JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGY
https://e-journal.uum.edu.my/index.php/jict

How to cite this article:
Abu-Shareha, A. A. (2022). TOPSIS-based regression algorithms evaluation. *Journal of Information and Communication Technology, 21*(4), 513-547. https://doi.org/10.32890/jict2022.21.4.3

TOPSIS-based Regression Algorithms Evaluation

Ahmad Adel Abu-Shareha
Department of Computer Science,
Faculty of Information Technology,
Al-Ahliyya Amman University, Amman, Jordan

a.abushareha@ammanu.edu.jo

Received: 12/4/2022 Revised: 26/5/2022 Accepted: 7/7/2022 Published: 20/10/2022

ABSTRACT

This paper developed a multi-criteria decision-making approach using the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) to benchmark the regression alternatives. Regression is used in diverse fields to predict consumer behavior, analyze business profitability, assess risk, analyze automobile engine performance, predict biological system behavior, and analyze weather data. Each of these applications has its own set of concerns, resulting in various metrics utilizations or those of similar measures but with diverse preferences. Multi-criteria decision-making analyzes, compares, and ranks a set of alternatives utilizing mathematical and logical processes with a complicated and contradictory set of criteria. The developed approach established the weights, which were the core of the evaluation process, to various values to mimic and address the regression's utilization in multiple applications with different concerns and using distinct datasets. The alternative judgment identified positive and negative ideal alternatives in the alternative space. The compared regression alternatives were scored and ranked based on their distance from these alternatives. The results showed that different preferences led to varying algorithm rankings, but top-ranked algorithms were distinguished using a specific dataset. Following that, using three datasets, namely Combined Cycle Power Plant, Real Estate, and Concrete, Voting using multiple classifiers (k-means-based classifiers) was the top-ranked in the Combined Cycle Power Plant and Real Estate datasets. In contrast, Decision Stump was the top-ranked in the Concrete dataset.

Keywords: Multi-criteria decision making, Regression, TOPSIS.

INTRODUCTION

Data mining applications are fast-growing with the increase in big data and the emergence of the so-called data science field (Provost & Fawcett, 2013). Data mining is implemented using one of the many available algorithms chosen based on the input data and the desired output. Data regression is a well-known and commonly used data mining application (Draper & Smith, 1998). Regression is a prediction task with continuous outputs trained using sample data. Statistical and logical methods are used to model and estimate the relationships between the dependent variable(s) and the independent variable(s). Although there are different regression analysis tasks, such as linear, nonlinear, and multiple, the regression algorithms all have the same fundamental concept and input and output formats (Fox & Weisberg, 2018). The existing regression algorithms can be classified into linear regression, Support Vector Machine (SVM), Nearest Neighbor, Decision Tree, Ensemble methods, and Neural Network (Draper & Smith, 1998).

Regression is used in various fields to predict consumer behavior, analyze business profitability, assess risk, analyze automobile engine performance, predict biological system behavior, and analyze weather data (Bates & Watts, 1988). Each of these applications has its own set of concerns, resulting in numerous metrics utilizations or those of similar measures but with diverse preferences (Baumann et al., 2019). Engineers and biologists, for example, are interested in convergence;

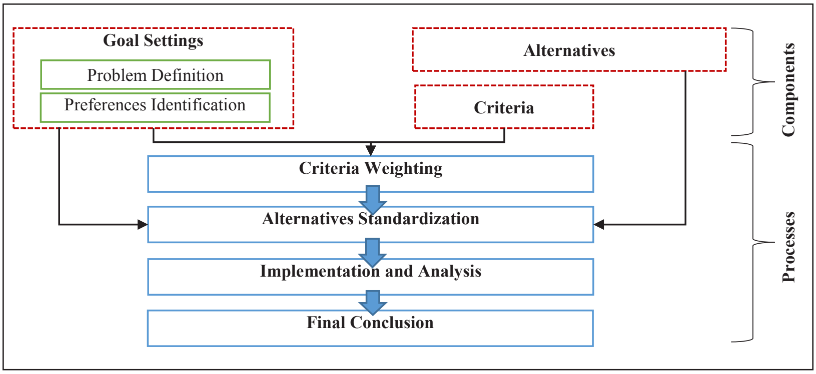
however, economists are more concerned with the accuracy of expected outputs (Pandey & Nguyen, 1999). Furthermore, different measurements attempt regression performance from different perspectives. Subsequently, benchmarking the regression algorithms in various fields, or even in a single field, is not trivial given the previously mentioned variations in concerns and measurements (Zorlu, 2012).

Multi-Criteria Decision Analysis/Making (MCDA/MCDM) analyzes, compares, and ranks a set of alternatives utilizing mathematical and logical processes with a complicated and contradictory set of criteria (Skilodimou et al., 2019). The components of the MCDM task are a problem with multiple alternatives with conflicting preferences and a ranking goal (Malczewski, 1999; Petrovic-Lazarevic & Abraham, 2004). As illustrated in Figure 1, MCDM takes a set of alternatives as input, each identified using values for the specified criteria. The criteria should be wisely selected to characterize the alternatives and reflect the desired goal. Performance benchmarking should consider, as an example, the error, model variance, and bias criteria. Furthermore, the set of utilized criteria should be well differentiated (Hwang & Yoon, 2012). The criteria are weighted based on the preferences. The alternatives are then standardized based on the weighted criteria. Finally, the alternatives are evaluated and ranked (Srisawat & Payakpate, 2016). As a result, MCDM can be used to rank the regression algorithms with different performance criteria. Moreover, the analysis can be used with multiple prospective evaluations from many evaluators, characterizing the regression algorithms' performance in various fields with countless concerns (Oliveira et al., 2014).

This paper develops a Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS)-based multi-criteria decision-making approach to evaluate various regression alternatives. The developed approach establishes the weights, which are the core of the evaluation process, to various values to mimic and address the regression's utilization in multiple applications with different concerns and using distinct datasets. The alternative judgment identifies positive and negative ideal alternatives in the alternative space. The compared regression alternatives are scored and ranked based on their distance from these alternatives.

Figure 1

MCDM's Processes and Components



RELATED WORKS

MCDM has been used to evaluate alternative algorithms in various fields, especially in data mining (i.e., classification and clustering). The widespread use of MCDM techniques can be traced back to several factors: 1) the number of data mining applications is rapidly increasing; 2) the outcomes of these applications are evaluated based on a group of preferences that is characterized by a set of criteria, which are mostly conflicted and overlapped (Kasim et al., 2011); and 3) the preferences differ from an application to another, which makes the alternatives evaluation complicated (Govindan & Jepsen, 2016).

MCDM Techniques

Various MCDM techniques have been developed, such as Multi-Attribute Utility Theory (MUAT) (Winterfeldt & Fischer, 1975), Simple Multi-Attribute Rating Technique (SMART) (Edwards, 1971), Analytical Hierarchy Process (AHP) (Wind & Saaty, 1980), Analytic Network Process (ANP) (Saaty, 2005), Case-Based Reasoning (CBR) (Li & Sun, 2008), Data Envelopment Analysis (DEA) (Belton & Vickers, 1993), TOPSIS (Shih et al., 2007), Elimination and Choice Translating Reality (ELECTRE) (Roy, 1968), Preference Ranking Organization Method for Enrichment of Evaluation (PROMETHEE) (De Keyser & Peeters, 1996), Weighted Sum Method (WSM) (MacCrimmon, 1968), and Grey Relational Analysis (GRA) (Julong, 1989) as summarized in Table 1.

The MCDM techniques are implemented using different mechanisms: pairwise comparison as implemented by AHP and ANP, individual alternative evaluation, such as SMART and WSM, or distance to some ideal solutions, such as TOPSIS. The characteristics of existing MCDM techniques include: 1) scalability; 2) applied in non-strict preferences with uncertainty; and 3) applied in limited data. Major disadvantages of the MCDM techniques are the inability to be implemented with multiple preferences in uncertain or diverse environments like the environment under which the regression is executed and the non-scalability (Gao & Xuan, 2019). Among the existing techniques, TOPSIS overcomes such problems by evaluating the alternatives based on their distance from positive and negative ideal alternatives.

Table 1

Multi-Criteria Decision-Making Techniques

Technique	Characteristics
MUAT	A scalable early approach for alternative ranking assigns a utility value for every alternative. The disadvantage of MUAT is that it is sensitive to strict preference identification.
SMART	Assigns a score to each alternative. It is scalable, yet it is also sensitive to strict preference identification.
AHP	Implements pairwise comparison of the alternatives, which allows for multiple preferences, but is non-scalable.
ANP	Similar to AHP, it allows for multiple preferences but is non-scalable.
CBR	Calculates the similarity between alternatives; it is applied for strict preference identification only and is non-scalable.
DEA	Compares alternatives to each other, and it is similar to CBR in that it is applied for strict preference identification only and is non-scalable.
TOPSIS	Scalable technique works with multiple preferences as it evaluates the alternatives based on their distances from ideal solutions, requiring extra data to produce accurate output.
ELECTRE	Similar to AHP and ANP, it implements pairwise comparison and allows for multiple preferences but is non-scalable.

(continued)

Technique	Characteristics
PROMETHEE	Calculate the similarity between alternatives; it is applied for strict preference identification only and is non-scalable.
WSM	Applied for simple problems with strict preference identification only.
GRA	Assigns a score to each alternative. It is scalable, yet it is also sensitive to strict preference identification.

