MULTIPLE CRITERIA DECISION MAKING

Vol. 18 2023

Rajiansyah*
Grzegorz Filcek**
David M. Ramsey***

EVALUATING A COMPUTER APPLICATION THAT AIDS MULTI-CRITERIA DECISION MAKING

DOI: 10.22367/mcdm.2023.18.03

Received: 31.07.2023 | Revised: 27.10.2023 | Accepted: 15.01.2024.

Abstract

In this article, we describe and assess the implementation of several methods of multi-criteria decision-making using a web-based computer application. Such an application makes it easier to determine the effectiveness of decisions. The methods adopted in this application are SMART (Simple Multi-Attribute Rating Technique), AHP (Analytic Hierarchy Process), and ANP (Analytic Network Process). Each of these methods has distinctive characteristics in determining the best alternative for the user. This study assesses the feasibility of each method in the application. The application is assessed based on functionality, reliability, efficiency, and usability. (1) Functionality is tested according to the appropriateness of the decisions made, (2) Reliability is assessed using stress testing, (3) Efficiency is assessed according to the computational effort, and (4) System usability is tested according to the user's answers to the Computer System Usability Questionnaire authored by J.R. Lewis. This research results in a decision support system based on SMART that has been appropriately tested and is ready for use.

Keywords: multi-criteria analysis, SMART, AHP, ANP, system usability.

* Wroclaw University of Science and Technology in Wrocław, Department of Computer Science and System Engineering, ul. Wybrzeże Stanisława Wyspiańskiego 27, 50-370 Wrocław, Poland, e-mail: rajiansyah@pwr.edu.pl, ORCID: 0000-0001-7379-6730.

** Wroclaw University of Science and Technology in Wrocław, Department of Computer Science and System Engineering, ul. Wybrzeże Stanisława Wyspiańskiego 27, 50-370 Wrocław, Poland, e-mail: grzegorz.filcek@pwr.edu.pl, ORCID: 0000-0002-9760-0656.

*** Wroclaw University of Science and Technology, Department of Computer Science and System Engineering, ul. Wybrzeże Stanisława Wyspiańskiego 27, 50-370 Wrocław, Poland, e-mail: david.ramsey@pwr.edu.pl, ORCID: 0000-0002-7186-1436.

1 Introduction

A decision support system (DSS) is a component of an organization's or enterprise's computer-based information system that incorporates a knowledge-based system (knowledge management) to assist in decision-making. In the 1970s, Michael S. Scott Morton coined the term 'management decision system' to characterize the concept of a decision support system. Decision support systems are designed to aid in all phases of the decision-making process, from identifying issues to accumulating relevant data and establishing the methodologies used in the decision-making process to evaluate alternatives.

Using decision-support systems, there are numerous ways to make a selection. When multiple criteria are relevant to making a decision, then multi-criteria analysis becomes practical. The primary objective of multi-criteria analysis is to address the difficulties that human decision-makers have demonstrated in consistently coping with large amounts of complex information. Multi-criteria analysis can be used to identify the single most desirable option, rank alternatives, shortlist a limited number of options for later in-depth evaluation, or simply distinguish acceptable from undesirable options. SMART, AHP, and ANP are three methodologies utilized in the process of multi-criteria analysis.

SMART is an adaptable decision-making strategy. The ease with which SMART can respond to the needs of decision-makers and analyze their responses has led to its widespread adoption. This technique provides a comprehensive understanding of a decision problem and is acceptable to the decision-maker due to its transparent analysis. To predict the worth of each option, an adaptive linear model is utilized. AHP is a method used to define appropriate weights for multiple criteria based on pairwise comparison. Inappropriate comparisons of the relative importance of criteria will lead to inconsistent judgments. AHP assumes that the criteria of choice are independent. Since this is not always the case, ANP was introduced as a generalization of AHP that considers the relationship between criteria.

The purpose of this paper is to describe how a prototype application that aids in making multi-criteria decision making can be practically tested. Our study group used this application to obtain purchase recommendations. Afterward, they answered a questionnaire regarding the appropriateness of the recommendations and the usability of each method. This prototype was developed by one of the authors as part of his studies. Unfortunately, the prototype is no longer available online due to maintenance costs. However, the results from this questionnaire can be used to adapt the prototype to the user's needs.

Section 2 gives an overview of the study, including the algorithms used and the aspects of feasibility tested. The methods used to assess aspects of the functioning of the application are described in Section 3. The results of the assessment are described in Section 4. Finally, in Section 5 we give some conclusions and directions for future research. For conciseness, the analysis of the results of the usability test is only described in detail for AHP, and ANP is only briefly described.

2 Overview of the study

2.1 Computer applications

A computer application is a fundamental, complete, and functional package consisting of all the necessary hardware and software installed in a computer to achieve a set purpose. A computer application processes data (receives inputs, processes them, executes commands, and outputs results). It includes hardware components and systems software that collaborate to facilitate the application's functioning. The computer acts as an interface between the user and the application allowing the user to input the required data and outputting the results from running the application to the user (Teeravarunyou and Poopatb, 2009). A practical computer application is a remarkable technological accomplishment that delivers exceptional swiftness, reliability, and adaptability (Herbert and Jones, 2004).

A computer application permits data input, manipulation, and storage. During the data processing phase, instruction sets, also known as programs, are provided to inform the system what to do with the input data. This form of application, known as a stored application, is the most prevalent in use today. It is highly adaptable because it can perform any task by importing a program from memory. Computer applications sometimes operate simply via the computer, but often also access external or interconnected devices.

2.2 Feasibility of a web-based application

Software quality can be assessed from the standpoint of the process of software development (process) or the product generated (product). The ISO 9001 standard may be used to assess processes of software development. In terms of product quality, device software may be assessed using the ISO 9126 standard or best practices defined by software practitioners. McCall Taxonomy is a well-known and widely accepted best practice that is described in a technical manual by MacCall and Matsumoto (1980).

The testing of web-based applications may essentially employ all the tools and techniques typically used in traditional software testing (Engels, Lohmann and Wagner, 2006). Some test methodologies and procedures must be changed and thoroughly described in Web applications. Furthermore, such testing will certainly necessitate the development of new testing methodologies and procedures to address those aspects that do not have a counterpart in traditional software testing (for example, the testing of hypertext structures). Testing can employ the quality attributes specified by ISO / IEC 9126-1 standards, namely representative aspects such as functionality, dependability, usability, and efficiency, as a general categorization (Olsina, Lafuente and Rossi, 2001). Testing or assessing the quality of web-based applications using the methods described above are separated into four primary aspects:

2.2.1 Functionality

Software functionality is a quality criterion associated with how well software achieves a user's goals. This covers the appropriateness of an application's functions to the user's objectives, the accuracy of the results, and the interoperability of software with other systems, and security software (Spinellis, 2006).

There are numerous measurement instruments that may be used to assess functionality based on the software quality criteria described by ISO/IIEC 9126-1 standards. Reviews and inspections are such instruments. Reviews and inspections are software life cycle checks that try to find and eradicate errors early in a product's development process. Furthermore, reviews and inspections might help to remedy faults and thus improve the future of a product. The correctness of results may sometimes be assessed by specialist software or an expert, e.g., if the correct result is objectively defined. However, since the accuracy (in the sense of applicability) of results in this study are user-dependent, a sample of users is required to assess this aspect of functionality under the category of usability.

- Link Examination: This test is used to confirm that no links in a web application are broken. Broken links are hypertext navigation structure links that lead to missing nodes (pages, pictures, etc.) or blanks. Testing these connections entails going from the first page to the last page via the possible routes (Engels, Lohman and Wagner, 2006).
- Browser Evaluation: A wide range of Web browsers are available. Each Web browser behaves differently depending on the manufacturer (for example, Microsoft, Mozilla, Opera), version (e.g., IE 8.0, 9.0), operating system (for example, Windows or Macintosh), hardware equipment (e.g., screen resolution and color depth) or configuration (e.g., cookie activation, language script, stylesheet). Browser testing seeks to identify faults in Web applications caused by browser incompatibilities (Engels, Lohman and Wagner, 2006).
- Security: The most critical feature of online applications is security. These features are used to regulate access to information, authenticate user identification, and encrypt private data. Testing a security feature (for example, en-

cryption) investigates, e.g., whether it is possible to show private data on the results page without logging in, or whether there are input characters that disable the security system.

2.2.2. Reliability/dependability

The capacity of software to sustain a specific degree of performance under certain circumstances is referred to as dependability. Three aspects of dependability are commonly tested: prevention, mitigation, and recovery. Reliability refers to the lack of faults in the software (i.e., prevention of errors). Dependability also involves fault tolerance, which covers the software's ability to recover data and restart operations after a failure, not necessarily of the application itself (Spinellis, 2006).

Subraya (ed., 2006) published *Integrated Approach to Web Performance Testing: A Practitioner's Guide*. Stress testing assists in determining the maximum load that a system can withstand before crashing or becoming substantially impaired. Concurrent use of an application is expected. Hence, the maximum number of users an online system may serve at one time should be specified. Negative testing refers to stress testing that attempts to destabilize the system being evaluated (Camciuc et al., 2005). Such a test assesses how a system recovers. Here are some examples of how stress testing may be used on a web-based system:

- Increasing the number of concurrent HTTP connections by doubling the number of ports on the network switch or routers connected to the server, for example, using a Simple Network Management Protocol (SNMP) command.
- Using an offline database to simplify restarts.
- Building a Redundant Array of Independent Disks (RAID) when the system is performing tasks on the Web server and database from which it takes resources (CPU, RAM, disk, network).

If the server is already suffering saturation or close to the maximum limit in managing the number of application users, it is deemed to only be able to comply with these constraints. This may be used to evaluate a server as a benchmark. If the load is sufficiently low enough to ensure efficient functioning, the Web application can be deemed to work properly.

