

Building an Expert System through Machine Learning for Predicting the Quality of a WEB Site Based on Its Completion

[Sastry Kodanda Rama Jammalamadaka](#)^{*}, Vishnu Priya B , Sasi Bhanu Jammalamadaka , [Bhupati CH](#) , Kamesh Balakrishna Duvvuri , [Raja Rao Budaraju](#)

Posted Date: 2 August 2023

doi: 10.20944/preprints202308.0161.v1

Keywords: Quality of website; Completeness of a website; Model learning to predict; Expert System; predicting the Quality of a website



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Building an Expert System through Machine Learning for Predicting the Quality of a WEB Site Based on Its Completion

B Vishnu Priya ¹, JKR Sastry ^{1,*}, J Sashi Bhanu ², Bhupati Ch ¹, DBK Kamesh ³ and Raja Rao Budaraju ⁴

¹ K L Deemed to be University, Vaddeswaram, Guntur

² CMR College of engineering technology, Hyderabad

³ Malla Reddy Engineering College for women, Secunderabad

⁴ Oracle America Inc

* Correspondence: drsastry@kluniversity.in

Abstract: The main channel for disseminating information is now the Internet. Users have different expectations for the calibre of websites regarding the posted and presented content. The WEB site's quality is influenced by up to 120 factors, each represented by two to fifteen attributes. A major challenge is quantifying the features and evaluating the quality of a website based on the feature counts. One of the aspects that determines a website's quality is its completeness, which focuses on the existence of all the objects and their connections with one another. It isn't easy to build an expert model based on feature counts to evaluate website quality, so this paper has focused on that challenge. Both a methodology for calculating a website's quality and a parser-based approach for measuring feature counts are offered. We provide a multi-layer perceptron model that is an expert model for forecasting website quality from the "Completeness" perspective. The accuracy of the predictions is 98%, whilst the accuracy of the nearest model is 87%.

Keywords: quality of website; Completeness of a website; Model learning to predict; expert system; predicting the Quality of a website

1. Introduction

Information is widely disseminated through websites. When it comes to providing users with high-quality material, the Quality of the websites is the most crucial concern. By considering every factor impacting the Quality of the website, more precise and pertinent information may be made available to the consumers. To assess the actual condition of the website, it is crucial to identify the variables that affect website quality, utilizing models that aid in quantitative website quality computation.

More than 42 parameters [1, 2] are utilized to calculate the website quality. A developed model is necessary for calculating a Sub-factor's Quality. To calculate the Quality of a factor, the Quality of all the sub-factors must be added and averaged.

Everyone now engages in the exchange of information via websites. The website's content is available in various formats, including sound, photographs, movies, audio, graphics, etc. Different material types are posted using multiple forms. Numerous new websites are launched every day. The real issue is the calibre of these websites. People can only force themselves to depend on the need for such content High Quality. The most challenging task is evaluating Quality. Customer happiness depends on how well-designed the website is that is searched or browsed for information.

For evaluating the Quality of websites, numerous aspects must be considered [3]. Among the factors are usability, connectivity, navigation, structure, safety, maintainability, dependability, functionality, privacy, portability, etc. Various tools, approaches, models, and procedures must be used to evaluate the calibre of a website. Most website businesses need many elements, such as a

logo, animated graphics, colour schemes, mouse-over effects, graphic art, database communication, and other requirements.

To assess a website's Quality, it is necessary to employ several characteristics [4], either directly connected or unrelated. Some techniques used to evaluate the Quality of websites are subjective and rely on user input. Statistical methods, usually called objective measurements, can be used to measure quality factors.

Different stakeholders have various quality standards. Other actors evaluate the Quality from various perspectives, including designers, management, consumers, developers, etc. End users think about usability, efficiency, creditability, and other things, whereas programmers think about maintainability, security, functionality, etc. Only after determining a website's demands from an actor's viewpoint can the quality criteria best satisfy those expectations be obtained.

Websites are becoming more complicated for those concerned with e-commerce, electronics, museums, etc. Due to the availability of more components and the interactions between the parameters choosing relevant quality factors and building quality evaluation procedures to quantify the variables is becoming more difficult. The relationship between traits, attributes, and websites must be considered. Websites must have an impartial, well-considered subjective inclusion and an objective evaluation. Given the criteria that necessitate those approaches for the evaluation, it is imperative to evaluate both procedures.

WEB sites have become the most important mechanisms to disseminate information on different aspects the user focuses on. However, the reliability of the information posted on the websites must be verified. The Quality of WEB sites is always the question that users want to know so that the dependability of the information posted on the site can be assessed.

Most of the assessment of the Quality of the WEB site is based on human-computer interaction through usability heuristics **Nelson et al.** [5], design principles **Tognazzini et al.** [6], and rules **Schneiderman et al.** [7]. Evaluating the Quality of a website has taken different dimensions. As such, the formal framework has yet to be arrived at, even though several quality standards have been prescribed [Law et al. [8], **Semeradova and Heinrich et al.** [9].

A website's quality has been described in various ways by different people. Most recently, many metrics have been used to determine a website's capacity to live up to its owners' and visitors' expectations [10] **Morale Vegas et al.** There is no common strategy because every solution suggested in the literature uses distinct evaluation techniques **Law et al.**, [11]. Evaluating the quality of a website is different from doing so for other systems. The website's quality is evaluated using multi-dimensional modelling [12] **Ecer et al.**

The Quality of a website is computed from three different perspectives, which include functional **Luieng et al.** [13], strategic **Maia and Furtado** [14] and experiential **Sambre et al.** [15]. All the methods focus on some measures, attributes, characteristics, dimensions etc. The methods are synonyms and explore distinctive features of the websites.

Different methodologies are used to compute the website's Quality, including experimental-quasi-experimental, descriptive-observational, associative, correlative, qualitative-qualitative, and subjective-objective; the methods are either participative (surveys, checklists) or non-participative (web analytics). The participative methods focus on user preferences and psychological responses **Bevan et al.**, [16]. The testing technique is frequently employed for computing quality indicators, including usability tests, A/B tests, ethnographic, think-aloud, diary studies, questioners, surveys, checklists, and interviews, to name a few **Rousala and Karause** [17].