Literature Review on MCDM in Data Mining

MCDM techniques have been used to evaluate options in various domains in the literature. Nakhaeizadeh and Schnabl (1997) proposed an approach for ranking 22 binary-classification algorithms using the DEA technique. The implemented approach aimed to ease the classification algorithms' evaluation and extend the previous evaluation, which was only based on accuracy. DEA was implemented with storage, training and testing times, and the training and testing error rate criteria. Osei-Bryson (2004) proposed an approach for ranking ten Decision Tree classification algorithms using the AHP MCDM method. The implemented approach evaluated these algorithms based on their accuracy, discrimination abilities, stability, number of leaves, and number of rules.

Lavesson and Davidsson (2007) used MCDM to evaluate the classification algorithms easily. DEA was used for ranking 18 classification algorithms. The classification algorithms were assessed based on training and testing accuracy, complexity, and true positive versus false positive criteria. Generally, for early utilization of the MCDM technique, as discussed earlier, the data mining algorithms are evaluated based on a single preference.

Peng et al. (2011) ranked seven multi-class classification algorithms using multiple MCDM techniques. The classification algorithms were evaluated using customized cost and benefit measures, which indicated a single preference in evaluating these algorithms. The criteria were weighted proportions of the misclassification samples, which depended on the dataset used. For example, misclassifying class-1 as class-2 might have a higher cost than misclassifying class-2 as class-1. The Multi-Criteria Optimization and Compromise Solution (VIKOR), WSM, PROMETHEE, and TOPSIS MCDM techniques were used for benchmarking and evaluation. As a result, each MCDM technique produced various ranks of the output. The disagreement problem among the utilized MCDM techniques was resolved by assigning a weight value to each of them.

Peng et al. (2012) ranked 38 binary-classification algorithms using multiple MCDM techniques. The proposed evaluation aimed to evaluate the classification algorithm for software defect prediction using 13 criteria, including various true positive, false positive, true negative, and false negative combinations, such as precision, recall, and accuracy. Area under receiver operative characteristic (AUC), mean square error (MSE), and training and testing times were also used as criteria for the evaluation. The benchmarking and evaluation techniques were DEA, ELECTRE, PROMETHEE, and TOPSIS MCDM. The disagreement in alternative rankings among these techniques was left unresolved. Similarly, Kou et al. (2012) used GRA, VIKOR, ELECTRE, PROMETHEE, and TOPSIS MCDM techniques to rank 17 classification algorithms in the risk analysis field. There were 11 criteria used, including accuracy, time, knowledge, attitudes, and practices (KAPs), and other measures based on true and false positive and negative portions. Spearman's rank correlation coefficient was used to adjust the pre-determined criteria weights and resolve the disagreement between the output of different MCDM techniques.

Kou et al. (2014) employed TOPSIS to rank five binary-classification algorithms in conjunction with three feature selection techniques: Principal Component Analysis (PCA), Independent Component Analysis (ICA), and Synthetic Minority Oversampling Technique (SMOTE). *The proposed evaluation used accuracy, sensitivity, specificity, type I error, type II error, and AUC criteria to assess the classification algorithm for the bank loan classification problem. K-Nearest Neighbor (KNN) was ranked first for bank loan classification among Naive Bayesian, Logistic, C4.5, and Classification and Regression Trees (CART).*

Peteiro-Barral et al. (2017) used GRA, VIKOR, and TOPSIS to rank five binary-classification algorithms in conjunction with a feature selection technique. *The proposed evaluation applied accuracy, true positive, true negative, precision, F-measure, and AUC criteria to assess the classification algorithm for detecting evaporative dry eye from eye images. Spearman's rank correlation coefficient was used to resolve the disagreement.* Song and Peng (2019) utilized TOPSIS to rank four binary-classification algorithms in conjunction with five oversampling techniques, including SMOTE. *The proposed evaluation used false positive, true false negative, F-measure, G-mean, AUC, and time criteria to assess the classification algorithm for* imbalanced financial risk assessment problems with a single preference setting.

Kou et al. (2020) used GRA, VIKOR, WSM, PROMETHEE, and TOPSIS to rank the SVM classifier in conjunction with ten different feature selection methods, information gain, Gini index, document frequency, distinguishing feature, expected cross-entropy, class discriminating measure, mutual information, odds ratio, Chi-square, and weighted log-likelihood. *The proposed evaluation applied running, training and testing times, stability, true positive, true negative, accuracy, MAE, and AUC criteria to assess the output for text classification problems with a single preference setting.*

In conclusion, a single preference is reflected when single or multiple MCDM methods are used, as presented in the literature. The weights of the criteria must be adjusted to achieve a consistent output using multiple MCDM techniques. On the other hand, adjusting the weights of the criteria alters preferences. The previously conducted benchmarking revealed that adjusting preference is implemented to achieve agreement across multiple techniques, resulting in inconsistent output, which necessitates changing preferences. Maintaining the original preferences and comprising multiple preferences in a single alternative ranking have not been investigated yet. Accordingly, TOPSIS is used in this paper, which allows for unifying the output and compromise ranking of the alternatives based on multiple preferences (e.g., criteria weighting). This paper evaluates and benchmarks regression algorithms in various applications via numerous datasets using TOPSIS and weight settings reflecting multiple preferences.

TOPSIS MCDM Technique

TOPSIS is characterized by its ability to compromise the ranking of the alternatives with internal and external decision grouping, allowing for negotiation among multiple decision-makers to reach a compromise solution and a combination of various preferences (Shih et al., 2007). TOPSIS is scalable; thus, new alternatives and decision-makers can be added at any stage of the decision-making process. Technically, positive and negative alternatives are identified in the alternative space, and the compared alternatives are ranked according to their distance from these alternatives. The top-ranked alternative has the shortest distance to the ideal positive alternative and the longest distance to the ideal negative alternative. Compared to other methods, the results can always be compromised (Adepoju et al., 2020).

TOPSIS has been used in various fields to evaluate algorithms and choices. Yap et al. (2019) reviewed MCDM techniques used to solve the common site selection problem in public services, logistics, energy generation, and retail facilities. The review found that AHP, ELECTRE, PROMTHEE, and TOPSIS were frequently used for site selection. TOPSIS's advantages, as stated in the review, are scalability and ease of use. Regardless of the number of alternatives and criteria, the method's implementation remains the same.

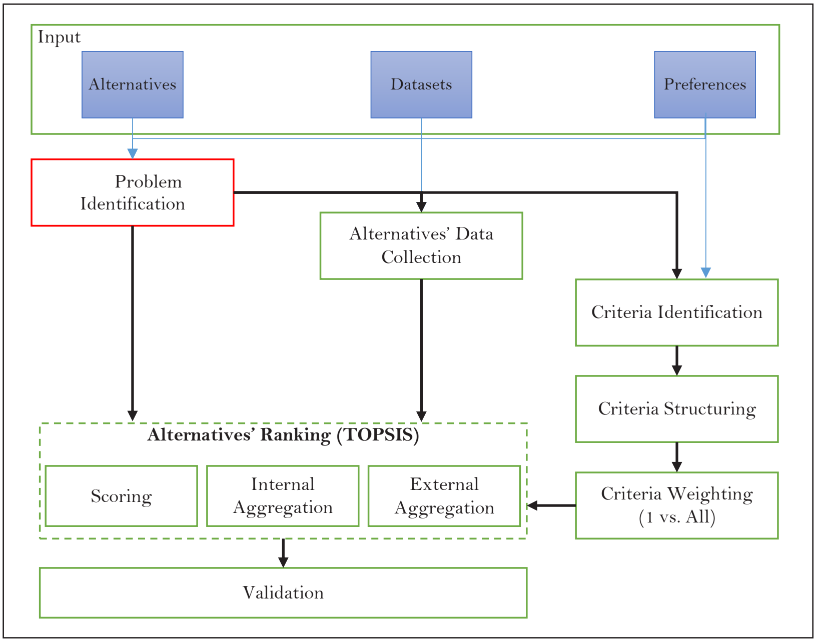
Song and Peng (2019) chose TOPSIS for ranking algorithms to assess financial risk imbalances. Compared to other techniques, this election was justified by its simplicity and widespread use. TOPSIS was used by Siregar et al. (2021) to select the most prominent class students based on scores, personality, and attitude. Although the selection was not contested, it was demonstrated that using TOPSIS simplified the selection of prominent class students. Chodha et al. (2022) used TOPSIS to find the best industrial robot for arc welding among the available options. TOPSIS was chosen because it allows for trade-offs between criteria and has a flexible weighting system. TOPSIS has been preferred in various alternative selection problems across multiple domains. Overall, TOPSIS stands out from other MCDM techniques because of its scalability and insensitivity.

PROPOSED WORK

A TOPSIS-based framework with weighted and structured criteria was proposed to solve the regression algorithms' evaluation, benchmarking, and ranking under multiple preferences. As illustrated in Figure 2, the proposed framework took the following inputs for evaluation and ranking: alternatives, datasets for the experiments, and preferences. The criteria were identified, structured, and weighted after formulating the problem. The scores of the alternatives that corresponded to the identified criteria were then collected (data collection). The alternatives were evaluated and ranked based on the scores and weights of the criteria. Multiple preferences were considered in the proposed framework; thus, multiple weight settings for the identified criteria were used. Internal and external aggregations were employed to resolve conflicts and unify the output results of multiple preferences. Within the TOPSIS ranking process, internal aggregation was linked to the distance calculation step. On the other hand, external aggregation was applied to the obtained ranks after the TOPSIS ranking had been applied. The results were finally validated using correlation testing.