2.2.3 Efficiency

Software efficiency is concerned with execution time and the use of resources (in particular, memory space and network resources). Hence, efficient applications carry out operations quickly without using a large amount of memory or loading the network. In some cases, an appropriate compromise needs to be found between these two aspects (Spinellis, 2006).

The Zone Research Group in the book *Post Test Execution Phase: Integrated Approach to Web Performance Testing: A Practitioner's Guide* (Subraya, ed., 2006) popularized the 8-second rule. This states that if a Web page is not downloaded within 8 seconds, the user is likely to depart. As illustrated in Table 1, the Zone Research Group also estimates the likelihood that an Internet surfer will wait a specified time for a website to open and the mean time for a website to load given the speed of a modem. These figures are based on a survey of 117 businesses (see Subraya, ed., 2006):

Table 1: The willingness of the user to wait for a website to load

Load time	Percentage of users waiting
10 seconds	84%
15 seconds	51%
20 seconds	26%
30 seconds	5%

Source: Subraya, ed. (2006).

Table 2: Expected load time based on connection speed

Modem speed	Expected load time
14.4 kilobyte modem	11.5 seconds
33.6 kilobyte modem	7.5 seconds
56 kilobyte modem	5.2 seconds
Cable/DSL modem	2.2 seconds
T1 and above	0.8 seconds

Source: Subraya, ed. (2006).

2.2.4 Usability

Usability criteria refer to the ease with which online applications can be used. Such criteria include being simple to comprehend, ease of achieving objectives, learnability, and operability, which involves the effort necessary to utilize the application. Although usability is a crucial aspect of quality, it cannot be assessed simply by an expert inspecting the web application code and how the application operates. Assessing usability requires a representative sample of the users of an application (Spinellis, 2006).

2.3 Multi-criteria analysis

The primary role of multi-criteria analysis is to address the difficulties that human decision-makers have demonstrated in dealing with enormous amounts of complex information in a consistent manner. Multi-criteria analysis can be used to identify a single most preferred option, rank options, shortlist a limited number of options for subsequent detailed evaluation, or simply differentiate acceptable alternatives from unacceptable ones. Multi-criteria analysis (MCA) can be applied in any field where a problem, alternatives, and criteria for selecting an alternative can be defined (Zlaugotne et al., 2020). There are various approaches to MCA, as evidenced by the expanding literature, and their number is continually growing. This variety results from the following: (1) there are many different types of decisions that fit the broad circumstances of MCA; (2) the time available to undertake the analysis may vary; (3) the amount or nature of data available to support the analysis may vary; (4) the analytical skills of those implementing the decision may vary; and (5) the culture and requirement of the organization/individual making the decision may vary.

Multi-criteria decision-making can be viewed as a complex and dynamic process with a managerial and a technical level. The managerial level specifies the goals, the preferences of the decision maker and ultimately selects an option (it may also supply data describing the alternatives and importance of criteria), whereas the technical level gathers data describing the alternatives (if required), implements method(s) for multi-criteria decision-making, and outputs the results of this analysis to the user, e.g., a ranking of the alternatives. Ultimately, decision-makers have the authority to approve or reject the solution proposed by the technical level. Typically, the decision-making process consists of five major stages: (1) describing the problem, generating alternatives, and setting criteria; (2) picking a suitable multi-criteria technique; (3) deriving weights for the criteria; (4) comparative assessment of alternatives; (5) ranking the options. In more detail, these steps involve:

- Step 1: Defining the problem, generating alternatives, and setting criteria. The following must be defined: The ultimate goal (e.g., purchase of a single alternative), the set of alternatives, the players, their objectives, their criteria, any areas of contention, the level of ambiguity, and the critical concerns. The problem might then be framed by providing data appropriate for assessing alternatives according to the given criteria.
- Step 2: Choose an acceptable approach. To rank alternatives, a multi-criteria method must be chosen and applied to the situation at hand. When deciding among different multi-criteria techniques, the decision-makers must consider the form of the data and the degree of uncertainty.
- Step 3: Determine the weights of the criteria: The next step is to assign weights to the criteria based on the approach used. Techniques such as AHP may be used to establish these weights based on input from the users. These weights describe the relative importance of criteria in the multi-criteria problem under study.

- Step 4: Assessment of the alternatives is based on the evaluation matrix: Based on the first three steps, an MCDM problem may be stated in matrix form as follows:

$$\begin{array}{ll} \text{Criteria} & C_1, C_2, \ldots, C_n \\ \text{Weights} & W_1, W_2, \ldots, W_n \\ \text{Alternatives} & \end{array}$$

$$\begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_m \end{bmatrix} \quad \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}$$

Figure 1: Comparison matrix

  where x_{ij} is the evaluation given to alternative i with respect to criterion j, w_j is the weight of criterion j, n is the number of criteria and m is the number of alternatives. These evaluations are transformed according to the method selected to obtain an overall evaluation of an alternative.
- Step 5: Ordering and/or classification of the alternatives: Finally, an ordering of the alternatives is outputted (possibly with a categorization of alternatives as acceptable or unacceptable). This can be translated into a recommendation, i.e., if the goal is to purchase a single alternative, the best-ranked alternative is suggested as a solution.

2.3.1 Simple Multi-Attribute Rating Technique (SMART)

Multi-Attribute Utility Theory (MAUT) is a commonly used approach to multi-criteria decision-making (see Keeney and Raiffa, 1976). It is based on the premise that each alternative is described by various numerical traits, each of which is assessed according to a given criterion. Each criterion has a weight that defines how significant it is compared to other criteria. Normally, these weights are defined so that their sum is equal to one (or 100). The overall score of an alternative is given by the appropriate weighted average of the scores based on the individual criteria. Introduced by Goodwin and Wright (1998), SMART (Simple Multi-Attribute Rating Technique) can be interpreted as a particular implementation of MAUT. This approach is a quantitative method of comparison used to integrate the assessment of costs, risk and each individual's or stakeholder's viewpoint. SMART is a technique of multi-criteria decision-making based on an analytical formula.

SMART is a very adaptable decision-making strategy that is becoming increasingly extensively utilized, due to the ease with which it reacts to decision--makers' demands and analyzes data. Because the analysis is transparent, this technique gives a thorough grasp of a situation and is user-friendly. SMART weighting employs a scale of 0 to 1, making it easier to calculate and compare the values of each alternative (Yunitarini, 2013, p. 46).

The Simple Multi-Attribute Rating Technique (SMART) is a complete decision-making paradigm that can consider qualitative and quantitative factors. It is easy to implement on a computer. SMART also enables easy interaction between an application and its environment. This, in turn, enables monitoring and regulating how the application works (Yulianti, 2015, p. 56).

2.3.1.1 Descriptions of the SMART procedure used in the application

Assume that *n* criteria are used to assess *m* options. The steps performed using the Simple Multi-Attribute Rating Technique (SMART) are as follows:

1. The users describe the weight of each criterion (factor weight), f_j, in a range between 1 and 10. The remaining steps are carried out automatically using a database describing the alternatives.
2. Calculate the normalized weight of each criterion, w_j, by dividing its weight by the sum of the factor weights, i.e.:

$$w_j = \frac{f_j}{\sum_{j=1}^{n} f_j} \tag{1}$$

3. Each of the traits used to assess the alternatives according to the criterion is normalized. This is done by dividing the absolute difference between the observed value of trait *j* for alternative *i*, $x_{i,j}$, and the "worst" value of trait *j*, x_j^W, by the difference between the maximum and minimum value of the trait. If large values of a trait are attractive (e.g., the duration of an electric battery), then the worst value is the smallest and the largest is the best, x_j^B. If large values of a trait are unattractive (e.g., price), then the worst value is the largest and the best value is the smallest. The normalized value of trait *j* for alternative *i* is given by $u_{i,j}$, where:

$$u_{i,j} = \frac{\left|x_{i,j} - x_j^W\right|}{\left|x_j^B - x_j^W\right|} \tag{2}$$

4. The overall score of alternative *i*, v_i, is obtained by calculating the appropriate weighted average of the normalized values according to each criterion, i.e.:

$$v_i = \sum_{j=1}^{n} w_j u_{i,j} \tag{3}$$

5. This score can be multiplied by 100 to obtain an overall value in the range from 0 to 100. The recommendations are given according to the predetermined goal. Most often, a ranking of the alternatives according to these scores is sufficient.

2.3.1.2 Implementation of the SMART method

We will follow the SMART approach to select a smartphone based on a ranking generated by the application. The criteria for choosing a smartphone are budget, memory, camera, feature, and battery. The procedure is as follows:

1. Users must select at least five alternatives to be compared by the application and specify the (unnormalized) weights for each criterion.
2. To initiate the procedure, the user must click 'execute'. The application will then automatically normalize the weights of the criteria and values of the traits.
3. The application ranks the alternatives according to the weighted scores.
4. The user decides which alternative to select.

2.3.2 Analytic Hierarchy Process (AHP)

The AHP, introduced by Saaty (1987), can tackle difficult issues regarding the weights to be ascribed to criteria. This difficulty may result from a decision problem's opaque structure, the ambiguity of a decision-maker's perceptions, and a lack of precise statistical data.

According to Yahya (Suryadi and Ramdhani, 2002, p. 131), decision problems often require immediate action. However, the characterization of the alternatives is so complex that the data are unlikely to be recorded numerically, but only qualitatively measured according to the perceptions, experience, and intuition of the decision-maker. Other models may, however, be adapted to the AHP during the decision-making process (see below).

