CHAPTER 1

INTRODUCTION AND OUTLINE

This chapter briefly introduces the research work proposed in the thesis. Section 1.1 gives an overview of the research undertaken. Section 1.2 sets out the research objectives. Section 1.3 illustrates the proposed framework and the main contributions arising from the work undertaken. Finally, section 1.4 presents an outline of the thesis describing the organisation of the chapters.

1.1. Introduction

All publicly available websites (documents and resources) together constitute the dynamic Web information space. Based on the recent statistics available on the “worldwidewebsize.com”, the currently indexed Web has at least 5.62 billion pages (Friday, 12 April, 2019). From read-only web to an ubiquitous tool for "e-activities” [1], Web has evolved into a pervasive knowledge discovery base reinforcing new sensing based and analytical applications. Website quality analytics is the method of measuring quality of websites considering user engagement and experience and improving its effectiveness and quality. Consequently, a Web quality model [2] provides an acceptance criterion defining the usability and accessibility of a website demonstrating its effectiveness in terms of user experiences (UX). Users have different perception of these websites and this perception is made while a user experiences the website when navigating and browsing through it. In this work, we implement three soft computing techniques namely Machine Learning, Evolutionary Computing, and Fuzzy Logic to determine quality of websites.

Various features contribute to the quality assessment of websites. These typify the quality attributes which help categorization of websites into good, average or poor. Building a learning model for quality assessment of websites enables understanding a site's worth and quality. The poor and average quality websites can be upgraded to the category of good quality websites by improving the quality considering the corresponding quality attributes. The Table 1.1 gives illustrations of some websites having low value of some of the quality attributes which further need to be improved.

In this work, we propose a website quality prediction model based on the relevant quality features using soft computing techniques. The remainder of this chapter sets out the research objectives, describes the main contributions of the work, and presents the outline of the thesis.

Table 1.1: Sample Websites

<table>
<tr><th>Website example</th><th>Indication of poor quality</th></tr>
<tr><td>

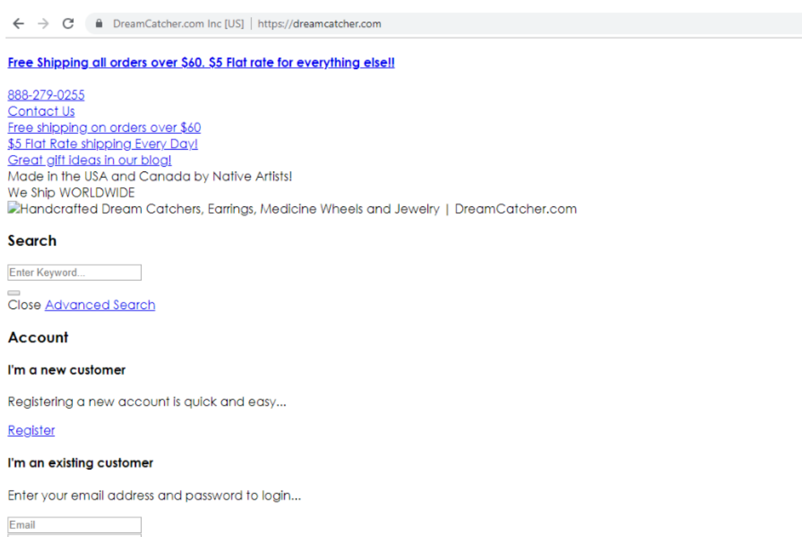
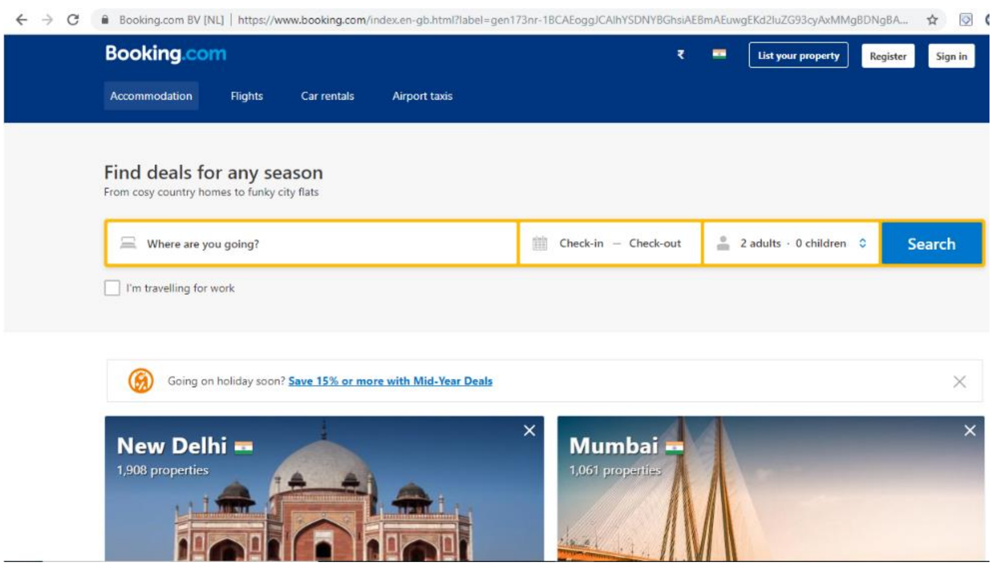


Fig. 1.1. www.dreamcatcher.com/

</td><td>Page speed score is less than 20% which increases the loading time of the page. The page fails to load the background image. Hence, it has a poor loading time.</td></tr>
</table>

 **Fig. 1.2.** www.arngren.net/	Readability is poor leading to a low score of the features typography and font, overall theme and colour scheme.
 **Fig. 1.3.** www.booking.com	Good overall theme, colour scheme as blue and grey are used as the background colour, but lacks social media connectivity as it is not connected to any of the social media sites such as Google, Facebook or Twitter.