Figure 2

The Proposed Framework



For multiple preferences, various settings for criteria weighting were considered. Each setting was formulated based on specific grouping and structuring of the criteria and assigning values in the form of a series. Therefore, the obtained output formed *k*-dimensional results, where *k* was the number of the underlying criteria. Accordingly, different applications with varied preferences could be projected over the *k*-dimensional preferences. Multiple weight settings do not require additional processing because TOPSIS allows for the late injection of decision-making preferences. The proposed approach obtained the optimized ranking of the alternatives in many applications characterized by applying flexible structuring of the weighting process's criteria and sequences.

Input Datasets

Three different datasets were used to evaluate and benchmark the regression algorithms. The first dataset was the Combined Cycle Power Plant (CCPP) dataset (Tufekci, 2014), which represented electricitygenerating gas turbines (GT), steam turbines (ST), and heat recovery steam. The dataset contained 9,568 full load samples collected over six years. Temperature (T), ambient pressure (AP), relative humidity (RH), and exhaust vacuum (V) were the independent variables, while the plant's net hourly electrical energy output (EP) was the dependent variable. The second dataset (Yeh & Hsu, 2018) was a Real Estate Market Valuation dataset with 414 samples collected from Sindian District, New Taipei City, and Taiwan. House age, distance to the MRT station, number of convenience stores nearby, geographic coordinates (latitude), and geographic coordinates (longitude) were independent variables, while price was the dependent variable. The last dataset was the Concrete Compressive Strength dataset, which included 173,370 samples of various concretes with different ingredients and ages (Yeh, 1998). Cement, blast furnace slag, fly ash, water, superplasticizer, coarse aggregate, fine aggregate (all measured as quantitative in kg for each m^3 mixture), and age were the independent variables. Concrete compressive strength was the dependent variable. Using multiple datasets expanded the outcomes to accommodate the ranking of alternatives and allowed for multiple preferences benchmarking.

Input Alternatives

Although the proposed approach evaluated and ranked a wide range of regression algorithms, it did not consider all of them. Any new algorithm can be evaluated and compared once the framework is established. Table 2 lists the algorithms that were evaluated and compared.

Input Preferences and Problem Identification

The proposed approach addressed the problem of evaluating and benchmarking the regression algorithms. Various preferences were considered by setting up several weight settings to mimic and address the regression's utilization in multiple applications with different concerns and using distinct datasets.

Step 1: Alternative Data Collection

As shown in Figure 3, the collected data were structured into a matrix, a decision matrix (DM), or an evaluation matrix (EM). Therefore, the alternative ranking process was conducted based on the inputs provided in the DM and the weights of the criteria.

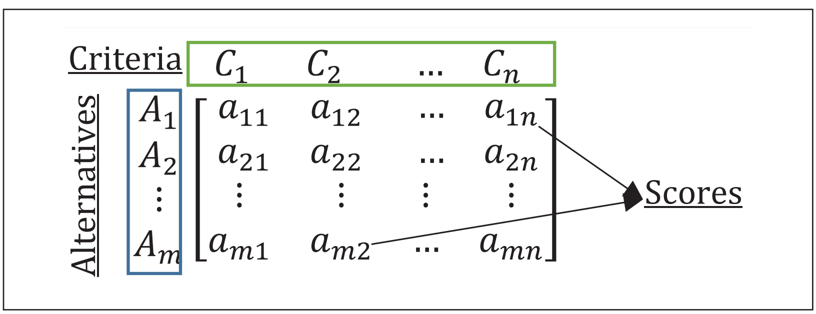
Table 2

Alternative Sets

#	Alternative	Category	Details
1.	Multiple Regression	Linear	Multiple regression is suitable for multiple independent variables.
2.	Polynomial SVM	SVM	Different kernels are used in various implementations of the SVM algorithm. The kernel transfers the input space into a higher dimension. Therefore, different kernels produce different results when used with SVM.
3.	Normalized SVM		
4.	Gaussian SVM		
5.	RBF-SVM		
6.	1-NN	KNN	Various implementations of the KNN algorithm consider different neighborhood numbers.
7.	3-NN		
8.	5-NN		
9.	7-NN		
10.	Decision Stump	Decision Tree	Different approaches to tree construction result in different decision tree algorithms.
11.	Random Tree		
12.	Random Forest		
13.	M5		
14.	Bagging	Ensemble Methods	Using well-known ensemble techniques to combine multiple classifications.
15.	Voting		
16.	Backpropagation 1-Layer NN	Neural Network	Different implementations of the NN algorithm with different numbers produce different results.
17.	Backpropagation 2-Layers NN		

Figure 3

The Decision Matrix Structure



Step 2: Criteria Identification

All known regression performance metrics were applied to evaluate and rank the regression algorithms. Prediction bias, model variance, and relative model variance were the three types of metrics applied in this study. Table 3 lists all of the metrics that were used.

Table 3

Criteria Sets

#	Metric	Category	Captured Aspect
1.	Mean Absolute Error (MAE)	Prediction Bias	These metrics capture the differences between predicted and actual values (average values for MAD).
2.	Mean Average Deviation (MAD)		
3.	Mean Square Error (MSE)	Model Variance	These metrics capture the global model differences between predicted and actual values.
4.	Root Mean Square Error (RMSE)		
5.	Relative Error (RERR)	Relative Model Variance	These metrics capture the global model differences between the predicted and actual values relative to the model's values.
6.	Normalized Root Mean Squared Error (NRMSE)		
7.	Correlation Coefficient (CC)		
8.	Signal-to-Noise Ratio (SNR)		
9.	Efficiency Coefficient (E)		

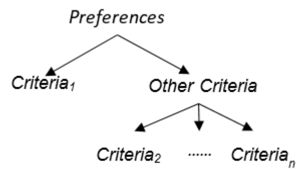
Step 3: Criteria Structuring and Weighting

The criteria are usually structured in a hierarchical representation to form groups and identify relative preferences. The proposed approach used a flexible structure to cover all possible applications with multiple settings. In each setting, one criterion was placed on one side of the created hierarchy, and the rest of the criteria were grouped on the other side, as illustrated in Figure 4.

Figure 4

Multiple Criteria Weighting using the Proposed Structuring

Setting 1	C_1	C_i		C_k	Sum
S_{11}	1.0	0	……..	0	1.0
S_{12}	0.9	x	……..	x	1.0
	:	:	:	:	:
S_{1n}	0.0	x	……..	x	1.0



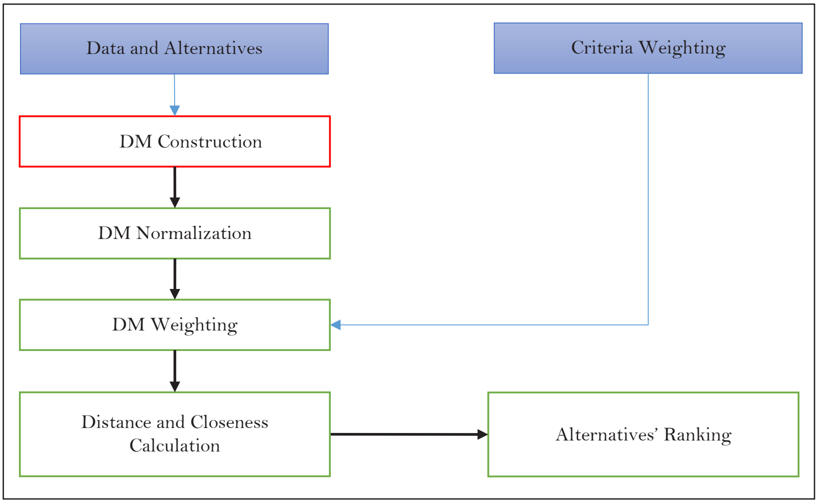
There were k different structures in the proposed approach, each formed around a single criterion. For each structure, there were n settings. The criterion subject matter in each structure was given a different value in each setting in the value range [0-1]. Then, at each structure, n different weights were assigned to the criteria set out of the k structures, leading to $k*n$ different ranking results, where n is the number of values in the set and k is the number of the criteria. Furthermore, for l datasets, $k*n*l$ rankings were generated. The weights that should be considered for the experiments were {0, 0.1, 0.2, 0.3, …1.0} or {0, 0.05, 0.1, 0.15, ….1.0}. A limited set of values, {0.2, 0.5, 0.8}, were utilized to reduce the number of the output ranks without affecting the results. Moreover, rather than creating structures based on individual criteria (i.e., nine criteria), the structure was created based on the three criteria groups. This structuring resulted in nine different rankings for each dataset.

Step 4: Alternative Ranking

The DM, which was created using the results of implementing the regression algorithms on the selected dataset, was used for decision-making. Once the DM was filled, the alternative judgment and ranking were implemented, as illustrated in Figure 5. The alternative ranking was implemented as given in the following steps. First, the DM was normalized so that the values within the matrix were mapped into a compact range. Each cell's value was divided by the square root of the value's sum in each column, whereby each column referred to a single criterion. Accordingly, the normalized value of each cell nc_{ij} was calculated based on the cells c_{ij} in the DM matrix, as given in Equation 1.