Suryadi and Ramdhani (2002, p. 131) claim that AHP is more effective than other approaches, based on the following factors:

- AHP can deal with criteria that form a hierarchical structure.
- Criteria are given a more prominent role and their weights can be derived using a more objective approach than adopted under SMART. Criteria and their weights can be developed using surveys, papers, and online material. The ability of AHP to derive quantitative weights from qualitative comparisons can be very beneficial to decision-makers who are not very mathematically literate.
- It is possible to monitor and regulate the output, both at the level of determining the weights of the criteria and at the level of comparing alternatives.

Furthermore, AHP can handle multi-objective and multi-criteria issues by comparing the preferences of each decision-maker involved. As a result, this model is a complete decision-making model.

2.3.2.1 Description of the AHP procedure used in the application

Although AHP can deal with hierarchical criteria, the procedure used assesses alternatives independently according to each criterion, i.e., the criteria have a horizontal structure (see Figure 2).

Hence, again assume that n criteria are used to assess m options. For the problem of selecting a smartphone, n = 5.

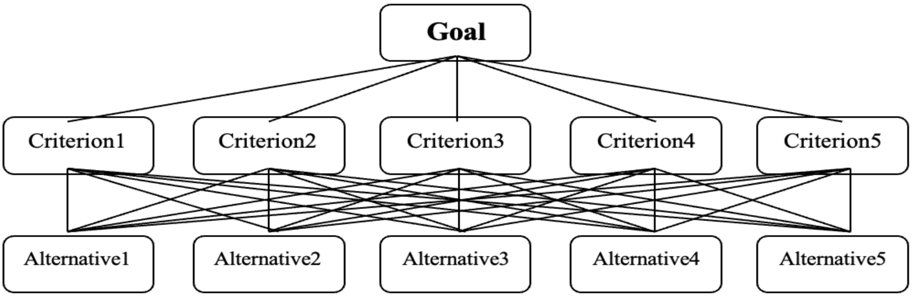


Figure 2: Structure of the hierarchy for AHP

1. The users compare the importance of criteria pairwise. For each of the $0.5n(n-1)$ pairs of criteria, the user must state a) which criterion is more important (unless they are equally important), b) the strength of the difference between the importance of criteria according to the scale given in Table 3. The value $c_{i,j}$ is ascribed to the comparison of the more important criterion, assumed to be criterion i, in a pair with the less important one, assumed to be criterion j. The reciprocal value, $c_{j,i} = \frac{1}{c_{i,j}}$, is ascribed to the comparison of the less important criterion with the more important one. It should be noted that an analogous procedure can be used to compare the attractiveness of pairs of alternatives according to each criterion. In this case, the value $c_{i,j}$ is ascribed to the more attractive one.

Table 3: The value $c_{i,j}$

Scale	Ascribed value $c_{i,j}$	Reciprocal $c_{j,i}$
Equally important/preferred	1	1
Equally to moderately more important/preferred	2	1/2
Moderately more important/preferred	3	1/3
Moderately to strongly more important/preferred	4	1/4
Strongly more important/preferred	5	1/5
Strongly to very strongly more important/preferred	6	1/6
Very strongly more important/preferred	7	1/7
Very to extremely strongly more important/preferred	8	1/8
Extremely more important/preferred	9	1/9

2. Using these comparisons, the application creates the following pairwise comparison matrix (for the case where there are five criteria) that describes the relative importance of the criteria. This matrix will be denoted by *M*.

Table 4: Pairwise Comparison Matrix

	Criteria 1	Criteria 2	Criteria 3	Criteria 4	Criteria 5
Criteria 1	1	$c_{1,2}$	$c_{1,3}$	$c_{1,4}$	$c_{1,5}$
Criteria 2	$c_{2,1}$	1	$c_{2,3}$	$c_{2,4}$	$c_{2,5}$
Criteria 3	$c_{3,1}$	$c_{3,2}$	1	$c_{3,4}$	$c_{3,5}$
Criteria 4	$c_{4,1}$	$c_{4,2}$	$c_{4,3}$	1	$c_{4,5}$
Criteria 5	$c_{5,1}$	$c_{5,2}$	$c_{5,3}$	$c_{5,4}$	1

3. The consistency index, *I*, is calculated using the following formula:

$$I = \frac{\lambda_{max} - n}{(n-1)} \tag{4}$$

where λ_{max} is the largest eigenvalue of the matrix *M*. Dividing this by the expected consistency of a randomly generated consistency matrix (for $n = 5$, this value is approximately 1.25), we obtain the consistency ratio, *R*. The pairwise comparisons are deemed to be consistent when $R \leq 0.1$. If this inequality is not satisfied, then the application asks the user to repeat the pairwise comparisons of the importance of the criteria.
4. The weight of criterion *i* is defined to be the sum of the entries in the *i*-th row of *M* divided by the sum of all the entries in *M*.
5. Once these weights have been calculated, the overall scores of the alternatives can be calculated as in Steps 3-5 of the SMART method using the weights derived in Step 4. Alternatively, scores for each alternative according to each criterion can be defined using the AHP. The user inputs pairwise comparisons of alternatives according to each criterion as described in Step 1. The overall score is calculated using the appropriate weighted average of the scores according to each criterion.

The application uses the AHP to derive both the weights of the criteria and the scores of the alternatives according to each criterion (in total this requires 60 pairwise comparisons). Given these weights and scores, the algorithm implements AHP in an analogous manner to SMART.

2.3.3 Analytic Network Process (ANP)

ANP is a novel approach to the decision-making process that provides a common framework for treating decisions without making any assumptions about the independence of criteria. The AHP assumes that "Sub-Criteria" at one level consider aspects related to only one criterion at the immediately higher level.

Hence, the relationship between these criteria can be defined by a hierarchal tree. Based on ANP, the dependencies between these criteria can vary according to a more general structure. In particular, a criterion at a given level in the hierarchy may be related to two or more criteria at a higher level. The weight of a criterion can be defined as a composite measure (Saaty, 2003). Hence, ANP is an approach that can capture and combine relations between criteria that AHP cannot analyze (Saaty, 2007). However, as always this comes with an increase in the complexity of the process. For conciseness, we omit a precise description of the algorithm used to implement ANP.

3 Methodology of the research

3.1 Research design

This study uses a research and development approach. Such methods are used to produce a specific product and test its effectiveness (Sugiyono, 2009). System development is directed at the effort to develop products that are ready for real use in the field. The object studied in this article is an application to aid in the selection of a smartphone, camera, or laptop by implementing the SMART, AHP, and ANP methods of multi-criteria analysis. The study was conducted in Wroclaw, Poland. The study began in February 2018. In performing data collection for this study, the authors made observations related to aspects of the functionality, reliability, and efficiency of the system being tested. In addition, the authors also collected data by using a questionnaire related to aspects of the usability of the application. Sampling was carried out using the purposive sampling technique. The minimum sample size in such experiments is, according to Gay (Rouse et al., 1999), 15, whereas Sugiyono (2009) recommended a sample size of at least 30. This study used a sample size of 30 students for aspects where the quality of an application cannot be assessed objectively by a small number of experts (e.g., usability, applicability of results). These users are not representative of the population as a whole but are expected to be reasonably representative of the users of such applications.

3.2 Research instruments

This research assessed the quality of this decision support system according to the aspects of its functionality, reliability, efficiency, and usability.

1. **Measuring Functionality.** The aspects of functionality considered are suitability, link functions, interoperability, and security. Assessing suitability was done by testing whether the application passes from the opening page to the closing page correctly and whether each operation runs correctly. Testing link

functioning was carried out by using the tool drlinkcheck.com to find out whether there is a broken link. Assessing interoperability was carried out by opening the application on leading browsers (Firefox, Chrome, and Opera) with the aid of the tool browsershots.org. Testing security was carried out by attempting to access pages that need authorization without going through the login. These aspects of functionality are assessed by calculating the elementary quality preference EP of each element tested. EP is calculated using the formula EP = max {0, (X_{max} − X) / X_{max}}, where X is the ratio of the number of system errors to the total number of system functions and X_{max} is the approved upper limit for this ratio. We assume that X_{max} = 0.04. For example, the X value for the aspect of suitability is X = #(link errors)/ #(total_links_used). EP is categorized into three levels: unsatisfactory (0-0.4), marginal (0.4-0.6), and satisfactory (0.6-1.0). After all the EP values have been obtained, a global evaluation, *P*, is calculated as a weighted average of the EP scores, i.e.:

$$P = \sum_{i=1}^{k} W_i EP_i \quad (5)$$

where W_i is the weight of aspect *i* and EP_i is the EP score with respect to aspect *i*. The sum of the weights must be equal to 1. Since the four aspects of functionality are of similar (high) importance, each was ascribed a weight of 0.25 (Olsina and Rossi, 2002).

2. **Measuring Reliability.** Testing of fault tolerance is carried out using stress testing to determine the condition of the system when the upper limit of system capability is attained. This stress testing is done with the help of the tools load.wpm.neustar.biz and Apache Bench with 200 concurrent connections and 10000 simultaneous requests from simulated visits. Reliability is also assessed using stress testing assuming the number of users is 120.
3. **Measuring Efficiency.** The run time is measured by load testing to see how fast a visitor can access the information system. Efficiency analysis is done with the aid of the following tools: webtoolhub.com, webpagetest.org, and tools.pingdom.com to measure the time it takes the user to access the system. The results obtained are then compared with the 8-second rule.
4. **Measuring Usability.** Usability is assessed using a web-based questionnaire adapted from the User Interface Usability Evaluation, described by Perlman (2009). This is a standard questionnaire based on IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use (Lewis, 1995). This questionnaire is displayed in Table 5.

