

Predicting Prices of Case Furniture Products Using Web Mining Techniques

Timucin Bardak *

This article presents a methodology based on web mining techniques for estimating furniture prices using e-commerce data. Data on different public e-commerce sites in the United States were collected and analyzed using web mining methods. Deep learning and random forest algorithms were used to predict the prices of different types of furniture. Bookcase and dresser type furniture, which are widely used in price estimation, were selected. The inquiry identified a collection of eight distinctive attributes linked to furniture items, spanning measurements such as width, depth, and height, alongside features encompassing frame material, partition count, drawer count, color, and price. In preparation for constructing predictive models, a dataset comprising 300 instances was compiled for comprehensive analysis. Models developed based on web mining to predict furniture prices gave satisfactory results. During the testing phase, the random forest algorithm outperformed deep learning, achieving high goodness of fit values of 0.89 and 0.94 for bookcase and dresser furniture, respectively. The results indicate that price estimation for dresser furniture was more accurate than for bookcases in all models. The findings demonstrate that web mining techniques can be used effectively in competitive furniture pricing, with potential to save time and cost in pricing for furniture purchasing.

DOI: 10.15376/biores.18.4.7412-7427

Keywords: Furniture industry; Price; Data mining; Prediction modeling

Contact information: Bartin University, Bartin Vocational School, Furniture and Decoration Program, 74200, Bartin, Turkey; *Corresponding author: timucinb@bartin.edu.tr

INTRODUCTION

The furniture industry is a critical sector in terms of creating both added value and employment in the economic development of many countries. When determining prices, furniture manufacturers consider many factors, such as profitability, competition, and market. For consumers, a product must be offered at a price that matches its value, which is an important determining factor when deciding to buy a product (Raposo *et al.* 2018). In a study by Lihra *et al.* (2012), factors influencing consumers' preferences in furniture shopping were examined. According to the results, approximately 50% of consumers consider price to be the most significant factor in product selection, 20% prioritize product customization, 20% consider delivery time, and 10% take into account the time required for product customization (Lihra *et al.* 2012). In the realm of online shopping, due to the inability of consumers to physically see and touch products, it is widely believed that elements such as pricing and product information play a more pronounced role in influencing consumer purchasing behavior. A study of online consumer behavior shows

that pricing plays an important role in influencing purchasing decisions (Zhao *et al.* 2021). Correct pricing for companies is a complex problem in competitive market conditions.

Data mining (DM) is used in many industries to solve challenging problems (He *et al.* 2022). DM can be defined as a method that allows the most valuable information to be obtained by analyzing data sets. This method includes data processing, model creation, feature extraction, and discovery studies (Pérez-Campuzano *et al.* 2022). Using important data sources and data mining algorithms helps reduce costs effectively by making it easier to determine the optimal price. Numerous investigations have demonstrated that data-driven models exhibit superior efficacy in predicting the prices of various products compared to conventional techniques (Chen *et al.* 2016; Li *et al.* 2019; Duan and Liu 2021). Studies have shown that Artificial Neural Network (ANN) cost models were developed using data from 300 building projects to evaluate the total construction cost for customers. As a result of the analysis, it was determined that the ANN models had a high prediction performance (Emsley *et al.* 2002; Hassim *et al.* 2018). In a separate investigation, Putri *et al.* (2019) devised predictive models utilizing monthly retail pricing data to forecast domestic and global beef prices in the Indonesian market. These models were efficacious in their predictive capacity (Putri *et al.* 2019).

Web mining (WM) is a field of research that seeks to discover valuable information, especially by referring to texts from web content analysis (Brauner *et al.* 2022). The WM technique has become extremely popular, as it offers the opportunity to discover new information in many fields. However, WM poses a considerable challenge due to the limited interpretability of its results, hindering the ability to draw meaningful inferences. Various machine learning (ML) techniques can be employed in web mining to extract valuable insights from web documents. Because of these techniques, the interpretability of web mining results can be increased (Lee and Lee 2011). Random forest (RF) and deep learning (DL) algorithms are widely used among ML methods and provide effective results for different data sets. The DL generates high-precision results and appropriate outputs using training datasets. The DL model consists of many layers, and each layer manages different stages of the data processing process. These layers are called the input layer, the output layer, and the hidden layer (Kim and Falcone 2017; Ghorbanzadeh *et al.* 2019; Saha *et al.* 2023). The RF is a commonly used method for developing predictive models. This method works by constructing many conditionally independent decision trees to solve the overfitting problem. These trees are created using different features and data points and are independent of each other (Breiman 2001; Orte *et al.* 2023). Both the DL and RF algorithms have their respective strengths within their domains. The RF algorithm offers significant advantages for medium to large-scale datasets, excelling in capturing non-linear relationships present in the data. It also simplifies tasks involving structured data and scenarios where interpretability holds importance. On the other hand, the DL algorithms prove particularly effective when dealing with extensive, unstructured data such as images, audio, and text. They also stand out for their ability to process diverse types of data. Thus, in the present study, a preference was given to these two distinct algorithm types. The fundamental distinction between The DL and The RF algorithms lies in the former's design to handle large, unstructured datasets, while the latter is more suitable for structured data and is often favored in cases emphasizing interpretability (Schonlau and Zou 2020; Sarker 2021).

