Classifying Scuba Diving Sites through Diver Reviews with a Web Scraping Based UTADIS Application

Yunus Yildirim[1], **Aydin Ulucan**[2] and **Kazim Baris Atici**[2,*]

[1] *Department of Business Administration, Afyon Kocatepe University, Afyonkarahisar, Turkey*
E-mail: ⟨*yunusyildirim@aku.edu.tr*⟩

[2] *Department of Business Administration, Hacettepe University, Ankara, Turkey*
E-mail: ⟨{*aulucan, kba*}*@hacettepe.edu.tr*⟩

Abstract. This research addresses the utilization of user-defined and web-based data for decision-making in recreational activities, focusing on the case of scuba diving, a globally significant recreational pursuit. While a wealth of user-generated information is available on the internet for various leisure activities, harnessing this data involves extensive data collection and organization efforts. Our proposed methodology involves the collection and organization of user reviews, both in verbal and quantitative forms, to create a pertinent dataset for applying multiple criteria decision methodologies to classify diving sites worldwide. An initial dataset containing over 14,000 diving sites worldwide is aggregated into 721 regions and these regions are classified using UTADIS methodology. The research showcases how user-generated review data can be transformed into valuable information, applying classification algorithms of multiple criteria decision analysis within the context of scuba diving. Furthermore, the proposed approach in this research holds the potential to serve as a model for leveraging user-generated data in decision-making processes across various service sectors such as hospitality and leisure that highly rely on customer experience by providing new insights on how more data-driven approaches can be utilized.

Keywords: leisure, multiple criteria decision analysis (MCDA), ordinal classification, scuba diving; UTADIS

Received: October 13, 2023; accepted: November 10, 2023; available online: December 19, 2023

DOI: 10.17535/crorr.2023.0012

---

1. Introduction

In today's digital era, user reviews play a pivotal role in shaping business decisions. Consumers worldwide generate vast amounts of online review data daily, making it a common practice to seek others' experiences when making purchasing decisions. This wealth of user-generated information also provides businesses with invaluable insights that were not available at this scale decades ago. Consequently, data science practices have become increasingly essential for organizing, classifying, and analyzing this data for continuous improvement. One domain profoundly influenced by user reviews is travel decisions. The internet is full of travel websites, blogs, and social media platforms where information is shared, leisure businesses market their offerings, and customers review and share their experiences. This rapid information exchange allows individuals to access detailed information about their travel destinations and specialized recreational activities for specific communities. Scuba diving is a prominent example.

Scuba diving has evolved into a billion-dollar global industry since the 1950s [7], with over 30 million licensed divers worldwide and approximately 6 million unlicensed individuals who

---

*Corresponding author.

have experienced scuba diving [16]. This sector significantly contributes to the tourism industry, creating numerous employment opportunities that keep an average of 30 million people employed every year [11]. Governments and organizations have taken notice of the significance of scuba diving, leading to initiatives like the 'Green Bubbles' project by the European Union, aimed at ensuring environmental, economic, and social sustainability [17].

With more than 14,000 scuba diving locations worldwide, site selection for diving practices presents an intriguing research challenge. This involves considering factors like diving type availability, depth, visibility, marine life, and expected experience levels. Stakeholders, including divers, businesses in scuba diving tourism, and tourism officials, would find value in comprehensive site rankings or classifications. The current paper addresses the classification of the world's scuba diving sites representing an order between both alternatives and classes in terms of preferability, using the UTADIS method from the field of Operational Research.

UTADIS (UTilities Additives DIScriminantes) method was initially introduced by Devaud et al. in 1980 [4], with subsequent refinements in its properties by Jacquet-Lagrèze & Siskos in 1982 [12]. Notably, Jacquet-Lagrèze [13] applied UTADIS to evaluate R&D projects, marking a significant early application. The late 1990s saw widespread use in financial decision problems, especially in research conducted by Zopounidis & Doumpos [29, 30, 31, 32]. UTADIS also found adaptation into multi-criteria decision (MCD) support systems like FINCLAS and PREFDIS [29, 32]. The following years have witnessed the method's application across diverse sectors.

The predominant area of the applications is finance. The method has been employed in the assessment of financial stability [5], risk at the institutional level [2], risk at the country level [26], failure [**?**], and fraud [22] as well as in portfolio management [6]. The method has also been applied to support strategic decisions in businesses in various sectors such as oil [21] and pharmaceutics [14]. UTADIS methodology is used for alternative selection such as renewable energy project selection [25] and supplier selection in the automotive sector [18]. With its variety of applications, well-established modeling, and unique feature of expecting minimal information from the decision-maker compared to other MCDA methodologies, the method has the potential for handling different problems from different areas. For instance, a different approach by Parimbelli et al. [19] shows the applicability of UTADIS in the healthcare sector by assessing a post-surgical complication risk based on the preferences of a spinal surgeon. Risk classification continues to be addressed in different areas such as food security with price volatility [15]. Recently, the method has become a subject of hybrid modelling with machine learning (ML) and clustering methods to cover both nominal and ordinal group classification representing similarities and rankings between alternatives [1, 10, 25].