The most recent method uses an expert system involving heuristic evaluation **Jainari et al.** [18]. Experts judge the Quality of a website concerning a chosen set of website features. The expert's methods are evaluated manually or through an automated software system. Few Expert systems also are developed using Artificial intelligence techniques **Jayanthi and Krishna Kumari** [19], Natural Language Processing **Nicolic et al.** [20]

According to a survey by **Morales-Vargas et al.** [21], one of the three dimensions—strategic, functional, and experimental—is used to judge the quality of websites. The necessary method for evaluating the quality of the website is expert analysis. They have looked at a website's

qualities and categorised them into variables used to calculate the website's quality. Different approaches have been put out to aid in quantifying the quality of a website.

Research Questions

1. What are the Features that together determine the Completeness of a WEB site?
2. How are the features Computed?
3. How did the features and the website's degree of excellence relate to one another
4. How do we predict the Quality of a WEB site given the code?

Research outcomes

1. This research presents a parser that computes the counts of different features given to a WEB site.
2. A model can be used to compute the Quality of a website based on the feature counts.
3. An example set is developed considering the code related to 100 WEB sites. Each website is represented as a set of features with associated counts, and the website quality is assessed through a quality model.
4. A Multi-layer perceptron-based model is presented to learn the Quality of a website based on the feature counts, which can be used to predict the Quality of the WEB site given the feature counts computed through a parser model.

2. Related work

By considering several elements, including security, usability, sufficiency, and appearance, **Fiaz Khawaja1 et al. [22]** evaluated the website's Quality. A good website is simple to use and offers the learning opportunity. A website's Quality increases when it is used more frequently. When consumers learn from high-quality websites, their experience might be rich and worthwhile. The factor "Appearance" describes how a website looks and feels, including how appealing it is, how items and colours are arranged, how information is organized meaningfully, etc. A method for calculating website quality based on user observations made while the website is being used has been presented by Kausar Fiaz Khawaja1.

Flexibility, safety, and usability are just a few of the elements that **Sastry et al. [23]** and Vijay Kumar Mantri et al. [24] considered while determining the website's Quality. "Usability" refers to the website's usefulness, enjoyment, and efficacy. The user presence linked to the website or browsing is called the "Safety" factor. There should never be public access to the user's connection to the website. The "Flexibility" aspect is connected to the capability included in a website's design that enables adjustments to the website even while it is being used. Users can assess a website's quality using PoDQA (Portal Data Quality Assessment Tool), which uses pre-set criteria.

Vassilis S. Moustakis et al.'s [25] presented that the assessment of the WEB quality needs to consider several elements, including navigation, content, structure, multimedia, appearance, and originality. The term "content" described the data published on a website and made accessible to visitors via an interface. The "content" quality factor describes how general and specialised a domain is fully possible.

Navigation refers to the aid created and offered to the user to assist in navigating the website. The ease of navigating a website, the accessibility of associated links, and its simplicity all affect its navigational Quality. The "Structure" quality aspect has to do with things like accessibility, organization of the content, and speed. The appearance and application of various multimedia and graphics forms can impact a website's overall feel and aesthetic. A website could be developed with a variety of styles in mind. A website's "Uniqueness" relates to how distinct it is from other comparable websites. A high-quality website must be unique, and users frequent such websites frequently. Vassilis proposed a technique known as AHP (Analytical Hierarchical Process), and they utilized it to determine the website's Quality. Numerous additional factors must also be considered to determine the website's Quality. Andrina Graniü et al.'s [26] assessment of the "portability" of

content—the capacity to move it from one site to another without requiring any adjustments on either end—has considered this.

Tanya Singh et al. [27] have presented an evaluation system that considers a variety of variables, such as appearance, sufficiency, security, and privacy. They portrayed these elements in their literal sense. A website's usability should be calculated to represent its Quality in terms of how easy it is to use and how much one can learn from it. The usability of a website refers to how easily users can utilize it. To a concerned user, some information published on the website can be private. The relevant information was made available to the qualified website users.

The exactness with which a user's privacy is preserved is the attribute "privacy"-related Quality. Only those who have been verified and authorized should have access to the content. Users' information communicated with websites must be protected to prevent loss or corruption while in transit. The security level used during the data exchange can be used to evaluate the website's Quality.

The "Adequacy" factor, which deals with the correctness and completeness of the content hosted on the website, was also considered. Anusha et al. [28] evaluated similar traits while determining the websites' quality. The most important factor they considered was "Portability." This is called portability when content and code may be transferred from one machine to another without being modified or prepared for the target machine. Another key element of websites is the dependability of the content users see each time they launch a browser window. When a user hits on a certain link on a website, it must always display the same content unless the content is continually changing. The dependability of a website is determined by the possibility that the desired page won't be accessible.

When improvements to the website are required, maintenance of the existing TA website is straightforward. The ease of website maintenance affects the Quality of the website. When evaluating the Quality of a website, taking the aspect of "maintainability" into account, several factors, such as changeability, testability, analysability, etc., must be considered. The capacity to modify the website while it is active is a crucial factor to consider regarding maintainability.

The website's capacity to be analysed is another crucial aspect that should be considered when evaluating a website's Quality. The ability to read the information, relate the content, comprehend it, and locate and identify the website's navigational paths all fall under a website's analysability category. When there is no unfinished business, changes to the user interface, and no disorganised presentation of the display, a website can be called to be stable. The reliable website stays mostly the same. When designing a website, the problem of testing must be considered. While the website is being used, updates should be able to be made. Nowadays, most websites still need to provide this feature.

Filippo Ricca et al. [29] have considered numerous other parameters to calculate website quality. The website's design, organization, user-friendliness, and organizational structure are among the elements considered. Web pages and their interlinking are part of the website's organization. The practical accessibility of the web pages is directly impacted by how they are linked. When creating websites, it is essential to consider user preferences. It's necessary to render the anticipated content.

According to Saleh Alwahaishi et al. [30], the levels at which the content is created and the playfulness with which the content is accessible are the two most important factors to consider when evaluating a website's quality. Although they have established the structure, more sophisticated computational methods are still required to evaluate the websites' quality. They contend that a broad criterion is required to assess the value of any website that provides any services or material supported by the website. The many elements influencing a website's quality have been discussed. In their presentation, Layla Hasan and Emad Abuelrub [31] stressed that WEB designers and developers should consider these factors in addition to quality indicators and checklists.