E-commerce platforms (EP) are an ideal data source for web mining. An EP encompasses a range of information concerning furniture products; for instance, details encompassing type, style, color variations, design attributes, and furniture prices serve as

only a few illustrative examples. These particulars constitute a subset of the comprehensive dataset concerning furniture items within such platforms. Furthermore, these websites also facilitate data acquisition about vastly diverse furniture pieces such as beds, sofas, tables, and chairs. Nevertheless, there exist constraints related to the information accessible through e-commerce platforms. To illustrate, e-commerce websites refrain from divulging personal or sensitive data about customers, such as credit card numbers or social security numbers. Moreover, detailed financial particulars concerning the company, such as revenue figures or profit margins, are generally not within the purview of e-commerce platforms. Instead, the central focus of e-commerce websites resides in furnishing customers with a secure and gratifying shopping encounter. The EP commerce sites are platforms that are becoming increasingly common today and remarkably transform the shopping experience for customers. Customer shopping behavior can be tracked and analyzed on these platforms (Massimino 2016; Zhou *et al.* 2021). In a study, online purchasing data of consumers from public e-commerce sites were collected for customer segmentation. These data were processed with web mining analysis to create profiles of customer purchasing behaviors (Zhou *et al.* 2021). A new approach using web mining techniques has shown effectiveness in accurately estimating unemployment rates. It was also emphasized that the proposed framework could help understand the factors underlying unemployment rates and can provide people with understandable qualitative clues (Li *et al.* 2014). In another study, a method is proposed based on the idea of developing a scenario using web mining for a specific topic. This method has been highlighted as having the potential to improve time-consuming desk research in scenario projects and has proven to be very useful (Kayser and Shala 2020).

This study aims to investigate the applicability of different ML algorithms to predict the prices of furniture products using a data set obtained from e-commerce sites. In addition to the furniture products' features, the prepared data set includes each product's price information. The price information utilized in our study is directly sourced from various e-commerce platforms themselves. The present analysis is founded upon the authentic price listings furnished by these platforms. The pricing data was gathered between the dates of January 21, 2023, and March 21, 2023. The investigation revealed a set of eight unique characteristics associated with furniture products. These attributes encompass a range of measurements, including width, depth, and height, as well as features such as frame material, partition count, drawer count, color, and price. To formulate predictive models, a dataset consisting of 300 instances was systematically compiled, facilitating a thorough analysis. The relationships between furniture products' features and price information in the data set were examined using data mining techniques. The results revealed that data mining techniques can be used successfully to predict the prices of furniture products. In addition, the study can provide an essential resource for the furniture industry to determine the correct pricing strategies and offer products that meet customer demands.

EXPERIMENTAL

Data Collection

In this study, price estimates were made for different types of furniture, such as bookcases and dressers. The data were obtained from three different public e-commerce sites and analyzed using ML models. The collected data were homogenized according to

criteria such as product categories and price ranges to eliminate inconsistencies arising from the different structures of the web pages. The analysis determined eight different features for the furniture: width, depth, height, frame material, number of partitions, number of drawers, color, and price. 300 pieces of data were gathered for two distinct furniture categories for predictive purposes. Of these datasets, 150 pertained to bookcases, while the remaining 150 were connected to dressers. Tables 1 and 2 summarize the attributes in the training dataset used for the price prediction of types of furniture.

Table 1. Summary of Training Dataset Used for Price Prediction of Bookcase Furniture

Attribute Name	Attribute Type	Attribute Description
Frame material	Nominal	Frame material of furniture (Particleboard, Fiberboard, or Particleboard and fiberboard)
Color	Nominal	Color of furniture (White, Brown, or Gray)
Width	Numeric	Width of furniture (range, 12 to 79 inch)
Depth	Numeric	Depth of furniture (range, 5 to 31 inch)
Height	Numeric	Height of furniture (range, 7 to 93 inch)
Number of partitions	Numeric	Number of partitions of furniture (range, 1 to 25)
Price	Numeric	The price of the furniture (range, 30 to 359 US Dollar)

Note: The term "inch" used in this table is approximately equivalent to 2.54 centimeters (cm).

Table 2. Summary of Training Dataset Used for Price Prediction of Dresser Furniture

Attribute Name	Attribute Type	Attribute Description
Frame material	Nominal	Frame material of furniture (Particleboard, Fiberboard, or Particleboard and fiberboard)
Color	Nominal	Color of furniture (White, Brown, or Gray)
Width	Numeric	Width of furniture (range, 14 to 64 inch)
Depth	Numeric	Depth of furniture (range, 13 to 57 inch)
Height	Numeric	Height of furniture (range, 13 to 57 inch)
Number of drawers	Numeric	Number of drawers of furniture (range, 1 to 12)
Price	Numeric	The price of the furniture (range, 59 to 710 US Dollar)

Data Preprocessing

In this study, the normalization process was applied to numerical values. This process was carried out to ensure the data were at the same scale and distribution.

Models

Deep learning (DL) and random forest (RF) algorithms were used for price estimates of different types of furniture. In the analysis, the dataset was divided into training and test datasets by the literature, and the training dataset covered 70% of the total dataset. The remaining 30% dataset was reserved as the test dataset (Velten *et al.* 2000; Furtney *et al.* 2022; Saha *et al.* 2023).

All models were developed with RapidMiner Studio Version 9.3 software (Boston, MA, USA), which has been widely used in many studies (Keet *et al.* 2015; Bardak *et al.* 2021; Mariano *et al.* 2022; Naser 2023). Figure 1 illustrates the workflow of the process used for model comparison.