There are various methods in the literature designed for classification tasks, which are done in two ways nominal and ordinal depending on the distinction level of the problem. The methods of ML are suited for nominal classification problems dealing with the categorization of data into distinct but non-ordered groups. On the other hand, MCDA encompasses various methods such as Analytic Hierarchy Process (AHP) Sort, ELECTRE-TRI and N–TOMIC, which are good for the ordinal classification tasks representing an inherent order and/or ranking between classes [10]. UTADIS is another MCDA method developed for MCD problems with ordinal group classification [9]. The method distinguishes itself from other MCDA methods by estimating the criteria weights, the marginal utility functions and the utility thresholds between groups based on rankings or predefined classifications done by experts or exogenous criteria. The UTADIS model aims to minimize the deviations from predefined classification in this estimation [31].

The study aims to collect and organize extensive user review data on scuba diving sites and subsequently apply the UTADIS classification methodology. The paper's contribution lies in that potential by applying the method in the scuba diving sector, demonstrating its versatility across research domains, and providing valuable insights for scuba diving tourism by classifying sites based on multiple criteria. Accordingly, the current attempt addresses an unexplored area in the literature, underlining its originality, and has the potential to influence data-driven research in similar areas. Additionally, it is important to note that the methodology of this research has the potential for broader applicability beyond the scuba diving context. The utilization of user-generated data and expert rankings with the MCD methods can be applied to other service industries that highly rely on customer experience such as hospitality and other recreational activities, allowing for the customization of travel itineraries and destination recommendations based on user feedback and expert evaluations.

The paper is structured as follows: Chapter 2 provides an examination of the UTADIS method's core principles and model. Chapter 3 discusses our model design and dataset. Chapter 4 presents the analysis findings, and Chapter 5 concludes the paper.

2. An Overview of the UTADIS Method

In a UTADIS application, the first step involves a predefined classification of alternatives, typically derived from expert opinions or exogeneous criteria. UTADIS aims to classify alternatives based on this predefined classification with minimal error. However, it can also identify misclassifications using additional criteria, making the initial classification less stringent. The method relies on a linear programming model.

Suppose a set of alternatives consisting of n alternatives ($A = \{a_1, a_2, \ldots, a_n\}$) and a set of criteria consisting of m criteria ($G = \{g_1, g_2, \ldots, g_m\}$). The predefined classes of alternatives are represented with $C_1 \succ C_2 \succ C_3 \ldots, C_{Q-1} \succ C_Q$, where $\succ$ presents dominance between the classes. In other words, C_1 is the group with the best alternatives and C_Q is the group with the worst alternatives compared to other groups. UTADIS aims to assign a utility score for each alternative between 0 and 1. To achieve that, normalization in terms of each criterion is required. For this purpose, the minimum and maximum values of each g_i criterion are obtained as $G_i = [g_{i*}, g_i^*]$ $(i = 1, 2, \ldots, m)$. Then every G_i range is divided into a_{i-1} number of sub-intervals at most. The number of sub-intervals is at the discretion of the decision-maker. The higher the number of sub-intervals, the more precise the model will be coming with an extra computational cost. Besides, determining the number of sub-intervals depends on how sensitive and reliable the data are. The next step is to determine the dividing points of sub-intervals which are calculated with the interpolation (j represents the index of sub-interval).

$$g_i^j = g_{i*} + \frac{j-1}{a_{i-1}}(g_i^* - g_{i*}) \tag{1}$$

Consider the value of alternative a for the criterion g_i is $g_i(a)$ and $g_i(a) \in [g_i^j, g_i^{j+1}]$. The marginal utility of $g_i(a)$ is shown as $u_i[g_i(a)]$ and calculated as below using linear interpolation:

$$u_i[g_i(a)] = u_i(g_i^j) + \frac{g_i(a) - g_i^j}{g_i^{j+1} - g_i^j}[u_i(g_i^{j+1}) - u_i(g_i^j)] \tag{2}$$

The value of $[u_i(g_i^{j+1}) - u_i(g_i^j)]$ cannot be negative because the utility score provided by a sub-interval cannot be less than the one at the preceding sub-interval. In other words, utility scores provided by the sub-intervals progress cumulatively. To fulfil this requirement, the function in (2) is transformed. The expression $[u_i(g_i^{j+1}) - u_i(g_i^j)]$ is replaced with a variable w_{ij} and the expression $u_i(g_i^j)$ is defined as the sum of the utility scores of the sub-intervals up to the current sub-interval. The marginal utility function becomes as below.

$$u_i[g_i(a)] = \sum_{k=1}^{j-1} w_{ik} + \frac{g_i(a) - g_i^j}{g_i^{j+1} - g_i^j} w_{ij} \tag{3}$$

With the above equation, the utility score of the alternative a from the criterion g_i is calculated and the global utility score for alternative a is the sum of the utility scores from all criteria and is shown as below.

$$U(a) = \sum_{i=1}^{m} u_i[g_i(a)] \tag{4}$$

The weights of all criteria in each sub-interval should be between 0 and 1, and the constraint (5) ensures this condition:

$$\sum_{i=1}^{m} \sum_{j=1}^{a_{i-1}} w_{ij} = 1 \tag{5}$$

A central concept in UTADIS methodology is the utility threshold. These thresholds set boundaries for predefined classes by separating the global utility scores of alternatives. The model's objective is to assign global utility scores to alternatives equal to or greater than their respective utility thresholds while optimizing the thresholds between classes. A global utility score below its predefined class's threshold indicates a prediction error. UTADIS aims to minimize two types of classification errors: over-prediction error σ^+ and under-prediction error σ^-. If the value σ^+ is greater than 0 for an alternative, the alternative is estimated in a higher (worse) class than the group it is pre-assigned. If the value σ^- is greater than 0, it indicates vice versa. The conditions are presented as below (u_i values represent utility thresholds and $U(a)$ represents the global utility score of alternative a).