The amount of information being shared over the Internet is frighteningly rising. Websites and web apps have expanded quickly. Websites must be of a high standard if they are often utilised for getting the information required for various purposes. Kavindra Kumar Singh et al. [32] have created a methodology known as WebQEM (Web Quality Evaluation Method) for computing the quality of websites based on objective evaluation. However, judging the quality of websites based on subjective

evaluation might be more accurate. They have quantitatively evaluated the website's quality based on an impartial review. They included the qualities, characteristics, and sub-characteristics in their method.

People communicate with one another online, especially on social media sites. It has become essential that these social media platforms offer top-notch user experiences. Long-Sheng Chen et al.'s [33] attempt to define the standards for judging social media site quality. They used feature selection methodologies to determine the quality of social networking sites. Metric evolution is required for the calculation of website quality.

According to metrics provided by Naw Lay Wah et al. [34], website usability was calculated using sixteen parameters, including the number of total pages, the proportion of words in body text, the number of text links, the overall website size in bytes, etc. Support vectors were used to predict the quality of web pages.

Sastry JKR et al. [35] assert that various factors determine a product's quality, including content, structure, navigation, multimedia, and usability. To provide new viewpoints for website evaluation, website quality is discussed from several angles [36–49]. However, none of the research has addressed how to gauge the websites' content's comprehensiveness regarding quality.

A hierarchical framework of several elements that allow quality assessment of WEB sites has been developed by Vassilis S. Moustakis et al. [50]. The framework considers the features and characteristics of an organisation. The Model does not account for computation measurements of either factors or sub-factors.

Oon-it, S [51] conducted a survey to find from the users the factors they consider reflect the Quality of health-related WEB sites. The users have opined that trust in health websites will increase when it is perceived that the Quality of the content hosted on web site is high considering the correctness, simplicity etc. **Allison R et al. [52]** have reviewed the existing methodologies and techniques relating to evaluating the Quality of web site and then presented a framework of factors and attributes. No computational methods have been covered in this presentation.

Barnes, Stuart et al. [53] conducted a questionnaire survey. He developed a method of computing the Quality based on responses by various participants' computed t-score. Based on the t-score, the questions were classified, with each class representing the WEB site quality.

Laayla et al. [54] have attempted to assess the Quality of the WEB sites based on the Quality of the services rendered through the WEB site. They have presented standard dimensions based on services rendered by the websites. The measurements they consider include accuracy, content, value-added content, reliability, ease of use, availability, speed downloading, customisation, effective internal search, s security, service support, and privacy. They collected criteria to measure each factor and then presented standards combining all the requirements to form an overall quality assessment. Combining different metrics into a single metric does not reflect the excellent quality of the website.

Sasi Bhanu et al. [55] have presented a manually defined expert model for assessing the WEB site quality. The factors related to "Completeness of a website have also been measured manually. The computations have been reflected considering a single WEB site. No Example set is used. The accuracy of prediction varied from website to website.

Rim Rekik et al. [56] have used a text mining technique to extract features of WEB sites and used a reduction technique to filter out the most relevant features of a website. An example set is constructed. Considering different sites. They have applied Apriori Algorithm to the example set to find associations between the criteria and find frequent measures. The Quality of the WEB site is computed considering the applicability of the association rule given a WEB site.

Several soft computing models have been used for computing the Quality of a Web site which include Fuzzy-hierarchical [57], Fuzzy Linguistic [[58][59], Fuzzy-E-MEANS [60], Bayesian [61], Fuzzy-neutral [62], **SVM** [63], Genetic Algorithm [64]. Most of these techniques focused on filtering out the website features and creating the example sets mined to predict the overall website quality. No generalized model is learned, which can be used for indicating the Quality of a WEB site given a new website.

Michal Kakol et al. [65] have presented a predictive model to find the credibility of content hosted on the website based on human evaluations considering the comprehensive set of independent factors. The classification is limited to a binary classification of the content. But scaled prediction is the need of the hour.

The expert systems referred to in the literature depend on Human beings who are considered experts and through experimentation which is a complicated process. No models as such existed that attempted to compute the website's features give the WEB application code. No model that learns the website quality is presented based on the measurement of website elements.

3. Preparing the Example set

The Quality of a website is computed considering 100 websites. The parameters considered include the Number of missing Images, videos, PDFs, Fields in the tables, columns in the forms and missing self-references. The WEB site quality is computed on a scale of 0.0 to 1.0. The grading of the WEB site is done considering as shown in **Table 1**:

Table 1. Quality Grading Table.

Starting Quality computed value	Ending Quality computed value	Quality Grading
0.81	1.0	Excellent
0.61	0.80	Very Good
0.41	0.60	Good
0.40	0.40	Average
0.00	0.39	Poor

The quality value of every parameter is computed based on the number of missing counts, which are calculated through a parser using the algorithm explained in the subsequent section. A sample computation is shown for each parameter-wise in **Table 2**.

Table 2. Sample Quality computation based on the counts computed through the parser.

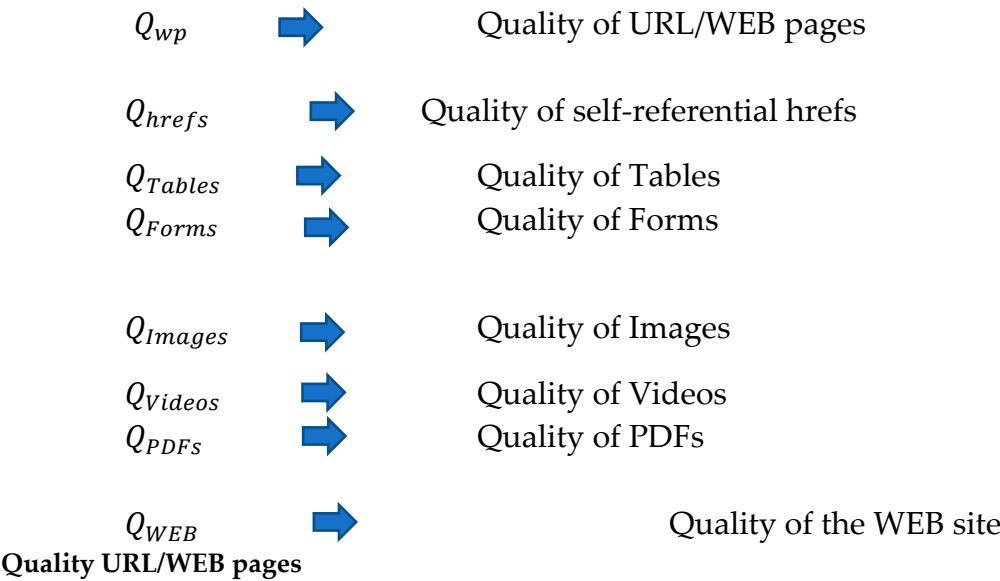
Missing Counts	0	1	2	3	4	Quality Value assigned
Quality Value	1.0	0.8	0.6	0.40	0	
Missing Images	4				✓	0.00
Missing Videos	3		✓			0.60
Missing PDFs	9				✓	0.00
Missing Columns in Tables	1	✓				0.80
Missing Fields in the forms	1	✓				0.80
Missing self-references	1	✓				0.80
The total Quality value assigned						3.00
Average quality value						0.50
Quality grade as per the Grading table above						Good