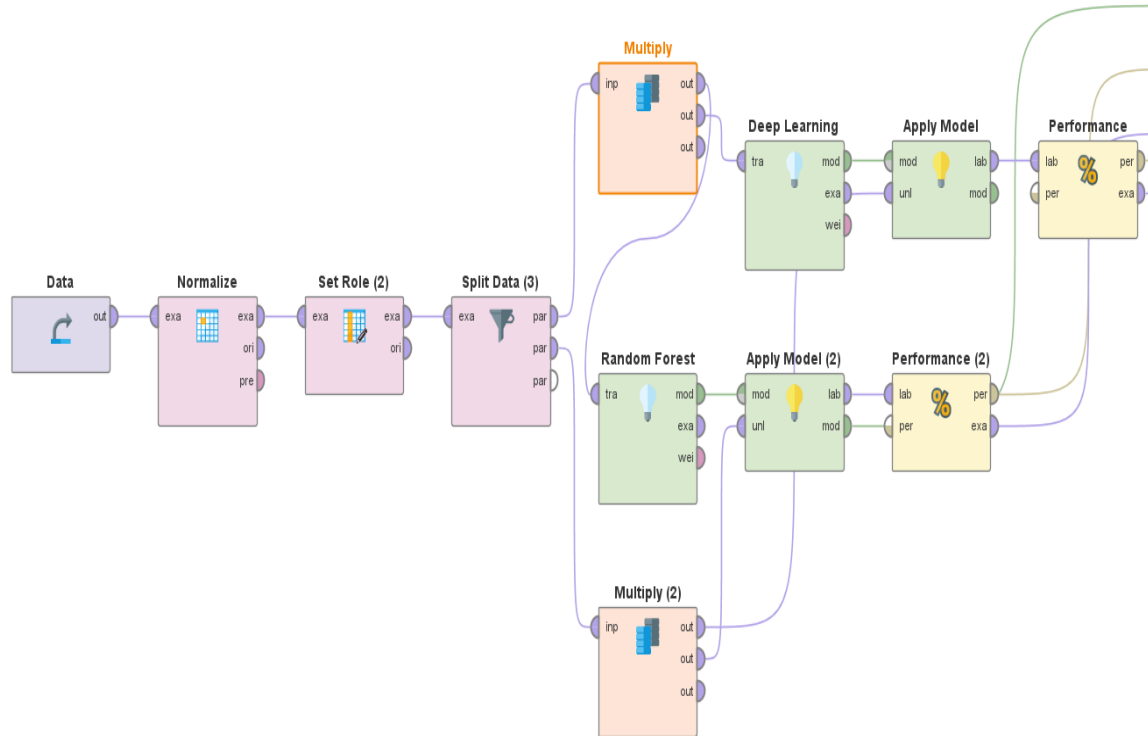


Fig. 1. The workflow of the process used for model comparison in bookcase furniture

The weights of attributes such as width, depth, height, frame material, number of partitions, number of drawers, and color were determined using the algorithm RF, which has the highest level of success

Model Evaluation

To evaluate the effectiveness of the predictions made by different prediction models, previously accepted performance criteria in the literature were used., the goodness of fit (R^2 , Eq. 1) and root mean square error (RMSE, Eq. 2) were used to measure the success of predictive models objectively (Razali and Al-Wakeel 2013; Li *et al.* 2019; Agwu *et al.* 2020),

$$R^2 = 1 - \frac{\sum_i (Y_i - \hat{Y}_i)^2}{\sum_i (Y_i - \bar{Y})^2} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_i (Y_i - \hat{Y}_i)^2}{n}} \quad (2)$$

where Y_i and \hat{Y}_i are the measured and predicted values, respectively and n represents the total number of samples (Pervez *et al.* 2023).

Optimization of the Parameters of the Models

In pursuing optimal prediction performance, adjusting various parameters with precision is essential. To this end, the grid operator is a useful tool for parameter optimization within the RapidMiner software platform. This approach enables identifying and selecting optimal parameters for model, thereby contributing to enhanced efficiency and accuracy in machine learning and other algorithmic applications. To determine the optimal parameters, the (Grid) operator was employed. The workflow prepared to determine the optimum parameters is shown in Fig. 2.

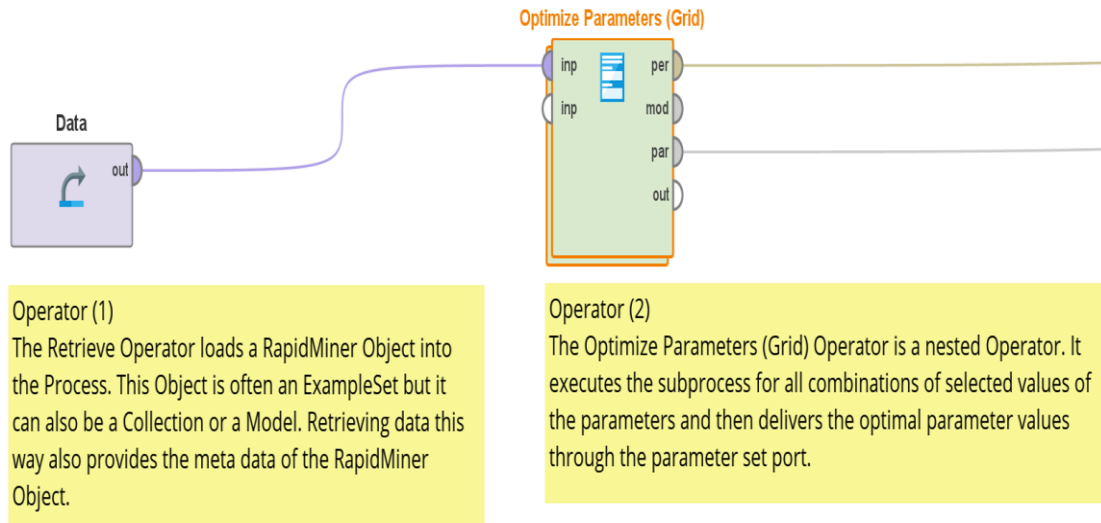


Fig. 2. The workflow prepared to determine the optimum parameters

To optimize model performance, significant parameters in DL and RF algorithms were adjusted. In the DL algorithm, parameters such as hidden layer sizes and train samples per iteration were selected. Hidden layer sizes determine the number of hidden layers and neurons in the model. For example, specifying “100,200,100” would create a model with 3 hidden layers, and the middle-hidden layer would have 200 neurons. This determines the amount of training data that is processed in each iteration. This parameter determines the rate at which scoring and model cancellation can occur. In the RF algorithm, parameters such as the number of trees and maximum depth were chosen. The number of trees specifies the count of random trees to be generated. For each tree, a subset of the example set is chosen through bootstrapping. Maximum depth is used to limit the depth of each random tree. The values of these parameters can vary based on the size and characteristics of the dataset (Rapidminer 2023).