Table 5: Usability satisfaction questionnaire

No.	Question	Score
1.	I am satisfied with the ease of using the whole system	1 2 3 4 5 6 7
2.	It is simple to use this system	1 2 3 4 5 6 7
3.	I can complete a task effectively when using this system	1 2 3 4 5 6 7
4.	I can complete a task quickly when using this system	1 2 3 4 5 6 7
5.	I can complete a task efficiently when using this system	1 2 3 4 5 6 7
6.	I feel comfortable using this system	1 2 3 4 5 6 7
7.	It is easy to learn how to use this system	1 2 3 4 5 6 7
8.	I am sure I will be more productive using this system	1 2 3 4 5 6 7
9.	If an error occurs, the system gives an error message and instructions on how to resolve it	1 2 3 4 5 6 7
10.	Whenever I make a mistake in using this system, I can return easily and quickly	1 2 3 4 5 6 7
11.	The information provided by this system is clear	1 2 3 4 5 6 7
12.	It is easy to find any necessary information	1 2 3 4 5 6 7
13.	The information provided by this system is easily understandable	1 2 3 4 5 6 7
14.	The information provided by the system helps me to effectively complete a task	1 2 3 4 5 6 7
15.	The layout of information on this system is clear	1 2 3 4 5 6 7
16.	This system display makes it easy for me to use	1 2 3 4 5 6 7
17.	I like the design of this system	1 2 3 4 5 6 7
18.	All the functions in this system are in line with my expectations	1 2 3 4 5 6 7
19.	Overall, I am satisfied with the performance of this system	1 2 3 4 5 6 7

Source: Based on Lewis (1995).

Analysis of usability begins by testing the internal consistency of the data obtained using Cronbach's α coefficient (see Cronbach, 1951):

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{k}\sigma_i^2}{\sigma^2}\right) \quad (6)$$

where k is the number of questions used to analyze an aspect, σ_i^2 the variance of the answers to question i and σ^2 the variance of all the answers to these questions. The internal consistency of the data is then assessed using Table 6.

Table 6: Internal consistency according to Cronbach's α

Cronbach's alpha	Internal consistency
$\alpha \geq 0.9$	*Excellent*
$0.7 \leq \alpha < 0.9$	*Good*
$0.6 \leq \alpha < 0.7$	*Acceptable*
$0.5 \leq \alpha < 0.6$	*Poor*
$\alpha < 0.5$	*Unacceptable*

The answers may be categorized into coherent aspects by selecting sets of answers for which Cronbach's α is maximized (see Section 4). Once this has been done, the aspects can be categorized according to the mean rating given by the users as a percentage score of the maximum score attainable (here 7).

4 Implementation

4.1 Application

The decision support system presented here can implement three methods of multi-criteria analysis (SMART, AHP, and ANP). Each method uses the appropriate algorithm to assess the alternatives based on the specified criteria and the weights (or comparisons) supplied by the user. The application also uses a library created to carry out multi-criteria analysis. This library is designed to make the system more flexible and can be used in all the modules. We developed a website application called 'YouDecide'. YouDecide is designed to help people who are having problems when making a purchase (of a smartphone, laptop, camera, etc.) or selection (e.g., bank, university, restaurant, etc.). The platform provides fast, simple, and free features to assist users in making a final selection by offering a ranking of the alternatives at the end of the process. However, the final decision ultimately rests with the user. Each method provides the user with graphical results and a ranking of the alternatives.

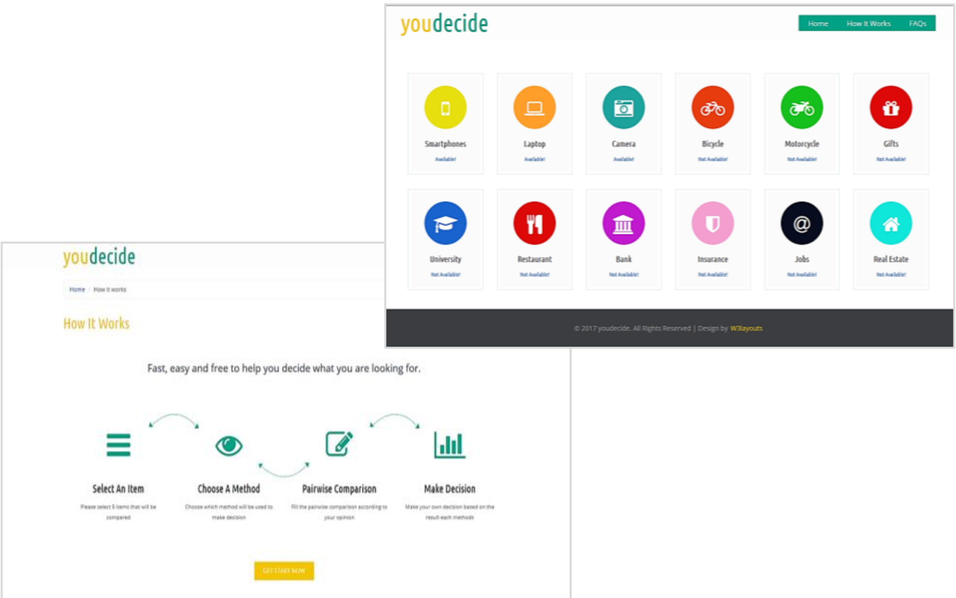
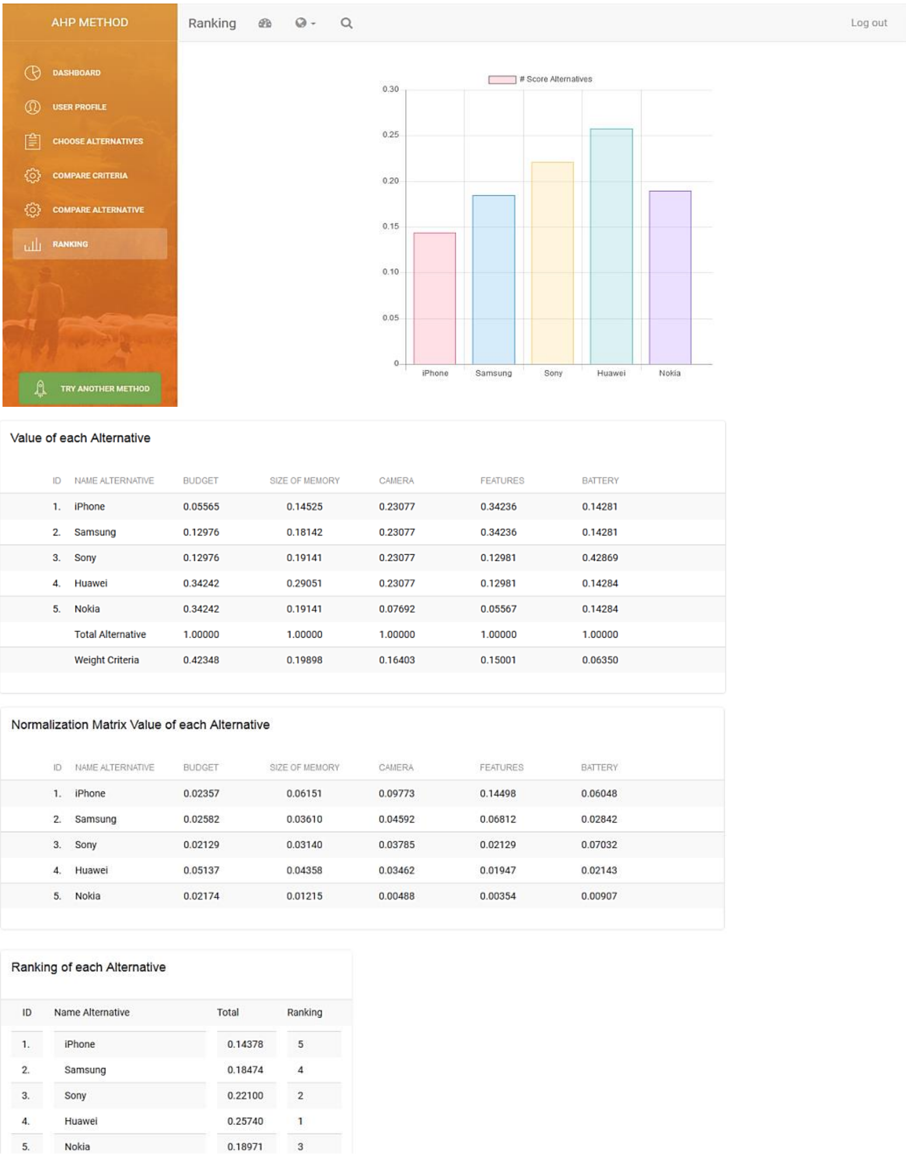


Figure 3: Home page of the application



Value of each Alternative

ID	NAME ALTERNATIVE	BUDGET	SIZE OF MEMORY	CAMERA	FEATURES	BATTERY
1.	iPhone	0.05565	0.14525	0.23077	0.34236	0.14281
2.	Samsung	0.12976	0.18142	0.23077	0.34236	0.14281
3.	Sony	0.12976	0.19141	0.23077	0.12981	0.42869
4.	Huawei	0.34242	0.29051	0.23077	0.12981	0.14284
5.	Nokia	0.34242	0.19141	0.07692	0.05567	0.14284
	Total Alternative	1.00000	1.00000	1.00000	1.00000	1.00000
	Weight Criteria	0.42348	0.19898	0.16403	0.15001	0.06350

Normalization Matrix Value of each Alternative

ID	NAME ALTERNATIVE	BUDGET	SIZE OF MEMORY	CAMERA	FEATURES	BATTERY
1.	iPhone	0.02357	0.06151	0.09773	0.14498	0.06048
2.	Samsung	0.02582	0.03610	0.04592	0.06812	0.02842
3.	Sony	0.02129	0.03140	0.03785	0.02129	0.07032
4.	Huawei	0.05137	0.04358	0.03462	0.01947	0.02143
5.	Nokia	0.02174	0.01215	0.00488	0.00354	0.00907

Ranking of each Alternative

ID	Name Alternative	Total	Ranking
1.	iPhone	0.14378	5
2.	Samsung	0.18474	4
3.	Sony	0.22100	2
4.	Huawei	0.25740	1
5.	Nokia	0.18971	3

Figure 4: Output using AHP

4.2 Testing the system

System testing is performed from the aspects of functionality, reliability, usability, and efficiency. These aspects were tested independently using several tools that provide facilities for testing web applications (drlinkcheck.com, webtoolhub.com, gtmetrix.com, uptrends.com, load.wpm.neustar.biz, and Apache Bench).