4. Methods and Techniques

4.1. Analysis of sub-factors relating to the factor "Completeness".

The information must be complete for consumers to understand the meaning of the content hosted on a website. The completion of the content that is hosted on the websites depends on a number of factors. Missing URLs or web pages, self-referential “hrefs” on the same page, tabular columns in tables, data items in input-output forms, context-appropriate menus, missing images, videos, and audios are just a few examples of the characteristics of the attribute "Completeness" that are taken into account and evaluated to determine the overall quality of the websites.

Connections, Usability, Structure, Maintainability, Navigation, Safety, functionality, Portability, Privacy, etc., are just a few of the numerous aspects of evaluating a website's quality. Each element has several related qualities or factors. Important supporting variables and features, such as the following, relate to completeness.



When evaluating the effectiveness of online sites, the number of missing pages is considered. If a “href” is present, but the matching WEB page is not in line with the directory structure indicated by the URL, the URL is deemed missing. The quantity of references missing from all web pages affects how well this sub-factor performs.

Quality of self-referential hrefs

“hrefs” can provide navigation in an HTML page, especially if the page size is huge. Self-refs are used to implement navigation on the same page. Self-hrefs are frequently programmed on the same page, but programme control is not transferred to those sub-refs. Additionally, although sub-herf is not defined in the code, control is still transferred. The quantity of missing sub-hrefs (not coded) or sub-hrefs that are coded but not referred to is used to determine the quality of this sub-factor.

Quality of Tables

Sometimes content is delivered by displaying tables. A repository of the data items is created, and linkages of those data items to different tables are established. The tables in the web pages must correspond to the Tables in the database, The columns code for defining a table must correspond to the Table attributes represented as columns. The number of discrepancies between HTML tables and RDBMS tables reveals the HTML tables' quality.

Quality of forms

Sometimes data is collected from the user using forms within the HTML pages. The forms are designed using attributes for which the data must be collected. The forms need to be better designed, not having the relevant fields or when no connection exists between the fields for which data is collected. The number of mismatches reveals the Quality of such forms coded into the website.

Missing images

The content displayed on websites is occasionally enhanced by incorporating multimedia-based items. The image quality needs to be improved. Some important considerations include the image's size and resolution. Every HTML page uses its URL to link to the images and normally specifies the size and resolution. Blank images are displayed when the images a "href" refers to don't exist. Second,

the attributes of the actual image must correspond to the HTML page's image code specifications. The number of images on each page may be counted, and each image's details, including its dimensions, resolution, and URL, can then be found. The quantity of blank or incorrectly sized or resized images can be used to gauge the quality of the images.

Missing videos

Video is another form of content found on websites with interactive content. Users typically have a strong desire to learn from online videos. The availability of the videos and the calibre of the video playback are the two sub-factors that impact video quality the most. An HTML or script file that mentions the videos contains the URL and player that must be used locally to play the videos.

The position, width, number of frames displayed per second, and other details are used to identify the videos. There must be a match between the movies' coded size and the videos' actual size stored at the URL location. Another crucial factor for determining the quality of the videos is whether they are accessible at the site mentioned in the code. The properties of the video, if it exists, can be checked. The HTML pages can be examined, and their existence at the supplied URL can be confirmed. The quantity of missing or mismatched videos reflects the quality of the websites.

Missing PDFs:

Consolidated, important, and related content is typically rendered using PDFs. Most websites offer the option to download PDFs. The material kept in PDFs is occasionally created by referring to the PDFs using hrefs. The quantity of missing PDFs can be used to calculate the quality of the PDFs.

4.2. Total Quality considering the factor "completeness."

When assessing the quality of a website, the component "Completeness" and its attributes, which are generated as an average of the Quality of the individual criteria, are considered. The more development is made regarding a characteristic, the more a feature's quality is diminished. When the "completeness" factor is considered, the corresponding sub-factors quality defines a website's overall quality. One sub-factors quality could be top-notch, while another could be below average. The sub-factors are independent as a result. Due to the equal importance of each sub-factor, there are no assigned weights. In equation (1), a simple measure of the factor's Quality may be the average quality of all the sub-factors.

$$Q_{WEB} = (Q_{wp} + Q_{hrefs} + Q_{Tables} + Q_{Forms} + Q_{Images} + Q_{Videos} + Q_{PDFs})/7 \quad (1)$$

4.3. Computing the counts of sub-factors through a parser.

A parser is developed which scans through the entire WEB site given as an example. The WEB site, as such, is a Hierarchical structure of files stored in the disk drive. An example of such a structure is shown in **Figure 1**. As such, the URL of the root of the structure is given as the beginning of the WEB site, which is taken as input by the parser.

