Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 11 NO: 2 PAGE: 564-573 YEAR: 2022
DOI: 10.17798/bitlisfen.1059772

Order Demand Forecast Using a Combined Approach of Stepwise Linear Regression Coefficients and Artificial Neural Network

Serdar GÜNDOĞDU[1*]

[1]*Department of Computer Technologies, Bergama Vocational School, Dokuz Eylül University, İzmir (ORCID: 0000-0003-2549-5284)*

Keywords: Feature selection, Forecast, Linear regression, Neural network.

Abstract
Nowadays, businesses' forecasts to meet the demands have become more critical. This study aimed to predict the fifteen-day order demand for an order fulfillment center using a Multilayer Perceptron Neural Network (MLPNN). The dataset used in the study was created from a real database of a large Brazilian logistics company and thirteen variables. Linear Regression Coefficients (LRC) were used as a feature selection method to reduce estimation errors. The study showed that among the variables, order type_A (A5), order type_B (A6), and order type_C (A7) had the most significant impact on total order forecasting. The effect of A6 was found to be greater than the effect of A7 and A5. The performance of the proposed model was evaluated using the mean absolute percent error (MAPE). LRC-MLPNN provided a MAPE of 2.97%. The results showed that better forecasting performance was obtained by selecting the independent variables to be used as input to the forecasting model with LRC. The proposed model can also be applied to different estimation problems.

1. Introduction

Companies operate in a business world where globalization, increased competition, and technology are rapidly developing. The agility of consumer responsiveness, key to the company's success, is critical for forecasting demands, effective supply chain management, and managing customer needs [1], [2]. Obtaining price and demand information has always been the most important commercial purpose. Demand and sales information affects production scheduling, inventory control, and delivery plans. It is an important factor in inventory management, with the objective of better matching supply and demand to reduce inventory costs and stock out [2]. Forecasting demands accurately is truly a challenging task for the supply chain [3]. Predicting using historical data is an important auxiliary tool for decision-making. In particular, the purpose of prediction in time series is to predict the behavior of complex systems by looking at past models of the phenomenon [4]. There is extensive literature on sales forecasting in commercial industries such as books [5], electronics [6], and textiles and clothing fashion [7], [8]. However, few studies focus on demand forecasting in the industrial valve sector, which is characterized by a combination of manufacturers and manufacturing industries [9].

Demand forecasting methods can be linear and nonlinear [10], [11]. Linear methods use univariate time series analysis, such as autoregressive integrated moving averages (ARIMA), multi-linear regression methods (MLR), and exponential smoothing [12], [13], [14]. Nonlinear methods use multiple nonlinear regression (MNLR), artificial neural networks (ANNs) [12], [15], [16], [17], fuzzy logic (FL) [18], [19], support vector machine (SVM) [20], genetic algorithm (GA), expert systems [21] and hybrid methods.

ANNs, widely used as forecasting tools, offer a way to make smart decisions today [22]. ANN algorithms are essential methods for demand forecasting due to their capabilities in nonlinear data and their superior forecasting performance [3], [23]. Many researchers' studies indicate the need for a demand forecast. Kuo and Xue [24] studied the ANN demand forecasting application in a beverage company [9]. The results of this study showed that the

[*]Corresponding author: *serdar.gundogdu@deu.edu.tr*
Received: 18.01.2022, Accepted: 03.06.2022

predictive ability of ANNs is better than ARIMA. Law used ANNs to demand forecasts in the tourism industry [25]. Ying and Hanbin [26] studied demand forecasting based on ANN and developed a three-layer ANN model for forecasting market demand [27]. Several examples of ANNs-based applications: financial failure [28], wind speed [29], foreign exchange rates [30], intraday electricity demand [31], and ATM cash demand [32], orders treatment center [33], semiconductor supply chain[34], heat demand forecasting at city level [35].

Feature selection (FS) plays an essential role in neural network (NN) based approaches. It is commonly associated with feature size, affecting machine learning and optimization problems run by backward induction. FS is a data preprocessing method that speeds up the learning process by reducing the number of features of the data set and facilitating the understanding and management of the data set, see [36]. It has been proven to be an effective initial step for different machine learning and data mining problems. FS aims to create simpler and more understandable models, improve data mining performance, and prepare clean data [37]. Acquiring different features requires additional costs such as money, time, and other resources. Too many features mean a high cost. Therefore, it is an important indicator in terms of reducing the number of selected features [38]. The advantages of FS can be summarized as follows: improving the prediction performance of the predictors and selecting faster and more efficient predictors. SVD (single value decomposition), PCA (principal component analysis), and Sammon's mapping are examples of size reduction methods [39].

This study aims to remove irrelevant features from the high-dimensional data set, identify the most relevant features, improve the accuracy of ANN estimation, and reduce the computational cost with LRC feature selection in daily order demand forecasting.