$$U(a) \geq u_1 \implies a \in C_1 \tag{6}$$

$$u_2 \leq U(a) < u_1 \implies a \in C_2 \tag{7}$$

$$u_k \leq U(a) < u_{k-1} \implies a \in C_k \tag{8}$$

$$U(a) < u_{Q-1} \implies a \in C_Q \tag{9}$$

σ^+ and σ^- are included in the model as decision variables and the above conditions are satisfied with the below constraints.

$$\sum_{i=1}^{m} u_i[g_i(a)] - u_1 + \sigma^+(a) \geq \delta \quad \forall a \in C_1 \tag{10}$$

$$\sum_{i=1}^{m} u_i[g_i(a)] - u_k + \sigma^+(a) \geq \delta \quad \forall a \in C_k \tag{11}$$

$$\sum_{i=1}^{m} u_i[g_i(a)] - u_{k-1} - \sigma^-(a) \leq -\delta \quad \forall a \in C_k \tag{12}$$

$$\sum_{i=1}^{m} u_i[g_i(a)] - u_{Q-1} - \sigma^-(a) \leq -\delta \quad \forall a \in C_Q \tag{13}$$

The parameter δ is a positive real number very close to 0 and it is used to ensure that the value of $U(a)$ is greater than the utility threshold of the group that alternative a belongs to. The linear programming model that minimizes the prediction error values subject to the constraints mentioned is given below.

$$\text{Min} \quad Z = \frac{\sum_{a \in C_1} \sigma^+(a)}{n_1} + \cdots + \frac{\sum_{a \in C_k} [\sigma^+(a) + \sigma^-(a)]}{n_k} + \cdots + \frac{\sum_{a \in C_Q} \sigma^-(a)}{n_Q} \tag{14}$$

$$\text{s.t.} \quad (10)\text{-}(13) \tag{15}$$

$$\sum_{i=1}^{m} \sum_{j=1}^{a_i - 1} w_{ij} = 1 \tag{16}$$

$$u_{k-1} - u_k \geq s \quad k = 2, 3, \ldots, Q-1 \tag{17}$$

$$w_{ij} \geq 0, \sigma^+ \geq 0, \sigma^- \geq 0 \tag{18}$$

The objective function is the summation of the averages of each group estimation error. n_i values in the objective function represent the number of alternatives in the corresponding group. The parameter s is the difference between the utility thresholds. This parameter is determined by the decision-maker. As the value increases, the criteria weights tend to be distributed more equally, but at the same time, the model's estimation power weakens.

3. Data Set & Methodology

The data is obtained from Scuba diving-related online web-based databases; Divetime [8] and Wannadive [27], which are online platforms that aim to provide divers with detailed information on all of the best scuba diving locations and destinations in the world. The platform presents user reviews as well as technical data about all diving sites all over the world. Due to the nature of the data in the given websites, the collection and organization of the data to be employed in UTADIS require special handling since the data mostly consists of verbal data distributed to thousands of pages of the website. The research design consists of five stages. In **Stage 1**, raw data is collected through web scraping, including user ratings and reviews, followed by the storage of both qualitative and quantitative data with the elimination of missing values. **Stage 2** involves converting verbal data into a quantitative format. In **Stage 3**, data is aggregated by grouping dive sites into regions, incorporating descriptive statistics and a new criterion—the number of dive sites in each region. **Stage 4** includes a pre-classification process established from a 'Best Dive Sites' ranking based on expert opinions. Finally, in **Stage 5**, UTADIS modelling is applied.

3.1. Data Collection

The experiences of scuba divers are reported on various websites. Among these websites, the abovementioned two are selected for comprehensiveness. The administrators and users of this site have provided both numerical and verbal information about the places where they have dived before. The data come in three forms: numerical values for the criteria that are measurable (depth, visibility, etc.), user ratings in discrete form, and or verbal comments for the non-measurable criteria. A total of 14,756 diving site data is available on this platform and they are distributed to different parts of the world (North America, Oceania and the Pacific, Europe, Asia, Middle East, Africa, Caribbean, South and Central America in descending order by number of dive sites).

As a first step of the research, Web Scraping is required to extract the data that come in the above-mentioned three different forms for all sites. For this purpose, *Python*'s web scraping features are utilized by using requests and *BeautifulSoup* [3]. packages and the extracted data are stored in Microsoft Excel datasheets. When missing data have been eliminated, it is observed that 12,613 diving sites exist with available data.

3.2. Data Organization

Once the data have been collected, an organization is required to prepare the data for multiple criteria analysis. This phase consists of two stages as the conversion of verbal data to numerical data and aggregation that is required to reduce to data to be handled by an MCDA methodology. Below, these two stages are explained:

Verbal Data Conversion Stage: Since many verbal expressions need to be searched and counted, the simple counting features of Python are used at this stage because of its ability to process big data with its simple coding structure [20]. There are two main criteria to be handled for 12,613 diving sites:

- *The "Current" Criterion:* This is related to the flow of water in a given diving site. It is expressed with standardized words by the divers (very weak, weak, strong, etc.). For this criterion, the rating is made on a scale of 0-5.