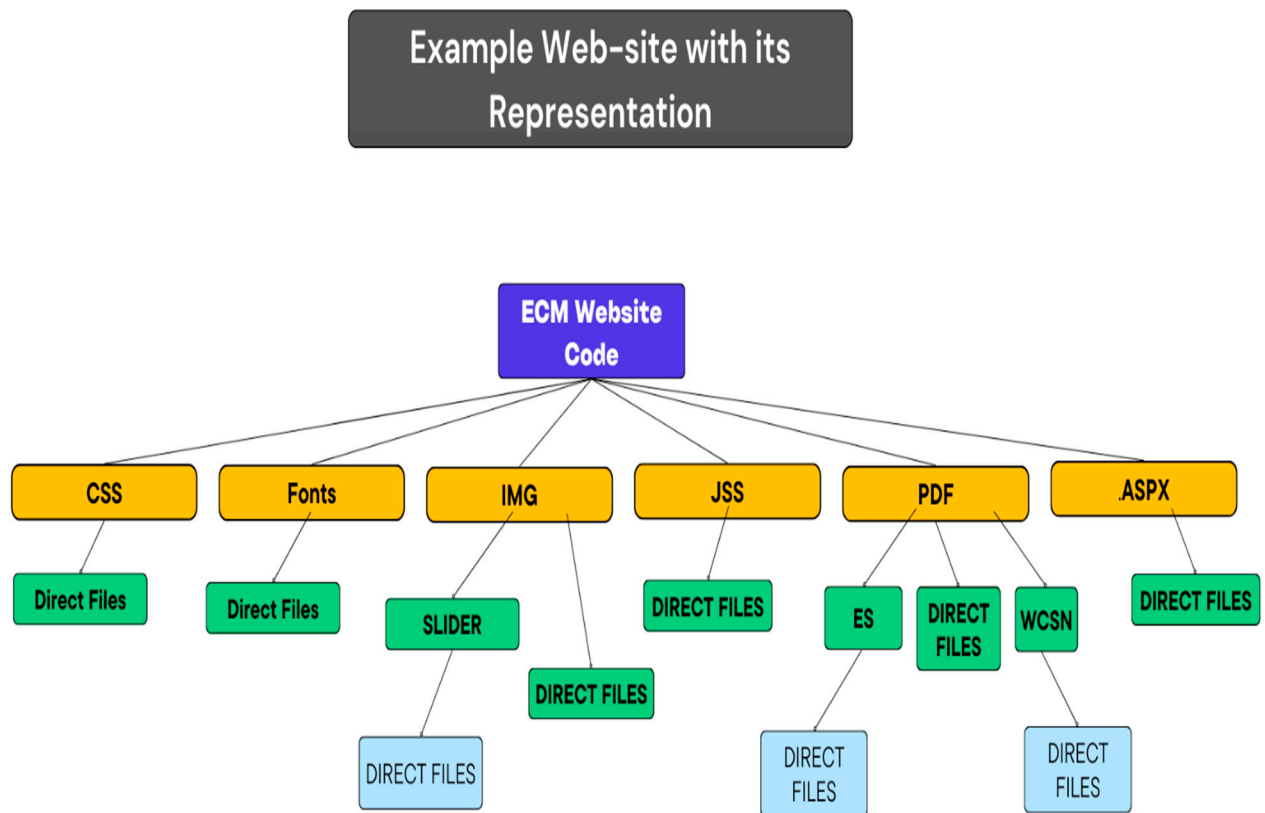


Figure 1. An example structure of a WEB site.

4.4. An algorithm for computing object counts.

The following algorithm represents a parser designed to compute the counts related to different objects described above. The counts include total referred objects, Total objects in existence, total number of missing objects and Total number of mismatched objects. This algorithm is used to build the example set and compute counts related to a new website using the predicted WEB site quality.

Input: WEB site structure

Outputs:

1. The list of all files
2. Number of the total, Existing and Missing Images,
3. Number of the capacity, Existing and Missing videos,
4. Number of the total, Existing and Missing PDFs,
5. Number of the total, Existing and Missing Fields in the tables,
6. Number of the total, Existing and Missing columns in the forms a
7. Number of total, Existing and Missing self-references

Procedure

1. Scan through the files in the structure and find the URLs of the Code files.

2. For each of the code file
 - a. Check for URLs of PDFS; if it exists, enter them into a PDF_URL array.
 - b. Check for URLs of Images; enter them into an Image URL array if it exists.
 - c. Check for URLs of Videos, and if it exists, enter a Video URL array.
 - d. Check for URLs of inner pages; if it exists, enter them into an inner-pages _URL array!
 - e. Check for the existence of tables and enter the description of the table as a string into an array.
 - f. Check for the existence of forms and enter the description of the forms as a string into an array.
3. For each entry in the PDF-Array
 - a. Add to Total-PDFS
 - b. Check for the existence of the PDF file using the URL.
 - i. If available, add to Existing-PDFS, Else add to Missing-PDFS.
4. For each entry in the Image-Array
 - a. Add to Total-Images
 - b. Check for the existence of the Image file using the URL.
 - i. If available, add to Existing-Images else, add to Missing Images.
5. For each entry in the Video-Array
 - a. Add to Total-Videos
 - b. Check for the existence of the Video file using the URL.
 - i. If available, add to Existing-Videos else, add to Missing- Videos.
6. For each entry in the Video-Array
 - a. Add to Total-Videos
 - b. Check for the existence of the Video file using the URL.
 - i. If available, add to Existing-Videos else, add to Missing- Videos.
7. For each entry in the inner-URL-Array
 - a. Add to Total-Inner-URLS
 - b. Check for the Existence of the Inner-URL.
 - i. If available, add to Existing-Inner-URLS, Else add to Missing-Inner-URLS
8. For each entry in Table-Desc-Array
 - a. Fetch the column's names in each of the entries.
 - b. Fetch the tables having the column names as Table fields.
 - c. If the column and file names are the same, add to Total-Tables and Existing Tables, And add to Missing Tables.
9. For each entry in Field-Desc-Array
 - a. Fetch the column's names in each of the entries.
 - b. Fetch the Forms having the field names as Form fields.
 - c. If the field name and the filed names are the same, add to Total-Forms and Existing Forms; else, add to Missing Tables.
10. Write all the counts in a CSV file.
11. Write all names of the program files to a CSV file.

12. Write all the Images URLs with associated properties to a CSV file
13. Write all the Video URLs with associated properties to a CSV file
14. Write all names of the PDF files into a CSV file.

The WEB sites are input to the parser, which computes the counts of different objects explained above.

4.5. Designing the Model Parameters.

A multi-Layer perceptron Neural network has been designed, which represents an expert model. The Model is used to predict the Quality of a website given the counts computed through the parser. **Table 3** details the design parameters used for building a NN network.

Table 3. Multi-Layer Perceptron Model parameters.

Type of Layer	Number of Inputs	Number of outputs	Type of activation function used	Type of kernel_INITIALIZER
Input Layer	6	6	RELU	Normal
Output layer	6	5	SIGMOID	Normal
Model Parameters	Loss Function	optimizers	Metrics	
	Cross Entropy	Adams	Accuracy	

4.6. Platform for Implementing the Model.

KERAS and TensorFlow have been used in the notebook interface run on Microsoft 11 operating system. The software system runs DELL LAPTOP, built on top of the 12th Generation with 8 cores and 2 Nvidia GPUS.