2. Material and Method

The dataset used was accessed from the UCI machine learning repository [40]. It was donated by Ricardo Pinto Ferreira et al. [33], Brazil. The purpose of the dataset is to predict daily demand orders for a company. The dataset consists of twelve predictors and a target (total orders) variable. Attributes are as follows (Table 1):

Table 1. Data used in the study

Attributes	Labels
Total orders (Target)	**A13**
Week of the month	**A1**
Day of the week	**A2**
Non-urgent order	**A3**
Urgent order	**A4**
Order type_A	**A5**
Order type_B	**A6**
Order type_C	**A7**
Fiscal sector orders	**A8**
Order from the traffic controller	**A9**
(1) Banking orders	**A10**
(2) Banking orders	**A11**
(3) Banking orders	**A12**

The proposed approach consists of three stages.

- Nine feature subsets (X1 to X9) were created for the DDFO (daily demand forecasting orders) dataset to construct the prediction model (Table 4). The LRC method was used for feature selection, and features (X9) were reduced to sizes smaller than twelve.
- Daily treatment orders were estimated by applying nine different input variations to the MLPNN system.
- The results were compared according to different inputs.

The block diagram of the proposed approach is shown in Figure 1.

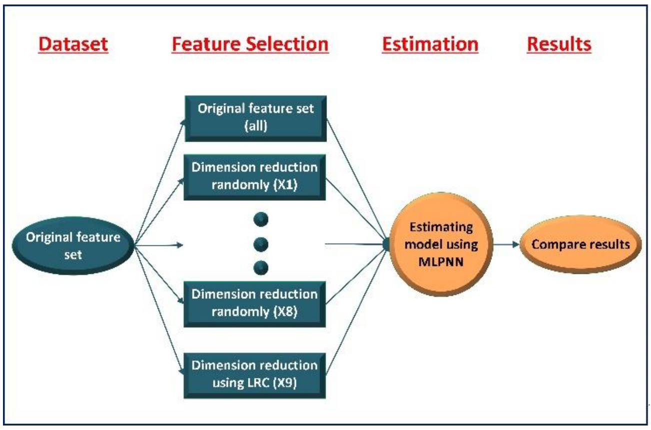

Figure 1. The block diagram of the proposed system

2.1. Multiple linear regression (MLR)

Multiple linear regression is a statistical technique that attempts to model the correlation between independent and dependent variables that depend on a linear equation [41]. The MLR model is:

$$y = c_0 + c_1x_1 + c_2x_2 + \ldots + c_kx_k + e \quad (1)$$

Here, y, c_0, x_k, c_k, and e denote the dependent variable, intersection point, independent variable, regression coefficients, and randomly measured errors.

Regression analysis evaluates the relationship of independent variables with a dependent variable. A linear regression equation with a stepwise method is sensitive to outliers to avoid inefficient measurement values. It is an extension of multiple linear regression analysis to estimate causal relationships that have been predetermined based on theory [42]. One of the methods used to find the correlation with the independent variables is the multicollinearity (MCL) test. MCL can be evaluated with variance inflation factors (VIFs), which determine the strength of the relationship between independent variables. Statistical software calculates a VIF value for each independent variable. VIF values higher than 4 indicate that there may be a problem [43].

Evaluating the regression beta coefficients (βs) is invaluable. These coefficients can be positive or negative, and a t-value associated with each can be calculated. The β is the degree of change in the dependent variable for changes in the independent variable. The t-test evaluates whether the β is different from zero. If the β is not statistically significant (if the t value is not significant), the relevant variable cannot significantly predict the outcome. If the β is significant, its sign is examined. If the coefficient is positive, there is an increase in the dependent variable as much as the β value for each unit increases in the independent variable [44].

The interactions of dependent and independent variables were analyzed and investigated using linear regression analysis by the IBM SPSS Statics 24 program.

2.2. Multiple-layer perceptron neural network

A typical MLPNN architecture consists of the input, hidden, and output layers. The output layer is fully connected to the hidden layers, and the hidden layers are fully connected to the input layer. Each layer has a weight matrix (w), a bias vector (b), and an output vector (y). If we examine the equation in Equation 2 used in problem-solving with ANNs; x_1, x_2,…..x_n, input data, w_1, w_2,… w_i, weights, the number of b is a bias, and σ is the activation function.

$$y = \sigma(\sum_{i=0}^{n} x_i \bullet w_i + b) \quad (2)$$

The neural network exposes the training data and calculates errors based on its outputs. These errors are used to adjust the biases and weights [45].

2.2.1. Design of MLP neural network

MLR aims to model the linear relationship between the independent variables and a dependent variable. The MLR equation can be written in terms of the independent variables (A1, A2,…A12) and dependent variable (A13) used in the study as follows (Equation 3).

$$A_{13} = c_0 + c_1A_1 + c_2A_2 + \ldots + c_{12}A_{12} + e \quad (3)$$

The MLR stepwise regression is a simple and popular feature selection method. It determines the suitable features used as input features to train the neural network [46].