4.2.1 Testing functionality

Testing is done on the component aspects of functionality as follows:

a) **Suitability.** The controller/library name and number of functions called for each step in a procedure are shown in Table 7. The Elementary Preference (EP) is calculated as the percentage of functions that run according to expectations. All the functions ran as expected. Hence, for each of the three algorithms, the EP score was 100%.

Table 7: The total number of functions for each method

Method	Controller/Library	Number of function calls
SMART	smart/alternative.php	1
	smart/inputAlternative.php	1
	smart/ranking.php	1
	smart/executeRanking.php	3
	Total	6
AHP	ahp/alternative.php	1
	ahp/inputAlternative.php	6
	ahp/includes/criteria.inc.php	8
	ahp/includes/weight.inc.php	19
	ahp/includes/alternative.inc.php	8
	ahp/includes/bobotk.inc.php	18
	ahp/includes/saaty.inc.php	8
	ahp/compareCriteria.php	1
	ahp/compareAlternative.php	1
	ahp/executeRanking.php	24
	Total	94
ANP	anp/alternative.php	1
	anp/inputAlternative.php	11
	anp/includes/criteria.inc.php	7
	anp/includes/weight.inc.php	18
	anp/includes/alternative.inc.php	8
	anp/includes/bobotk.inc.php	18
	anp/includes/saaty.inc.php	8
	anp/compareCriteria.php	1
	anp/compareAlternative.php	1
	anp/executeRanking.php	135
	Total	198

b) **Accuracy.** Testing the system using drlinkcheck.com showed that all 54 links (48 internal and 6 external) contained in the application were functional. Hence, the value of EP is 100% for each algorithm. The output from this analysis is shown in Figure 5.

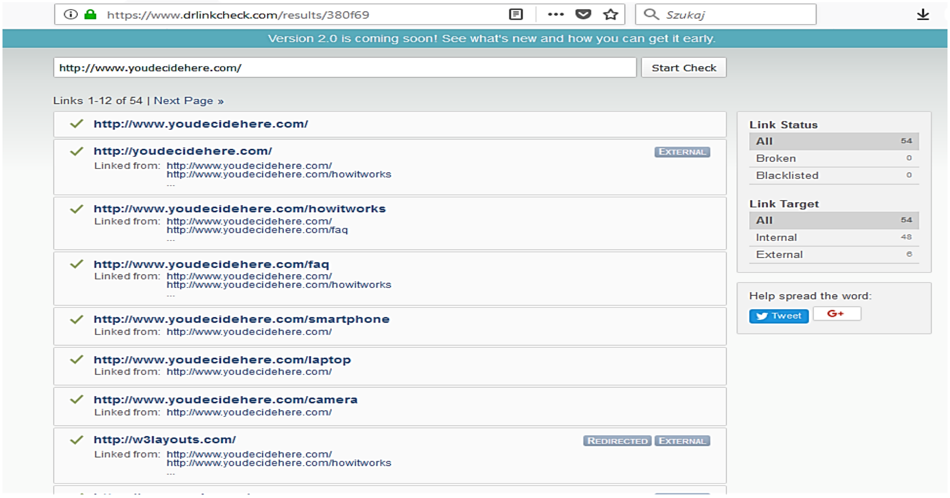


Figure 5: Testing broken links

c) **Interoperability.** Interoperability was tested using browsershots.org. The results are shown in Table 8.

Table 8: Results from interoperability testing

No.	Operating System	Browser	Result
1.	Debian 6.0	Firefox 57.0	√
2.	Debian 6.0	Chrome 38.0	√
3.	Debian 6.0	Opera 12.14	√
4.	Windows 8	Firefox 35.0	–
5.	Windows 8	Opera 15.0	–
6.	Windows 8	Chrome 18.0	√
7.	Windows 7	Firefox 56	√
8.	Windows 7	Opera 15.0	√
9.	Windows 7	Chrome 18.0	√
10.	Linux Ubuntu 9.10	Firefox 35.0	√
11.	Linux Ubuntu 9.10	Opera 15.0	√
12.	Linux Ubuntu 9.10	Chrome 18.0	√

The test results show that the system runs properly on ten of the twelve configurations of the operation system and browser. Hence, the EP value for the interoperability component is 83.33% for each algorithm.

d) **Security.** None of the features in the application are accessible to unauthorized users. Hence, the value of EP is 100% for each algorithm. Having obtained the EP values for the 4 component aspects of functionality for each algorithm, the overall value, P, from that aspect can be calculated. It should be noted that these scores were identical for each of the three algorithms. Assuming that the weights of the components are equal, the P value for each algorithm is:

$$P = 0{,}25 * 100 + 0{,}25 * 100 + 0{,}25 * 83{,}33 + 0{,}25 * 100 = 95{,}83\% \quad (7)$$

Such a value indicates that the functioning of the algorithm is satisfactory.

4.2.2 Testing reliability

Test results using the tool load.wpm.neustar.biz can be seen in Figure 6. These results show that the system can serve 15 users per minute. This means that the system has no problem in being used, e.g. for small laboratory groups.

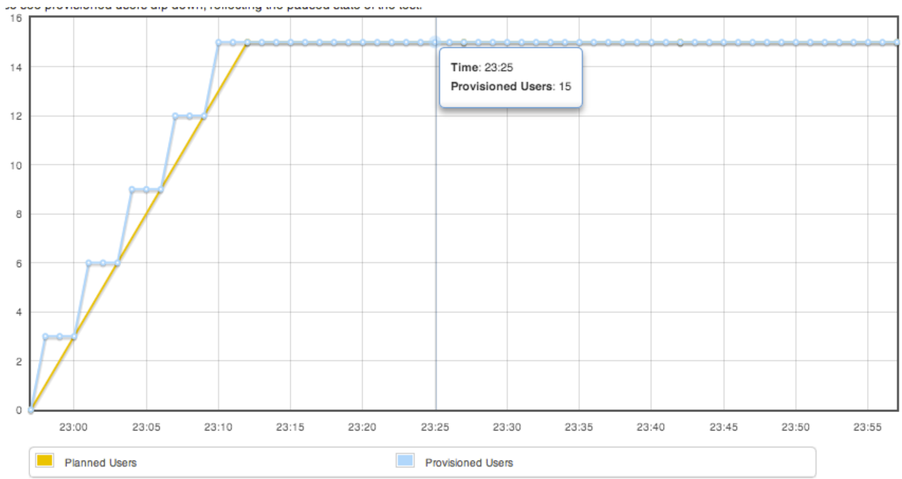
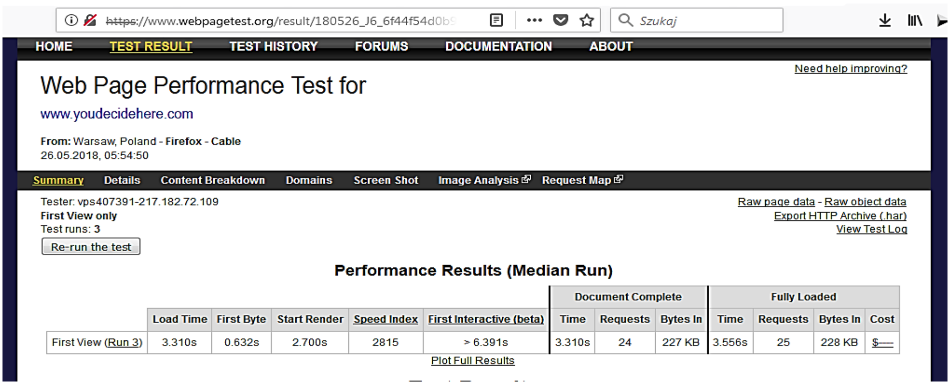


Figure 6: Result reliability testing

4.2.3 Testing efficiency

The results of the test using the webtoolhub.com, gtmetrix.com, and uptrends.com can be seen in Figure 7 below:



