest result in the in the estimation study performed on number of overnight stays for the province of Province of Mugla and Linear analyses have followed it respectively. In conclusion, it has been found in this study performed by multivariate data and different tourism destinations that different regression methods have been used for estimated values to yield the results closest to the actual values and regression analyses have shown variance by tourism destinations and determining the regression method according to the tourism destination planned to be estimated has yielded the estimated values closest to the actual values.

Different regression analysis methods gave the best results for different tourism destinations in forecasting studies. Determined that the regression analysis that gives the best result for the destination of the forecasting study is determined and the closest values can be reached by the regression analysis which gives the result suitable for the destination. In future, more research will be focused to apply more sophisticated forecasting techniques using latest technology. Other advanced data mining techniques to predict visitor arrivals.

REFERENCES

1) Wu, D.,Song, H., &Shen, S. (2016). New Developments in Tourismand Hotel DemandModelingandForecasting. *International Journal of ContemporaryHospitality Management*.

2) Koupriouchina, L.,van der Rest, J. P., &Schwartz, Z. (2014). On revenuemanagementandtheuse of occupancyforecastingerrormeasures. *International Journal of Hospitality Management*, *41*, 104-114.

3) Lim, C.,Chang, C., &McAleer, M. (2009). Forecasting h (m) otel guestnights in New Zealand. *International Journal of Hospitality Management*, *28*(2) , 228-235.

4) Cortés-Jiménez, I.,&Blake, A. (2011). Tourismdemandmodelingbypurpose of visitandnationality. *Journal of Travel Research*, *50*(4), 408-416.

5) Smeral, E. (2010). Impacts of theworldrecessionandeconomiccrisis on tourism: Forecastsandpotentialrisks. *Journal of Travel Research*, *49*(1), 31-38.

6) Athanasopoulos, G.,&Hyndman, R. J. (2008). ModellingandforecastingAustraliandomestictourism. *Tourism Management*, *29*(1), 19-31.

7) Baggio, R.,&Sainaghi, R. (2016). Mapping time seriesintonetworks as a tooltoassessthecomplexdynamics of tourismsystems. *Tourism Management*, *54*, 23-33.

8) Cuhadar, M.,Cogurcu, I., &Kukrer, C. (2014). Modellingandforecastingcruisetourismdemandtoİzmirbydifferentartificialneural network architectures. *International Journal of Business andSocialResearch*, *4*(3), 12-28.

9) Guizzardi, A.,&Stacchini, A. (2015). Real-time forecastingregionaltourismwithbusinesssentimentsurveys. *Tourism Management*, *47*, 213-223.

10) Falk, M. (2014). Impact of weatherconditions on tourismdemand in thepeaksummerseasonoverthelast 50years. *Tourism Management Perspectives*, *9*, 24-35.

11) Song, H., Lin, S., Witt, S. F., &Zhang, X. (2011). Impact of financial/economiccrisis on demandfor hotel rooms in Hong Kong. *Tourism Management*, *32*(1), 172-186.

12) Wu, E. H.,Law, R., &Jiang, B. (2010). Data miningfor hotel occupancy rate: an independentcomponentanalysisapproach. *Journal of Travel &Tourism Marketing*, *27*(4), 426-438.

13) Chen, C. F.,Lai, M. C., &Yeh, C. C. (2012). Forecastingtourismdemandbased on empiricalmodedecompositionandneural network. *Knowledge-BasedSystems*, *26*, 281-287.

14) Claveria, O.,&Torra, S. (2014). ForecastingtourismdemandtoCatalonia: Neuralnetworks vs. time seriesmodels. *EconomicModelling*, *36*, 220-228.

15) Lin, C. J.,Chen, H. F., & Lee, T. S. (2011). Forecastingtourismdemandusing time series, artificialneuralnetworksandmultivariateadaptiveregressionsplines: EvidencefromTaiwan. *International Journal of Business Administration*, *2*(2), 14.

16) Turkish State Institute of Statisticsweb page (www.die.gov.tr)

17) Central Bank of Republic of Turkeyweb page(http://evds.tcmb.gov.tr)

18) European Central Bank (ECB) web page(https://www.ecb.europa.eu)

19) TURSAB web page (www.tursab.org.tr)

20) Naik, A., & Samant, L. (2016). Correlation review of classification algorithm using data mining tool: WEKA, Rapidminer, Tanagra, Orange and Knime. *Procedia Computer Science*, *85*, 662-668.

21) Duriqi, R., Raca, V., & Cico, B. (2016, June). Comparative analysis of classification algorithms on three different datasets using WEKA. In *Embedded Computing (MECO), 2016 5th Mediterranean Conference on* (pp. 335-338). IEEE.

22) Cang, S. (2014). A Comparative Analysis of Three Types of Tourism Demand Forecasting Models: Individual, Linear Combination and Non-linear Combination. *International Journal of TourismResearch*, *16* (6), 596-607.