Predictors of total orders (A13) were assessed using stepwise linear regression analysis. All independent variables listed in Table 1 were entered into the model as potential predictors (A1, A2,.....A12). A value of $p < .001$ was considered statistically significant. After stepwise linear regression analysis, A5 (Order type_A), A6 (Order type_B), and A7 (Order type_C) were found to be three independent predictors of the dependent variable (β_{A5}, β_{A6}, β_{A7}=0.566, 0.463, 0.210; $p<0.001$). A new formula (Equation 4) has emerged by updating Equation 3 between the three significant independent and dependent variables.

$$A_{13} = c_0 + c_5A_5 + c_6A_6 + c_7A_7 + e \quad (4)$$

The proposed method determined the most accurate model inputs to get the best performance. The resulting three-input ANN equation is presented below (Equation 5).

$$A_{13} = \sigma(\sum_{i=5}^{7} A_i \bullet w_i + b) \quad (5)$$

The ANN in the study is a fully connected multilayer perceptron neural network (MLPNN), whose structure is shown in Figure 2. The proposed model is constructed to forecast total orders (A13) using the MatLab R2021a. The research model of this study consists of a structure with three hidden layers with LRC-selected feature subsets in a network.

Figure 2 shows the interconnections between the input and output parameters of the MLPNN to predict the fifteen-day order demand for an order fulfillment center. Neurons present in the input and output layers equal the number of input and output parameters, respectively [47]. Finding the optimum number of hidden layers and neurons helps design the best architecture [48]. The training datasets are used to obtain the optimum architecture of the model with the appropriate number of hidden layers, neurons in the hidden layers, momentum term, learning rate, and the number of iterations. The dataset is split into training, validation, and test datasets to perform model training. Training continues until the Sum Squared Error (SSE) value reaches the minimum value. The necessary steps during model training are as follows: (i) selection of hidden layers and neurons in the hidden layers, (ii) selection of learning rate and momentum constant, and (iii) selection of iteration numbers.

The data used to test the models was obtained from an order fulfillment center. The parameters optimization of the proposed MLPNN model was performed using a random search and trial and error approach. The result showed that the architecture of the best MLPNN model has an optimal parameter on a momentum constant of 0.99, a learning rate of 0.95, and three hidden layers. The training dataset is divided into 70%, 15%, and 15%, respectively, for each model's training, validation, and testing data. There are 30, 60, and 15 neurons in the created network structure for the first, second, and third hidden layers. Neural networks were trained using the Levenberg-Marquardt algorithm. The trained MLPNN's validity was evaluated using SSE. It has been shown that the smallest error gives more accurate results in the estimation, especially helping the senior manager to make an effective decision for the operation of the business by considering the orders.

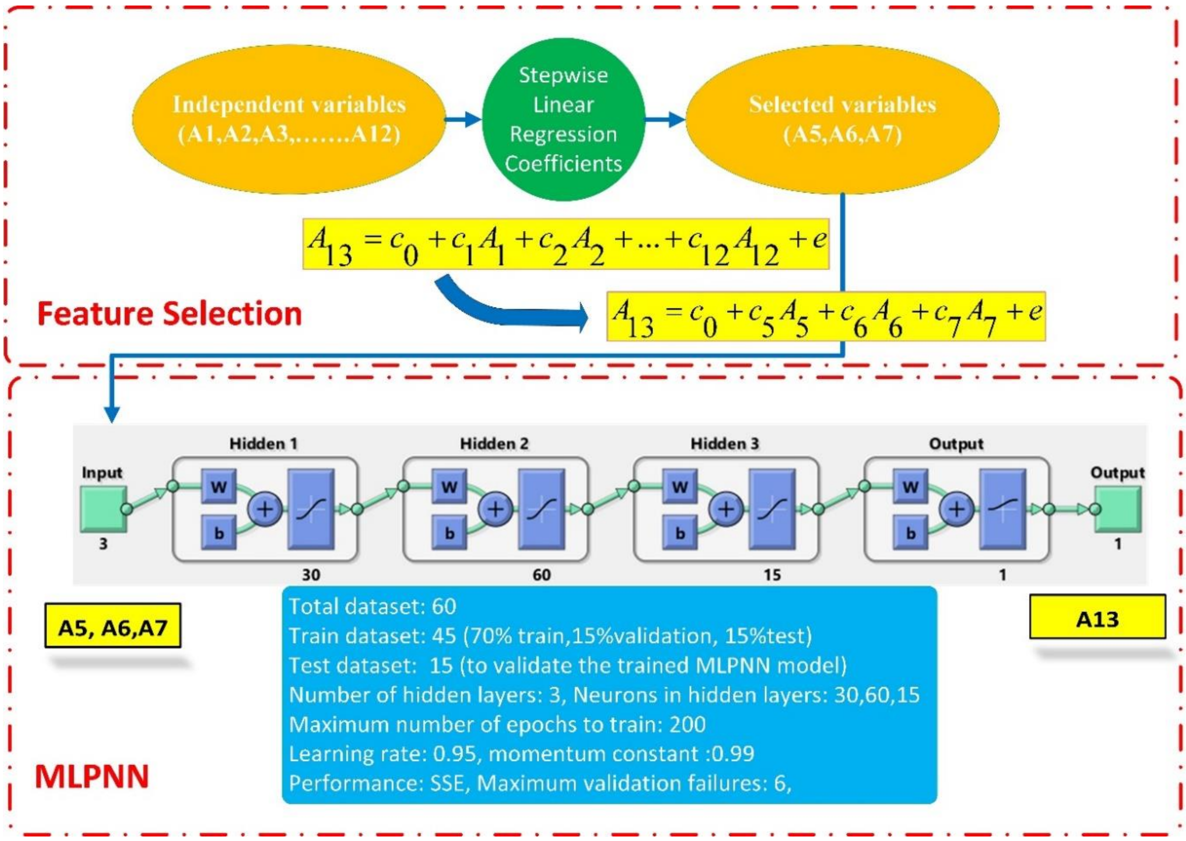

Figure 2. Schematics of the proposed MLP neural network

During the design of the MLPNN model, the data were divided into two sets of training and test subsets. While the training dataset was used to develop the MLPNN model, the test dataset not included in the training group was used to validate and examine the performance of the trained MLPNN model. This model has been proven that it can be used to predict future demand based on historical data.