1.2. Research Objectives

Statement of Research Question

"Is it possible to predict the quality of a website and automate a predictive model which classifies a website quality as good, average and poor?"

This unifying research question can be broken down into the following five questions, each of which will be addressed by this research.

- How can we predict the quality of a website?
- Which features of a website are important for website quality prediction?
- Is it possible to get an optimal feature set which predicts a website quality more accurately?
- Which soft computing technique is best for the website quality prediction?
- Which metaheuristic-based feature selection method and machine learning technique give best results when combined?

Consequently, the three main research objectives of the work undertaken are:

i) **Research Objective I** – To identify the attributes which contribute to website quality.

ii) **Research Objective II –** To propose a feature-based predictive model for website quality classification.

iii) **Research Objective III –** To find out the best soft computing technique for website quality prediction.

The objective of this thesis is to generate a website quality dataset based on the relevant features and then use soft computing techniques to automatically determine the quality of website accurately.

1.3. Proposed Model

Identifying relevant attributes of user satisfaction and relevance is difficult due to the subjectivity and preference of user. The identification and analysis of website quality attributes is done based on pertinent studies done in this domain [3-13]. Websites are collected from seven domains, namely, .com, .net, .org, .int, .gov, .edu and .mil and 13 quality attributes (i.e. Relevance, Total Size, Broken Links, Communication, Loading Time, Social Media Connectivity, Overall Theme, Compatibility, Global Audience, Resolution, Typography and Font, Colour Scheme, and Keyword matching or Page Rank) are used to label the websites within each domain into good, average and poor. These attributes are evaluated using online tools for each website and score-based system is used to prepare a dataset of 700 websites and their corresponding 13 quality attributes.

Moreover, which attributes enhance the prediction accuracy is imperative for improved model performance. Motivated by this, in this work, we propose an optimal model of Web quality analytics. Optimal feature selection which quantifies the quality of unstructured, hyperlinked, heterogeneous websites are imperative for superior classification results. Swarm search methods have been proposed in various studies for feature selection as an alternative to the conventional search methods [14,15]. Swarm search extends local search by implementing both the local search and global exploration. We validate the advantage of implementing feature selection optimization process using “Particle Swarm Optimization” (PSO) [16] “Elephant Search Algorithm” (ESA) [17] and “Wolf Search Algorithm” (WSA) [18] with the website quality analytics task to improve the classifiers’ performance. This defines the optimal quality attribute subset to classify websites into three categories: good, average, and poor. Metaheuristic search algorithms such as PSO, ESA, and WSA are implemented to select the best feature subset and Six fundamental supervised learning algorithms have been implemented to analyze the website quality. These are Naive Bayesian (NB), Support Vector Machine (SVM), Multilayer Perceptron (MLP: Neural Network), K-Nearest Neighbor (K-NN), Decision Tree (DT), and Random Forest (RF).

Fuzzy logic facilitates human decision making that includes a range of possibilities between YES and NO. It is employed to handle the concept of partial truth, where the truth value may range between completely true and completely false. This is in contrast to a conventional logic block that a computer can understand which takes precise input and produces a definite output as TRUE or FALSE, which is equivalent to human's YES or NO. The fuzzy logic works on the levels of possibilities of input to achieve the definite output which makes it able to deal with uncertainties in engineering. Fuzzy inference (reasoning) is the actual process of mapping from a given input to an output using fuzzy logic. Fuzzy Inference Systems (FIS) also known as Fuzzy Expert System has been successfully applied in fields such as automatic control, data classification, decision analysis, expert systems, and computer vision. In this study we implement a fuzzy inference system for the task of Website Quality Quantification and propose a $QualScore_{site}$ which determines Quality Score of a site. Website Quality Quantification is a novel method which quantifies the website quality as a website quality score using a fuzzy expert system. Fuzzy inference system with 13 inputs and 64 rules is implemented to generate a website quality score which is indicative of the quality of website (good, average or poor).

1.4. Organisation of Thesis

This thesis is structured into 5 chapters followed by references

Chapter 1 presents the research problem, research objectives, justifies the need for and outlines the main contributions arising from the work undertaken.

Chapter 2 provides essential background and context for this thesis and provides a complete justification for the research work described in this thesis.

Chapter 3 provides the details of methodology employed and outlines the Website Quality Prediction Model that constitutes the proposed approach of the research.

Chapter 4 describes the experimental results obtained using various performance evaluators.

Chapter 5 presents the conclusion based on the contributions made by this thesis and highlights the future research avenues.

CHAPTER 2
LITERATURE REVIEW

This chapter discusses the concepts of the techniques used in our work. Also, it discusses the background work in the research domains of website quality prediction and analytics.

2.1. Background Concepts

"Soft Computing consists of various techniques and is described as a *blanket term* leveraging computational intelligence, comprising of several methodologies which are themselves inter-related to one other in varied forms". The first level categorisation of soft computing includes the following:

- Machine Learning
- Neural Networks
- Probabilistic Reasoning
- Evolutionary Computing
- Fuzzy Logic

In this work, we implement three soft computing techniques including Machine Learning, Evolutionary Computing, and Fuzzy Logic.

- Supervised Machine Learning methods have been implemented for website quality classification into good, average and poor.
- Swarm Intelligence based algorithms which come under the category of Evolutionary computing have been implemented to select an optimised feature subset to improve the performance of baseline classifiers.
- Fuzzy Logic based inference engine has been implemented for website quality quantification to generate a website quality score which determines how good is website quality.