Figure 5

Alternative Ranking using TOPSIS



$$nc_{ij} = c_{ij} \Big/ \sqrt{\sum_{i=1} c_{ij}^2} \quad (1)$$

Second, the normalized matrix was weighted with the previously determined weights, as given in Equation 2. Each weight set produced a one-column matrix, which was then multiplied by the weight matrix to produce the weighted matrix. Accordingly, the weighted matrix combined the criteria values and the weights.

$$\begin{bmatrix} nc_{11} & nc_{12} & \dots & nc_{1n} \\ nc_{21} & nc_{22} & \dots & nc_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ nc_{m1} & nc_{m2} & \dots & nc_{mn} \end{bmatrix} * \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \quad (2)$$

Third, the ideal positive and negative alternatives were created by selecting each column's best and worst values as a component of the positive and negative alternatives, respectively, from the weighted normalized matrix. Accordingly, the maximum value was selected as the positive alternative component for the benefit criterion, while the minimum value was selected for the ideal negative alternative. On the other hand, the minimum value was selected as the positive alternative component for the cost criteria, while the maximum value was selected for the negative alternative.

Fourth, the Euclidean distances between each alternative and the positive and the negative ideal alternatives were calculated. These values were then used to calculate a single closeness value for each alternative, as given in Equation 3.

$$\boldsymbol{C_i = D_{i^-}/(D_{i^-} + D_{i^*}),\ \ 0 < \ C_i < 1} \tag{3}$$

where C_i is the closeness of the alternative i, D_i^- is the distance of the alternative i from the ideal negative alternative, and D_i^+ is the distance of the alternative i from the ideal positive alternative.

Fifth, the ranking process was implemented based on the closeness values. The greater the value, the higher the rank of the alternative.

Step 5: Internal and External Aggregations

Internal aggregation was accomplished by averaging each alternative's positive and negative distances. The average negative distance D_i^- was calculated by averaging the distances of the alternative i from all the ideal negative alternatives in various settings, and D_i^+ was calculated by averaging the distances of the alternative i from all the ideal positive alternatives. The external aggregation was achieved by averaging the closeness values calculated for each alternative.

Step 6: Validation

The results were validated using Pearson's correlation, one of the most widely used statistical methods for correlation identification. The purpose of the validation was to ensure that the obtained results were applicable under various settings. Accordingly, if the correlation was satisfied, the conclusions about the ranking algorithms could be generalized and accepted for regression algorithm ranking.

EXPERIMENTAL RESULTS

The experiments were conducted based on the performance of the regression algorithms over the selected datasets, i.e., the CCPP, Real State, and Concrete datasets, and were based on the pre-determined measures. The results of the regression algorithms were used to fill in the DMs, as given in Tables 4, 5, and 6. As noted in the DMs, the algorithms' scores varied depending on the utilized datasets, and their ranks differed from one criterion to another.

Table 4

Results of the Regression Algorithms using the CCPP Dataset

	Criteria								
	MAE	MAD	MSE	RMSE	RERR	NRMSE	CC	SNR	E
A_1	19.43403	14.43947	567.7044	23.82655	1.248509	0.315583	0.01065	5.62E-10	169.1346
A_2	19.60503	14.71364	577.9304	24.04018	1.271656	0.318413	0.01061	5.57E-10	170.3858
A_3	19.39061	14.32943	565.0255	23.77026	1.238867	0.314838	0.01000	5.63E-10	163.532
A_4	19.66724	14.96634	585.8304	24.20393	1.283326	0.320582	0.01113	5.56E-10	175.6967
A_5	19.50998	14.60721	572.393	23.92474	1.259163	0.316884	0.01078	5.60E-10	170.8642
A_6	19.68414	14.8138	587.0185	24.22846	1.28652	0.320907	0.00814	5.55E-10	150.6141
A_7	19.56646	14.72562	579.5454	24.07375	1.270077	0.318858	0.00961	5.58E-10	162.4711
A_8	19.53067	14.70703	577.2778	24.02661	1.265371	0.318233	0.01044	5.59E-10	168.9335
A_9	19.50785	14.69953	575.6564	23.99284	1.261886	0.317786	0.01064	5.60E-10	170.2581
A_{10}	18.1243	14.26623	503.7362	22.44407	1.104438	0.297272	0.00527	6.03E-10	111.7116
A_{11}	19.66084	14.84474	586.8431	24.22485	1.286074	0.320859	0.00713	5.56E-10	141.0118
A_{12}	19.48426	14.67835	575.2648	23.98468	1.260897	0.317678	0.0083	5.61E-10	150.4973
A_{13}	19.46074	14.64874	573.424	23.94627	1.25712	0.317169	0.01116	5.61E-10	174.0266
A_{14}	19.46053	14.67349	574.1352	23.96112	1.25853	0.317366	0.00887	5.61E-10	155.4509
A_{15}	14.83645	0.0499	291.2848	17.06707	0.641083	0.226054	0.01514	7.36E-10	11.90703
A_{16}	19.46694	14.60663	574.7838	23.97465	1.260223	0.317545	0.01091	5.61E-10	172.2393
A_{17}	19.49763	14.72349	575.9377	23.9987	1.262825	0.317864	0.01108	5.60E-10	173.7875

Table 5

Results of the Regression Algorithms using the Real Estate Dataset

	Criteria								
	MAE	MAD	MSE	RMSE	RERR	NRMSE	CC	SNR	E
A_1	13.32916	7.993705	277.485	16.65788	11.35884	0.151573	0.04798	4.40E-07	52.75095
A_2	13.31934	7.821003	277.5082	16.65858	11.72238	0.151579	0.05213	4.40E-07	54.8414
A_3	13.72776	8.723152	287.4174	16.95339	9.436571	0.154262	0.03346	4.27E-07	44.63133
A_4	14.06036	9.365502	301.0966	17.35214	9.966936	0.15789	0.04805	4.17E-07	55.6485
A_5	12.85725	7.255946	259.2109	16.10003	9.303319	0.146497	0.05542	4.56E-07	53.7682
A_6	15.00159	10.54343	337.6709	18.37582	10.94713	0.167205	0.04213	3.91E-07	55.41497
A_7	14.39233	9.957382	315.9843	17.77595	9.925125	0.161747	0.04492	4.07E-07	55.2664
A_8	14.14982	9.680929	303.0854	17.40935	9.450445	0.158411	0.06655	4.14E-07	66.53686
A_9	14.24369	9.709482	305.5909	17.48116	9.521151	0.159064	0.05672	4.12E-07	61.34422
A_{10}	13.08475	9.108823	268.9413	16.39943	8.026656	0.149221	0.0252	4.48E-07	36.75445
A_{11}	15.10666	10.77552	343.8447	18.54305	11.20602	0.168727	0.06551	3.88E-07	70.63635
A_{12}	14.30725	9.699291	308.1619	17.55454	9.592677	0.159732	0.0552	4.10E-07	60.75921
A_{13}	13.63737	9.012325	283.9876	16.85193	9.1053	0.153339	0.06797	4.30E-07	64.65814
A_{14}	13.8606	9.258524	294.0098	17.14671	9.012219	0.156021	0.04996	4.23E-07	56.00375
A_{15}	10.72183	0.264804	185.1483	13.60692	4.874314	0.123812	0.04343	5.47E-07	8.679745
A_{16}	16.56863	12.68808	443.2339	21.05312	13.9848	0.191566	0.00138	3.54E-07	10.95708
A_{17}	15.94012	11.96351	405.6578	20.14095	10.76768	0.183266	0.02382	3.68E-07	44.44491

Table 6

Results of the Regression Algorithms using the Concrete Dataset

	Criteria								
	MAE	MAD	MSE	RMSE	RERR	NRMSE	CC	SNR	E
A_1	17.01126	10.68257	461.96	21.49325	14.74882	0.267762	0.02254	5.55E-08	71.43227
A_2	17.7249	11.66235	505.9242	22.49276	15.38437	0.280214	0.02090	5.33E-08	72.65793
A_3	17.75952	12.42747	501.2203	22.38795	19.08918	0.278908	0.00800	5.32E-08	45.51912
A_4	18.10662	12.67169	523.7138	22.88479	20.48301	0.285098	0.00456	5.22E-08	35.11889
A_5	16.03029	9.001416	410.018	20.2489	12.32207	0.25226	0.02306	5.89E-08	66.28174
A_6	18.60244	13.40853	545.9089	23.36469	26.26925	0.291076	0.00720	5.08E-08	45.06739
A_7	17.23615	11.81444	477.342	21.84816	17.12625	0.272183	0.02860	5.48E-08	84.65201
A_8	16.77933	11.08559	452.2751	21.26676	16.09855	0.26494	0.02785	5.63E-08	80.76323
A_9	16.85113	10.9036	452.5762	21.27384	15.50967	0.265029	0.01091	5.60E-08	50.06903
A_{10}	15.2317	7.726542	358.7189	18.93988	11.32238	0.235952	0.03820	6.20E-08	73.95369
A_{11}	18.92505	13.36973	565.7072	23.7846	23.78257	0.296307	0.01693	4.99E-08	69.64417
A_{12}	18.01294	12.06291	513.4392	22.6592	19.82032	0.282287	0.01180	5.24E-08	55.4037
A_{13}	18.26766	12.08031	523.8025	22.88673	22.2922	0.285122	0.02551	5.17E-08	81.75175
A_{14}	17.90409	11.81737	501.1061	22.3854	19.19237	0.278876	0.00903	5.27E-08	47.89071
A_{15}	13.4922	0.1322	280.3355	16.74322	7.838864	0.208586	0.27007	7.00E-08	27.76732
A_{16}	19.11713	13.52201	576.8698	24.01811	11.3994	0.299217	0.02042	4.94E-08	76.51818
A_{17}	18.84437	13.4473	557.8767	23.61941	26.11079	0.29425	0.02118	5.01E-08	76.96097

As previously mentioned, different settings were used for the criteria weighting. Table 7 lists the set of criteria weights that were utilized. Although the weight settings were created depending on the criteria group, not an individual criterion, it was clear that each criterion was given different weights in different settings. Accordingly, the goal of evaluating and benchmarking the alternatives using various weights was achieved. The criteria's weights under different settings are illustrated in Figure 6.

