Forecasting with Multilayer Perceptron Algorithm the Occupancy Rate of Accommodation Establishments in Turkey

Adem Golcuk[1], Mahmut Sami Yasak[2*], Muhammed Bakir Dalmizrak[3]

Abstract: Artificial neural networks (ANN), which is one of the applications of artificial intelligence, is the information processing technology that analyzes the existing data by mimicking the working structure of the human brain and creates new information with different learning algorithms. In recent years, ANN have become popular in scientific and business fields. In the hotel industry, researchers have recently focused on the classification of tourist segments of neural networks and predicting visitor behavior. However, it was requested to include ANN in the hotel occupancy rate forecast. In this paper, forecasted occupancy rate of the hotel in Turkey by using ANN a class that Multilayer Perceptron (MLP) algorithm. Accommodation Statistics" of monthly data between 2000-2018 years obtained from the Republic of Turkey Ministry of Culture and Tourism was used in this study. 3 input values and 1 output value were used in the training of MLP developed for hotel occupancy rate forecast. As a result of the study, a forecast that occupancy rate close to the actual occupancy rate was obtained. İt has been found to have low error rates. The success rate of the algorithm is 91.85 %.

Keywords: Artificial Neural Networks, Hotel Occupancy Rate, Multilayer Perceptron

1. Introduction

In modern societies, it is imperative to predict the future trends and directions of development of social events and to determine them scientifically. Predicting the developments in tourism with the methods based on scientific basis also facilitates the decision-making of managers. Therefore, it will be possible to prepare development methods in the economy and the wastefulness of the financial resources to be allocated by the investors will be prevented and necessary measures will be taken. The tourism sector, like other sectors in the country, needs accurate estimates based on reliable data. In the tourism industry, the fact that requests for tourist services are sensitive to factors and this makes forecast in the industry more important. Hotel businesses are also faced with a changing over time desire. These requests vary on a monthly basis. In addition, accommodation demands vary according to political, social, and economic conditions [1].

The occupancy rate indicates the success of the hotel in the sale of rooms, which make up the bulk of the revenue. In addition, the occupancy rate is an important element in the management and supervision of the hotel. Planning works begin with estimating the income to be obtained and the number of rooms occupied. The estimated number of workers, estimated lighting and heat costs are then determined based on the occupancy rate. This makes it easier to find the cost and estimated revenue of the hotel [2]. Room occupancy rates are calculated annually and monthly. In this study, hotel accommodation statistics obtained from the Republic of Turkey Ministry of Culture and Tourism were used as occupancy rate. The purpose of this study is forecast with using ANN that occupancy rate of business certified hotels of The Ministry of Culture and Tourism in Turkey.

The handleable of the study can be discussed in three parts. First, modeling with neural networks of the demand for tourism and hotels in Turkey. Secondly is determine the model providing the highest accuracy by comparing the occupancy rates of forecast accuracy retroactively applied the model.

Third is forecasts for 2018 year and 2019 first 2 months as monthly of demand for tourism and hotels in Turkey through the specified model. Also, to create the infrastructure for the touristic development plans prepared with the forecasted made, to facilitate the making it easier decisions of the hotel managers in the annual planning are the aims of this study.

As a result of research done on the subject, in development countries has related works such as the total of people coming to the hotels, the number of overnight stays, average stay and occupancy rate forecast using ANN. But, seen to be the scarcity of works on this subject in Turkey. For example, in a study performed by Çuhadar and Kayacan, they estimated the occupancy rate in accommodation establishments by using ANN [1]. In foreign countries works Law worked occupancy rates of hotel establishments operating in Hong Kong [3], Law and Au worked request travel of Japanese for the travel to Hong Kong with ANN [4], In another study by Law, an extrapolation time series forecasting technique predicts future hotel occupancy rates [5], Palmer et. al., worked forecast the tourism time series with ANN [6], Youn and Gou worked occupancy rates of hotels in Korea with ANN [7], Zakhary et al., worked forecast hotel occupancy rate using Monte Carlo simulation [8], Yang et. al., designed a website and tried to estimate the most appropriate hotel assessment according to the wishes of the customers with ANN [9].