2.1.1. Meta-Heuristics

Meta-Heuristics, is a type of two-tier search algorithm that has an upper level of logics controlling how a population of search agents who are programmed to do local heuristic search scout the search space collectively. Many of these metaheuristics are inspired by nature are called nature-inspired optimization algorithms or sometimes called "bio-inspired optimization algorithms" (if it imitates some biological entity) as they are used to solve optimization problems. Some of these metaheuristics which are designed with the intend of discovering the best solution are called metaheuristic search algorithms [15 17']. These metaheuristics classically include Artificial Bee Colony Algorithm (ABC), Particle Swarm Optimization (PSO) [16], Firefly Algorithm (FA) [19,]Ant Colony Optimization (ACO), the Bat Algorithm (BAT) [20], Artificial Immune System Algorithm (AIS), Flower Pollination Algorithm (FPA), the Bacterial Foraging Algorithm (BFO), the Cuckoo Search Algorithm (CS), Elephant Search Algorithm (ESA) [17], Wolf Search Algorithm (WSA) [18], etc. which fall under the category of population-based metaheuristic swarm search methods.

Metaheuristics lookout the search space in distributed-fashion for a best solution with the aim to improve it iteratively exploring the regions yet to be explored. Fig. 2.1 represents the classification of population-based, nature inspired metaheuristics with the domain of this study marked as a red box. The circle in the figure encapsulates algorithms used in this study. PSO, ESA and WSA have been used in this work for the task of feature selection.

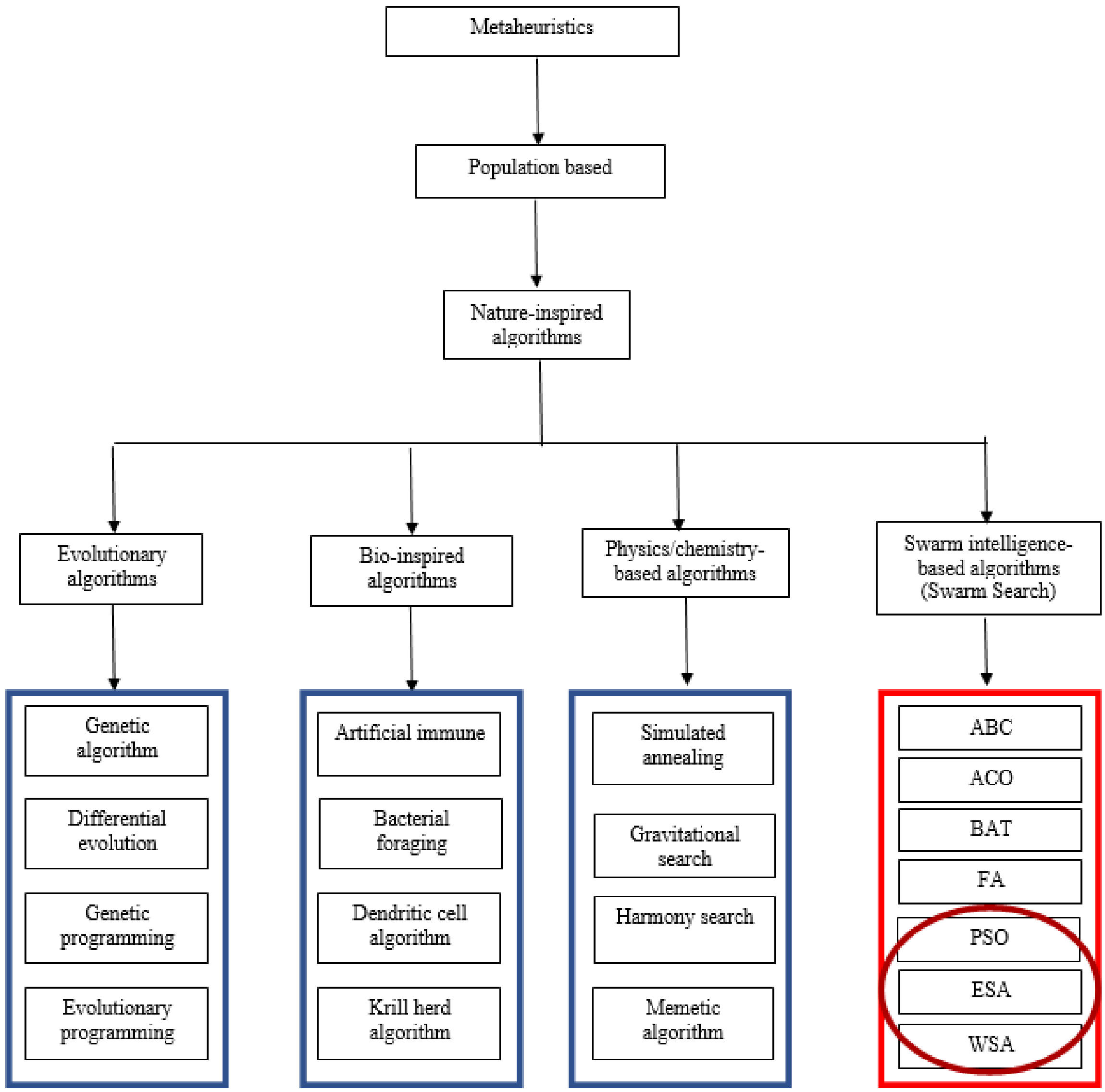


Fig. 2.1. Classification of Metaheuristics.

2.1.2. Classification Algorithms

"Classification is the process of predicting the class of given data points. Classes are sometimes called as targets/ labels or categories". "Classification-predictive-modelling is the task of approximating a mapping function (f) from input variables (X) to discrete output variables (Y)".

The following Table 2.1 describes shallow classifiers used in this research.

Table 2.1: Classification Techniques.