The optimal parameters for each model and furniture type were determined separately and are presented in Tables 3 and 4. For the DT model, the optimal parameters for achieving the highest level of performance for bookcase furniture were 90 the hidden layer sizes and -2 train samples per iteration. The most suitable parameters for achieving effective results for dresser furniture type were 90 hidden layer sizes and -2 train samples per iteration. As for the RF model, to attain the best performance for bookcase furniture, the number of trees was set to 20 and maximal depth to 7. On the other hand, for dresser furniture, the number of trees was set as 20 and the maximal depth was set as 7 to obtain optimal outcomes.

Table 3. Optimal Model Parameters for DL Algorithm

Bookcase Furniture			Dresser Furniture		
Hidden layer sizes	Train samples per iteration	RMSE	Hidden layer sizes	Train samples per iteration	RMSE
30	-2	31.320	30	-2	65.944
60	-2	31.658	60	-2	67.446
90	-2	31.080	90	-2	60.190
120	-2	31.839	120	-2	62.988
150	-2	32.282	150	-2	62.998
30	-1	31.320	30	-1	65.944
60	-1	31.658	60	-1	67.446
90	-1	31.080	90	-1	60.190
120	-1	31.839	120	-1	62.988
150	-1	32.282	150	-1	62.443
30	0	31.320	30	0	65.944
60	0	31.658	60	0	67.446
90	0	31.080	90	0	60.190
120	0	31.839	120	0	62.988
150	0	32.282	150	0	62.443
30	1	48.557	30	1	102.593
60	1	49.647	60	1	130.332
90	1	52.129	90	1	102.279
120	1	60.789	120	1	113.562
150	1	38.001	150	1	69.170

Table 4. Optimal Model Parameters for RF Algorithm

Bookcase Furniture			Dresser Furniture		
Number of trees	Maximal depth	RMSE	Number of trees	Maximal depth	RMSE
20	2	50,072	20	2	121,532
60	2	48,285	60	2	116,297
100	2	48,196	100	2	115,127
140	2	48,171	140	2	114,084
20	4	32,105	20	4	69,321
60	4	29,278	60	4	72,455
100	4	29,020	100	4	69,691
140	4	28,680	140	4	68,869
20	7	21,845	20	7	51,715
60	7	21,875	60	7	53,186
100	7	21,899	100	7	53,792
140	7	22,910	140	7	54,938

RESULTS AND DISCUSSION

Two different ML models were developed using the data obtained from web pages, and each model was trained and tested on the same data sets. Tables 5 and 6 present the prediction performances for furniture. A total of 300 data samples from three different e-commerce websites were collected for this study: 114 samples were provided by the first website, 97 samples were provided by the second website, and 89 samples were provided by the third website.

Table 5. The Prediction Performances for Bookcase Furniture

Model	Testing Phase		Training Phase	
	R ²	RMSE	R ²	RMSE
DL	0.797	31.080	0.710	33.185
RF	0.892	21.845	0.934	16.581

Table 6. The Prediction Performances for Dresser Furniture

Model	Testing Phase		Training Phase	
	R ²	RMSE	R ²	RMSE
DL	0.874	60.190	0.827	51.648
RF	0.939	51.715	0.969	23.106

In terms of performance measurements, the RF algorithm demonstrated the highest R² (0.939) for price estimations of dresser furniture, while the DL algorithm exhibited the lowest R² (0.797) for library furniture. The RF algorithm outperformed the DL algorithm in all tests, demonstrating higher performance. It has been determined that all models were suitable for predicting furniture prices. In the literature, an R² value greater than 0.70 is accepted as a satisfactory result (Wadie *et al.* 2006; Heng and Suetsugi 2013). These results show that models can be used for price estimation in the furniture industry. Figure 3 illustrates the correlation between the actual and predicted prices of bookcase furniture during the testing phase, utilizing the DL and RF models.

Figure 4 illustrates the correlation between the actual and predicted prices of dresser furniture during the testing phase, utilizing the DP and RF models.

The predicted results of the two data mining models for bookcase and dresser furniture are presented in Tables 7 and 8 during the testing phase. The data presented in Tables 7 and 8 assess the predictive power of our models. These data demonstrate that the present models are aligned with actual prices and generally exhibit a low error rate. These findings indicate that these models could be an effective tool in forecasting future price trends.

For dresser furniture, the percentages of correct predictions obtained by DL and RF models were determined as 79.62% and 83.06%, respectively. For bookcase furniture, the correct prediction percentages of the DL and RF models were calculated as 74.75% and 82.60%, respectively. It was found that the prediction success of dresser furniture was higher than bookcase furniture. This is attributed to the fact that bookcase furniture shows more variety in the end product.