The advantages of ANN can be summarized like that. ANN can learn from previous experience, and once they are trained, they can give quick and accurate answers to missing and incorrect data. Failure of one or more cells of the ANN does not prevent the system from producing output. A network undergoes slow and relative deterioration over time. In this case, the network does not

[1] Biomedical Eng., Faculty of Tech., Selcuk University, Konya – 42002, TURKEY, ORCID ID: 0000-0002-3178-9118
[2] The Degree of Master of Science in Mechatronics Eng., Selcuk University, Konya – 42002, TURKEY, ORCID ID: 0000-0003-4444-161X
[3] The Degree of Master of Science in Mechatronics Eng., Selcuk University, Konya – 42002, TURKEY ORCID ID: 0000-0002-3828-4577
Corresponding Author Email: mhmtsmyasak@gmail.com

break down as soon as the problem occurs. They can learn the ANN events and comment on similar events. ANN can perform multiple tasks simultaneously and produce fast solutions [10].

Apart from the advantages mentioned above, ANN has disadvantages which are not suitable for some situations. The disadvantages of ANN can be summarized like that. The most important problem for ANN is that when creating a solution for a problem and it cannot explain the why and how. This reduces trust in the network. There is no specific rule in determining the structure of ANN. The proper network structure is achieved by trial and error and experience. ANN can be applied in many areas but it is always wrong to think that it works correctly. In some cases, they may produce incorrect results depending on the educational data.

2. Material and Method

In this study, between 2000 and 2018 years hotel accommodation data obtained from Republic of Turkey Ministry of Culture and Tourism were used. To reach these data, needs to be create a request on the official website of the Republic of Turkey Ministry of Culture and Tourism. Then, these data was sent to our e-mail address specified on the request screen by the Republic of Turkey Ministry of Culture and Tourism. The total number of people coming to the hotel that certified by Republic of Turkey Ministry of Culture and Tourism can be shortened as follows - "Total Number of People Incoming to Facility" (TNPIF). The total number of overnight stays of the people staying in the hotels is expressed as follows - "Total Overnight" (TO). The average monthly stay of the people staying in hotels is expressed as follows - "Average Stay" (AS). Monthly occupancy rate of hotels is expressed as follows - "Occupancy Rate" (OR). Some of the data used for training in the study are shown in Table 1.

Table 1. Accommodation statistics between 2000-2018 years

MONTHS	TNPIF	TO	AS	OR
JANUARY	1162865	2116716	1.82	19.3
FEBRUARY	1376687	2372260	1.72	21.64
MARCH	1552444	2847586	1.83	24.68
APRIL	1797620	3833062	2.13	26.32
MAY	2081302	5028740	2.42	29.48
JUNE	2442872	6406214	2.62	33.99
JULY	2796218	8087654	2.89	42.8
AUGST	3059047	8665771	2.83	45.88
SEPTEMBER	2615011	7013605	2.68	37.25
OCTOBER	2513745	6499280	2.59	37.72
NOVEMBER	1968000	3994404	2.03	28.44
DECEMBER	1664844	3340404	2.01	24.51
JANUARY	1470269	2684955	1.83	22.98
FEBRUARY	1570544	2918513	1.86	25.04
MARCH	1736408	3617893	2.08	29.37
APRIL	1977159	4773867	2.41	31.60
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
JULY	7782271	21743431	2.79	67.04
AUGST	7944599	22841211	2.88	70.37
SEPTEMBER	6882227	19175508	2.79	58.75
OCTOBER	5414224	15078774	2.79	46.89
NOVEMBER	4101468	8465296	2.06	33.73
DECEMBER	4010187	7961341	1.99	34.92

Table 1 shows the 12 months of 2000, the first 4 months of 2001 and the last 6 months of 2017. There are 192 months of data between January 2000-December 2006, January 2008-December 2017 for training, (Data for 2007 is not available at the Republic of Turkey Ministry of Culture and Tourism.) 12 months data of 2018 and 2 months data of 2019 for validation, and totally 206

months data are available. Thus, hold-out technique is used to in order to make split the data. These data has split by 93.2 % for training and 6.8 % for validation. The reason of these split criteria is the real data for 2018 much should be compare with predicted values of 2018 because of this study to making forecasting for Occupancy Rate only 2018 year. Also 2018 year and first two month belongs of 2019 data constitutes are 6.8 % of all data. TNPIF and TO data units are millions, AS data units are daily, OR is expressed as a percentage. The input values of MLP, which is developed for hotel occupancy rate forecast are made up of data from TNPIF, TO and AS. Output values consist of OR data.