Technique	Description
Naïve Bayes (NB)	"Naïve Bayes is a basic probabilistic model based on the Bayes Theorem which needs trivial amount of training data for estimating the necessary parameters". [21,22].
Support Vector Classifier (SVM)	"The Support Vector Machine (SVM) is described by a decision plane which builds decision boundaries for separating group of instances as different class members by buildings a hyper-plane or a set of hyper-planes that is further utilized for classification" [23,24]. The function of kernel is to take data as input and simulate the projection of the initial data in a feature space with higher dimension.
K-Nearest Neighbor (KNN)	"K-Nearest Neighbor is defined as a non-parametric method that relies on the category labels where an output is a class membership". K-NN identifies the class of an item based on the majority votes of its neighbors [25].
Multi Layer Perceptron (MLP)	"MLP is a type of neural networks which consists of input, hidden and output nodes and is a self-adaptive and data driven technique which can adjust itself with different types of data even if the specification of functional or distributional form for the underlying model is not described". [26].
Decision Tree (DT)	"Decision Trees consists of branches, root and leaf nodes where the corpus is broken down into smaller subsets and consequently builds an associated decision tree with the input features associated with the non-leaf nodes and the classes are represented by the leaf nodes". [27].
Random Forest (RF)	"The Random Forest is an ensemble classifier based on Decision Trees which makes a prediction

	about the class, not simply based on one decision trees, but by an (almost) unanimous prediction, made by 'K' decision trees" [28].

2.1.3. Fuzzy Inference System (FIS)

"A fuzzy Inference System abbreviated as FIS is a system that uses fuzzy set theory to map inputs to outputs". If intended to do fuzzy classification, the inputs are features and the outputs are the classes [29]. The process of fuzzy inference involves *membership functions, fuzzy logic operators*, and *IF-THEN rules* [30].

***Membership functions*:** Membership functions characterize fuzziness or the degree of truth in fuzzy logic. It represents a curve that describes how each point in the input space is mapped to a membership value also called degree of membership between 0 and 1. The Fuzzy Logic Toolbox includes 11 built-in membership function types. These functions are built from some basic functions as:

- piecewise linear functions,
- quadratic and cubic polynomial curve,
- the sigmoid curve and
- the Gaussian distribution functions.

***Fuzzy logic operators**:* Fuzzy logical reasoning is related to standard Boolean logic. Standard logical operations will hold when the fuzzy values are kept at boundaries like 0 (completely false), and 1 (completely true). In fuzzy logic, AND operator is implemented with minimum, OR operator is implemented using maximum and NOT as negation.
i.e.

$$A\ AND\ B = Min(A, B) \tag{2.1}$$

$$A\ OR\ B = Max(A, B) \tag{2.2}$$

$$NOT\ A = 1 - A \tag{2.3}$$

***IF-ThEN rules*.** Usually the knowledge involved in fuzzy reasoning is expressed as rules. A fuzzy IF-THEN rule is of the following form:

$$IF\ Input_1\ is\ A_1\ AND\ Input_2\ is\ A_2 \ldots\ldots\ldots..AND\ Input_n\ is\ A_n\ THEN\ Z = B$$

where, $Input_i$ and Z are fuzzy variables and A_i and B are fuzzy terms (fuzzy sets). The “IF” part is known as antecedent or premise, whereas the “THEN” part is termed as a consequence or conclusion. Statements in the antecedent (or consequent) parts of the rules may well involve fuzzy logical connectives such as ‘AND’ and ‘OR’.

MATLAB implements two types of fuzzy inference systems that can be using the Fuzzy Logic Toolbox [31]:

- Mamdani-type
- Sugeno-type.

The most commonly used fuzzy inference method is the Mamdani's fuzzy inference system.in this system the output membership functions are fuzzy sets. After the aggregation process, there is a fuzzy set for each output variable that needs defuzzification. Sugeno fuzzy inference systems on the other hand are used to model a inference system in which the output membership functions are either linear or constant making this the key difference between Sugeno and Mamdani. This fuzzy inference system was introduced in 1985 and also is called Takagi-Sugeno-Kang. Also, another fundamental difference between Mamdani-type FIS and Sugeno-type FIS is the way the crisp output is generated from the fuzzy inputs. While Mamdani-type FIS uses the technique of defuzzification of a fuzzy output, Sugeno-type FIS uses weighted average to compute the crisp output. In the Sugeno FIS the consequents of the rules are not fuzzy.

2.2. Related Work

The success with website’s discoverability on the web and visitor engagement is fundamentally related to the “quality” of website. Pertinent studies propose web quality models to facilitate website evaluation and assessment based on various quality parameters/attributes. WebQual is one such model developed by Loiacono et al. [4] which measures website quality based on 12 core dimensions. The instrument was developed and validated using 3 samples. In 2009, Sørum et al. [5], broadly categorised quality attributes as information quality, system quality, and service quality and address the norms used by national web awards to measure the website quality in Scandinavian countries. Hartmann et al. [3] have explained how framing of website affects the user judgement of website quality and explored the contextual factors that influence user experience. The authors also focussed on how the presentation of information about a website influence user experience and gave a framework of user judgement and decision making. Sobecki and Żatuchin [6] have described methods for website usability evaluation, quality of website interface, and models for organizing data in frames further to be used in an expert system. WA—Web Assessment tool was also developed based on the three phases of a transaction (information, agreement, and settlement) by the Schubert and Selz [32,33]. An extended model of the same, that is, an Extended Web Assessment Model (EWAM) which includes elements of TAM, social influence, and reviewing four practitioner reports on Web evaluation was also proposed by Schubert and Dettling [34].

The model proposed in this work is better than previously existing models [4-6, 32-34] as the models considers every feature from the visual aspect of website to technical aspect of the website in the features set. Also, as per the best of our knowledge, this is the first work including the google page rank for website quality analytics.