4.7. Inputting the Data into the Model.

The example set is broken into batches, each of 10 Examples, and the records are shuffled in each epoch. Each split is converted to a batch of 5 for updating the weights and the bios.

5. Results and Discussion

5.1. Sub-Factor counts, for example, WEB sites.

Table 4 shows the counts obtained for 10 WEB sites calculated using the parser explained in the above section.

Table 4. Counts and grading, for example, websites.

ID	# missing Images	# Missing Videos	# Missing PDFS	# Missing Tables	# Missing forms	# Missing internal Hrefs	Quality of the WEB site
1	4	3	9	1	1	1	average
2	1	2	0	1	0	0	very good
3	2	4	1	2	1	1	good

4	2	2	1	2	1	2	very good
5	2	3	1	2	1	3	good
6	1	2	1	2	1	4	average
7	1	1	1	1	1	1	very good
8	2	2	2	2	2	2	good
9	3	3	3	3	3	3	average
10	4	4	4	4	4	4	poor

5.2. Weight computations for the NN model

The weights learnt for achieving optimum accuracy are shown in **Table 5** while fixing the bias to be zero for each output. The number of Epochs has been varied from 100 to 1000, and the batch size from 2 to 10, and it has been found that optimum results have been obtained when 300 epochs and batch size = 5 have been used.

Table 5. Learned weights for optimum accuracy.

Weight Code	Weight Value	Wight Code	Weight Value
W111	0.0005345	W121	-0.03049852
W112	-0.02260396	W122	-0.05772249
W113	0.10015791	W123	0.0124933
W114	-0.00957603	W124	0.05205087
W115	0.0110722	W125	-0.02575279
W116	-0.07497691]	W126	0.06270903
W131	0.03905119	W141	[-0.02733616
W132	0.04710016	W142	0.02808586
W133	-0.01612358	W143	-0.03189966
W134	-0.00248795	W144	0.07678819
W135	-0.06121466	W145	-0.05594458
W136	-0.0188451	W146	-0.04489214]
W151	0.000643	W161	0.02465009
W152	0.0143626	W162	0.02291734
W153	-0.00590346	W163	0.06510213
W154	-0.05017151	W164	0.0216038
W155	0.00431764	W165	0.02364654
W156	-0.04996827	W166	0.04817796
W211	0.00150367	W221	0.05486471
W212	-0.02436988	W222	-0.0747726
W213	-0.04478416	W223	-0.03751294
W214	-0.0215895	W224	-0.00753696
W215	-0.01126576	W225	-0.16550754
W231	-0.02882431	W241	-0.0697467
W232	0.09704491	W242	-0.00334867
W233	0.00701219	W243	0.00892285

Weight Code	Weight Value	Wight Code	Weight Value
W234	0.05021231	W244	0.08749642
W235	-0.12358224	W245	0.08793346
W251	-0.0181542	W261	-0.06405859
W252	-0.09880255	W262	-0.07070417
W253	-0.00041602	W263	0.01609092
W254	0.02695975	W264	0.00031056
W255	-0.03195139	W265	-0.10547637

5.3. Accuracy Comparisons

The average accuracy of the MLP model is 98.2%, with a standard deviation of 0.2%. Jayanthi et al. [66] have provided a comparison of the accuracy of their Extreme Learning Machine (ELM) method with the accuracy of SVM [67] and the Ivory method [68]. Table 6 shows the comparison of all the methods.

Table 6. Comparison of accuracy of methods used to compute the Quality of the WEB sites.

Method	Reference	Average Accuracy	Standard Deviation
SVM	[67]	78.00%	0.3%
Ivory	[68]	75.00%	0.7%
ELM	[66]	89.00%	1.2%
MLP	This paper	98.20%	0.2%

5.4. Discussion

A learning deep learning model can compute the Quality of the WEB site by considering two layers which are the input layer and the output layers. For the Model, 6 Inputs (Number of missing Images, videos, PDFs, Fields in the tables, columns in the forms and missing self-references) and 5 outputs (Excellent, very good, good, average and poor) have been considered. It has been witnessed that the accuracy model is optimum when 500 Epochs with a batch size = 5 are used.

The example data is prepared in .csv format by considering 100 different websites and running the parser. The counts related to the quality factor "completeness" have been determined, and the same is fed as input to the multi-Layer perceptron model. The data is split into 10 slots, each containing 10 examples, and the examples in each are shuffled so that linearity, if any exists, is removed. The Model is trained by iterating the number of epochs commencing from 100 and effecting an increment of 50 for each iteration. An optimum accuracy of 98.2% is achieved.

6. CONCLUSIONS

A poorly quality website is never used; even if used, it will lead to adverse information dissemination. A hierarchy of factors, sub-factors, and sub-sub-Factors could characterize a website. Knowledge based on a multi-stage expert system is required to compute the overall WEB site quality. The deficiencies of a website can be known through visualization of the Quality of different factors. Using MLP based NN model accuracy of predicting the Quality of a WEB is achieved to the extent of 98.2%

References

1. K. F. Khawaja and R. H. Bokhari, "Exploring the Factors Associated With Website Quality," *Department of Technology Management, International Islamic University, Islamabad, Pakistan*, vol. 10, pp. 37-45, 2010.