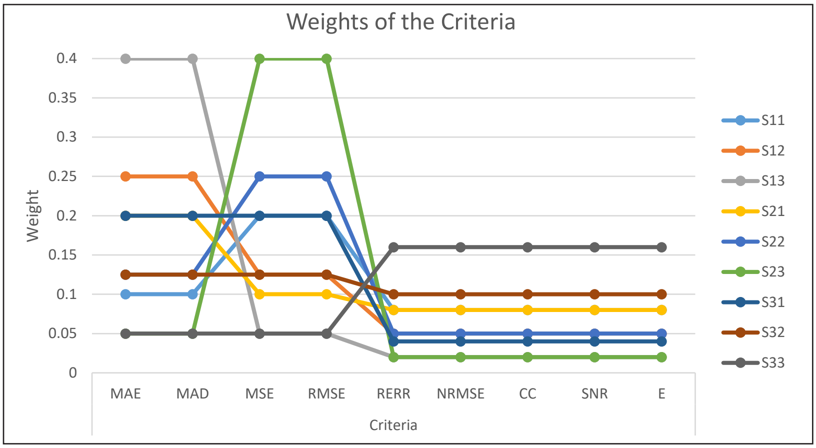
Table 7

Criteria Weighting Values

#	Criteria									
	MAE	MAD	MSE	RMSE	RERR	NRMSE	CC	SNR	E	Sum
S_{11}	0.1	0.1	0.2	0.2	0.08	0.08	0.08	0.08	0.08	1.0
S_{12}	0.25	0.25	0.125	0.125	0.05	0.05	0.05	0.05	0.05	1.0
S_{13}	0.4	0.4	0.05	0.05	0.02	0.02	0.02	0.02	0.02	1.0
S_{21}	0.2	0.2	0.1	0.1	0.08	0.08	0.08	0.08	0.08	1.0
S_{22}	0.125	0.125	0.25	0.25	0.05	0.05	0.05	0.05	0.05	1.0
S_{23}	0.05	0.05	0.4	0.4	0.02	0.02	0.02	0.02	0.02	1.0
S_{31}	0.2	0.2	0.2	0.2	0.04	0.04	0.04	0.04	0.04	1.0
S_{32}	0.125	0.125	0.125	0.125	0.1	0.1	0.1	0.1	0.1	1.0
S_{33}	0.05	0.05	0.05	0.05	0.16	0.16	0.16	0.16	0.16	1.0

Figure 6

Distribution of the Criteria Weights



The first three settings, S_{11}, S_{12}, and S_{13}, corresponded to the first structure. The prediction bias, MAE and MAD, were combined in one branch of the tree and the rest of the criteria in the other. The second set of settings, which included S_{21}, S_{22}, and S_{23}, tallied with the second structure, in which the model variances MSE and RMSE were combined into the same branch of the tree. The last set, which comprised S_{31}, S_{32}, and S_{33}, and the rest of the criteria in the relative model variance group, were combined into one branch.

Results of Alternative Ranking

Tables 8, 9, and 10 show the ranking values for the alternatives obtained using the developed TOPSIS approach in various settings for the three datasets. Each alternative's ranking and indexing values were calculated using all settings. Note that all the utilized criteria were cost-based (i.e., as opposed to benefit-based). Therefore, the lower the value, the better the result. The ranking scores of the alternatives in all settings for the three datasets are shown in Figures 7, 8, and 9.

Table 8

Results of the Alternative Ranking using the CCPP Dataset

	S_{11}		S_{12}		S_{13}		S_{21}		S_{22}		S_{23}		S_{31}		S_{32}		S_{33}	
A_1	0.192	10	0.095	9	0.044	4	0.155	10	0.125	9	0.074	4	0.090	9	0.217	10	0.289	10
A_2	0.188	11	0.089	12	0.030	15	0.152	11	0.117	13	0.051	16	0.083	13	0.215	11	0.289	11
A_3	0.211	8	0.106	7	0.051	3	0.172	8	0.138	7	0.084	3	0.101	7	0.239	8	0.320	8
A_4	0.172	17	0.080	17	0.023	17	0.139	17	0.105	17	0.039	17	0.074	17	0.198	17	0.266	17
A_5	0.186	13	0.089	11	0.035	9	0.150	13	0.118	11	0.062	8	0.084	11	0.211	13	0.283	13
A_6	0.251	5	0.120	5	0.036	8	0.205	4	0.155	5	0.058	9	0.110	5	0.289	4	0.397	4
A_7	0.214	7	0.102	8	0.033	11	0.174	7	0.133	8	0.054	13	0.094	8	0.246	7	0.333	7
A_8	0.192	9	0.091	10	0.031	13	0.156	9	0.120	10	0.053	14	0.085	10	0.220	9	0.297	9
A_9	0.187	12	0.089	13	0.031	14	0.152	12	0.118	12	0.055	12	0.083	12	0.214	12	0.288	12
A_{10}	0.387	2	0.193	2	0.099	2	0.295	2	0.287	2	0.276	2	0.206	2	0.412	2	0.562	2
A_{11}	0.277	3	0.134	3	0.040	6	0.227	3	0.173	3	0.065	7	0.123	3	0.320	3	0.441	3
A_{12}	0.251	4	0.120	4	0.040	5	0.204	5	0.157	4	0.069	5	0.112	4	0.288	5	0.394	5
A_{13}	0.175	16	0.084	15	0.032	12	0.142	16	0.111	15	0.058	10	0.079	15	0.199	16	0.266	16
A_{14}	0.236	6	0.113	6	0.038	7	0.191	6	0.148	6	0.067	6	0.105	6	0.270	6	0.368	6
A_{15}	0.699	1	0.849	1	0.955	1	0.750	1	0.807	1	0.925	1	0.861	1	0.659	1	0.554	1
A_{16}	0.181	14	0.087	14	0.035	10	0.147	14	0.114	14	0.056	11	0.082	14	0.207	14	0.276	14
A_{17}	0.176	15	0.084	16	0.029	16	0.142	15	0.110	16	0.052	15	0.078	16	0.201	15	0.269	15

Table 9

Results of the Alternative Ranking using the Real Estate Dataset

	S_{11}		S_{12}		S_{13}		S_{21}		S_{22}		S_{23}		S_{31}		S_{32}		S_{33}	
A_1	0.471	5	0.413	4	0.396	4	0.397	5	0.519	4	0.618	4	0.461	4	0.399	6	0.333	11
A_2	0.460	6	0.420	3	0.408	3	0.396	6	0.518	5	0.618	5	0.467	3	0.385	7	0.297	13
A_3	0.497	4	0.375	7	0.339	5	0.397	4	0.500	6	0.582	7	0.424	6	0.455	3	0.486	5
A_4	0.415	9	0.317	9	0.287	9	0.322	10	0.439	9	0.528	9	0.370	9	0.357	9	0.343	9
A_5	0.508	3	0.469	2	0.457	2	0.441	2	0.572	2	0.686	2	0.518	2	0.428	5	0.337	10
A_6	0.341	15	0.226	14	0.186	14	0.256	15	0.334	14	0.392	14	0.269	14	0.314	15	0.356	7
A_7	0.388	10	0.276	13	0.240	13	0.293	11	0.396	13	0.473	13	0.326	13	0.343	11	0.361	6
A_8	0.370	14	0.290	10	0.263	10	0.281	14	0.414	11	0.518	10	0.348	10	0.299	16	0.253	16
A_9	0.383	11	0.290	11	0.259	11	0.290	13	0.415	10	0.510	11	0.346	11	0.319	14	0.293	14
A_{10}	0.559	2	0.382	5	0.327	6	0.424	3	0.544	3	0.650	3	0.443	5	0.516	2	0.595	3
A_{11}	0.272	17	0.195	16	0.167	15	0.197	17	0.296	15	0.367	15	0.241	15	0.223	17	0.208	17
A_{12}	0.381	12	0.289	12	0.259	12	0.290	12	0.410	12	0.501	12	0.342	12	0.320	13	0.299	12
A_{13}	0.412	8	0.343	8	0.319	7	0.325	9	0.469	7	0.589	6	0.402	7	0.331	12	0.265	15
A_{14}	0.431	7	0.330	6	0.298	8	0.334	7	0.458	8	0.554	8	0.385	8	0.369	8	0.353	8
A_{15}	0.773	1	0.885	1	0.966	1	0.804	1	0.861	1	0.951	1	0.897	1	0.731	1	0.636	1
A_{16}	0.373	13	0.199	15	0.062	17	0.325	8	0.237	16	0.088	17	0.179	16	0.441	4	0.618	2
A_{17}	0.302	16	0.158	17	0.076	16	0.251	16	0.204	17	0.150	16	0.154	17	0.350	10	0.497	4