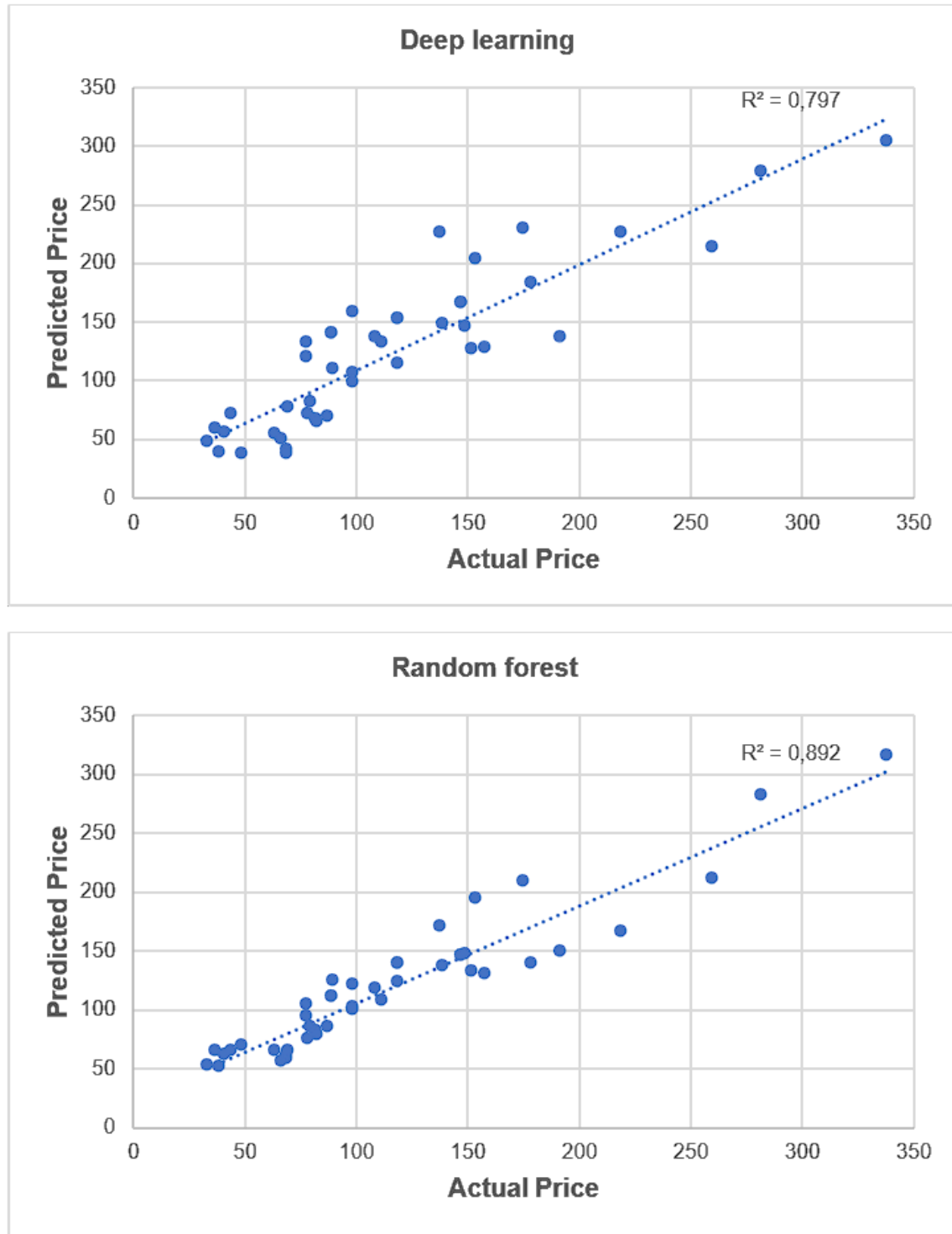


Fig. 3. The DL and RF models' actual and predicted price results for bookcase furniture

In literature, the assessment of model performance based on prediction percentage values can be classified as follows: a prediction percentage value exceeding 90% indicates a highly accurate model, while a prediction percentage value falling within the range of 89% to 80% is a good prediction. Additionally, a prediction percentage value ranging from 79% to 50% is considered a reasonable prediction model (Lewis 1982).