2. J. K. R. Sastry and T. S. Lalitha, "A framework for assessing the quality of a WEB SITE, PONTE," *International Journal of Science and Research*, vol. 73, 2017.
3. V. K. Mantri, "An Introspection of Web Portals Quality Evaluation," *International Journal of Advanced Information Science and Technology*, vol. 5, pp. 33-38, 2016.
4. V S Moustakis, "Website quality assessment criteria," *Proceedings of the Ninth International Conference on Information Quality*, 2004.
5. Nielsen, J. (2020), "10 usability heuristics for user interface design", Nielsen Norman Group, available at: <https://www.nngroup.com/articles/ux-research-cheat-sheet/> (accessed 3 May 2021).
6. Tognazzi, B. (2014), "First principles of interaction design (revised and expanded)", *asking*.
7. Shneiderman, B. (2016), "The eight golden rules of interface design", Department of Computer Science, University of Maryland
8. Law, R., Qi, S. and Buhalis, D. (2010), "Progress in tourism management: a review of website evaluation in tourism research", *Tourism Management*, Vol. 31 No. 3, pp. 297-313.
9. Shneiderman, B., Plaisant, C., Cohen, M.S., Jacobs, S., Elmqvist, N. and Diakopoulos, N. (2016), *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, 6th ed., Pearson Higher Education, Essex.
10. Morales-Vargas, A., Pedraza-Jimenez, R. and Codina, L. (2020), "Website quality: an analysis of scientific production", *Profesional de la Información*, Vol. 29 No. 5, p. e290508.
11. Law, R. (2019), "Evaluation of hotel websites: progress and future developments", *International Journal of Hospitality Management*, Vol. 76, pp. 2-9.
12. Ecer, F. (2014), "A hybrid banking websites quality evaluation model using AHP and COPRAS-G: a Turkey case", *Technological and Economic Development of Economy*, Vol. 20 No. 4, pp. 758-782.
13. Leung, D., Law, R. and Lee, H.A. (2016), "A modified model for hotel website functionality evaluation", *Journal of Travel and Tourism Marketing*, Vol. 33 No. 9, pp. 1268-1285.
14. Maia, C.L.B. and FUIiado, E.S. (2016), "A systematic review about user experience evaluation", in Marcus, A. (Ed.), *Design, User Experience, and Usability: Design Thinking and Methods*, Springer International Publishing, Cham, pp. 445-455.
15. Sanabre, C., Pedraza-Jimenez, R. and Vinyals-Mirabent, S. (2020), "Double-entry analysis system (DEAS) for comprehensive quality evaluation of websites: case study in the tourism sector", *Profesional de la Información*, Vol. 29 No. 4, pp. 1-17, e290432.
16. Bevan, N., Carter, J. and Harker, S. (2015), "ISO 9241-11 Revised: what have we learnt about usability since 1998?", in Kurosu, M. (Ed.), *Human-Computer Interaction: Design and Evaluation*, Springer International Publishing, Cham, pp. 143-151.
17. Rosala, M. and Krause, R. (2020), *User Experience Careers: If that a Career in UX Looks Like Today*, Fremont, CA
18. Jainari, MH., Baharum, A., Deris, F.D., Mat Noor, N.A., Ismail, R. and Mat Zain, NH. (2022), "A standard content for university websites using heuristic evaluation", in Arai, K. (Ed.), *Intelligent Computing. SA! 2022. Lecture Notes in Networks and Systems*, Springer, Cham, Vol. 506, doi: 10.1007/978-3-031-10461-9_19.
19. Jayanthi, B. and Krishnakumari, P. (2016), "An intelligent method to assess webpage quality using extreme learning machine", *International Journal of Computer Science and Network Security*, Vol. 16 No. 9, pp. 81-85.
20. Nikolic, N., Grljevic, O. and Kovacevic, A. (2020), "Aspect-based sentiment analysis of reviews in the domain of higher education", *Electronic Library*, Vol. 38, pp. 44-64.
21. Morales-Vargas, A., Pedraza-Jimenez, R. and Codina, L. (2023), "Website quality evaluation: a model for developing comprehensive assessment instruments based on key quality factors", *Journal of Documentation*
22. K. F. Khawaja and R. H. Bokhari, "Exploring the Factors Associated With Quality of Website," *Department of Technology Management, International Islamic University, Islamabad, Pakistan*, vol. 10, pp. 37-45, 2010.
23. J. K. R. Sastry and T. S. Lalitha, "A framework for assessing the quality of a WEB SITE, PONTE," *International Journal of Science and Research*, vol. 73, 2017.
24. V. K. Mantri, "An Introspection of Web Portals Quality Evaluation," *International Journal of Advanced Information Science and Technology*, vol. 5, pp. 33-38, 2016.
25. V S Moustakis, "Website quality assessment criteria," *Proceedings of the Ninth International Conference on Information Quality*, 2004.
26. A. Graniü, et al., "Usability Evaluation of Web Portals," *Proceedings of the ITI 2008, 30th Int. Conf. on Information Technology Interfaces*, 2008.
27. T. Singh, et al., "E-Commerce Website Quality Assessment based on Usability," *Department of Computer Science & Engineering, Amity University, Uttar Pradesh, Noida, India*, pp. 101-105.
28. R. Anusha, "A Study on Website Quality Models," *Department of Information Systems Management, MOP Vaishnav College for Women (Autonomous), Chennai*, vol. 4, pp. 1-5, 2014.