2.1. Dataset Normalization

The input and output values were normalized in the range 0 and 1 to shorten training time and reduce the number of correlations of output neurons before giving them to the designed MLP model. Normalization operation is calculated using the following formula [11].

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

In equation (1), where x is represents the input and output values, x_{min} is minimum of the input and output values, x_{max} is maximum of the input and output values and x' is represents the obtained normalized values. The data normalized according to equation (1) is shown in Table 2.

Each normalized input block was prepared in 1x204 resolution by making it a single dimension or in other words a horizontal array. For this purpose, matrix transposing process was applied to the vertical blocks of data. As a result, the size space of the input data is set to a resolution of 3x204.

Table 2. Normalized data

MONTHS	TNPIF	TO	AS	OR
JANUARY	0	0	0.05235602	0
FEBRUARY	0.03019469	0.01197387	0	0.04581946
MARCH	0.05501407	0.03424593	0.05759162	0.10534560
APRIL	0.08963640	0.08042179	0.21465968	0.13745839
MAY	0.12969632	0.13644695	0.36649214	0.19933424
JUNE	0.18075513	0.20099042	0.47120418	0.28764440
JULY	0.23065259	0.27977665	0.61256544	0.46015273
AUGST	0.26776777	0.30686513	0.58115183	0.52046211
SEPTEMBER	0.20506360	0.22945058	0.50261780	0.35147836
OCTOBER	0.19076340	0.20535116	0.45549738	0.36068141
NOVEMBER	0.11369647	0.08798169	0.16230366	0.17897004
DECEMBER	0.07088655	0.05733761	0.15183246	0.10201684
JANUARY	0.04340980	0.02662563	0.05759162	0.07205796
FEBRUARY	0.05757005	0.03756932	0.07329842	0.11239475
MARCH	0.08099240	0.07033974	0.18848167	0.19718034
APRIL	0.11498985	0.12450452	0.36125654	0.24084589
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
JULY	0.93475396	0.91963719	0.56020942	0.93479537
AUGST	0.95767698	0.97107521	0.60732984	1
SEPTEMBER	0.80765499	0.79931357	0.56020942	0.77246916
OCTOBER	0.60035215	0.60735536	0.56020942	0.54023888
NOVEMBER	0.41497240	0.29747160	0.17801047	0.28255335
DECEMBER	0.40208223	0.27385808	0.14136125	0.30585470

2.2. Multilayer Perceptron Model

The MLP model that is used to solve multi-class pattern recognition problems, where the output of each neuron in the previous layer feeds the input of each neuron in the next layer, this is an ANN class with feedforward [12]. MLP is a feedforward neural network with one or more layers between the input and

output layers. Feedforward means data flows from the input to the output layer in the forward direction [13][14]. MLP is widely used for pattern classification, recognition, forecast, and approach. MLP can solve problems that are not linearly separated. An example MLP model is shown in Figure 1.
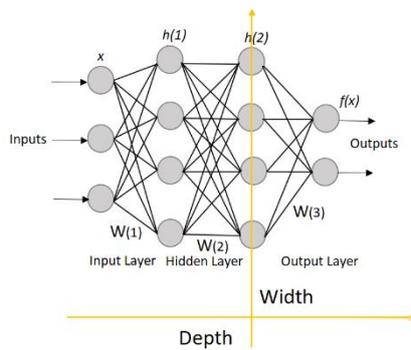

Figure 1. General structure of Multilayer Perceptron models [15]

In Figure 1 shown, expression x refers to the input neurons in the input layer, $f(x)$ the output neurons in the output layer, and h denotes hidden neurons in the hidden layers. W refers to the weight values of neurons obtained. The values obtained from the output neurons consist of the sum of the values and weights of the previous neurons [16]. The resulting neuron output values are multiplied by activation functions to introduce nonlinear properties, enabling the model to learn more powerfully and efficiently [17].