To check the performance of the MLPNN model with different input sets, the mean absolute percentage error (MAPE) was measured. Using the NN approach, this performance criterion was used to measure how close the measured values were to the predicted values. This criterion helps evaluate the model's capabilities with different input sets for the A13. MAPE is computed as follows (Equation 6):

$$MAPE = \frac{100\%}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \qquad (6)$$

where y_i and $\hat{y}_i$ are the predicted values and the observed A13 values based on the proposed model and N is the number of days in the test data, respectively.

Model performances for each input set were compared based on the estimation of 15 samples (days).

3. Results and Discussions

Using the stepwise method, multiple linear regression was conducted to compare the independent influences of A1, A2, A3, A4…A11, and A12 on the A13.

Descriptive statistics are used to summarize and describe the basic features of data. Table 2 shows descriptive statistics for variables of interest.

Table 2. Descriptive statistics for variables of interest

	Descriptive Statistics		
	Type of Variables	**Mean**	**Std. Deviation**
A13	Dependent	300.87	89.60
A1	Independent	3.02	1.28
A2	Independent	4.03	1.40
A3	Independent	172555	69.51
A4	Independent	118921	27.17
A5	Independent	52112	18.83
A6	Independent	109.230	50.74
A7	Independent	139.53	41.44
A8	Independent	77.40	186.50
A9	Independent	44504.35	12197.91
A10	Independent	46640.83	45220.74
A11	Independent	79401.48	40504.42
A12	Independent	23114.63	13148.04

The stepwise regression procedure selects variables in a step-by-step manner [49]. This method adds or subtracts independent variables one by one by evaluating the statistical significance of the variable. It removes the least significant variable to reduce the size of the features. When the algorithm ends, a single regression model is produced. While the t statistical test is used to determine the importance of each independent variable, beta coefficients are used to determine the significance level for each independent variable. The dependent variable will increase by the number of beta coefficients [44]. Table 3 shows the results of the stepwise regression analysis.

Unstandardized regression coefficients estimated from linear regression are presented in the second column of Table 3. Standardized coefficients, also called Beta coefficients, are given in the fourth column of the same table. Based on the descriptive statistics above, it is seen how each of the independent variables affects the dependent variable (A13) while keeping all other variables constant. Beta coefficients can be used to evaluate the sensitivity of the factors [50]. A higher absolute value of the standardized coefficient indicates a stronger effect on the dependent variable (A13). As Table 3 shows, the A5, A6, and A7 have the most important effect on the A13. The effect of A6 (Beta = 0.566, $p<0.001$) is greater than the effects of A7 (Beta=0.463, $p<0.001$) and A5 (Beta=0.210, $p<0.001$).

The dimensions of attributes were reduced to 10 (A3-A12), 7 (A3, A4, A8-A12), 5 (A3, A4, A6, A7, A11), 5 (A3-A7), 4 (A3, A5-A7), 4 (A5-A8), 3 (A3, A6, A7) predictive attributes from 12, randomly. In addition, the dimensions of attributes of datasets were reduced to three (A5, A6, A7) from twelve using the LRC method. Thus, nine feature subsets (X1 to X9) for the DDFO dataset were constructed to build the forecasting model (Table 4).

Table 3. Stepwise linear regression analysis for the effect of independent variables on A13 (Target-Total orders)

	Unstandardized Coefficients		**Standardized Coefficients**			**Collinearity Statistics**	
	B	**Std. Error**	**Beta**	**t statistic**	**p**	**Tolerance**	**VIF**
Cons.	5.05E-13	.000		.000	1.000		
A3	1.49E-14	.000	.000	.000	1.000	.078	12.816
A4	1.29E-14	.000	.000	.000	1.000	.219	4.570
A6	1.000	.000	.566	75.09E+6	<0.001	.208	4.811
A7	1.000	.000	.463	44.91E+6	<0.001	.111	8.971
A5	1.000	.000	.210	35.91E+6	<0.001	.345	2.897
Cons.	6.47E-13	.000		.000	1.000		
A3	9.74E-15	.000	.000	.000	1.000	.124	8.072
A6	1.000	.000	.566	88.63E+6	<0.001	.284	3.517
A7	1.000	.000	.463	78.78E+6	<0.001	.337	2.969
A5	1.000	.000	.210	46.62E+6	<0.001	.571	1.751
Cons.	1.99E-13	.000		.000	1.000		
A6	1.000	.000	.566	13.16E+7	<0.001	.615	1.625
A7	1.000	.000	.463	11.67E+7	<0.001	.726	1.378
A5	1.000	.000	.210	55.93E+6	<0.001	.807	1.239

(Table title row: **Coefficients**)