- *The "Marine Life" Criterion:* Under this title, there exist some standardized expressions (various, very little, wonderful, etc.) that can be treated as in the Current criterion. However, there also exist non-standardized expressions which provide a list of plants and animals that can be seen around the diving sites. To fully capture the information, the ratings should be assigned in a manner that covers both standardized and non-standardized expressions. To determine which marine plants and animals are significant compared to others, a conventional approach based on the sign language developed by professional scuba divers is followed. The purpose of this sign language is to alert other divers when an interesting creature is detected on the spot [?]. The animals and plants that are considered interesting in this language are *Shrimp, Fire Coral, Nudibranch, Unicorn Seahorse, Garden Eel, Octopus, Pelagic Fish, Grouper, Pufferfish, Angelfish, Jellyfish, Seahorse, Boxfish, Lionfish, Shark, Clown 'Anemone' Fish, Lobster, Stingray, Crab, Ghost Stingray, Turtle, Barracuda Fish, Moray Eels, Whale, Dolphin, Napoleon Fish.* After rating the standard expressions between 0-5, a dive site is rated again between 0-5 for the interesting marine life it has with a total of 10 ratings at maximum. This is achieved by counting the number of animals and plants in the list and normalizing it to a 0-5 scale.

Aggregation Stage: After the conversion, it is observed that there are still missing data (either for a set or the whole criteria). Besides, the data consists of 12,613 alternatives which are too big to be analyzed with any MCDA method. To reduce the number of alternatives and to minimize the negative effects of missing data, an aggregation approach to group the dive sites in their respective regions is utilized. By this approach, the alternatives become diving regions instead of diving sites and the missing data of each site is absorbed within their dive region. In other words, big data has become metadata that contains original information while pursuing an acceptable size for the analysis. As a result, 721 diving region alternatives were obtained by aggregating the data of 12,613 diving sites.

At the aggregation stage, the number of diving sites in each region was counted with *Python*'s ability to read, edit and save data in Excel with the openpyxl package [24]. Similarly, in each dive region, the number of sites suitable for each type of diving and each experience level is counted. This information has also become new criteria for the new alternative set.

The extracted data also include the *Type of Diving* (Reef, Rock, Wreck, Wall, Deep, Drift, Cave, Shark, Night) in binary form and *Experience* (Novice, Intermediate, Advanced) criteria. With the aggregation, these data have also become functional. In terms of Type of Diving, each type is counted in each region (e.g. number of sites available in each region for reef diving). For Experience, each type of experience is counted in each region (e.g. the number of sites in a region suitable for novices). For the rest of the criteria (depth, visibility, current, and marine life), average values were taken for each region to represent all the dive sites in it. Also, the minimum and maximum values that criteria have in each region are added as new criteria (e.g. max. depth, min. depth, etc.). As a result, the data is organized into 721 diving regions with 25 criteria in 7 groups. When the 25 criteria used in the analysis are examined (see Table 3 under Section 3.4), it can be seen that predominantly the count data reflecting variety in suitability for different types of diving and experience are used as criteria. In terms of depth, visibility, current and marine life, minimum average and maximum values are employed as new criteria, which represent the extremes as well as standards across dive sites in regions.

3.3. UTADIS Application Design

To apply UTADIS to the organized data, a predefined classification of the alternatives is required. If it does not readily exist, the classification can be based on expert opinion or can be established by practising a certain rule.

Pre-Classification stage: About scuba diving, it is possible to find different rankings on the web. However, those rankings are mostly for the sites instead of regions. To obtain a predefined classification for our scuba diving regions, The World's Best 100 Diving Spots ranking of Scuba Travel [23] website is utilized, which is the most prominent ranking that comes out of a search with the keywords 'best dive sites' on the web. The task here is to match our regions with the best site data in this new resource.

To integrate two data sets, it is necessary to adapt the dive site ranking into a dive region ranking. The steps taken in this transformation process are as follows:

- Every diving site in the Best 100 Diving Spots list is induced to the region list. This is done by matching the diving sites in the top 100 list with 721 diving region alternatives obtained at the data organization stage. 53 unique regions are obtained with this operation. These 53 regions are ranked based on the best rank of the corresponding diving site in the top 100 list. For instance, if there are two sites from California ranked 1^{st} and 25^{th} Best 100 Diving Spots list, the rank of California is assigned as 1^{st}. Since the region covers more than one diving site, there are repeating regions in the top 100 list as in the California example. For the cases of diving regions repeating in the list, a frequency value for each diving region is also calculated since it gives useful information about that region's popularity. The frequency values reveal how many sites that belong to a certain region are included in the Best 100 Diving Spots list. For example, there are 12 dive sites from the Red Sea region in the list, therefore, the frequency value of the Red Sea region is 12.