29. F. Ricca and P. Tonella, "Analysis and Testing of Web Applications," *Centro per la Ricerca Scientifica e Tecnologica*, I-38050 Povo (Trento), Italy, 2001.
30. S. Alwahaishi and V. Snášel, "Assessing the LCC Websites Quality," *Springer-Verlag Berlin Heidelberg*, in F. Zavoral NDT 2010, SI, CCIS 87, pp. 556-565, 2010.
31. L. Hasan and E. Abuelrub, "Assessing the Quality of Web Sites," *Applied Computing and Informatics*, vol. 9, pp. 11-29, 2011.
32. Singh KK et al., "Implementation of a Model for Websites Quality Evaluation – DU Website," *International Journal of Innovations & Advancement in Computer Science*, vol. 3, 2014.
33. L. S. Chen and P. Chung, "Identifying Crucial Website Quality Factors of Virtual Communities," *Proceedings of the International Multi-Conference of Engineers and computer scientists, IMECS*, vol. 1, 2010.
34. N. L. Wah, "An Improved Approach for Web Page Quality Assessment," *IEEE Student Conference on Research and Development*, 2011.
35. Sastry J. K. R et al., "Quantifying quality of WEB sites based on content," *International Journal of Engineering and Technology*, vol. 7, pp. 138-141, 2018.
36. Sastry J. K. R et al., "Quantifying quality of websites based on usability," *International Journal of Engineering and Technology*, vol. 7, pp. 320-322, 2018.
37. Sastry J. K. R., et al., "Structure-based assessment of the quality of WEB sites," *International Journal of Engineering and Technology*, vol. 7, pp. 980-983, 2018.
38. Sastry J. K. R., et al., "Evaluating quality of navigation designed for a WEB site," *International Journal of Engineering and Technology*, vol. 7, pp. 1004-1007, 2018.
39. N. P. Kolla et al., "Assessing the quality of WEB sites based on Multimedia content," *International Journal of Engineering and Technology*, vol. 7, pp. 1040-1044, 2018.
40. J. S. Babu et al., "Optimizing webpage relevancy using page ranking and content-based ranking," *International Journal of Engineering and Technology (UAE)*, vol. 7, pp. 1025-1029, 2018.
41. K. S. Prasad et al., "An integrated approach towards vulnerability assessment & penetration testing for a web application," *International Journal of Engineering and Technology (UAE)*, vol. 7, pp. 431-435, 2018.
42. M. V. Krishna et al., "A framework for assessing the quality of a website," *International Journal of Engineering and Technology (UAE)*, vol. 7, pp. 82-85, 2018.
43. R. B. Babu et al., "Analysis on visual design principles of a webpage," *International Journal of Engineering and Technology (UAE)*, vol. 7, pp. 48-50, 2018.
44. S. S. Pawar and Y. Prasanth, "Multi-Objective Optimization Model for QoS-Enabled Web Service Selection in Service-Based Systems," *New Review of Information Networking*, vol. 22, pp. 34-53, 2017.
45. B. Bhavani et al., "Review on techniques and applications involved in web usage mining," *International Journal of Applied Engineering Research*, vol. 12, pp. 15994-15998, 2017.
46. K. K. Durga and V. R. Krishna, "Automatic detection of illegitimate websites with mutual clustering," *International Journal of Electrical and Computer Engineering*, vol. 6(3), pp. 995-1001, 2016.
47. T. Y. Satya and Pradeepini G., "Harvesting deep web extractions based on hybrid classification procedures," *Asian Journal of Information Technology*, vol. 15, pp. 3551-3555, 2016.
48. S. J. S. Bhanu, et al., "Implementing dynamically evolvable communication with embedded systems through WEB services," *International Journal of Electrical and Computer Engineering*, vol. 6(1), pp. 381-398, 2016.
49. L. Prasanna et al., "Profile-based personalized web search using Greedy Algorithms," *ARPJ Journal of Engineering and Applied Sciences*, vol. 11, pp. 5921-5925, 2016.
50. Vassilis S. Moustakis, Charalambos Litos, Andreas Dalivigas, and Loukas Tsironis, WEB site Quality Assessment Criteria, *Proceedings of the Ninth International Conference on Information Quality (ICIQ-04)*, pp. 59-73
51. oon-itt, S. Quality of health websites and their influence on perceived usefulness, trust and intention to use: an analysis from Thailand. *J Innov Entrep* 8, 4 (2019). <https://doi.org/10.1186/s13731-018-0100-9>
52. Allison R, Hayes C, McNulty CAM, Young V. A Comprehensive Framework to Evaluate Websites: Literature Review and Development of GoodWeb. *JMIR Form Res*. 2019 Oct 24;3(4):e14372. doi: 10.2196/14372. PMID: 31651406; PMCID: PMC6914275.
53. Barnes, Stuart & Vidgen, Richard. (2000). WebQual: An Exploration of Website Quality. 298-305.
54. Layla, Hasan Emad Abuelrub, Assessing the Quality of the web sites, *Applied Computing, and Informatics* (2011), 9, 11-29
55. J. Sasi Bhanu1, DBK Kamesh, JKR Sastry, Assessing Completeness of a WEB site from Quality Perspective, *International Journal of Electrical and Computer Engineering (IJECE)* Vol. 9, No. 6, December 2019, pp. 5596-5603 ISSN: 2088-8708, DOI: 10.11591/piece.v9i6.pp5596-5603
56. Rim Rekik*, Ilhem Kallel, Jorge Casillas, Adel M. Alimi, *International Journal of Information Management* 38 (2018) 201–216
57. Lin, H.-F. (2010). An application of fuzzy AHP for evaluating course website quality, *Computers and Education*, 54, 877–888. (Fuzzy hierarchical)

58. Heradio, R., Cabrerizo, F. J., Fernández-Amorós, D., Herrera, M., & Herrera-Viedma, E. (2013). A fuzzy linguistic model to evaluate the Quality of Library, *International Journal of Information Management*, 33, 642–654. (*Fuzzy Linguistics*)
59. Esteban, B., Tejeda-Lorente, Á., Porcel, C., Moral-Muñoz, J. A., & Herrera-Viedma, E. (2014). Aiding in the treatment of low back pain by a fuzzy linguistic Web system. *Rough sets and current trends in computing, lecture notes in computer science (Including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*, (*Fuzzy Linguistics*)
60. Cobos, C., Mendoza, M., Manic, M., León, E., & Herrera-Viedma, E. (2013). Clustering of web search results based on an iterative fuzzy C-means algorithm and Bayesian information criterion. *2013 joint IFSA world congress and NAFIPS annual meeting, IFSA/ NAFIPS 2013*, 507–512. (Fuzzy e-means)
61. Dhiman, P., & Anjali (2014). Empirical validation of website quality using statistical and machine learning methods. *Proceedings of the 5th international conference on confluence 2014: The next generation Information technology summit*, 286–291. (Bayesian)
62. Liu, H., & Krasnoproschin, V. V. (2014). Quality evaluation of E-commerce sites based on adaptive neural fuzzy inference system. *neural networks and artificial intelligence, communications in computer and information science*, 87–97. (Fuzy Neutral)
63. Vosecky, J., Leung, K. W.-T., & Ng, W. (2012). Searching for quality microblog posts: Filtering and ranking based on content analysis and implicit links. *Database systems for advanced applications, lecture notes in computer science (Including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*, 397–413. (SVM)
64. Hu, Y.-C. (2009). Fuzzy multiple-criteria decision-making in the determination of critical criteria for assessing service quality of travel websites. *Expert Systems with Applications*, 36, 6439–6445. (Genetic Algorithms)
65. Michal Kakol, Radoslaw Nielek *, Adam Wierzbicki, Understanding and predicting Web content credibility using the Content Credibility Corpus, *Information Processing and Management* 53 (2017) 1043–1061
66. Jayanthi, B. and Krishnakumari, P. (2016), "An intelligent method to assess webpage quality sing extreme learning machine", *International Journal of Computer Science and Network Security*, Vol. 16 No. 9, pp. 81-85.
67. Huang, Guang-Bin, Xiaojian Ding, and Hongming Zhou. "Optimization method based extreme learning machine for classification." *Neurocomputing* 74, no. 1 (2010): 155-163.
68. Huang, Guang-Bin, Hongming Zhou, Xiaojian Ding, and Rui Zhang. "Extreme learning machine for regression and multiclass classification." *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on* 42, no. 2 (2012): 513-529.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.