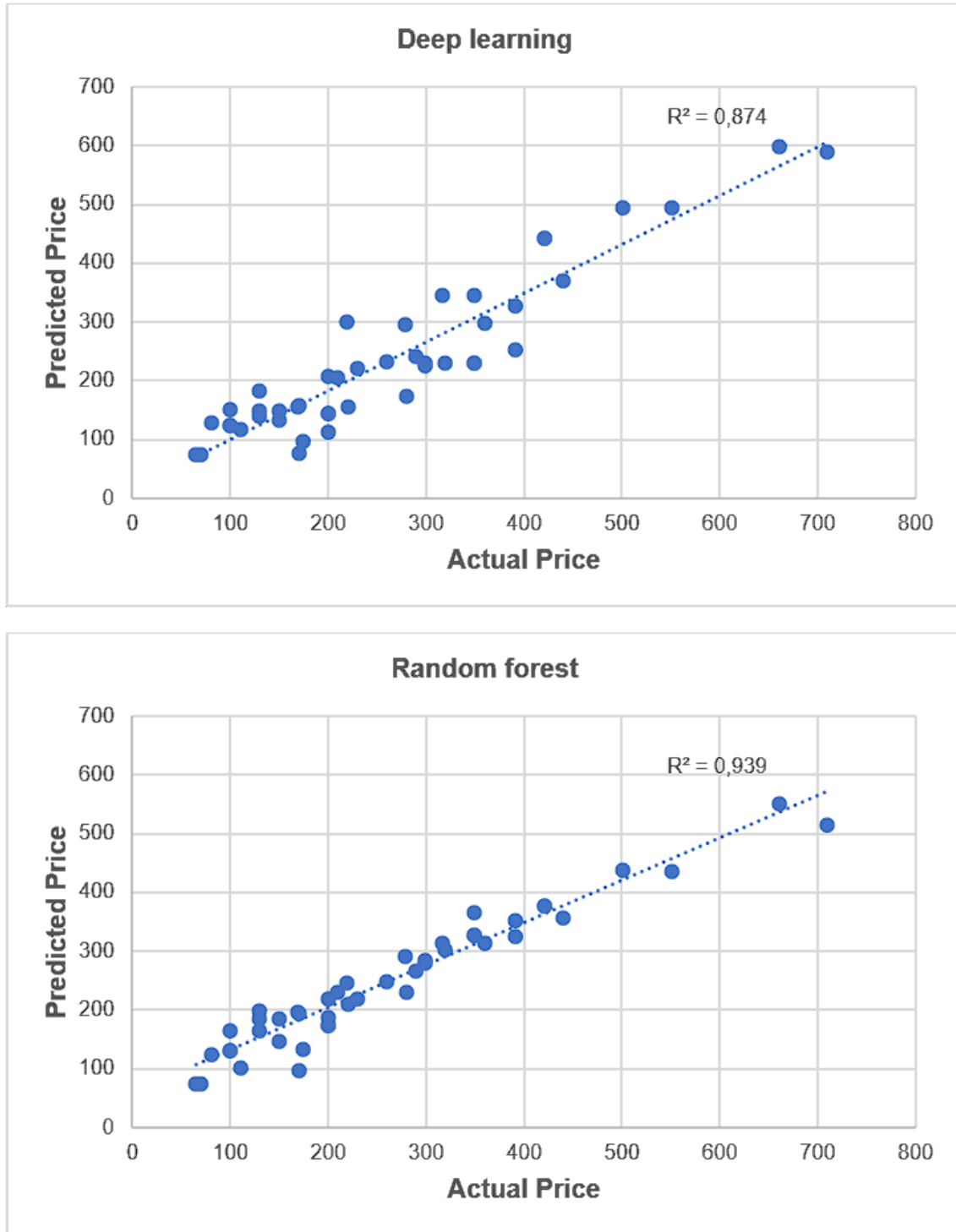


Fig. 4. The DL and RF models' actual and predicted price results for dresser furniture

Table 7. Predicted Results of the Two Machine Learning Models Tested for Bookcase Furniture

Sample	Actual Price	DL		RF	
		Predicted Price	Error (%)	Predicted Price	Error (%)
1	34	48	41	52	54
2	37	58	56	64	73
3	39	38	2	51	30
4	41	55	34	61	48
5	44	71	61	65	48
6	49	37	24	69	40
7	64	54	15	65	2
8	67	49	26	56	17
9	69	40	42	58	16
10	69	37	46	61	11
11	70	77	10	65	7
12	78	132	70	105	35
13	78	119	53	93	19
14	79	70	12	75	6
15	80	81	2	85	6
16	82	66	20	81	1
17	82	66	20	81	1
18	83	64	23	79	5
19	88	69	21	85	4
20	89	139	56	111	25
21	89	139	56	111	25
22	90	109	21	124	38
23	99	158	59	121	22
24	99	106	7	100	1
25	99	99	0	101	2
26	109	136	25	118	9
27	112	132	18	108	4
28	119	152	28	139	17
29	119	152	28	139	17
30	119	114	4	122	3
31	138	225	63	170	23
32	139	148	7	136	2
33	147	165	12	146	0
34	147	165	12	146	0
35	149	146	2	147	2
36	152	126	17	131	14
37	154	202	31	195	27
38	158	128	19	130	18
39	176	228	30	208	18
40	179	183	2	140	22
41	192	136	29	149	22
42	219	226	3	166	24
43	260	214	18	211	19
44	282	278	1	281	0
45	338	304	10	314	7

Table 8. Predicted Dresser Furniture Prices of the Two Machine Learning Models for Testing Phase

Sample	Actual Price	DL		RF	
		Predicted Price	Error (%)	Predicted Price	Error (%)
1	65	76	17	76	17
2	70	76	9	76	9
3	80	128	60	124	54
4	99	125	26	131	33
5	99	152	54	164	66
6	99	125	26	131	33
7	110	118	7	104	6
8	129	184	43	200	55
9	129	141	9	185	43
10	129	149	15	164	27
11	149	150	1	186	25
12	150	135	10	147	2
13	169	156	8	198	17
14	170	159	7	195	15
15	170	78	54	97	43
16	173	96	44	135	22
17	199	208	4	220	11
18	199	145	27	175	12
19	199	114	43	187	6
20	199	145	27	175	12
21	209	206	2	233	11
22	218	302	39	246	13
23	219	156	29	210	4
24	229	222	3	219	4
25	258	235	9	249	4
26	279	297	6	292	5
27	280	175	37	231	17
28	289	242	16	268	7
29	299	233	22	284	5
30	299	226	24	280	6
31	315	346	10	314	0
32	320	232	28	303	5
33	349	229	34	326	7
34	349	345	1	367	5
35	349	229	34	326	7
36	360	299	17	314	13
37	390	327	16	353	9
38	390	254	35	325	17
39	421	443	5	377	11
40	421	443	5	377	11
41	440	371	16	357	19
42	500	496	1	440	12
43	550	497	10	437	20
44	660	599	9	551	16
45	710	592	17	515	27

The relative importance of attributes was ascertained using Random Forest (RF) algorithms. These attributes encompassed depth, height, frame material, number of partitions, number of drawers, and color. The corresponding weights were assigned to each

attribute, signifying their individual feature significance. The attribute weights obtained through the RF algorithms are depicted for bookcase and dresser in Table 9.

Table 9. Attribute Weights Obtained through the RF Algorithms Depicted for Bookcase and Dresser

Bookcase		Dresser	
Attribute	Weight	Attribute	Weight
Width	0.291	Height	0.244
Height	0.195	Width	0.208
Depth	0.142	Color	0.189
Color	0.137	Depth	0.146
Number of partitions	0.125	Frame Material	0.108
Frame Material	0.110	Number of drawers	0.105

According to the RF algorithm (Table 9), the bookcase and dresser furniture's width, height, depth, and color are the key factors influencing the price. This situation suggests that a product that deviates from standard measurements and specifications can be regarded as customized, which may result in a higher price. As suggested by Zavadskas and colleagues (2010), data mining algorithms serve as valuable methods for determining the weights of different attributes in complex decision-making scenarios.