- With the above operation, data consists of ranking and frequency values for 53 regions. After ranking is obtained, the alternatives are divided into three classes to represent the predefined classification to be used in UTADIS. In the predefined classifying process, the frequency values obtained for the diving regions are taken as a basis. The alternatives and their predefined classes are presented in Table 1 under Section 3.4. It can be observed that the frequency values of the regions in the upper ranks are mostly greater than 2, the frequencies of the regions in the medium ranks are mostly 2, and the regions in the lower ranks mostly have a frequency value of 1.

UTADIS stage: Following the data organization and predefined classification processes, the resulting data set consists of 53 alternatives and 25 criteria. The data have predefined classifications and are ready for UTADIS application. At this point, one may ask what will happen to the remaining regions. The answer lies behind the methodological aspect of UTADIS in the optimal weights for criteria that fit the data are also obtained. This property of the method enables to calculation of the utility score of any additional alternative. Comparing the utility score with the optimal thresholds the class of the alternative can be identified. In summary, a 53-alternative set is used to identify the classification and the classes of the remaining alternatives are determined using the optimal solution for this set.

During the UTADIS application, the number of sub-intervals for initial partitioning is selected according to the heuristic (HEUR1) defined by Doumpos and Zopounidis [9], which accepts a number of sub-intervals such that there is at least one alternative belonging in each subinterval. According to Doumpos and Zopounidis [9], large numbers of sub-intervals make the criteria more precise however, they also make the model more rigid which makes it difficult to estimate the classes with accuracy. Following that, the number of sub-intervals in our model is selected as 4 which complies with HEUR1. The s parameter in the model (14) is taken as 0.2 after several trials. It is observed that when the value is less than 0.2, the weights do not distribute well between the criteria. On the other hand, the values over 0.2, the misclassification increases. The linear programming model (14) is then solved with OpenSolver add-in of Microsoft Excel. The findings are summarized in the following section.

3.4. Results

With UTADIS, 53 diving regions are classified into three groups according to 25 criteria and predefined classes. The utility scores, error values, and threshold values are all presented in Table 1. The utility thresholds for each class represented as u_1 and u_2 are obtained as 0.5725 and 0.3725, respectively. The alternatives with utility scores over 0.5725 belong to Class 1, the alternatives with utility scores below 0.3725 belong to Class 3, whereas the utility scores between these values represent Class 2. The utility scores and thresholds provide the class of an alternative, based on the performance regarding its criteria values. Therefore, it is possible to observe the deviations from the predefined classification. This information is provided by the error values. An error value different than 0 indicates the alternative that does not belong to its predefined class. For instance, the utility scores of the Borneo region in Malaysia, and the West Side region in Palau are estimated to be lower than u_1. These alternatives were assigned to Class 1 in predefined classification; however, they actually belong to Class 2 based on their utility values obtained in UTADIS. Similarly, the Visayas region in the Philippines was in Class 2 in the predefined classification but its utility score is estimated to be greater than u_1, meaning that this region actually belongs to Class 1 based on its utility value obtained in UTADIS.

Table 1: The Pre-Classification and the Findings of UTADIS

Rank	**Country, Region**	**# of Dive Sites**	**Frequency**	**Scores**	**Error Values sigma+**	**sigma-**	**Thresholds**
1	Egypt, The Red Sea	408	12	0.6357	0	0	
2	Malaysia, Borneo	93	5	0.5564	0.0162	0	
3	Australia, Queensland	386	5	0.6355	0	0	
4	Palau, Western Side	4	2	0.4893	0.0834	0	
5	United States, Hawaii	306	2	0.6278	0	0	
6	Belize, Lighthouse reef	31	2	0.5726	0	0	
7	Thailand, Similan Islands	42	3	0.5726	0	0	
8	Indonesia, Lesser Sunda Islands	132	3	0.6350	0	0	
9	Ecuador, Galapagos Islands	57	4	0.6438	0	0	**u1**
10	Australia, Western Australia	69	1	0.3462	0.0264	0	0.5725
11	Philippines, Visayas	262	2	0.6056	0	0.0332	
12	Vanuatu, Espiritu Santo	22	1	0.4233	0	0	
13	French Polynesia, Tuamotu Archipelago	62	2	0.5724	0	0	
14	South Africa, Kwa Zulu Natal	93	3	0.4492	0	0	
15	Cayman Islands, Little Cayman	58	1	0.3726	0	0	
16	Fiji, Taveuni	41	1	0.5684	0	0	
17	Mozambique, Inhambane	8	1	0.3213	0.0513	0	
18	Sudan, Port Sudan	11	2	0.3726	0	0	
19	New Zealand, Northland Region	73	1	0.5085	0	0	
20	Mexico, Cozumel	30	2	0.3726	0	0	
21	Philippines, Mindoro	42	2	0.4264	0	0	
22	Indonesia, Sulawesi	108	2	0.5333	0	0	
23	Indonesia, Western New Guinea	12	2	0.3726	0	0	
24	Maldives, Medhu-Uthuru Province	145	2	0.5724	0	0	
25	Mexico, Playa del Carmen	20	1	0.3083	0.0643	0	
26	Australia, New South Wales	153	2	0.5724	0	0	**u2**
27	Mexico, Baja California Sur	45	1	0.3724	0	0	0.3725
28	Dominica, Scott's Head	3	1	0.4185	0	0.046	
29	Tanzania, Zanzibar Archipelago	26	1	0.3240	0	0	
30	Cyprus, Larnaca Bay	9	1	0.1914	0	0	
31	Scotland, Orkney	57	1	0.3724	0	0	