Table 10

Results of the Alternative Ranking using the Concrete Dataset

	S_{11}		S_{12}		S_{13}		S_{21}		S_{22}		S_{23}		S_{31}		S_{32}		S_{33}	
A_1	0.675	4	0.449	5	0.261	4	0.595	4	0.546	5	0.418	6	0.442	5	0.706	4	0.802	7
A_2	0.645	10	0.415	8	0.200	7	0.571	10	0.501	10	0.310	10	0.399	9	0.686	9	0.795	8
A_3	0.660	7	0.411	10	0.174	11	0.574	7	0.510	8	0.326	9	0.398	10	0.699	6	0.827	2
A_4	0.651	8	0.407	11	0.164	13	0.571	9	0.497	12	0.283	12	0.390	12	0.695	7	0.824	4
A_5	0.727	2	0.519	3	0.375	3	0.646	2	0.623	2	0.567	3	0.523	3	0.746	1	0.826	3
A_6	0.618	12	0.384	14	0.142	14	0.544	12	0.469	14	0.244	14	0.365	14	0.661	12	0.770	11
A_7	0.637	11	0.407	12	0.198	8	0.556	11	0.508	9	0.373	7	0.399	8	0.666	11	0.755	13
A_8	0.661	6	0.433	6	0.242	6	0.577	6	0.540	6	0.441	5	0.431	6	0.685	10	0.770	12
A_9	0.701	3	0.457	4	0.255	5	0.610	3	0.566	3	0.447	4	0.452	4	0.732	3	0.850	1
A_{10}	0.740	1	0.569	2	0.465	2	0.665	1	0.680	1	0.717	2	0.587	2	0.741	2	0.787	10
A_{11}	0.597	15	0.371	16	0.135	15	0.530	15	0.447	15	0.215	16	0.350	15	0.644	15	0.753	14
A_{12}	0.648	9	0.412	9	0.182	10	0.571	8	0.499	11	0.298	11	0.395	11	0.690	8	0.807	6
A_{13}	0.610	13	0.392	13	0.172	12	0.543	13	0.471	13	0.269	13	0.374	13	0.648	14	0.737	15
A_{14}	0.663	5	0.423	7	0.195	9	0.583	5	0.515	7	0.326	8	0.408	7	0.704	5	0.824	5
A_{15}	0.407	17	0.623	1	0.862	1	0.469	17	0.552	4	0.785	1	0.646	1	0.363	17	0.270	17
A_{16}	0.604	14	0.372	15	0.134	16	0.535	14	0.445	16	0.209	17	0.348	16	0.658	13	0.793	9
A_{17}	0.587	16	0.365	17	0.132	17	0.520	16	0.443	17	0.219	15	0.345	17	0.629	16	0.725	16

Figure 7

Alternative Ranking using the CCPP Dataset

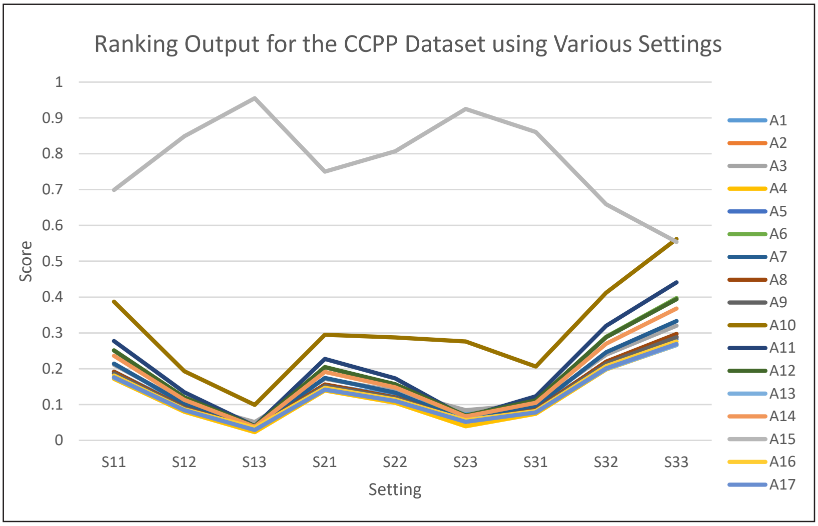


Figure 8

Alternative Ranking using the Real Estate Dataset

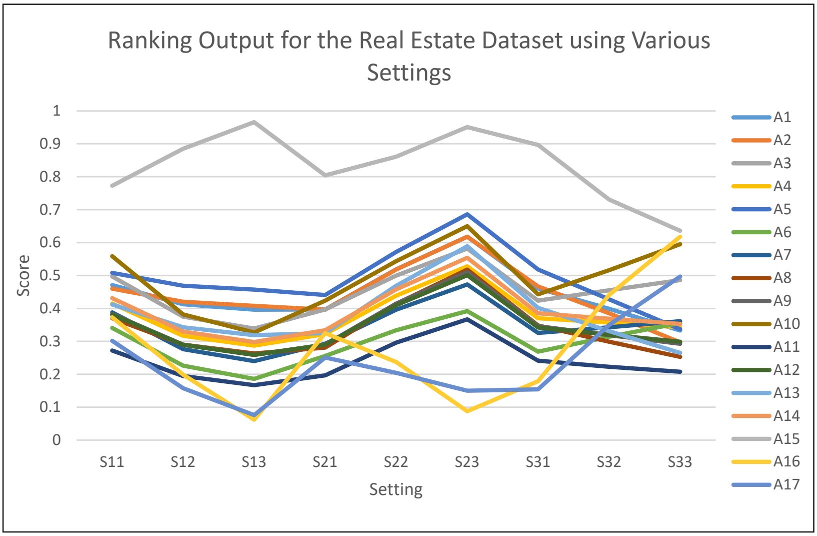
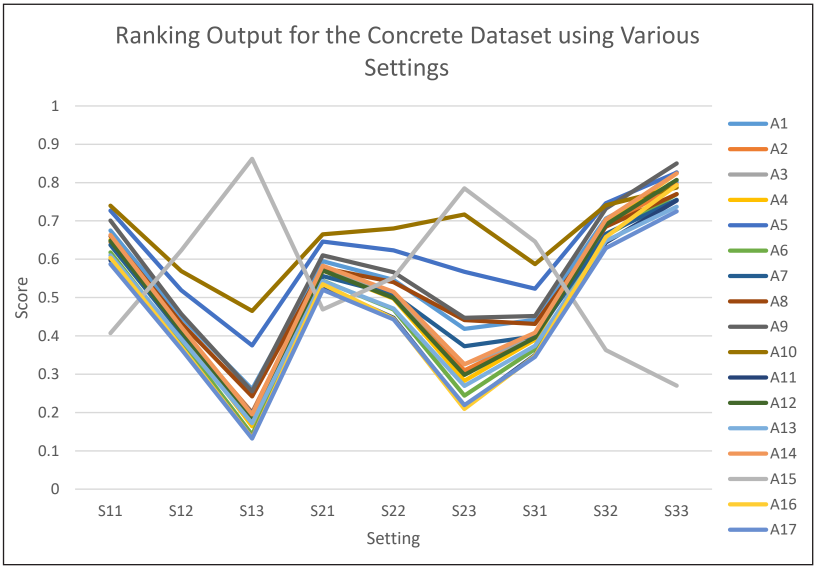


Figure 9

Alternative Ranking using the Concrete Dataset



As given in the results, different preferences led to different rankings. Despite these differences, the top-ranked algorithms could be identified in each field, more specifically, using a specific dataset. The results showed that Bagging was the best algorithm in the CCPP dataset. The worst algorithm was SVM for all settings. The top-ranked algorithm in the Real Estate dataset was also Bagging, and the worst ones varied based on the settings, with Decision Tree ranking last in some of the settings and Neural Network ranking last in others. For the Concrete dataset, the Bagging algorithm ranked first in various settings, especially when the model variance criteria (MSE, RMSE) were given the highest weights. In other settings, however, Decision Stump, KNN (with K=5), and RBF-SVM were ranked first. Surprisingly, the Bagging algorithm ranked last in some of these settings when the model variance criteria (MSE, RMSE) were given equal or lower weights than the others.

Results of Internal and External Aggregations

Group aggregations combined the results of the same alternative in different settings. The internal and external aggregation results, as shown in Figures 10, 11, and 12, showed that the rankings were similar in both cases. It should also be noted that the external aggregation always produced higher scores than the internal aggregation using the CCPP dataset. For the other datasets, however, there was a match between the internal and external aggregation scores. In the Real Estate dataset, Bagging was the best algorithm according to the group decision using the internal and external approaches. Decision Stump was the best algorithm for the Concrete dataset, followed by the RBF-SVM algorithm according to the group decision-making using the internal and external approaches.