Table 4. The nine feature subsets for DDFO dataset

Feature Subset	No of selected features	Features
X1	1	A1,A2,A3,A4,A5,A6,A7,A8,A9,A10,A11,A12
X2	2	A3,A4,A5,A6,A7,A8,A9,A10,A11,A12
X3	3	A3,A4,A8,A9,A10,A11,A12
X4	4	A3,A4,A6,A7,A11
X5	5	A3,A4,A5,A6,A7
X6	6	A3,A5,A6,A7
X7	7	A5,A6,A7,A8
X8	8	A3,A6,A7
X9*	9	A5,A6,A7

* *Obtained subset by feature selection using LRC*

The MAPE values of models for the prediction of A13 with different inputs (subsets) are shown in Figure 3a. As a result, better model performances were obtained with the X9 feature subset (Figure 3a and Figure 4f). Figure 3b illustrates the regression performance of the proposed model obtained using the X9 feature subset. A strong relationship was found between the predicted and the observed variable. According to the results, the correlation coefficient (R^2) is approximately 0.992, and this value is quite high.

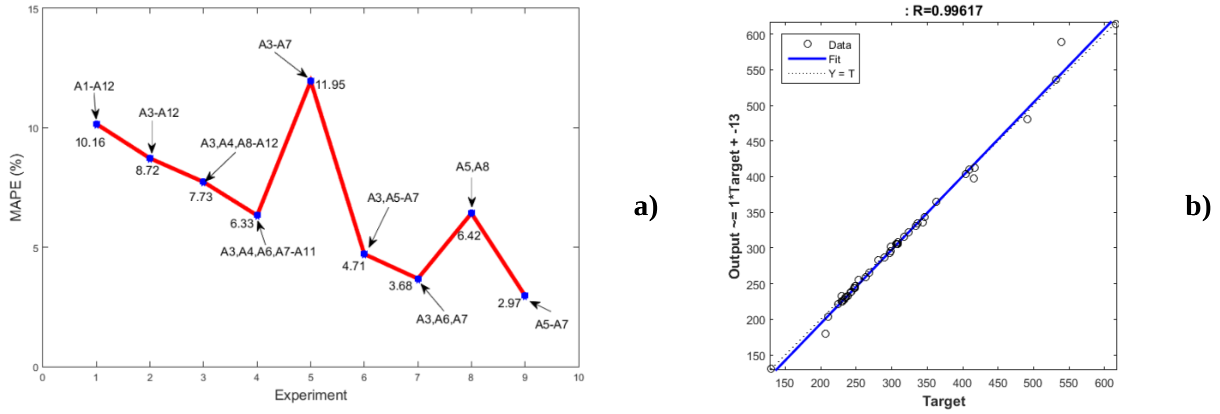

Figure 3. a) Mean absolute percentage error on demand forecast with different subsets of selected attributes **b)** Regression performance of the model for X9 feature subset.