CHAPTER 3
PROPOSED MODEL

"A Website quality model essentially consists of a set of criteria used to determine if a website reaches certain levels of fineness". The application of soft computing facilitates automatic prediction of a website quality within seconds which takes quality attributes as the input and gives results in the form of website quality prediction.

This chapter illustrates the novel techniques that constitute the proposed model. Section 3.1 describes the process of dataset preparation. Section 3.2 describes the step by step process of optimized website quality analytics and section 3.3 describes the proposed $QualScore_{site}$ model.

3.1. Dataset Preparation

As there is no dataset available throughout the web containing the attributes of websites for website quality prediction, the foremost and primary step is the preparation of a dataset. The figure 3.1 illustrates the steps follows for the preparation of dataset to carry out the work for this thesis.

Fig. 3.1. Dataset Preparation

3.1.1. Collection of Websites

In this work, 100 websites from each of the following Top-Level-Domains (total of 700 websites) have been considered for evaluation of the quality using optimized classification techniques. These Top-Level are- .com (commercial), .net (network), .org (organization), .gov (government), .int (international), .edu (educational), .mil (military).

3.1.2. Feature Identification (Quality Attributes)

To implement the quality predictive model, it is imperative to extract features which contribute to the learning process, thus the quality attributes are extracted and evaluated for the 700 websites which forms the dataset of this research. The effectiveness of a website highly depends on how the user is experiencing a website, i.e. its user experience (UX). User behaviour, actions, perceptions and satisfaction all collectively define UX. UX metrics typically revolve around the following factors:

• *Ease of use* which is broadly divided into *ease of understanding* and *ease of navigation between the webpages* [4].

• *Usefulness* which may be

1. *Usefulness in gathering information* which directly depends upon information quality (accuracy, relevance, completeness, and frequent updating of information)

2. *Usefulness in carrying out a transaction* which includes the factors such as trust and response (quick response to emails, online completeness of a transaction), the extent of on-line support or customer service [7-9] may encourage or detract a consumer to do the task online [3]. However, to evaluate customer service, multiple interactions are required.

• *Visual appeal* has three main factors, colour scheme, font size and style, and the overall design. Visual appeal also constitutes of website's design creativity and originality [3]. Some knowledge regarding interface design can be such as Gnome 2.0, the Sun Java Look & Feel, Apple or MS Windows Guidelines [6].

• *Emotional appeal* [10] encourages users to continue browsing [11]. Visual appeal and emotional appeal are together called as *look and feel* [5].

• The capability of a website to provide *personalized communications* to match the customer's needs [12,13]. Functions such as search-fields, support users to search for information online or contact information such as number, email must be provided on the website to provide personal communication & interaction facility.

Other than User Experience, other important quality attributes include

• *Technical assessment* which is used to compute technical features of the websites such as the presence of broken links and the total size of the websites [5].

• *System quality* (usability, reliability, adaptability, availability, functionality) which is related to the Internet environment, and help to measure the desired characteristics of a website [5].

• *Page Rank* or the importance of a website.

Based on the studies of pertinent literature, in this research we identified 16 attributes which gauzed the website quality. To simplify and accurately assess the websites for these quality attributes, they were categorized into 6 major groups. The following figure 3.2 illustrates the 6 groups and 16 quality attributes identified which determine the quality of a website. Out of the 16 attributes, 13 are quantifiable, which can be evaluated and rated based on scoring criteria. Although, the attributes accuracy, completeness, and last update of a web page determine Information Quality but cannot be evaluated as the quantitative analysis of these parameters is time-variant and user-dependent. So, these evaluation criteria of these three attributes is not considered in this research, though this can be a topic of future work.

Fig. 3.2. Website Quality Attributes

3.1.3. Evaluation of Quality Attributes

The following Table 3.1 enlists 6 major groups, 16 attributes with their description, and the evaluation criteria for 13 attributes.

Table 3.1: Quality Attributes and their Evaluation Criteria

<table>
<tr><th>Domain</th><th>Attributes</th><th>Description</th><th>Evaluation</th></tr>
<tr><td rowspan="4">Information Quality</td><td>Relevance</td><td>The extent to which information of interest is relevant.</td><td>Crowd sourcing is done to rate the websites on a scale of 1-10 depending on the relevance of data and advertisement displayed on web page.</td></tr>
<tr><td>Accuracy</td><td>The extent to which information on a webpage is accurate.</td><td rowspan="3"></td></tr>
<tr><td>Completeness</td><td>The extent to which information is specified.</td></tr>
<tr><td>Updation</td><td>The recency of the update of information.</td></tr>
<tr><td>Technical assessment</td><td>Total size</td><td>Total size of web page in KB or MB.</td><td>For measurement of webpage size, an online tool <u>GTMETRIX</u> has been used and scores are given as follows:
<table>
<tr><th>Web page size (in MB)</th><th>Score</th></tr>
<tr><td>< 1 MB (in KB)</td><td>10</td></tr>
<tr><td>1MB-3MB</td><td>9</td></tr>
<tr><td>3 – 5 MB</td><td>8</td></tr>
<tr><td>5 – 7 MB</td><td>7</td></tr>
<tr><td>7 – 9 MB</td><td>6</td></tr>
<tr><td>9 – 11 MB</td><td>5</td></tr>
</table></td></tr>
</table>

			11 – 13 MB | 4 13 – 15 MB | 3 15 – 25 MB | 2 25 – 50 MB | 1 > 50 MB | 0
	Broken Links	The presence of broken links throughout the website.	For the assessment of broken links an online tool *DEADLINKCHECKER* is used which crawls through the website to identify presence of broken links. The scores have been given as follows: **Percentage of broken links** | **Score** 0 | 10 0 – 0.010 | 9 0.010 – 0.020 | 8 0.020 – 0.030 | 7 0.030 – 0.040 | 6 0.040 – 0.050 | 5 0.050 – 0.080 | 4 0.080 – 0.100 | 3 0.100 – 0.200 | 2 0.200 – 0.500 | 1 > 0.500 | 0
	Page Speed	It is inversely proportional to the time taken by a web page to load completely.	An online tool GTMETRIX is used which gives us the page speed score in percentage. We have converted the percentage to score out of 10 as per its ceiling value. For example: 81% - 90% has been scored as 9.