According to the results before and after the aggregation using internal and external group decision-making, different weight settings affected the performance in some datasets but had no effect in others. Despite these differences, the performance could be noted and concluded, indicating that the proposed approach has established a benchmark for regression algorithm ranking that can be applied to any dataset and regression algorithm.

Figure 10

Group Aggregation for Alternative Ranking using the CCPP Dataset

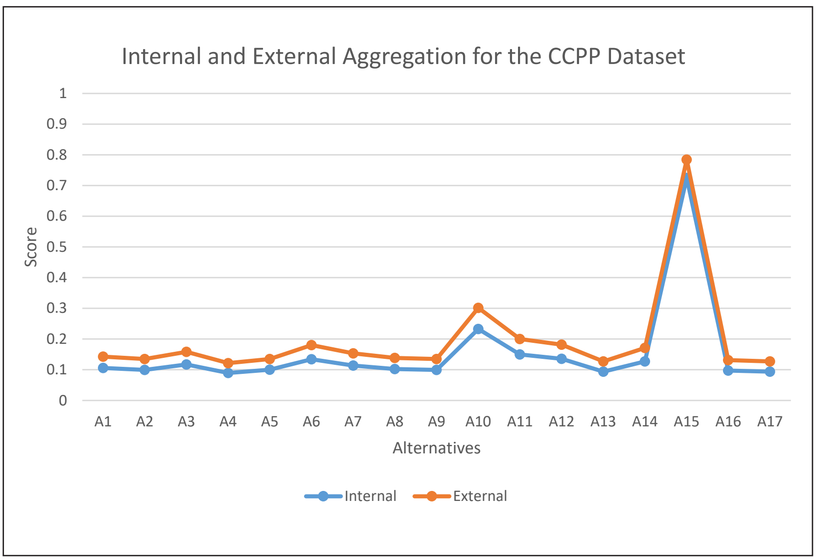


Figure 11

Group Aggregation for Alternative Ranking using the Real Estate Dataset

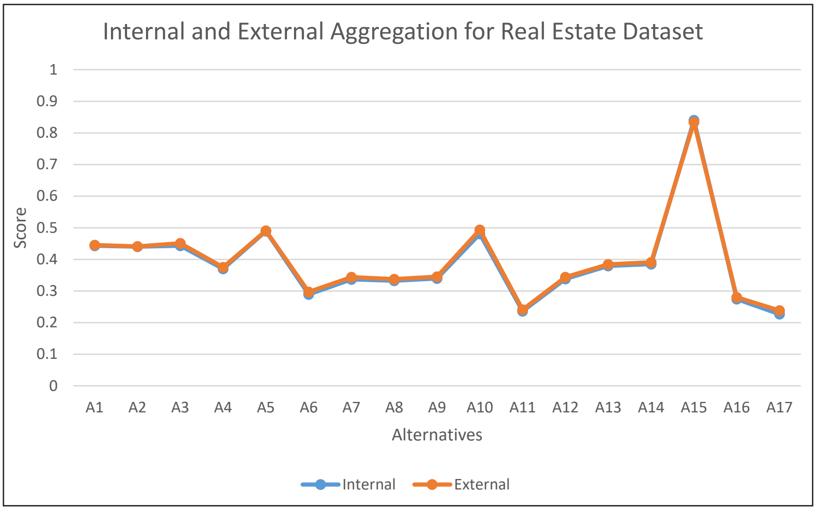
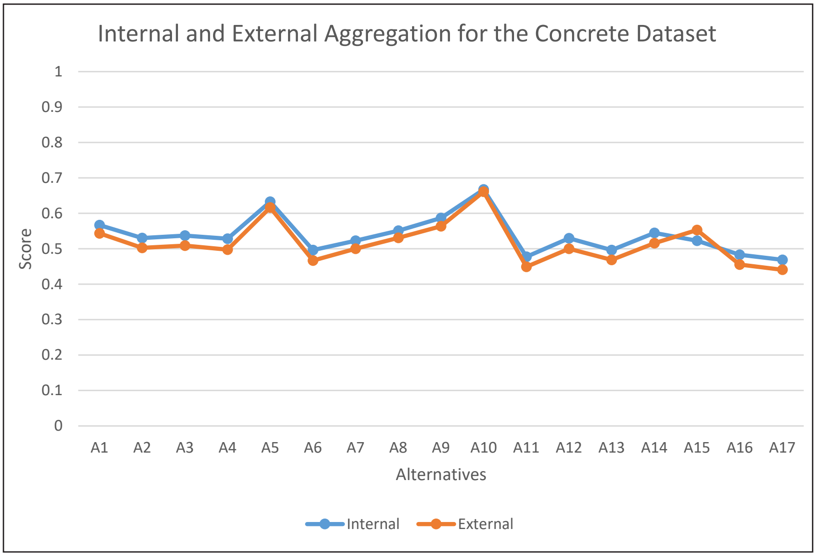


Figure 12

Group Aggregation for Alternative Ranking using the Concrete Dataset



Results of the Validation

The paired comparative correlation values for the Pearson's correlation test using the CCPP dataset are listed in Table 11. The results showed in most cases, the correlation values were close to 1, indicating a true correlation between the measured values. As a result, with a few exceptions, the obtained results can be generalized for the utilized dataset, despite the weight variations. Table 12 lists the correlation values between the comparative samples using the Real Estate dataset. Although the correlation values were generally lower than those in the first dataset, the results can be generalized for the utilized dataset, despite the significant weight variations. As given in Table 13, the results under the Concrete dataset cannot be generalized because the correlation varied. A low correlation existed between different settings obtained in this dataset, which explains why the ranking results varied in different settings. Accordingly, the generalization can be made using the aggregation approaches, in which the Decision Stump algorithm was proven to be the top-ranked algorithm.

Table 11

Results of Pair-Correlation of Different Settings using the CCPP Dataset

	S_{11}	S_{12}	S_{13}	S_{21}	S_{22}	S_{23}	S_{31}	S_{32}	S_{33}
S_{11}									
S_{12}	0.961								
S_{13}	0.933	0.996							
S_{21}	0.988	0.991	0.975						
S_{22}	0.984	0.994	0.982	0.998					
S_{23}	0.969	0.990	0.984	0.986	0.995				
S_{31}	0.964	1.000	0.995	0.992	0.996	0.993			
S_{32}	0.996	0.935	0.900	0.973	0.964	0.942	0.938		
S_{33}	0.854	0.678	0.610	0.769	0.746	0.706	0.685	0.895	

Table 12

Results of Pair-Correlation of Different Settings using the Real Estate Dataset

	S_{11}	S_{12}	S_{13}	S_{21}	S_{22}	S_{23}	S_{31}	S_{32}	S_{33}
S_{11}									
S_{12}	0.951								
S_{13}	0.918	0.993							
S_{21}	0.973	0.970	0.937						
S_{22}	0.940	0.965	0.968	0.903					
S_{23}	0.823	0.868	0.898	0.752	0.961				
S_{31}	0.940	0.990	0.993	0.934	0.990	0.929			
S_{32}	0.933	0.843	0.780	0.945	0.761	0.567	0.786		
S_{33}	0.555	0.363	0.268	0.573	0.242	0.007	0.270	0.804	

Table 13

Results of Pair-Correlation of Different Settings using the Concrete Dataset

	S_{11}	S_{12}	S_{13}	S_{21}	S_{22}	S_{23}	S_{31}	S_{32}	S_{33}
S_{11}									
S_{12}	-0.137								
S_{13}	-0.446	0.947							
S_{21}	0.932	0.227	-0.097						
S_{22}	0.450	0.818	0.595	0.734					
S_{23}	-0.071	0.985	0.917	0.279	0.856				
S_{31}	-0.130	0.999	0.945	0.231	0.825	0.991			
S_{32}	0.970	-0.368	-0.646	0.822	0.221	-0.311	-0.363		
S_{33}	0.883	-0.568	-0.799	0.665	-0.013	-0.522	-0.566	0.970	

CONCLUSION

A TOPSIS-based approach for regression alternative ranking with multiple criteria weighting settings was proposed in this paper. Accordingly, a flexible criteria structure was utilized with multiple criteria weights to enable regression evaluation from multiple preferences, to mimic the varied preferences found in different science and engineering fields that employ the regression algorithms for different applications. According to the results before and after the aggregation using internal and external group decision-making, different weight settings affected the performance in some datasets but had no effect in other datasets. Despite these differences, the performance can be noted and concluded, indicating that the proposed approach has established a benchmark for regression algorithm ranking that can be applied to any dataset and regression algorithm. The concluded results were validated using Pearson's correlation tests on pair-series (pair settings).