Table 1 – *The Pre-Classification and the Findings of UTADIS (cont.)*

32	The Solomon Islands, Gizo	52	1	0.3798	0	0.0074
33	Costa Rica, Cocos Islands	19	2	0.3724	0	0
34	Mozambique, Quirimbas	22	1	0.3724	0	0
35	Malta, Cirkewwa	2	1	0.1470	0	0
36	Australia, Christmas Island	7	1	0.4936	0	0.1211
37	Thailand, Krabi	50	1	0.4930	0	0.1205
38	Palau, Peleliu	6	1	0.3724	0	0
39	Cayman Islands, Grand Cayman Island: North Wall	48	1	0.2644	0	0
40	Micronesia, Federated States of Micronesia	96	2	0.5224	0	0.15
41	Fiji, Kadavu	35	1	0.4417	0	0.0692
42	Mozambique, Ponto do Barra	11	1	0.3417	0	0
43	Iceland, Thingvellir National Park	2	1	0.3724	0	0
44	Cayman Islands, Grand Cayman Island: East End	14	1	0.1684	0	0
45	Brazil, Pernambuco	28	1	0.3724	0	0
46	Venezuela, Los Roques	12	1	0.3724	0	0
47	Seychelles, Outer Islands	1	1	0.3724	0	0
48	New Caledonia, Grande Terre	12	1	0.3447	0	0
49	Papua New Guinea, Milne Bay Province	50	1	0.3365	0	0
50	Bonaire, Bonaire	94	1	0.3724	0	0
51	Papua New Guinea, Kavieng and New Ireland	33	1	0.3558	0	0
52	Thailand, Surat Thani Province	56	1	0.4417	0	0.0693
53	Malta, Gozo	12	1	0.2297	0	0

Relying on the error values, 13 of the 53 regions in total were misclassified, which means that 75.5% of the alternatives have been classified in their actual classes in predefined classification. The model classified 7 out of 9 alternatives in Class 1, 13 out of 17 in Class 2 and 20 out of 27 in Class 3 in coherence with their predefined classification, showing 77.8 %, 76.5 % and 74.1 % accuracy respectively as seen in Table 2. It is possible to say that the UTADIS classification is quite harmonious with the predefined classification, which is based on The World's Best 100 Diving Spots ranking of Scuba Travel (2019).

Table 2: Confusion Matrix and Pre-Classification Accuracy

		Pre-Classification		
		C1	**C2**	**C3**
Estimated by UTADIS	**C1**	7	1	0
	C2	2	13	7
	C3	0	3	20
Accuracy		77.8%	76.5%	74.1%

As stated, the original dataset includes 721 diving regions and predefined classification consist of only 53 of them. To obtain the classes of the remaining 668 regions, the crucial information is the optimal weights of criteria obtained through the linear programming model of UTADIS presented in (14). The optimal weights provide the weightings of each criterion in obtaining the utility score of any given alternative. Table 3 presents the weights of each criterion that form the basis for utility score calculation.

Table 3: Optimal Weights of the Criteria

	Group	Criteria	Weight
1	Dive Site	Total Number of Dive Sites	0.017299
2	**Type of Diving**	Number of dive sites suitable for Reef diving	0
3		Number of dive sites suitable for Rock diving	0
4		Number of dive sites suitable for Wreck diving	0
5		Number of dive sites suitable for Wall diving	0.12508

Table 3 – *Optimal Weights of the Criteria (cont.)*

6		Number of dive sites suitable for Deep diving	0.019587
7		Number of dive sites suitable for Drift diving	0
8	**Type of Diving**	Number of dive sites suitable for Cave diving	0.017868
9		Number of dive sites suitable for Shark diving	0.064168
10		Number of dive sites suitable for Night diving	0.003673
11		Number of dive sites suitable for Novice divers	0
12	**Experience**	Number of dive sites suitable for Intermediate divers	0.017827
13		Number of dive sites suitable for Advanced divers	0
14		Minimum Depth	0.042529
15	**Depth (in meters)**	Average Depth	0
16		Maximum Depth	0.107993
17		Minimum Visibility	0.204958
18	**Visibility (in meters)**	Average Visibility	0
19		Maximum Visibility	0
20		Minimum Current	0
21	**Current**	Average Current	0.120868
22		Maximum Current	0.076141
23		Minimum Marine Life rating	0.096383
24	**Marine Life**	Average Marine Life rating	0.033455
25		Maximum Marine Life rating	0.052172

Table 4: Class 1 Dive Regions

Country	Region	# of Dive Sites	Original*	Misclassified?**	Class	Score
United States	California	308			1	0.6780
Ecuador	Galapagos Islands	57	X		1	0.6438
Egypt	The Red Sea	408	X		1	0.6357
Australia	Queensland	386	X		1	0.6355
Indonesia	Lesser Sunda Islands	132	X		1	0.6350
Philippines	Luzon	92			1	0.6305
Fiji	Mamanuca Islands	78			1	0.6289
United States	Hawaii	306	X		1	0.6278
Honduras	Utila Island	91			1	0.6257
Fiji	Lomaiviti group	34			1	0.6256
United States	Florida	600			1	0.6087
Philippines	Sulu Sea	19			1	0.6079
Philippines	Visayas	262	X	X	2	0.6056
Australia	Victoria	201			1	0.5978
Bahamas	San Salvador	34			1	0.5953
Virgin Islands	US Virgin Islands	79			1	0.5918
Spain	Canary Islands	123			1	0.5883
Thailand	Similan Islands	42	X		1	0.5726
Belize	Lighthouse reef	31	X		1	0.5726

*This column shows if the region is in the original data set the weights are calculated via UTADIS (53 alternatives). The utility scores of the rest are calculated using the optimal weights from UTADIS.