Interaction	Communication	Search field	We have given score according to the four subfields giving 2.5 for the presence of each and a total of 10 if all of the four are present.
		Address	
		Contact	
		Email	
	Social Media Connectivity	Google	A good website should be allied to various social media sites budding today out of which Google, Facebook, and Twitter are most trending. So, we have given a score of 9 to the websites which are connected with three of these social media platforms, i.e. Google, Facebook, and Twitter (a score of 3 for each) and a score of 10 if other social media platforms like YouTube, LinkedIn are also linked.
		Facebook	
		Twitter	
		Others	
Usage	Browser Compatibility	The compatibility of a website with different browsers.	We have tested websites with different browsers which are Internet Explorer, Google Chrome, and Mozilla Firefox. Different websites showed different behaviours and scores have been given in accordance with that behaviour.
	Resolution	The appearance of website with different screen sizes.	On experiencing a website on different screen sizes some showed varied behaviour with different screen sizes and some showed same behaviour. Scores have been given according to the behaviours.
	Global Audience	The language of information on the website.	Websites having information in English and used globally have been scored 10. Websites in a regional language have been given a score of 1. Websites having data in English but used regionally have been scored between 2 – 9.

Visual Appeal	Typography & Font	Font style and font size of the data on the website.	We have given highest score to 12-point Arial font as it is considered to be most legible and sharpest font and is most preferred by the users of websites [35]. We have given second highest score to 12-point Times New Roman as it is fastest to read [35]. The other fonts lagging behind in the queue are 12-point anti-aliased Arial, 10-point Arial, 12-point anti-aliased Times New Roman, 10-point times New Roman. Other fonts styles and size have been scored on the word of their readability.
	Colour Scheme	The colour combinations used to display the website.	Visually appealing colours on the websites likely lead to user satisfaction [36]. We have evaluated websites on the basis of colour combinations used and how these colours strike our eyes. Websites with blue and grey colour schemes have been scored highest as blue is associated with wealth, security and trust and is more aesthetically appealing [37]. Websites having cool colours like blue and green have been given higher score as compared to websites with warm colours such as red, yellow. as blue is more relaxing than red or yellow [38]. Unlike other warm colours, Orange is preferred by users [39] and we scored it higher as compared to pink, green or magenta. Websites with brown or black colour schemes have been scored the least as these colours are associated with sad and stale moods [36]. Websites with yellow background colour have been scored the least as it is distracting colour [36].
	Overall Theme	The association of data, images, tables used in the website.	The visual appeal people experience while visiting a web page via placement and association of its data, tables, and images have been scored.

Page Rank		The rank rendered to a website for its importance.	We used and online tool (https://www.checkpagerank.net/) that provided us with the page rank of all websites.

3.2. Optimized Website Quality Analytics

In this work, the models have been proposed for website quality evaluation: PSO+classification, ESA+classification, and WSA+classification and the comparative performance of the models have been illustrated. Figure 3.3 explains the system architectures of the model giving insight to the step by step process of website quality analytics using feature selection and supervised classification techniques.

The data collection and feature extraction processes have already been explained in section 3.1. the following subsections explain the further steps in detail.