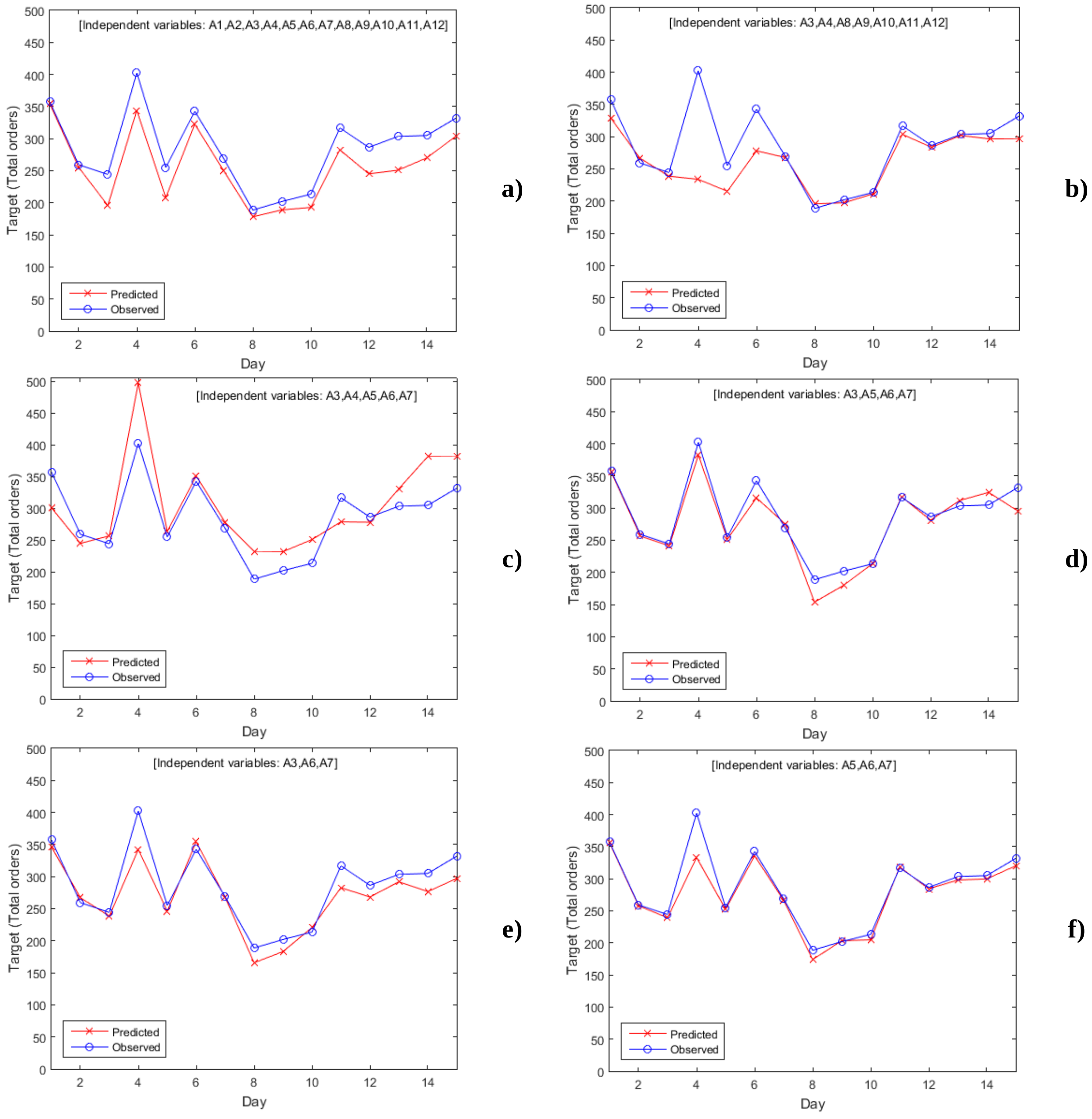

Figure 4. Comparison of predictions with observed orders according to **a)** X1 **b)** X3 **c)** X5 **d)** X6 **e)** X7 **f)** X9 feature subsets

Figure 4 shows the performance of the models (different subsets of attributes selected) over the test data. The experimental results are shown in Figures 4a, 4b, 4c, 4d, 4e, and 4f; the marker curve (circle) represents observed data, and the marker curve (cross) represents the output of MLPNN. Figures 4a, 4b, 4c, 4d, and 4e were obtained using randomly selected features as model input. Figure 4f shows the performance of the proposed 3-input model. Compared to the others, the graph obtained for the proposed model shows a strong correlation between the predicted and observed values.

The data set used in this study was also used by Ferreira et al. in a previous study. Ferreira et al. found valuable results by using 12 variables in artificial neural network inputs. This study obtained successful results (R^2 with 0.992, MAPE with 2.97%) using only three variables, including order type_A (A5), order type_B (A6), and order type_C.

4. Conclusion

Demand forecasting is the field of predictive analytics dedicated to understanding consumer demand. This process needs to be accurate and timely in a trading company. This paper aims to forecast the daily demand orders of a valve manufacturing industry using the MLP neural network model and reduce the forecast error using LRC-based feature selection. The goal of feature selection is to reduce the number of attributes in the dataset. Using the linear regression coefficients and their significance, features decreased from twelve to three (A5, A6, A7). The MAPE value for the model was found to be less than 3%. The best results were obtained with the A5 (order type_A), A6 (order type_B), and A7 (order type_C) attributes. The proposed feature selection method can be considered as a good tool to improve forecast accuracy and shorten forecast time in forecasting daily demand. In future research, the feature selection technique using linear regression coefficients will be tested in many different sectors.

The Declaration of Publishing Ethics

The author declares that this study complies with Research and Publication Ethics.

References

[1] E. Eckhaus, “Consumer Demand Forecasting: Popular Techniques, Part 1: Weighted and Unweighted Moving Average,” 2010. [Online]. Available: http://www.purchasesmarter.com/articles/118. (Accessed: Nov. 22, 2021).

[2] S. P. Sethi, H. Yan, and H. Zhang, *Inventory and Supply Chain Management with Forecast Updates*. Chapter 1. New York: USA/Springer, 2005.

[3] R. Fildes, K. Nikolopoulos, S. F. Crone, and A. A. Syntetos, “Forecasting and operational research: a review,” *Journal of the Operational Research Society*, vol. 59 (9), pp. 1150–1172, September 2008.

[4] J. P. Donate, P. Cortez, G. G. Sánchez, and A.S. Miguela, “Time series forecasting using a weighted cross-validation evolutionary artificial neural network ensemble,” *Neurocomputing*, vol. 109, pp. 27–32, June 2013.

[5] K. Tanaka, “A sales forecasting model for new-released and nonlinear sales trend products,” *Expert Systems with Applications*,” vol. 37, pp. 7387–7393, November 2010.

[6] P. C. Chang, and Y.W. Wang, “Fuzzy Delphi and backpropagation model for sales forecasting in PCB industry,” *Expert Systems with Applications*, vol. 30, pp. 715–726, May 2006.

[7] Y. Ni, and F. Fan, “A two-stage dynamic sales forecasting model for the fashion retail,” *Expert Systems with Applications*, vol.38, pp. 1529–1536, March 2011.

[8] Z. L. Sun, T. M. Choi, K. F. Au, and Y. Yu, “Sales forecasting using extreme learning machine with applications in fashion retailing,” *Decision Support Systems*, vol. 46, pp. 411–419, December 2008.