**This column reveals the regions that are misclassified in the predefined classification and yet assigned to the correct class by UTADIS. Note that such a case is valid only for regions that are in the original data set. For instance, the Visayas region in the Philippines was in Class 2 in the predefined classification but its utility score is estimated to be greater than u1, which means that this region actually belongs to Class 1.

With the weight information, it is possible to estimate the utility score and classify any alternative that fits with the relevant criteria. The classification of all 721 regions is provided as a supplement to this paper. To give a sense of the process, Table 4 presents Class 1 with the alternatives belonging to the 53-alternative UTADIS data set and the alternatives classified as Class 1 using the weights. Observe that Table 4 consists of alternatives with a utility score of over 0.5725, which is the threshold value for Class 1. 11 additional regions belong to Class 1. As for the other classes, 118 and 566 additional regions are found to be belonging to Classes 2 and 3, respectively.

4. Conclusion

Many real-world decision problems involve multiple criteria, but the data may not be initially suitable for quantitative analysis. This research demonstrates how user review data can be processed and analyzed by MCDA classification algorithms. The study systematically collects and processes user-defined scuba diving experiences from web-based sources by converting verbal data into a quantitative dataset. Using the UTADIS method, scuba diving sites are classified based on properly organized and aggregated data. The research integrates end-user perspectives into decision-making, presenting a synthesis of MCDA with a novel web-based data organization method for scuba diving. This approach broadens the applicability of the UTADIS method beyond financial and economic problems. The proposed methodology can be applied to other service sectors where the customer experience and ratings are crucially important for future sales. The instances can include the hospitality sector to select, rank or sort tourist destinations as well as healthcare tourism to analyze the best hospitals in a specific region by criteria such as wait times, quality of care, and patient satisfaction. Therefore, the proposed methodology and the resulting classification, therefore the ranking, has the potential to provide insight for various stakeholders from customers to policymakers as well as the researchers who are interested in utilizing user review data for quantitative decision algorithms. Future research directions may include adaptation of the user-defined data into other MCDA methodologies to rank or sort alternatives in other domains. Furthermore, the scalability of the UTADIS method across various domains also remains an intriguing prospect. Investigating its adaptability to larger datasets and more complex decision-making scenarios can provide a deeper understanding of its practicality and effectiveness.

References

[1] Asl, F. R., Sinaei, H., and Neisi, A. (2021). Forming a Portfolio of Exchange Traded Funds with the Clustering and UTADIS Hybrid Models. Quarterly Journal of Quantitative Economics (Former Economic Studies), 18(1), 79-96. doi: 10.22055/jqe.2020.32330.2204.

[2] Baourakis, G., Conisescu, M., van Dijk, G., Pardalos, PM, Zopounidis, C. (2007). A multicriteria approach for rating the credit risk of financial institutions. Computational Management Science, 6, 347-356. doi: 10.1007/s10287-007-0050-3.

[3] Python Package Index (2004). Beautiful Soup HTML Parser. https://pypi.org/project/beautifulsoup4/ [Accessed 01/10/21].

[4] Devaud, J.M., Groussaud G., Jacquet-Lagrèze E. (1980). UTADIS: Une Méthode de Construction de Fonctions d'utilité Additives Rendant Compte de Jugements Globaux. Germany, Bochum: European working group on MCDA.

[5] Dimitras, AI. (2002). Evaluation of Greek Construction Companies' Securities Using UTADIS Method. European Research Studies, 5 (1-2), 95-107. doi: 10.35808/ersj/77.

[6] Dimitras, AI, Sagka, IV. (2012). Forming a portfolio using multi-criteria method UTADIS. Contributions to accounting research I: scientific studies in honor of the professor George Kafoussis, 1, 116-132.

[7] Dimmock, K., Musa, G. (2015). Scuba Diving Tourism System: A framework for collaborative management and sustainability. Marine Policy, 54, 52-58. doi: 10.1016/j.marpol.2014.12.008.

[8] Divetime (2019). 14756 Scuba Diving Sites: Best Scuba Locations Top Destinations. http://divetime.com/divesites/ [Accessed 01/10/21].

[9] Doumpos, M., Zopounidis, C. (2002). Multicriteria Decision Aid Classification Methods. New York: Kluwer Academic Publishers. doi: 10.1007/b101986.

[10] Esmaelian, M., Shahmoradi, H., and Vali, M. (2016). A novel classification method: A hybrid approach based on extension of the UTADIS with polynomial and PSO-GA algorithm. Applied Soft Computing, 49, 56-70. doi: 10.1016/j.asoc.2016.07.017.

[11] Graver, D. K. (2016). Scuba Diving, 5th Edition. United States: Human Kinetics.