ACKNOWLEDGMENT

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

Adepoju, S. A., Oyefolahan, I. O., Abdullahi, M. B., & Mohammed, A. A. (2020). Multi-criteria decision-making based approaches in website quality and usability evaluation: A systematic review. *Journal of Information and Communication Technology*, *19*(3), 399–436. https://doi.org/10.32890/jict2020.19.3.5

Bates, D. M., & Watts, D. G. (1988). *Nonlinear regression analysis and its applications* (Vol. 2). Wiley New York. https://doi.org/10.1002/9780470316757

Baumann, P., Hochbaum, D. S., & Yang, Y. T. (2019). A comparative study of the leading machine learning techniques and two new optimization algorithms. *European Journal of Operational Research*, *272*(3), 1041–1057. https://doi.org/10.1016/j.ejor.2018.07.009

Belton, V., & Vickers, S. P. (1993). Demystifying DEA-A visual interactive approach based on multiple criteria analysis. *Journal of the Operational Research Society*, *44*(9), 883–896. https://doi.org/10.1057/jors.1993.157

Chodha, V., Dubey, R., Kumar, R., Singh, S., & Kaur, S. (2022). Selection of industrial arc welding robot with TOPSIS and Entropy MCDM techniques. *Materials Today: Proceedings*, *50*, 709–715. https://doi.org/10.1016/j.matpr.2021.04.487

De Keyser, W., & Peeters, P. (1996). A note on the use of PROMETHEE multicriteria methods. *European Journal of Operational Research*, *89*(3), 457–461. https://doi.org/10.1016/0377-2217(94)00307-6

Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (3rd ed., Vol. 326). John Wiley & Sons. https://doi.org/10.1002/9781118625590

Edwards, W. (1971). Social utilities. *Engineering Economist*, *6*(1), 119–129.

Fox, J., & Weisberg, S. (2018). *An R companion to applied regression* (3rd ed.). Sage Publications.

Gao, Z., & Xuan, T. (2019). VIKOR method for ranking concrete bridge repair projects with target-based criteria. *Results in Engineering*, *3*, 1–9. https://doi.org/10.1016/j.rineng.2019.100018

Govindan, K., & Jepsen, M. B. (2016). ELECTRE: A comprehensive literature review on methodologies and applications. *European Journal of Operational Research*, *250*(1), 1–29. https://doi.org/10.1016/j.ejor.2015.07.019

Hwang, C.-L., & Yoon, K. (2012). *Multiple attribute decision making: Methods and applications a state-of-the-art survey* (Vol. 186). Springer Science & Business Media. https://doi.org/10.1007/978-3-642-48318-9

Julong, D. (1989). Introduction to grey system theory. *The Journal of Grey System*, *1*(1), 1–24.

Kasim, M. M., Ibrahim, H., & Bataineh, M. S. (2011). Multi-criteria decision making methods for determining computer preference index. *Journal of Information and Communication Technology*, *10*, 137–148. https://doi.org/10.32890/jict.10.2011.8115

Kou, G., Lu, Y., Peng, Y., & Shi, Y. (2012). Evaluation of classification algorithms using MCDM and rank correlation. *International Journal of Information Technology & Decision Making*, *11*(01), 197–225. https://doi.org/10.1142/S0219622012500095

Kou, G., Peng, Y., & Lu, C. (2014). MCDM approach to evaluating bank loan default models. *Technological and Economic Development of Economy*, *20*(2), 292–311. https://doi.org/10.3846/20294913.2014.913275

Kou, G., Yang, P., Peng, Y., Xiao, F., Chen, Y., & Alsaadi, F. E. (2020). Evaluation of feature selection methods for text classification with small datasets using multiple criteria decision-making methods. *Applied Soft Computing*, *86*(1), 1–14. https://doi.org/10.1016/j.asoc.2019.105836

Lavesson, N., & Davidsson, P. (2007). Evaluating learning algorithms and classifiers. *International Journal of Intelligent Information and Database Systems*, *1*(1), 37–52. https://doi.org/10.1504/IJIIDS.2007.013284

Li, H., & Sun, J. (2008). Ranking-order case-based reasoning for financial distress prediction. *Knowledge-based systems*, *21*(8), 868–878. https://doi.org/10.1016/j.knosys.2008.03.047

MacCrimmon, K. R. (1968). *Decisionmaking among multiple-attribute alternatives: A survey and consolidated approach*. R. C. S. M. Ca.

Malczewski, J. (1999). *GIS and multicriteria decision analysis*. John Wiley & Sons.

Nakhaeizadeh, G., & Schnabl, A. (1997, August). Development of multi-criteria metrics for evaluation of data mining algorithms. In *International Conference on Knowledge Discovery and Data Mining (KDD), Newport Beach CA* (pp. 37–42).

Oliveira, M., Fontes, D. B., & Pereira, T. (2014). Multicriteria decision making: A case study in the automobile industry. *Annals of Management Science*, *3*(1), 109. https://doi.org/10.24048/ams3.no1.2014-109

Osei-Bryson, K.-M. (2004). Evaluation of decision trees: A multi-criteria approach. *Computers & Operations Research*, *31*(11), 1933–1945. https://doi.org/10.1016/S0305-0548(03)00156-4

Pandey, G., & Nguyen, V.-T.-V. (1999). A comparative study of regression based methods in regional flood frequency analysis. *Journal of Hydrology*, *225*(1), 92–101. https://doi.org/10.1016/S0022-1694(99)00135-3

Peng, Y., Kou, G., Wang, G., & Shi, Y. (2011). FAMCDM: A fusion approach of MCDM methods to rank multiclass classification algorithms. *Omega*, *39*(6), 677–689. https://doi.org/10.1016/j.omega.2011.01.009

Peng, Y., Wang, G., & Wang, H. (2012). User preferences based software defect detection algorithms selection using MCDM. *Information Sciences*, *191*(1), 3–13. https://doi.org/10.1016/j.ins.2010.04.019

Peteiro-Barral, D., Remeseiro, B., Méndez, R., & Penedo, M. G. (2017). Evaluation of an automatic dry eye test using MCDM methods and rank correlation. *Medical & Biological Engineering & Computing*, *55*(4), 527–536. https://doi.org/10.1007/s11517-016-1534-5

Petrovic-Lazarevic, S., & Abraham, A. (2004). Hybrid fuzzy-linear programming approach for multi criteria decision making problems. *International Journal of Neural, Parallel & Scientific Computations*, *11*(1&2), 53–68.

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data*, *1*(1), 51–59. https://doi.org/10.1089/big.2013.1508

Roy, B. (1968). Classement et choix en presence de points de vue multiples. *Revue Française D'informatique et de Recherche Opérationnelle*, *2*(8), 57–75. https://doi.org/10.1051/ro/196802V100571

Saaty, T. L. (2005). *Theory and applications of the analytic network process: Decision making with benefits, opportunities, costs, and risks*. RWS publications. https://doi.org/10.1007/978-1-4614-7279-7

Shih, H.-S., Shyur, H.-J., & Lee, E. S. (2007). An extension of TOPSIS for group decision making. *Mathematical and Computer Modelling*, *45*(7–8), 801–813. https://doi.org/10.1016/j.mcm.2006.03.023

Siregar, V. M. M., Sonang, S., Purba, A. T., Sugara, H., & Siagian, N. F. (2021 February). Implementation of TOPSIS algorithm for selection of prominent student class. In *Annual Conference on Science and Technology Research (ACOSTER) 2020, Medan, North Sumatera, Indonesia*. *Journal of Physics: Conference Series* (vol. 1783, pp. 1–6).

Skilodimou, H. D., Bathrellos, G. D., Chousianitis, K., Youssef, A. M., & Pradhan, B. (2019). Multi-hazard assessment modeling via multi-criteria analysis and GIS: A case study. *Environmental Earth Sciences*, *78*(2), 47. https://doi.org/10.1007/s12665-018-8003-4

Song, Y., & Peng, Y. (2019). A MCDM-based evaluation approach for imbalanced classification methods in financial risk prediction. *IEEE Access*, *7*, 84897–84906. https://doi.org/10.1109/ACCESS.2019.2924923

Srisawat, C., & Payakpate, J. (2016). Comparison of MCDM methods for intercrop selection in rubber plantations. *Journal of Information and Communication Technology*, *15*(1), 165–182. https://doi.org/10.32890/jict2016.15.1.8

Tufekci, P. (2014). Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning methods. *International Journal of Electrical Power & Energy Systems*, *60*, 126–140. https://doi.org/10.1016/j.ijepes.2014.02.027

Wind, Y., & Saaty, T. L. (1980). Marketing applications of the analytic hierarchy process. *Management Science*, *26*(7), 641–658.

Winterfeldt, D. V., & Fischer, G. W. (1975). Multi-attribute utility theory: Models and assessment procedures. *Utility, Probability, and Human Decision Making*, *11*(1), 47–85. https://doi.org/10.1007/978-94-010-1834-0_3

Yap, J. Y. L., Ho, C. C., & Ting, C.-Y. (2019). A systematic review of the applications of multi-criteria decision-making methods in site selection problems. *Built Environment Project and Asset Management*, *9*(4), 548–563. https://doi.org/10.1108/BEPAM-05-2018-0078

Yeh, I.-C. (1998). Modeling of strength of high-performance concrete using artificial neural networks. *Cement and Concrete Research*, *28*(12), 1797–1808. https://doi.org/10.1016/S0008-8846(98)00165-3

Yeh, I.-C., & Hsu, T.-K. (2018). Building real estate valuation models with comparative approach through case-based reasoning. *Applied Soft Computing*, *65*, 260–271. https://doi.org/10.1016/j.asoc.2018.01.029

Zorlu, K. (2012). A comparative study of using the methods of multiple linear regression and artificial neural networks in organizational correlations for the fields of management and organization. *Uluslararası Yönetim İktisat ve İşletme Dergisi*, *8*(17), 1–25. https://doi.org/10.11122/IJMEB.2014.8.17.343