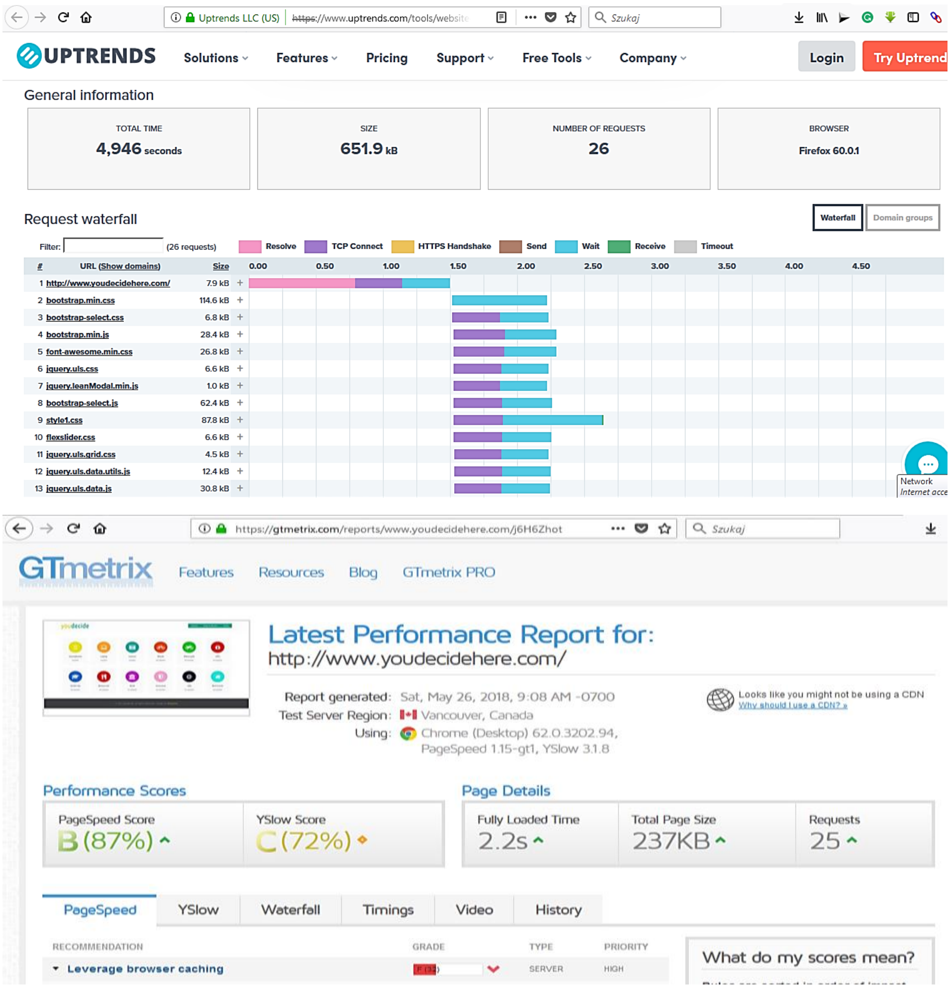


Figure 7: The results of testing the efficiency aspect

The test results indicate that the user's waiting time does not exceed 8 seconds. Hence, the system can be said to run well (see Subraya, ed., 2006).

4.2.4 Testing usability

The results of testing the usability of the AHP method based on the sample of 30 users who wanted to buy a smartphone, camera, or laptop are presented in Tables 9 and 10 (for the SMART and ANP methods, these results will be given in the Appendix). Table 9 illustrates the distribution of answers to each question. Table 10 illustrates the answers given by each user.

The consistency of the answers to all of the questions is described by a Cronbach's α score of 0.57715. This is somewhat low, but this is unsurprising since the questions above relate to different aspects of usability. For example, the answers to Questions 1-8 and 10 seem to be related to the ease of using the AHP algorithm. These aspects are rated relatively lowly. On the other hand, Questions 9 and 11-17 seem to be more related to the output of the system, which is rated more highly. Finally, the final question asks for an overall impression, which does not seem to fall into either of these two categories. In order to appropriately group these questions, Cronbach's α was maximized by sequentially removing a question from the set questions of considered. Starting from the full set of questions, the following questions were successively removed: Q5, Q13, Q10, Q9, Q6, Q7, Q3, Q2 and Q4. Cronbach's α based on the remaining set of variables (Q1, Q8, Q11-12, Q14-19) is 0.60135, which indicates a reasonable level of coherence. This set of questions may be interpreted as indicating users' level of satisfaction with the system as a whole, in particular its layout and output. Except for the overall appraisal of the system, these ratings were generally good. We then searched for another set of coherent answers from the questions that were removed in the procedure described above. After removing questions Q5, Q4, Q10, Q13, and Q9 in turn, Cronbach's α based on the remaining set of variables (Q2-3, Q6-7) is 0.68815, which indicates a satisfactory level of coherence. These questions can be interpreted as indicating the ease of using the algorithm (particularly since the ratings for the SMART method were clearly better in this aspect).

This aspect is rated relatively poorly. It may be concluded that the users' lack of satisfaction with the application resulted from the complexity of the algorithm rather than the application's interface or output. Since the overall assessments of the application are mediocre (the mean rankings for Q6-7 were less than 50% of the maximum possible), we may conclude that the usability of the AHP algorithm is "**Unacceptable**". The results obtained for the ANP model are similar, but generally worse.

Table 9: Results of the usability test for the AHP method
(Distribution of answers to each question)

No.	Question	Score							Average Score
		1	2	3	4	5	6	7	
1.	I am satisfied with the ease of using the whole system	0	2	5	13	10	0	0	4.033
2.	It is simple to use this system	0	0	12	17	0	0	1	3.700
3.	I can complete a task effectively when using this system	0	0	5	15	5	5	0	4.333
4.	I can complete a task quickly when using this system	0	0	25	5	0	0	0	3.166
5.	I can complete a task efficiently when using this system	0	0	23	7	0	0	0	3.233
6.	I feel comfortable using this system	0	0	21	9	0	0	0	3.300
7.	It is easy to learn how to use this system	0	0	21	9	0	0	0	3.300
8.	I am sure I will be more productive using this system	0	1	10	19	0	0	0	3.600
9.	If an error occurs, the system gives an error message and instructions on how to resolve it	0	0	0	0	0	0	30	7.000

Table 9 cont.

No.	Question	Score							Average Score
		1	2	3	4	5	6	7	
10.	Whenever I make a mistake in using this system, I can return easily and quickly	0	0	16	13	1	0	0	3.533
11.	The information provided by this system is clear	0	0	10	15	5	0	0	3.833
12.	It is easy to find any necessary information	0	0	0	0	9	9	12	6.100
13.	The information provided by this system is easily understandable	0	0	0	14	12	4	0	4.667
14.	The information provided helps me complete a task effectively on this system	0	0	0	0	5	15	10	6.167
15.	The layout of information on this system is clear	0	0	0	0	0	5	25	6.833
16.	The system display makes it easy for me to use	0	0	0	0	5	10	15	6.333
17.	I like the design of this system	0	0	0	0	2	5	23	6.666
18.	All the functions in this system are in line with my expectations	0	0	10	10	10	0	0	4.000
19.	Overall, I am satisfied with the performance of this system	2	0	20	8	0	0	0	3.133

Table 10: Results of the usability test for the AHP method (Answers given by each individual)

User No.	Question No.																			Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
1.	5	4	6	3	4	3	3	3	7	5	5	6	5	7	7	6	7	4	4	90
2.	5	4	6	3	4	3	3	2	7	3	5	6	5	6	7	6	7	4	4	86
3.	5	4	6	3	3	3	3	3	7	4	5	6	5	6	7	6	7	4	4	87
4.	5	4	6	3	3	4	4	3	7	4	3	5	4	5	6	5	6	3	3	80
5.	2	4	3	3	3	3	3	4	7	4	3	5	4	5	6	5	6	3	3	73
6.	3	3	3	3	4	3	3	4	7	4	3	5	4	5	6	5	5	3	1	73
7.	3	3	3	3	4	3	3	4	7	3	4	6	5	6	7	6	6	3	3	79
8.	4	4	5	3	3	4	4	4	7	4	4	5	4	5	6	5	6	3	3	80
9.	4	4	5	3	3	4	4	3	7	3	4	7	6	6	7	6	6	3	3	85
10.	4	4	5	3	3	4	4	4	7	4	4	7	6	6	7	6	6	3	3	87
11.	4	4	5	3	3	4	4	4	7	3	3	5	4	6	7	7	7	4	4	85
12.	4	3	4	3	3	4	4	4	7	4	4	7	6	6	7	7	7	4	4	88
13.	2	3	3	3	3	3	3	3	7	3	3	7	6	5	6	5	5	3	1	73
14.	3	3	3	3	3	3	3	3	7	3	4	6	5	7	7	7	7	3	3	80
15.	4	4	4	4	4	3	3	3	7	3	3	7	5	6	7	7	7	5	3	86
16.	5	7	6	3	3	3	3	3	7	3	3	5	4	6	7	7	7	5	3	87
17.	4	3	4	3	4	3	3	3	7	3	3	5	4	6	7	7	7	5	3	81
18.	4	4	5	3	3	4	4	4	7	3	4	6	5	7	7	7	7	5	3	89
19.	3	4	4	4	3	3	3	4	7	3	4	6	5	7	7	7	7	3	3	84
20.	3	3	4	4	4	3	3	4	7	3	3	5	4	6	7	6	7	5	3	81
21.	4	3	4	4	3	3	3	4	7	4	3	5	4	6	7	6	7	5	3	82
22.	5	3	4	3	3	3	3	4	7	4	4	6	4	6	7	7	7	5	3	85
23.	4	3	4	3	3	3	3	4	7	3	4	6	4	6	7	6	7	4	3	81
24.	5	4	4	3	3	3	3	4	7	4	4	7	5	7	7	6	7	4	3	87
25.	4	4	4	3	3	3	3	4	7	4	4	7	5	6	7	7	7	5	4	87
26.	5	4	4	3	3	3	3	4	7	3	4	7	4	7	7	7	7	5	4	87
27.	4	4	4	3	3	4	4	4	7	3	4	7	5	7	7	7	7	5	4	89
28.	5	4	4	3	3	4	4	4	7	4	4	7	5	7	7	7	7	4	3	90
29.	4	3	4	4	3	3	3	4	7	4	5	7	4	7	7	7	7	4	3	87
30.	5	3	4	3	3	3	3	3	7	4	5	7	4	7	7	7	7	4	3	86

Table 11: Comparison of the three methods

Method	Functionality	Reliability	Efficiency	Usability	Result
SMART	Good	Good	Good	Good	√
AHP	Good	Good	Good	Unacceptable	**x**
ANP	Good	Good	Good	Unacceptable	**x**

Based on the results given in Table 11, the tests on the implementation of the SMART, AHP, and ANP methods indicate that the SMART method is „Good" and should be recommended to users rather than the other two methods to make decisions such as to buy a smartphone, camera, laptop, etc. For such purposes, the other two methods should either be simplified or not given as an option.