[12] Jacquet-Lagrèze, E., Siskos, Y. (1982). Assessing A Set of Additive Utility Functions for Multicriteria Decision Making, the UTA Method. European Journal of Operational Research, 10, 151–164. doi: 10.1016/0377-2217(82)90155-2.

[13] Jacquet-Lagrèze, E. (1995). An Application of the UTA Discriminant Model for the Evaluation of R & D Projects. In: Pardalos, P.M., Siskos, Y., Zopounidis, C. (Eds) Advances in Multicriteria

Analysis. Nonconvex Optimization and Its Applications, vol 5. Boston: Springer. doi: 10.1007/978-1-4757-2383-0_13.

[14] Laghrabli, S., Benabbou, L., Berrado, A. (2016). Strategic Decision Processes Classification Framework Using UTADIS. 11th International Conference on Intelligent Systems: Theories and Applications (SITA), Morocco. doi: 10.1109/SITA.2016.7772317.

[15] Laryea, R., Carling, K., and Cialani, C. (2020). A decision support system that incorporates price volatility in risk classifying regional food security. International Journal of Risk Assessment and Management, 23(3/4), 223–235. doi: 10.1504/IJRAM.2020.114368.

[16] Lew, AA, (2013). Where to Dive? Ranking the World's top Scuba Diving Location. Tourism, Leisure and Global Change Conference, 1. Malaysia.

[17] Lucrezi, S., Milanese, M., Markantonatou, V., Cerrano, C., Sarà, A., Palma, M., Saayman, M. (2017). Scuba diving tourism systems and sustainability: Perceptions by the scuba diving industry in two Marine Protected Areas. Tourism Management, Elsevier, 59(C), 385-403. doi: 10.1016/j.tourman.2016.09.004.

[18] Manshadi, ED, Mehregan, MR, Safari H. (2015). Supplier Classification Using UTADIS Method Based on Performance Criteria. International Journal of Academic Research in Business and Social Sciences, 5 (2), 31-45, doi:10.6007/IJARBSS/v5-i2/1457.

[19] Parimbelli E, Szymon W, O'Sullivan D, Kingwell S, Michalowski W, Michalowski M. (2020) How Do Spinal Surgeons Perceive The Impact of Factors Used in Post-Surgical Complication Risk Scores? AMIA Annual Symposium Proceedings 2019, 699-706.

[20] Porcu, P. (2018). Python for Data Mining Quick Syntax Reference. Apress. doi: 10.1007/978-1-4842-4113-4.

[21] Rangel, LAD, (2015). Multicriteria Classification of the Organizational Commitment Factors: Application of the UTADIS Method. Systems and Management, 10, 623-632. doi: 10.20985/1980-5160.2015.v10n4.638.

[22] Ruicheng, Y., Rongrong, G., Qing, S., (2016). Detecting Fraudulent Financial Data Using Multicriteria Decision Aid Method. International Conference on Information Science and Control Engineering, 3, 321-324. doi: 10.1109/ICISCE.2016.78.

[23] SCUBA Travel (2019). World's 100 Best Dive Sites. https://www.scubatravel.co.uk/topdiveslong.html [Accessed 01/10/21].

[24] Sweigart, A., (2015). Automate the Boring Stuff with Python: Practical Programming for Total Beginners. USA: No Starch Press, Inc.

[25] Ulucan, A., Atıcı, KB (2009). Multi Criteria Decision Analysis Approaches in Energy Projects Evaluation Process and Turkey Applications, Hacettepe University Journal of Economics and Administrative Sciences, 27 (2), 141-159.

[26] Ulucan, A., Atıcı, K.B. (2013). A Multiple Criteria Sorting Methodology With Multiple Classification Criteria and an Application to Country Risk Evaluation. Technological and Economic Development of Economy, 19(1), 93-124. doi: 10.3846/20294913.2012.763070.

[27] Wannadive (2019). A Dive Site Atlas Made by Divers for Divers. https://wannadive.net/ [Accessed 01/10/21].

[28] Yang, R., and Jiang, Q. (2020). Detecting Falsified Financial Statements Using a Hybrid SM-UTADIS Approach: Empirical Analysis of Listed Traditional Chinese Medicine Companies in China. Discrete Dynamics in Nature and Society, 1-15. doi: 10.1155/2020/8865489.

[29] Zopounidis, C., Doumpos M., (1998). Developing A Multicriteria Decision Support System for Financial Classification Problems: The FINCLAS System. Optimization Methods and Software, 8, 277-304. doi: 10.1080/10556789808805680.

[30] Zopounidis, C., Doumpos M., (1999). Business failure prediction using the UTADIS multicriteria analysis method. Journal of the Operational Research Society, 50, 1138-1148. doi: 10.2307/3010084

[31] Zopounidis, C., Doumpos M., (2000). PREFDIS: A Multicriteria Decision Support System For Sorting Decision Problems. Computers And Operations Research, 27(7-8), 779-797. doi: 10.1016/S0305-0548(99)00118-5

[32] Zopounidis, C., Doumpos M., Pasiouras, F. (2007). Multicriteria Framework for the Prediction of Corporate Failure in the UK. Frontiers in Finance and Economics, 4 (1), 65-90. ssrn: abstract=982181

Supplementary Material: Click Here.