The utilization of web mining as an approach to enhance furniture price determination enables companies to analyze competitor prices effectively. Through conducting thorough analyses, companies can establish price ranges that align with the market constraints, resulting in considerable time and cost savings. Moreover, the web mining approach could enhance customer satisfaction and engagement. Consequently, incorporating web mining techniques in the furniture industry's pricing strategies is valuable for gaining a competitive edge. In future research, it is planned to examine persistent disagreements among different websites based on specific furniture features, such as price. These analyses can provide valuable insights to stakeholders and relevant parties in the industry, enabling them to make better decisions and anticipate potential conflicts in advance.

CONCLUSIONS

1. The different furniture prices were reasonably yielded and estimated through web mining techniques.
2. Two distinct prediction models were implemented utilizing the deep learning (DL) and random forest (RF) algorithms, and their performances were evaluated by employing R^2 and root mean squared error (RMSE) metrics. In the testing phase, the RF algorithm exhibited superior performance, achieving an R^2 value of 0.892 and 0.939 for bookcase and dresser furniture, respectively.
3. In the predictions made by DL and RF algorithms, a higher success rate was determined in dresser furniture than in bookcase furniture.

4. According to the RF algorithm, the main factors that affect the prices of furniture are width, height, depth, and color.
5. Furniture companies can enhance their pricing optimization by conducting competitor analysis. In this regard, web mining emerges as a promising alternative to conventional methodologies, offering significant time and cost savings for these companies.

REFERENCES CITED

- Agwu, O. E., Akpabio, J. U., and Dosunmu, A. (2020). "Artificial neural network model for predicting the density of oil-based muds in high-temperature high-pressure wells," *Journal of Petroleum Exploration and Production Technology* 10(3), 1081-1095. DOI: 10.1007/s13202-019-00802-6
- Bardak, S., Bardak, T., Peker, H., Sözen, E., and Çabuk, Y. (2021). "Predicting effects of selected impregnation processes on the observed bending strength of wood, with use of data mining models," *BioResources* 16(3), 4891-4904. DOI: 10.15376/biores.16.3.4891-4904
- Brauner, T., Heumann, M., Kraschewski, T., Prahlow, O., Rehse, J., Kiehne, C., and Breitner, M. H. (2022). "Web content mining analysis of e-scooter crash causes and implications in Germany," *Accident Analysis and Prevention* 178, article ID 106833. DOI: 10.1016/j.aap.2022.106833
- Breiman, L. (2001). "Random Forests," *Machine Learning* 45(1), 5-32. DOI: 10.1023/A:1010933404324
- Chen, Y., Zou, Y., Zhou, Y., and Zhang, C. (2016). "Multi-step-ahead crude oil price forecasting based on grey wave forecasting method," *Procedia Computer Science* 91, 1050-1056. DOI: 10.1016/j.procs.2016.07.147
- Duan, H., and Liu, Y. (2021). "Research on a grey prediction model based on energy prices and its applications," *Computers & Industrial Engineering* 162, article ID 107729. DOI: 10.1016/j.cie.2021.107729
- Emsley, M. W., Lowe, D. J., Duff, A. R., Harding, A., and Hickson, A. (2002). "Data modelling and the application of a neural network approach to the prediction of total construction costs," *Construction Management and Economics* 20(6), 465-472. DOI: 10.1080/01446190210151050
- Furtney, J. K., Thielsen, C., Fu, W., and Le Goc, R. (2022). "Surrogate models in rock and soil mechanics: integrating numerical modeling and machine learning," *Rock Mechanics and Rock Engineering* 55(5), 2845-2859. DOI: 10.1007/s00603-021-02720-8
- Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, S. R., Tiede, D., and Aryal, J. (2019). "Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection," *Remote Sensing* 11(2), article 196. DOI: 10.3390/rs11020196
- Hassim, S., Muniandy, R., Alias, A. H., and Abdullah, P. (2018). "Construction tender price estimation standardization (TPES) in Malaysia," *Engineering, Construction and Architectural Management* 25(3), 443-457. DOI: 10.1108/ECAM-09-2016-0215
- He, Y., Chu, Y., Song, Y., Liu, M., Shi, S., and Chen, X. (2022). "Analysis of design strategy of energy efficient buildings based on databases by using data mining and