In this study, the design criteria of the MLP model were determined by 3 different input and 1 output values that vary with time and depend on different variables between each other. The designed model has 3 dimensional input space, one dimensional output space. There are 2 hidden neuron layers and 1 output layer. While there are 100 neurons in the hidden layers, there are 1 neurons in the output layer to give the desired OR values. Figure 2 shows the MLP model, w indicates the weight values of neurons obtained after the procedure, b indicates the bias value.
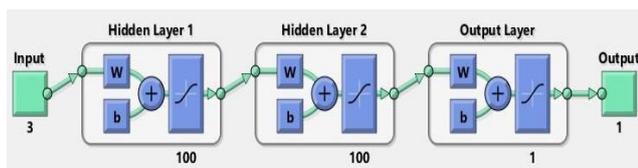

Figure 2. Designed MLP model

TNPIF, TO, AS data, which are the input values of the model, are given to the input layer of the 3x204 dimension model, and the output value of OR is 1x204 dimension of the model's output layer, in other words, output data (OR) is made ready for training. The training parameters of the model occurs which are Learning-rate is 0.01, momentum value is 0.9, detrimental value is 0.7, the activation function that is multiplied by the neuron output values of the layers has been determined as Sigmoid function [18] Stochastic Gradient Descent (SGD) function [19], the number of epoch 1000 is determined. After the specified parameter settings, the model was started to train. According to the input data, the weighted values and neuron values calculated by the following equation (2) that calculates the output of the neurons in the output layer.

$$d(x) = \sum_{i=1}^{n} w_i x_i + w_{n+1} \tag{2}$$

In equation (2), where $d(x)$ is output value of neuron, x_i is neuron value, w_i is weight value, w_{n+1} is refer to weight value of bias. The resulting neuron output value is multiplied by the activation function to produce output data by MLP. The output data were compared with the OR data which is the output data and error calculations were performed. These error calculations, Mean Square Error (MSE), Root Mean Square Error (RMSE), Percentage of Absolute Error Rates (MAPE) these were used in the calculation of performance values of the forecast results are given. These error metrics are calculated by the following equations.

$$MSE = \frac{\sum_{i=1}^{n}(y_i - y'_i)^2}{n} \tag{3a}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - y'_i)^2}{n}} \tag{3b}$$

$$MAPE = 100 \left(\frac{\sum_{i=1}^{n} \left| \frac{y_i - y'_i}{y_i} \right|}{n} \right) \tag{3c}$$

In equation (3a), (3b) and (3c), where y_i is actual values, y'_i is predicted values, i is months to forecast, n is refer to the number of months to forecast. The calculated error metrics were determined above equations as MSE=24.52, RMSE=4.95, MAPE=8.15%. In stated that accuracy of forecast rate as "high rate" in cases where MAPE value is less than 10%, and "correct rate" in cases between 10%-20% [1][20][21]. In the light of the obtained findings, whole 12 months of 2018 and first two months of 2019 were predicted and the results are shown in Table 3 as the actual OR and the forecasted OR of the months of 2018 and 2019. Also, Table 3 data is graphed and shown in Figure 3.

Table 3. Actual and forecasted hotel occupancy rates for the 12 months of 2018 and the first 2 months of 2019

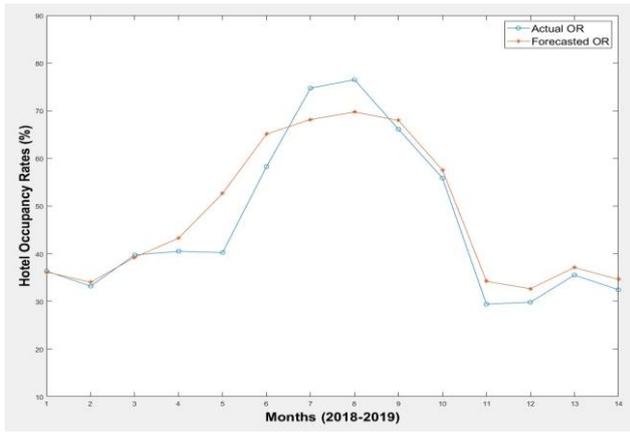| MONTHS	ACTUAL OR (%)	FORECASTED OR (%)
JANUARY	36.33	36.15
FEBRUARY	33.18	33.99
MARCH	39.73	39.23
APRİL	40.50	43.27
MAY	40.23	52.64
JUNE	58.27	65.11
JULY	74.73	68.16
AUGUST	76.49	69.72
SEPTEMBER	66.11	67.97
OCTOBER	55.86	57.55
NOVEMBER	29.41	34.20
DECEMBER	29.81	32.65
JANUARY	35.51	37.10
FEBRUARY	32.43	34.61