[9] P. Kumar, M. Herbert, and S. Rao, “Demand forecasting using Artificial Neural Network Based on Different Learning Methods: Comparative Analysis", *IJRASET*, vol. 2, pp. 364-374, April 2014.

[10] G. S. Groppo, M. A. Costa, and M. Libânio, “Predicting water demand: a review of the methods employed and future possibilities”, *Water supply*, vol. 19, pp. 2179-2198, August 2019.

[11] G. P. Zhang, “An investigation of neural networks for linear time-series forecasting,” *Computers&Operations Research*, vol. 28, pp. 1183–1202, October 2001.

[12] J. Adamowski, and C. Karapataki, “Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: evaluation of different ANN learning algorithms," *Journal of Hydrologic Engineering*, vol. 15, pp. 729–743, October 2010.

[13] J. Caiado, "Performance of combined double seasonal univariate time series models for forecasting water demand," *Journal of Hydrologic Engineering*, vol. 15, pp. 215–222, March 2010.
[14] J. Adamowski, H. F. Chan, S. O. Prasher, B. Ozga-Zielinski, and A. Sliusarieva, "Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada," *Water Resources Research*, vol. 48, pp. 1-14, January 2012.
[15] M. Ghiassi, D. K. Zimbra, and H. Saidane, "Urban water demand forecasting with a dynamic artificial neural network model", *Journal of Water Resources Planning and Management*, vol. 134, pp. 138–146, March 2008.
[16] M. Fırat, M. A. Yurdusev, and M. E. Turan, "Evaluation of artificial neural network techniques for municipal water consumption modeling," *Water Resources Management*, vol. 23, pp. 617–632, March 2009.
[17] M. Fırat, M .E. Turan, and M. A. Yurdusev, "Comparative analysis of neural network techniques for predicting water consumption time series," *Journal of Hydrology*, vol. 384, pp. 46–51, April 2010.
[18] A. Altunkaynak, M. Özger, and M. Çakmakcı, "Water consumption prediction of Istanbul city by using fuzzy logic approach," *Water Resources Management*, vol. 19, pp. 641–654, October 2005.
[19] M. Fırat, M. E. Turan, and M. A. Yurdusev, "Comparative analysis of fuzzy inference systems for water consumption time series prediction," *Journal of Hydrology*, vol. 374, pp. 235–241, August 2009.
[20] C. Peña-Guzmán, J. Melgarejo, and D. Prats, "Forecasting water demand in residential, commercial, and industrial zones in Bogotá, Colombia, using Least-Squares Support Vector Machines," *Mathematical Problems in Engineering*, vol. 2016, pp. 1-10, December 2016.
[21] M. Nasseri, A. Moeini, and M. Tabesh, "Forecasting monthly urban water demand using Extended Kalman Filter and Genetic Programming," *Expert Systems with Applications*, vol. 38, pp. 7387–7395, June 2011.
[22] A. Chawla, A. Singh, A. Lamba, N. Gangwani, and U. Soni, "Demand Forecasting using Artificial Neural Networks – A case study of American Retail Corporation," *in Advances in Intelligent Systems and Computing: Applications of Artificial Intelligence Techniques in Engineering*, H. Malik, S. Srivastava, Y. Sood., A. Ahmad, Eds. Singapore: Springer, 2019. pp. 79-89.
[23] A. Öztekin, R. Kizilaslan, S. Freund, and A. Iseri "A data analytic approach to forecasting daily stock returns in an emerging market," *European Journal of Operational Research*, vol. 253, pp. 697-710, September 2016.
[24] R. J. Kuo, and K. C. Xue, "A decision support system for sales forecasting through fuzzy neural networks with asymmetric fuzzy weights," *Decision Support Systems*, vol. 24, pp. 105–126, December 1998.
[25] R. Law, "Backpropagation learning in improving the accuracy of neural network-based tourism demand forecasting," *Tourism Management*, vol. 21, pp. 331–340, August 2000.
[26] Z. Ying, and X. Hanbin, "Study on the model of demand forecasting based on artificial neural network," in *2010 Ninth International Symposium on Distributed Computing and Applications to Business, Engineering and Science, Hong Kong, China, August 10-12, 2010*, pp. 382-386.
[27] S. Bhadouria, and A. Jayant, "Development of ANN Models for Demand Forecasting," *American Journal of Engineering Research (AJER)*, vol. 6, pp. 142-147, December 2017.
[28] P. du Jardin, and E. Séverin, "Forecasting financial failure using a Kohonen map: A comparative study to improve model stability over time," *European Journal of Operational Research*, vol. 221, pp. 378-396, September 2012.
[29] Q. Cao, B. T. Ewing, and M. A. Thompson, "Forecasting wind speed with recurrent neural networks," *European Journal of Operational Research*, vol. 221, pp. 148-154, August 2012.
[30] G. Sermpinis, K. Theofilatos, A. Karathanasopoulos, E. F. Georgopoulos, and C. Dunis, "Forecasting foreign exchange rates with adaptive neural networks using radial-basis functions and particle swarm optimization", *European Journal of Operational Research*, vol. 225, pp. 528-540, March 2013.
[31] M. S. Kim, "Modeling special-day effects for forecasting intraday electricity demand," *European Journal of Operational Research*, vol. 230, pp. 170-180, October 2013.
[32] K. Venkatesh, V. Ravi, A. Prinzie, and D. V. Poel, "Cash demand forecasting in ATMs by clustering and neural networks," *European Journal of Operational Research*, vol. 232, pp. 383-392, January 2014.
[33] R. P. Ferreira, A. Martiniano, A. Ferreira, A. Ferreira, and R. J. Sassi, "Study on Daily Demand Forecasting Orders using Artificial Neural Network," *IEEE Latin America Transactions*, vol. 14, pp. 1519-1525, March 2016.