- statistical metrics approach,” *Energy and Buildings* 258, article ID 111811. DOI: 10.1016/j.enbuild.2021.111811
- Heng, S., and Suetsugi, T. (2013). “Using artificial neural network to estimate sediment load in ungauged catchments of the Tonle Sap river basin, Cambodia,” *Journal of Water Resource and Protection* 5, 111-123. DOI: 10.4236/jwarp.2013.52013
- Kayser, V., and Shala, E. (2020). “Scenario development using web mining for outlining technology futures,” *Technological Forecasting and Social Change* 156, article ID 120086. DOI:10.1016/j.techfore.2020.120086
- Keet, C. M., Ławrynowicz, A., D’Amato, C., Kalousis, A., Nguyen, P., Palma, R., Stevens, R., and Hilario, M. (2015). “The data mining optimization ontology,” *Journal of Web Semantics* 32, 43-53. DOI: 10.1016/J.WEBSEM.2015.01.001
- Kim, P. T., and Falcone, R. A. (2017). “Nonaccidental trauma in pediatric surgery,” *Surgical Clinics* 97(1), 21-33. DOI: 10.1016/j.suc.2016.08.002
- Lee, K. C., and Lee, S. (2011). “Interpreting the web-mining results by cognitive map and association rule approach,” *Information Processing and Management* 47(4), 482–490. DOI: 10.1016/j.ipm.2010.11.005
- Lewis, C. D. (1982). *International and Business Forecasting Methods*, Butterworths, London, UK.
- Li, J., Zhu, S., and Wu, Q. (2019). “Monthly crude oil spot price forecasting using variational mode decomposition,” *Energy Economics* 83, 240-253. DOI: 10.1016/j.eneco.2019.07.009
- Li, Z., Xu, W., Zhang, L., and Lau, R. Y. K. (2014). “An ontology-based web mining method for unemployment rate prediction,” *Decision Support Systems* 66, 114-122. DOI: 10.1016/j.dss.2014.06.007
- Lihra, T., Buehlmann, U., and Graf, R. (2012). “Customer preferences for customized household furniture,” *Journal of Forest Economics* 18(2), 94-112. DOI: 10.1016/j.jfe.2011.11.001
- Mariano, A. M., Ferreira, A. B. de M. L., Santos, M. R., Castilho, M. L., and Bastos, A. C. F. L. C. (2022). “Decision trees for predicting dropout in engineering course students in Brazil,” *Procedia Computer Science* 214, 1113-1120. DOI: 10.1016/j.procs.2022.11.285
- Massimino, B. (2016). “Accessing online data: Web-crawling and information-scraping techniques to automate the assembly of research data,” *Journal of Business Logistics* 37(1), 34-42. DOI: 10.1111/jbl.12120
- Naser, M. Z. (2023). “Machine learning for all! Benchmarking automated, explainable, and coding-free platforms on civil and environmental engineering problems,” *Journal of Infrastructure Intelligence and Resilience* 2(1), article ID 100028. DOI: 10.1016/j.iintel.2023.100028
- Orte, F., Mira, J., Sánchez, M. J., and Solana, P. (2023). “A random forest-based model for crypto asset forecasts in futures markets with out-of-sample prediction,” *Research in International Business and Finance* 64, article ID 101829. DOI: 10.1016/j.ribaf.2022.101829
- Pérez-Campuzano, D., Rubio Andrada, L., Morcillo Ortega, P., and López-Lázaro, A. (2022). “Visualizing the historical COVID-19 shock in the US airline industry: A data mining approach for dynamic market surveillance,” *Journal of Air Transport Management* 101, article ID 102194. DOI: 10.1016/j.jairtraman.2022.102194
- Pervez, Md. N., Yeo, W. S., Mishu, Mst. M. R., Talukder, Md. E., Roy, H., Islam, Md. S., Zhao, Y., Cai, Y., Stylios, G. K., and Naddeo, V. (2023). “Electrospun nanofiber

- membrane diameter prediction using a combined response surface methodology and machine learning approach,” *Scientific Reports* 13(1), article 9679. DOI: 10.1038/s41598-023-36431-7
- Putri, R. T., Sukiyono, K., and Sumartono, E. (2019). “Estimation of Indonesian beef price forecasting model,” *Agritropica: Journal of Agricultural Science* 2(1), 46-56. DOI: 10.31186/j.agritropica.2.1.46-55
- Rapidminer. (2023). *RapidMiner Documentation*, (<https://docs.rapidminer.com/>).
- Raposo, A., Frade, S., Alves, M., and Marques, J. F. (2018). “The neural bases of price estimation: Effects of size and precision of the estimate,” *Brain and Cognition* 125, 157-164. DOI: 10.1016/j.bandc.2018.07.005
- Razali, A. M., and Al-Wakeel, A. A. (2013). “Mixture Weibull distributions for fitting failure times data,” *Applied Mathematics and Computation* 219(24), 11358-11364. DOI: 10.1016/j.amc.2013.05.062
- Saha, S., Bera, B., Shit, P. K., Bhattacharjee, S., and Sengupta, N. (2023). “Prediction of forest fire susceptibility applying machine and deep learning algorithms for conservation priorities of forest resources,” *Remote Sensing Applications: Society and Environment* 29, article ID 100917. DOI: 10.1016/j.rsase.2022.100917
- Sarker, I. H. (2021). “Machine learning: Algorithms, real-world applications and research directions,” *SN Computer Science* 2(3), 160. DOI: 10.1007/s42979-021-00592-x
- Schonlau, M., and Zou, R. Y. (2020). “The random forest algorithm for statistical learning,” *The Stata Journal* 20(1), 3-29. DOI: 10.1177/1536867X20909688
- Velten, K., Reinicke, R., and Friedrich, K. (2000). “Wear volume prediction with artificial neural networks,” *Tribology International* 33(10), 731-736. DOI: 10.1016/S0301-679X(00)00115-8
- Wadie, B. S., Badawi, A. M., Abdelwahed, M., and Elemabay, S. M. (2006). “Application of artificial neural network in prediction of bladder outlet obstruction: A model based on objective, noninvasive parameters,” *Urology* 68(6), 1211-1214. DOI: 10.1016/J.UROLOGY.2006.08.1079
- Zavadskas, E. K., Turskis, Z., Ustinovichius, L., and Shevchenko, G. (2010). “Attributes weights determining peculiarities in multiple attribute decision making methods,” *Engineering Economics* 66(1), 32-43. DOI: 10.3846/transport.2010.52
- Zhao, H., Yao, X., Liu, Z., and Yang, Q. (2021). “Impact of pricing and product information on consumer buying behavior with customer satisfaction in a mediating role,” *Frontiers in Psychology* 12, article 720151. DOI: 10.3389/fpsyg.2021.720151
- Zhou, J., Wei, J., and Xu, B. (2021). “Customer segmentation by web content mining,” *Journal of Retailing and Consumer Services* 61, article ID 102588. DOI: 10.1016/j.jretconser.2021.102588

Article submitted: May 25, 2023; Peer review completed: August 5, 2023; Revised version received and accepted: August 31, 2023; Published: September 11, 2023.
DOI: 10.15376/biores.18.4.7412-7427