Figure 3. The graph of actual hotel OR data and forecasted OR data (Table 3. graph)

The variation of the MSE values obtained during the training of the model, train, validation and test data according to the number of 1000 training epochs is shown in Figure 4. In the training process, the best MSE = 0.0020 value was obtained at 240th epoch. This result was obtained according to normalized input and output values.
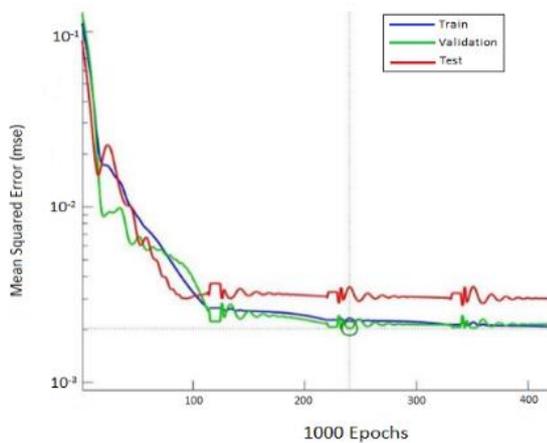

Figure 4. Decreasing MSE performance graph by number of training epochs

The performance values of the forecast results obtained in this study and the performance values of the other studies used in the study are given in Table 4. In Çuhadar's study, there are 5 input values and 1 output value for the training of the model [1], while in Law's study there are 6 input values and 1 output value for the training of the model [4]. In Table 4, the MSE values of the other studies are lower than proposed method due to the higher training data sets and consequently, the high correlation between the input data of the neurons for the MLP used.

Table 4. Comparison with other methods

METHODS	MSE	RMSE	MAPE (%)
OURS	24.52	4.95	8.15
ÇUHADAR	3.6	1.9	4
LAW	20.28	5.28	4.77

3. Discussion and Conclusion

In this study, the occupancy rate of hotels certified by the ministry enterprises in Turkey were forecasted by ANN. The data obtained from Republic of Turkey Ministry of Culture and Tourism were used to model of ANN. The data were obtained by selecting the tourism statistics demand over the official website of the Republic of Turkey Ministry of Culture and Tourism under the title of accommodation statistics demand.

These data were to the training dataset into between 2000-2017 to train the artificial neural network and to the validation data to examine the predictive accuracy level of the 2018 data. The output of the MLP model is compared with the Republic of Turkey Ministry of Culture and Tourism data, the forecasted results were found to be efficient. This study demonstrated the applicability of the ANN model estimated hotel occupancy rates in Turkey. To forecasting the occupancy rate of hotels in Turkey, the source for the business manager of the hotel will help to make a spending plan. Also, in future studies, it is seen that occupancy rates of hotels operating within a certain region or province limits, tourist spending in hotels, number of overnight stays and average lengths of stay can be estimated by using more variables with ANN.

The general conclusion to be drawn from the study; If the values of the independent variables are known and model have enough data to training, it is seen that MLP can make very good predicted. Although the forecasted OR values contain a high degree of accuracy, further works can be performed to test the applicability of the model. In this study, 3 independent variables were used for input value in forecasting hotel occupancy rate. In future studies, if more input values and more years' statistical data are obtained and used for the training of MLP model, higher forecasted, in other words, lower MSE and MAPE performance values could be obtained.

Training and forecasting of the model are performed in Matlab R2018a software.

References

[1] M. Çuhadar, and C. Kayacan, "Yapay Sinir Ağları Kullanılarak Konaklama İşletmelerinde Doluluk Oranı Tahmini: Türkiye'deki Konaklama İşletmeleri Üzerine Bir Deneme", *Anatolia: Turizm Araştırmaları Dergisi.*, vol. 16, no. 1, 2005.