[34] W. Fu, C.- F. Chien, and Z.- H. Lin, "A hybrid forecasting framework with neural network and time-series method for intermittent demand in semiconductor supply chain," in *Advances in Production Management Systems*: *IFIP WG 5.7 International Conference, APMS 2018, Seoul, Korea, August 26-30, 2018*, I. Moon, G. M. Lee, J. Park, D. Kiritsis G. Cieminski, Eds. Springer, Cham., 2018. pp. 65-72.

[35] P. Hietaharju, M. Ruusunen, and K. Leiviska, "Enabling Demand Side Management: Heat Demand Forecasting at City Level," *Materials*, vol. 12, pp. 1-17, January 2019.

[36] H. Peng, C. Ding, and F. Long, "Minimum redundancy maximum relevance feature selection," *IEEE Intelligent Systems*, vol. 20, pp. 70–71, December 2015.

[37] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu, "Feature Selection: A Data Perspective," *ACM Computing Surveys*, vol. 50, pp. 1-45, November 2018.

[38] Y. Zhang, D. Gong, X. Gao, T. Tian, and X. Sun, "Binary differential evolution with self-learning for multi-objective feature selection," *Information Sciences*, vol. 507, pp. 67-85, January 2020.

[39] R.A. Chinnathambi, M. Campion, A. S. Nair, and P. Ranganathan, "Investigation of Price-Feature Selection Algorithms for the Day-Ahead Electricity Markets," in *2018 IEEE Electrical Power and Energy Conference (EPEC), Toronto, ON, Canada, October 10-11, 2018*, pp. 1-6.

[40] UCI Machine Learning Repository, "Daily Demand Forecasting Orders Data Set," 2017. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Daily+Demand+Forecasting+Orders (Accessed: Nov. 22, 2021).

[41] K. Abrougui, K. Gabsi, B. Mercatoris, C. Khemis, R. Amami, and S. Chehaibi, "Prediction of organic potato yield using tillage systems and soil properties by artificial neural network (ANN) and multiple linear regressions (MLR)," *Soil and Tillage Research*, vol. 190, pp. 202-208, July 2019.

[42] N. Susijawati, A. Setiawan, G. M. Putri, S. Maryam, A. Firasati, and M. Alwi, "The Effect of Organizational Commitment and Organizational Support as Intervening Variables to Turnover Intention of Employees," in *International Symposium on Social Sciences, Education, and Humanities (ISSEH 2018),* in Advances in Social Science, Education and Humanities Research, vol. 306, March 2019, pp. 283-285.

[43] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate Data Analysis A Global Perspective*. England/Pearson Education Limited, 2010.

[44] Statistics Solutions website, "Regression". [Online]. Available: https://www.statisticssolutions.com/directory-of-statistical-analyses-regression-analysis/regression/. [Accessed: Dec. 05, 2021).

[45] M. Elbisy, and F. Osra, "Application of Group Method of Data Handling Type Neural Network for Significant Wave Height Prediction," *American Journal of Neural Networks and Applications*, vol. 5, pp. 51-57, November 2019.

[46] R. J. Kuo, M. C. Shieh, J.W. Zhang, and K.Y. Chen, "The application of an artificial immune system-based backpropagation neural network with feature selection to an RFID positioning system," *Robotics and Computer-Integrated Manufacturing*, vol. 29, pp. 431-438, December 2013.

[47] M. Lashkarbolooki, B. Vaferi, A. Shariati, and A. Z. Hezave, "Investigating vapor–liquid equilibria of binary mixtures containing supercritical or near-critical carbon dioxide and a cyclic compound using cascade neural network," *Fluid Phase Equilibria*, vol. 343, pp. 24-29, April 2013.

[48] P. L. Narayana, J. H. Kim, A. K. Maurya, C. H. Park, J.- K. Hong, J.- T. Yeom, and N. S. Reddy, "Modeling Mechanical Properties of 25Cr-20Ni-0.4C Steels over a Wide Range of Temperatures by Neural Networks," *Metals*, 10, pp. 1-12, February 2020.

[49] J. FrosT, "Guide to Stepwise Regression and Best Subsets Regression," 2020. [Online]. Available: https://statisticsbyjim.com/regression/curve-fitting-linear-nonlinear-regression/. (Accessed: Dec. 05, 2021).

[50] S. B. Zhou, L. Shengjie, and X. Yiming, "Effects of Filler Characteristics on the Performance of Asphalt Mastic: A Statistical Analysis of the Laboratory Testing Results," *International Journal of Civil Engineering*, vol. 16, pp. 1175–1183, September 2018.