5 Conclusions and future work

5.1 Conclusions

Based on the results of the research, our conclusions can be summarized as follows:

1. The application fulfills the feasibility criteria of functionality, regardless of the algorithm used. The global evaluation of feasibility gave a score of 95.83%, indicating that the system is satisfactory.
2. The application also fulfills the reliability criteria of functionality, regardless of the algorithm used. The system can handle 15 users every minute.
3. Testing the efficiency aspect generated an average waiting time of 3.48 seconds. This result is better than required by the 8-second rule.
4. However, in testing the usability aspect, only the SMART method meets the feasibility criteria of usability based on the survey results. For each question, the mean ratings of the implementation of the SMART algorithm were high. In addition, the minimum average score for any of the questions was 6.0 (85.71%). This result indicates that users rate SMART highly over the whole range of aspects of usability (ease of use, attractiveness of interface, and clarity of the output). Analysis of the ratings for the AHP algorithm indicates that there were two relatively coherent sets of answers. One of these sets describes the users' level of satisfaction with the interface and output. These aspects tended to be rated fairly highly. The other set of answers describes how easy the method is to use. This aspect was generally poorly rated. The ratings of the ANP algorithm showed a similar pattern with the ease of use being rated even lower than for AHP. Hence, for the purposes of purchasing objects such as smartphones based on numerical, the algorithm that we recommend is the SMART method. This is due to SMART being easy to understand, simple, quick in making a decision, efficient, and effective. These results are in line with those obtained by Bottomley, Doyle and Green (2000),

who state that the direct rating approach is both simple and effective. However, Wachowicz and Roszkowska (2023) find that the method of directly giving a score to the weight of a criterion might be problematic, since decision-makers avoid ascribing extreme weights. This might result from the fact that it might be more natural to compare the weights of the criteria.

5.2 Future work

Based on the comments above, it would be reasonable to simplify AHP, so that only the weights of criteria are derived using AHP. Once these weights have been derived, the scores of the alternatives according to each criterion can be defined using the same procedure as used by SMART. This would reduce the number of pairwise comparisons that need to be made, especially when there are many alternatives. In the problem considered here, the number of comparisons needed would be reduced from 60 to 10.

Future work could be aimed at developing and testing the use of other methods, such as Simple additive Weighting, PROMETHEE, and ELECTRE, in a computer application. For the purposes of purchasing, e.g., a smartphone, it is recommended that the amount of input from the user should be at most moderate. However, a large number of methods can be implemented as long as the weights of the criteria are set appropriately. Hence, it could be useful to implement even a relatively complex mathematical method based on raw numerical data and "user-defined" weights, as long as the results from its use are intuitively pleasing. The choice of method depends heavily on the decision to be made and the data that are available. The selection of a smartphone on the basis of an application supporting multi-criteria decisions and several numerical variables seems to be a reasonable approach. However, the choice of a costly, unique good, such as a flat, or second-hand car would require a more advanced approach (see Ramsey, 2020).

Note from the authors

Unfortunately, the application is no longer available online. However, it is hoped that the analysis carried out above is illustrative in showing how such an application can be tested before being made generally available.

Appendix

		Functions	Content and Structure	Infrastructure and Environment
Functionality	**Suitability**	Reviews and inspections, Test-driven development	Checklists, Lexical testing, Style guides, Reviews	
	Accuracy	Capture/Replay, Test-driven development	Static analysis, Link testing, Lexical testing, Reviews	Static analysis, Link testing
	Interopera-bility	Cross-browser and cross-platform compatibility testing	Test printing, Checklists, Reviews, Compatibility testing	Cross-browser and cross-platform compatibility testing
	Compliance	Compatibility testing, Style guides, Test-driven development	Checklists, Compatibility testing, Style guides, Reviews	Cross-browser and cross-platform compatibility testing
	Security	Analysis of common attacks, Reviews and inspections		Analysis of common attacks, Forced-error testing, Ethical hacking
Reliability	**Maturity**	Endurance testing		Endurance testing
	Fault Tolerance	Forced-error testing, Stress testing		Forced-error testing, Low-resource testing, Stress testing
	Recoverability	Forced-error testing, Fail-over testing		Fail-over testing, Forced-error testing, Low-resource testing
Usability	**Understanda-bility**	Usability studies, Heuristic evaluation	Static readability analysis, Usability studies	
	Learnability	Usability studies, Heuristic evaluation		
	Operability	Usability studies, Heuristic evaluation		Heuristic evaluation
	Attractiveness		Publicity testing	
Efficiency	**Timing Behavior**	Load and Stress testing,Monitoring		Load and Stress testing, Monitoring
	Resource Utilization	Endurance testing	Load testing	Endurance testing, Monitoring

Figure 8: Web-based application testing methods

Source: Olsina, Lafuente, Rossi (2001).

Table 12: Results of the usability test for the SMART method
(Distribution of answers to each question)

No.	**Question**	**Score**							**Average Score**
		1	2	3	4	5	6	7	
1.	I am satisfied with the ease of using the whole system	0	0	0	0	3	6	21	6.666
2.	It is simple to use this system	0	0	0	0	0	3	27	6.900
3.	I can complete a task effectively when using this system	0	0	0	0	1	2	27	6.867
4.	I can complete a task quickly when using this system	0	0	0	0	0	3	27	6.900
5.	I can complete a task efficiently when using this system	0	0	0	0	0	3	27	6.900
6.	I feel comfortable using this system	0	0	0	0	0	0	30	7.000
7.	It is easy to learn how to use this system	0	0	0	0	0	0	30	7.000
8.	I am sure I will be more productive by using this system	0	0	0	5	5	5	15	6.167
9.	If an error occurs, the system gives an error message and instructions on how to resolve it	0	0	0	0	0	0	30	7.000
10.	Whenever I make a mistake in using this system, I can return easily and quickly	0	0	0	0	1	1	28	6.900
11.	The information provided by this system is clear	0	0	0	5	0	15	10	6.000
12.	It is easy to find any necessary information	0	0	0	0	0	18	12	6.367
13.	The information provided by this system is easily understandable	0	0	0	0	2	4	24	6.733
14.	The information provided helps me effectively complete a task on this system	0	0	0	0	0	5	25	6.833
15.	The layout of information on this system is clear	0	0	0	0	0	3	27	6.900
16.	The system display makes it easy for me to use	0	0	0	0	0	1	29	6.967
17.	I like the design of this system	0	0	0	0	2	5	23	6.700
18.	All the functions in this system are in line with my expectations	0	0	0	0	0	10	20	6.667
19.	Overall, I am satisfied with the performance this system	0	0	0	0	0	0	30	7.000

Table 13: Results of the usability test for the SMART method
(Answers given by each individual)

User No.	**Question No.**																			Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
1	7	7	5	7	7	7	7	7	7	7	7	6	7	7	7	7	6	6	7	128
2	7	7	7	7	7	7	7	7	7	7	7	6	7	7	7	7	7	7	7	132
3	7	7	7	7	7	7	7	7	7	7	7	6	7	7	6	7	7	7	7	131
4	7	7	7	7	7	7	7	7	7	7	6	6	7	7	7	7	7	7	7	131
5	7	7	7	7	7	7	7	7	7	7	6	6	7	7	7	7	7	7	7	131
6	7	7	7	7	7	7	7	7	7	7	6	6	7	7	7	7	7	7	7	131
7	6	6	6	6	7	7	7	4	7	7	4	6	5	6	6	6	5	6	7	114
8	7	7	7	6	7	7	7	4	7	7	4	6	5	6	6	7	6	6	7	119
9	7	7	7	7	7	7	7	7	7	7	6	6	7	6	7	7	6	6	7	128
10	6	6	6	6	7	7	7	4	7	7	4	6	6	6	7	7	5	6	7	117
11	6	6	7	7	7	7	7	4	7	7	4	6	6	6	7	7	6	6	7	120
12	7	7	7	7	7	7	7	7	7	7	6	6	6	7	7	7	6	7	7	129
13	7	7	7	7	6	7	7	4	7	5	4	6	6	7	7	7	7	7	7	122
14	5	7	7	7	6	7	7	5	7	6	6	7	7	7	7	7	7	6	7	125
15	7	7	7	7	6	7	7	5	7	7	6	7	7	7	7	7	7	7	7	129
16	5	7	7	7	7	7	7	7	7	7	6	7	7	7	7	7	7	7	7	130

Table 13 cont.