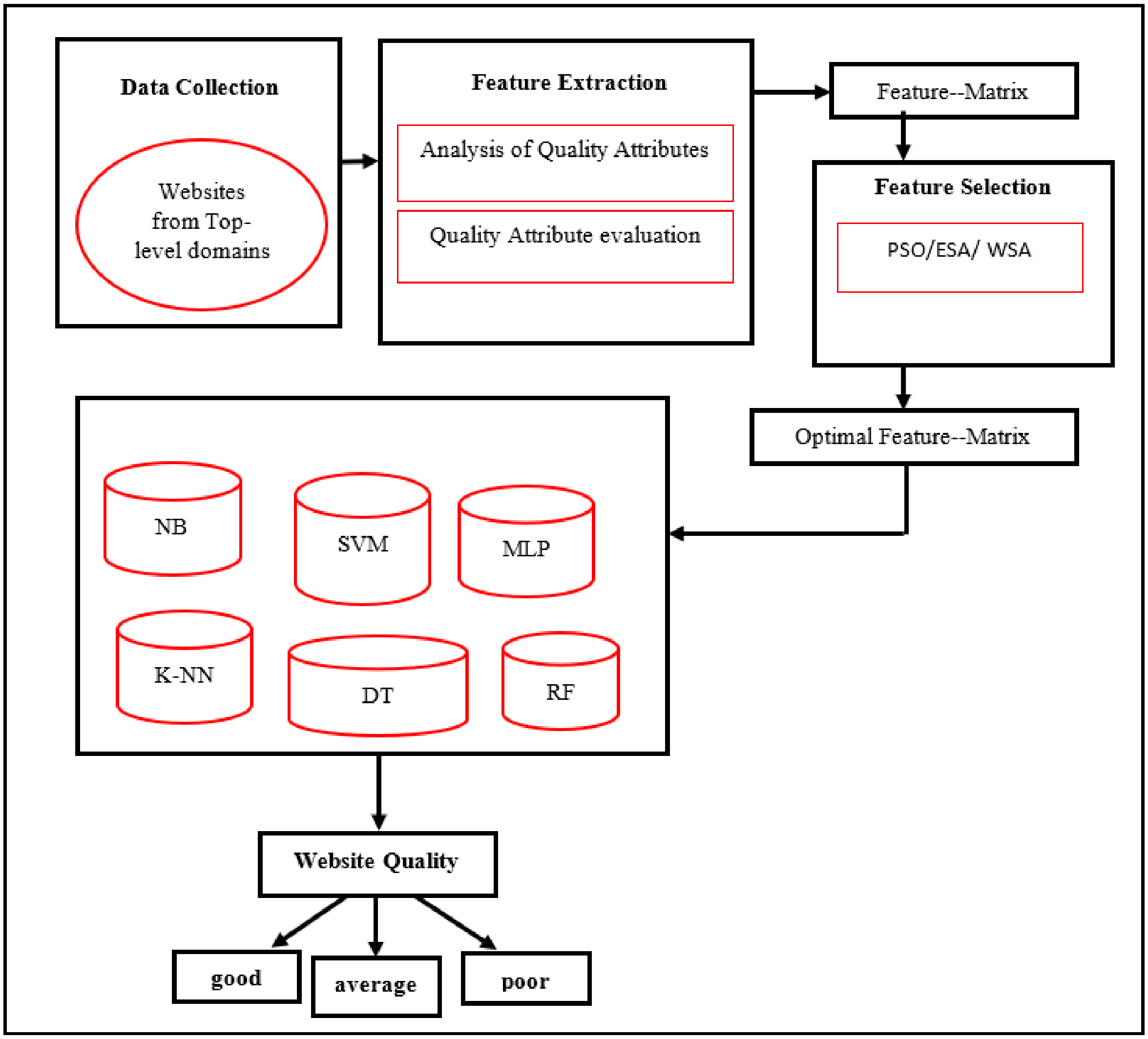


Fig. 3.3. The Optimized Website Quality Analytics

3.2.1. Feature Subset Selection using Meta Heuristic based Optimization Algorithms

"Feature selection is the task of electing a subset of features from the original set of features forming patterns in a given dataset" [40]. If the size of a problem is reduced, it leads to easy learning for classification algorithms which may also lead to improvement in accuracy. This is because computational requirements are reduced and speed of classification is increased due to reduction in the size of dataset which is used to train the classifier.

Particle Swarm Optimization (PSO)

The algorithm is initiated with a population of random particles representing solutions which is a point in search space. Every particle has a memory to store its own best position. Also, the location of particle with the best position amongst all the particles is called the global best. The concept of a flying particle is illustrated in figure. 3.4.

where,

$x_i(t)$ is the position of the particle i at the time t.

$v_i((t)$ is the velocity of the particle i at the time t.

$P_i(t)$ is the best position of particle i at the time t.

$g(t)$ is the global best experience of any particle at the time t.

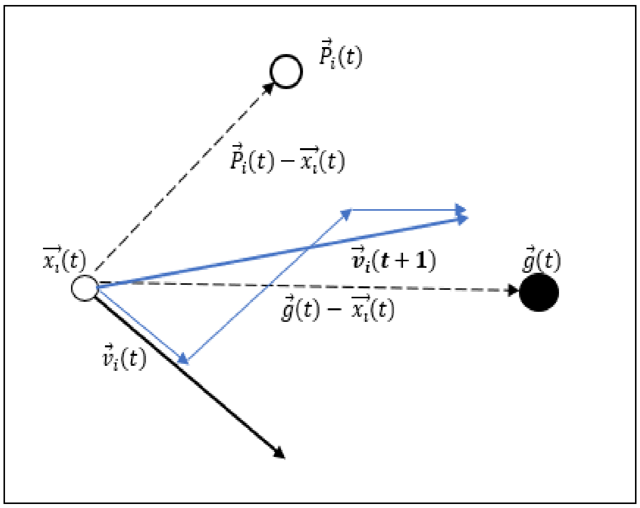


Fig. 3.4. The Concept of a Flying Particle

The Standard PSO pseudo code is given in figure. 3.5.

For each particle i

For each dimension d

Initialize position X_{id} randomly within range

Initialize velocity V_{id} randomly within range

End-For

End-For

Iteration t = 1

Do

For each particle i

Calculate Fitness Value

If Fitness ($X_{id}(t)$) > Fitness ($P_{id}(t)$)

$P_{id} = X_{id}$

End-If

End-For

Chose the particle having the best Fitness value as the $g_d(t)$

For each particle i

For each dimension d

Calculate velocity according to the equation

$V_{id}(t+1) = w.V_{id}(t) + r_1.c_1.(P_{id}(t) - X_{id}(t)) + r_2.c_2.(g_d(t) - X_{id}(t))$

Update particle position according to the equation

$X_{id}(t+1) = X_{id}(t) + V_{id}(t+1)$

End-For

End-For

t = t + 1

While maximum iterations or minimum error criteria are not attained

Fig. 3.5. Pseudo Code for PSO

The following are the considerations for the standard PSO model.

- $r_1$and $r_2$are uniformly distributed random functions with the values in the range of 0 and 1.
- d = 1, 2, 3, …… , S is the dimension.
- The term $w.V_{id}(t)$ is the inertia term, with w being a coefficient known as the inertia coefficient or the inertia weight.
- $c_1$and c_2 are the acceleration coefficients.

- The term $r_1 . c_1 . \left(P_{id}(t) - X_{id}(t)\right)$ is called cognitive component which is responsible for thinking of an individual particle.
- The term $r_2 . c_2 . \left(g_d(t) - X_{id}(t)\right)$ is called social component which allows the particles to collaborate among themselves.
- The cognitive component pulls the particles to the personal best positions while the social component has the role of pulling every particle towards the global best position.
- A new velocity vector $V_{id}(t+1)$ is created by combining the three components: inertia term, cognitive component, and social component.
- This new velocity vector is responsible for changing a particle to an updated position (better location) in the search space $X_{id}(t+1)$.