[2] M. Emeksiz, Beş Yıldızlı Otel İşletmeleri İçin Getiri Yönetimi Uygulama Modeli. Eskişehir: Anadolu Üniversitesi, 2002.

[3] R. Law, "Room occupancy rate forecasting: a neural network approach," *International Journal of Contemporary Hospitality Management.,* vol. 10, no. 6, pp. 234-239, Nov. 1998.

[4] R. Law, and N. Au, "A neural network model to forecast Japanese demand for travel to Hong Kong," *Tourism Management.*, vol. 20, no. 1, pp. 89-97, Feb. 1999.

[5] R. Law., "Initially testing an improved extrapolative hotel room occupancy rate forecasting technique," *Journal of Travel & Tourism Marketing*, vol. 16, no. 2-3, pp. 71-77, 2004.

[6] A. Palmer, J. J. Montano, and A. Sesé., "Designing an artificial neural network for forecasting tourism time series," *Tourism management,* vol. 27, no. 5, pp. 781-790, Oct. 2006.

[7] H. Youn, and Z. Gu., "Predicting Korean lodging firm failures: An artificial neural network model along with a logistic regression model," *International Journal of Hospitality Management.*, vol. 29, no. 1, pp. 120-127, Mar. 2010.

[8] A. Zakhary, A. F. Atiya, H. El-Shishiny, and N. E. Gayar, "Forecasting hotel arrivals and occupancy using Monte Carlo simulation," *Journal of Revenue and Pricing Management.*, vol. 10, no. 4, pp. 344-366, July. 2011.

[9] Y. Yang, J. Tang, H. Luo, and R. Law, "Hotel location evaluation: A combination of machine learning tools and web GIS," *International Journal of Hospitality Management.,* vol. 47, pp. 14-24, May. 2015.

[10] M. M. Saritas, and A. Yasar, "Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification", *International Journal of Intelligent Systems and Applications in Engineering.,* vol. 7, no. 2, pp. 88-91, May. 2019.

[11] Guray Sonugur, Turkey [Online]. Available:https://guraysonugur.aku.edu.tr/2017/03/07/normalizasyon-islemi-yapan-fonksiyon/, Accessed on: Jun. 29, 2019

[12] J. Mao, and A. K. Jain, "Artificial neural networks for feature extraction and multivariate data projection." *IEEE transactions on neural networks*, vol. 6 no. 2, pp. 296-317, Mar. 1995.

[13] D. E. Rumelhart, E. H. Geoffrey, and J. R. Williams, "Learning internal representations by error propagation," *California Univ San Diego La Jolla Inst for Cognitive Science*, 1985.

[14] R. P. Lippmann, "An introduction to computing with neural nets," *IEEE Assp magazine*, vol. 4, no. 2, pp. 4-22, July 1988.

[15] Discover, Share, Learn, USA [Online]. Available: https://www.slideshare.net/xavigiro/multilayer-perceptron-dlai-d1l2-2017-upc-deep-learning-for-artificial-intelligence?fbclid=IwAR2xE8VutlbRYrNpcg5dpHalgIXDk_yczbv_X_zvShljzWSdQHRfGi8zNg, Accessed on: Jun. 26, 2019

[16] E. Alpaydın, "Introduction to Machine Learning", 2nd ed. vol. 4., E. Kocabıyık, Ed., Istanbul, Turkey, Bogazici University Press, 2018, pp. 200.

[17] S. Hashem, "Sensitivity analysis for feedforward artificial neural networks with differentiable activation functions." [Proceedings 1992] *IJCNN International Joint Conference on Neural Networks,* vol. 1. pp. 419-424, IEEE, June 1992.

[18] S. Saruchi, Adaptive sigmoid function to enhance low contrast images. *International Journal of Computer Applications*, vol.55 no.4, pp. 45-49, Oct. 2012.

[19] N. Qian, "On the momentum term in gradient descent learning algorithms." *Neural networks* vol. 12, no.1, pp. 145-151, Jan.1999.

[20] S. F. Witt, and C. A. Witt. "Modeling and forecasting demand in tourism," Academic Press Ltd, 1992, pp. 25.

[21] C. D. Lewis, "Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting," Butterworth Heinemann, 1982, pp. 509.