User No.	**Question No.**																			Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
17	7	7	7	7	7	7	7	7	7	7	6	7	7	7	7	7	7	7	7	132
18	7	7	7	7	7	7	7	7	7	7	6	6	7	7	7	7	7	6	7	130
19	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	133
20	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	133
21	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	133
22	6	7	7	7	7	7	7	5	7	7	6	7	7	7	7	7	7	7	7	129
23	7	7	7	7	7	7	7	7	7	7	6	7	7	7	7	7	7	7	7	132
24	5	7	7	7	7	7	7	7	7	7	7	6	7	7	7	7	7	6	7	129
25	7	7	7	7	7	7	7	7	7	7	7	6	7	7	7	7	7	6	7	131
26	7	7	7	7	7	7	7	7	7	7	7	6	7	7	7	7	7	7	7	132
27	6	7	7	7	7	7	7	5	7	7	7	6	7	7	7	7	7	7	7	129
28	7	7	7	7	7	7	7	7	7	7	6	6	7	7	7	7	7	7	7	131
29	7	7	7	7	7	7	7	7	7	7	6	7	7	7	7	7	7	7	7	132
30	6	7	7	7	7	7	7	5	7	7	6	7	7	7	7	7	7	7	7	129

Table 14: Results of the usability test for the ANP method
(Distribution of answers to each question)

No.	**Question**	**Score**							**Average Score**
		1	2	3	4	5	6	7	
1.	I am satisfied with the ease of using the whole system	23	6	1	0	0	0	0	1.267
2.	It is simple to use this system	27	3	0	0	0	0	0	1.100
3.	I can complete a task effectively when using this system	20	10	0	0	0	0	0	1.333
4.	I can complete a task quickly when using this system	30	0	0	0	0	0	0	1.000
5.	I can complete a task efficiently when using this system	17	10	3	0	0	0	0	1.533
6.	I feel comfortable using this system	28	2	0	0	0	0	0	1.067
7.	It is easy to learn how to use this system	30	0	0	0	0	0	0	1.000
8.	I am sure I will be more productive using this system	29	1	0	0	0	0	0	1.033
9.	If an error occurs, the system gives an error message and instructions on how to resolve it	10	12	8	0	0	0	0	1.933
10.	Whenever I make a mistake in using this system, I can return easily and quickly	0	0	10	8	12	0	0	4.066
11.	The information provided by this system is clear	5	10	10	5	0	0	0	2.500
12.	It is easy to find any necessary information	0	5	5	5	15	0	0	4.000
13.	The information provided by this system is easily understandable	25	5	0	0	0	0	0	1.167
14.	The information provided effectively helps me to complete a task on this system	0	0	7	8	15	0	0	4.300
15.	The layout of information on this system is clear	0	0	11	9	10	0	0	3.967
16.	The system display makes it easy for me to use	0	15	15	0	0	0	0	2.500
17.	I like the design of this system	0	0	10	5	15	0	0	4.167
18.	All the functions in this system are in line with my expectations	20	10	0	0	0	0	0	1.333
19.	Overall, I am satisfied with the performance this system	27	3	0	0	0	0	0	1.100

Table 15: Results of the usability test for the ANP method
(Answers given by each individual)

User No.	Question No.																			Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
1	3	2	2	1	3	2	1	2	3	5	4	5	2	5	5	3	5	2	2	55
2	1	1	1	1	1	1	1	1	1	3	1	2	1	5	5	3	5	2	2	36
3	1	1	1	1	1	1	1	1	1	3	3	5	2	5	5	3	5	2	2	42
4	2	2	2	1	3	2	1	1	3	5	4	5	2	5	5	3	5	2	1	53
5	1	1	1	1	1	1	1	1	1	3	1	2	1	5	5	3	5	2	1	36
6	2	2	2	1	2	1	1	1	3	5	4	5	2	5	5	3	5	2	1	51
7	2	1	2	1	3	1	1	1	3	5	4	5	2	5	5	3	5	2	1	51
8	1	1	1	1	1	1	1	1	2	5	4	5	1	3	3	2	4	1	1	38
9	1	1	1	1	1	1	1	1	1	3	1	2	1	3	3	2	4	1	1	29
10	2	1	2	1	2	1	1	1	3	5	3	5	1	3	3	2	4	1	1	41
11	1	1	1	1	1	1	1	1	1	3	1	2	1	5	5	3	5	2	1	36
12	1	1	1	1	1	1	1	1	2	5	3	5	1	4	4	3	5	2	1	42
13	2	1	1	1	2	1	1	1	3	5	3	5	1	4	4	3	5	2	1	45
14	2	1	1	1	2	1	1	1	3	5	3	5	1	3	3	2	3	1	1	39
15	1	1	1	1	1	1	1	1	1	3	1	2	1	4	4	3	5	1	1	33
16	1	1	1	1	1	1	1	1	2	5	3	5	1	3	3	2	3	1	1	36
17	1	1	1	1	1	1	1	1	2	5	3	5	1	3	3	2	3	1	1	36
18	1	1	1	1	1	1	1	1	2	5	3	5	1	3	3	2	5	1	1	38
19	1	1	1	1	1	1	1	1	1	3	2	4	1	5	4	2	3	1	1	34
20	1	1	1	1	1	1	1	1	2	4	2	4	1	5	5	3	5	1	1	40
21	1	1	2	1	2	1	1	1	3	4	2	4	1	5	5	3	5	1	1	43
22	1	1	2	1	2	1	1	1	2	4	2	4	1	5	4	2	3	1	1	38
23	1	1	1	1	2	1	1	1	2	4	2	3	1	4	4	3	4	1	1	37
24	1	1	1	1	1	1	1	1	1	3	2	4	1	5	4	2	3	1	1	34
25	1	1	1	1	1	1	1	1	2	4	3	5	1	5	4	2	3	1	1	38
26	1	1	1	1	1	1	1	1	1	3	2	3	1	4	3	2	3	1	1	31
27	1	1	1	1	1	1	1	1	1	3	2	3	1	4	3	2	4	1	1	32
28	1	1	2	1	2	1	1	1	2	4	2	3	1	4	3	2	3	1	1	35
29	1	1	2	1	2	1	1	1	2	4	3	5	1	5	4	3	5	1	1	43
30	1	1	2	1	2	1	1	1	2	4	2	3	1	5	3	2	3	1	1	36

References

Bottomley P.A., Doyle J.R., Green R.H. (2000), *Testing the Reliability of Weight Elicitation Methods: Direct Rating versus Point Allocation*, Journal of Marketing Research, 37(4), 508-513.

Camciuc A., Gheorghiu H., Constantin N., Gavan M. (2005), *Aspects Concerning the Implementation of the Probabilistic Approach in Evaluating the Reliability of the Structures*, Transilvania University Press of Braşov.

Cronbach L.J. (1951), *Coefficient Alpha and the Internal Structure of Tests*, Psychometrika, 16(3), 297-334.

Diomidis S. (2006), *Code Quality: The Open-Source Perspective* (Effective Software Development Series), Adobe Press, Hoboken, NJ.

Engels G., Lohmann M., Wagner A. (2006), *The Web Application Development Process* [in:] G. Kappel, B. Proll, S. Reich, W. Retschitzegger (eds.), *Web Engineering: The Discipline of Systematic Development of Web Applications*, John Wiley & Sons, England.

Goodwin P., Wright G. (1998), *Decision Analysis for Management Judgment,* John Wiley & Sons, Hoboken, NJ.

Herbert A.J., Jones K.I.S. (2004), *Computer Systems: Theory, Technology, and Applications* (Vol. 1), Springer Science & Business Media.

Keeney R.L., Raiffa H. (1976), *Decison with Multiple Objectives: Preferences and Value Tradeoffs*, Wiley, New York.

Lewis J.R. (1995), *IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use*, International Journal of Human-Computer Interaction, 7(1), 57-78.

Olsina L., Lafuente G., Rossi G. (2001), *Specifying Quality Characteristics and Attributes for Websites* [in:] *Web Engineering: Managing Diversity and Complexity of Web Application Development*, Springer, Berlin, Heidelberg, 266-278.

Olsina L., Rossi G. (2002), *Measuring Web Application Quality with WebQEM*, IEEE Multimedia, 9(4), 20-29.

Perlman G. (2009), *User Interface Usability Evaluation with Web-Based Questionnaire*, http://hcibib.org/perlman/questiOn.html.

Ramsey D.M. (2020), *A Game Theoretic Model of Choosing a Valuable Good via a Short List Heuristic*, Mathematics, 8(2), 199.

Rouse M.W., Borsting E., Hyman L., Hussein M., Cotter S.A., Flynn M., Scheiman M., Gallaway M., De Land P.N. (1999), *Frequency of Convergence Insufficiency among Fifth and Sixth Graders*, Optometry and Vision Science, 76(9), 643-649.

Saaty T.L. (1987), *The Analytic Hierarchy Process − What It Is and How It Is Used*, Mathematical Modelling, 9(3-5), 161-176.

Saaty T.L. (2003), *Decision-making with the AHP: Why Is the Principal Eigenvector Necessary?* European Journal of Operational Research, 145(1), 85-91.

Saaty T.L. (2007), *Time-dependent Decision-making. Dynamic Priorities in the AHP/ANP: Generalizing from Points to Functions and from Real to Complex Variables*, Mathematical and Computer Modelling, 46(7-8), 860-891.

Spinellis D. (2006), *Code Quality: The Open Source Perspective*, Adobe Press, Hoboken, NJ.

Subraya B.M., ed. (2006), *Integrated Approach to Web Performance Testing: A Practitioner's Guide*, IGI Global, Hershey, PA.

Sugiyono (2009), *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif, dan R&D*, Alfabeta, Bandung.

Suryadi K., Ramdhani M.A. (2002), *Sistem Pendukung Keputusan: Suatu Wacana Struktural Idealisasi dan Implementasi*, Remaja Rosdakarya, Bandung.

Teeravarunyou S., Poopatb B. (2009), *Computer-based Welding Training System*, International Journal of Industrial Engineering, 16(2), 116-125.

Wachowicz T., Roszkowska E. (2023), *How Well May the Direct Linguistic Declarations Substitute AHP in Defining Accurate Criteria Weights?* International Transactions in Operational Research, Wiley.

Yulianti E. (2015), *Sistem Pendukung Keputusan Pemilihan Mobil Dengan Metoda Simple Multi Attribute Rating (SMART)*, Jurnal Momentum, 17(1).

Yunitarini R. (2013), *Sistem Pendukung Keputusan Pemilihan Penyiar Radio Terbaik*, Jurnal Mikrotek, 1(1), 43-52.

Zlaugotne B., Zihare L., Balode L., Kalnbalkite A., Khabdullin A., Blumberga D. (2020), *Multi-Criteria Decision Analysis Methods Comparison*, Environmental and Climate Technologies, 24(1), 454-471.