WEKA implements "PSO Search" for feature selection which is based on Geometric PSO [14]. The Geometric PSO algorithm is different from the standard PSO algorithm [41]. It does not have any velocity and contains mutation. The equation of position update is the convex combination. Moreover, parameters *w*1, *w*2, and *w*3 are non-negative and add up to one. The following figure 3.6 gives the Geometric PSO algorithm.

For all particle *i* do
initialise position *xi* at random in the search space
end For
While stop criteria not met do
For all particle *i* do
set personal best ˆ*xi* as best position found so far by the particle
set global best *g* as best position found so far by the whole swarm
end For
For all particle *i* do
update position using a randomized convex combination
$x_i = CX(x_i, w_1), (g, w_2), (x_i, w_3)$
mutate *xi*
End for
End while

Fig. 3.6. Geometric PSO Algorithm

Elephant Search Algorithm (ESA)

Elephant Search Algorithm simulates the behavior of elephants followed by elephants for their survival. The search agents are divided into two groups each group searching in specified patterns. The male elephants perform search by outreaching to different dimensions of search space afar and hang out solitary in nature. The female elephants form groups to perform local search at certain close proximities forming concentric circles which includes nuclear family unit of females. The herd constitutes of a female leader which is the eldest of all which is also called matriarch, the adult daughters of the leader, the calves of the leader's daughters, other relatives. When female calves grow-up they may remain in the same group or join other groups whereas when males grow up, they will separate from the female group and pursue their own search far away. On the other hand, the male elephants are solitary and do not have a specific group, and hence they randomly roam around looking for females for reproduction.

The ESA algorithm works on following assumptions [17]:

- Each search agent has a visual range which is generally greater for male elephants than for female elephants and is calculated by Euclidean Distance. The elephants move in random walk to search for food within its visual range.
- When two male elephants meet each other, they compare their current fitness-values followed by the backing off of the elephant with a lower fitness value. It will divert its search in another direction.
- The elephants update their current positions at intervals and the best information is carried forward in the future generation through some evolutions.

The figure 3.7 gives the pseudo code of Elephant Search Algorithm.

```
Input: E_m - male elephant population size, E_f - female elephant population size, r_m- radius of male elephant search range, r_f -radius of female elephant search range, A- age limit of elephant;
output: b^global
Initialise ();                                            // initialise all elephants and parameters
While (t<epoch && b^global not satisfactory) do
For each male elephant e_m in E_m do
e_m = Generate_new_location (e_m,t−1);              // global search
If Dist (e_mi, e_mj) < r_m                          // male elephants avoid being too crowd
Weak_elephant_escape (e_m,weakest);
End-if
                                                          // update all male elephant's fitness
                                                          // update best fitness f_min and best location g_best
Update_all_male_elephants_fitness (e_m);
Keep_best (e_m);
End-for
For each female elephant e_f in E_f do
Female_group_leader_move (e_f);                      // local search
E_f = Generate_new_location (e_f,t−1);
                                                          // update all female elephant's fitness
                                                          // update best fitness f_min and best location g_best
Update_all_female_elephants_fitness (e_f);
Keep_best (e_f);
End-for
For all elephants (e_m in E_m) and (e_f in E_f) do
IF life_test(e) > A then
Remove_elephant (e)
Baby_elephant_born (e')
End-for
                                                          // update the global best position from all elephants
If fitness (b_e^local) > = fitness (b^global) then
b^global = b_e^local
End-if
t + +
End-while
```

Fig. 3.7. Pseudo code of Elephant Search Algorithm.

Wolf Search Algorithm (WSA)

The Wolf Search Algorithm is formulated by simulation of the preying behaviour of wolves. A wolf hunts independently and rarely joins its peer provided the peer has inhabited a better terrain. WSA can be visualized as multiple individual wolves gathering from various directions towards the optimal solution, instead of a single herd searching for best solution in one direction at a time. The natural behaviours of wolves are simulated in WSA as follows [18]:

- In nature, wolves have an unparalleled memory which stores food in caches and track prey. This unparalleled memory is simulated in WSA where each wolf has memory to store the positions which are previously visited by it.
- Wolves search for prey during hunting and at the same time they watch out for threats coming towards them. WSA is supplied with a threat probability mechanism that imitates wolves' episodes encountering their enemies. In this situation a wolf moves away in a random direction by a large distance from its position which prevents getting stuck in local optima.
- In nature, wolves have an outstanding judgement of smell which helps them to locate prey. WSA simulates this by enabling each wolf to have a "sensing distance" that creates a coverage area which is called visual distance. While searching when a wolf is not able to find food (the global optimum) or a better terrain than its current position, the wolves move in Brownian motion.

The WSA follows some rules which are given below:

1. The wolves have a visual distance with a radius as v and X as a set of continuous possible solutions. In hyper-plane, this distance would be estimated by Minkowski distance as in (1):

$$v \leq d(x_i, x_c) = \left(\Sigma_{k=1}^{n} \left|x_{i,k} - x_{c,k}\right|^{\lambda}\right), x_c \in X \tag{3.1}$$

where $x_i$is the current position

$x_c$are the positions all the wolves near x_i

λ is the order of the hyper space.

2. The quality of a wolf's current position is given by the fitness of the objective function. The wolf always tries to move to better terrain inhabited by a companion and will finally choose best terrain in case of multiple better terrains. Else, the wolf will move arbitrarily in Brownian Motion.

3. When a wolf senses an enemy, it will escape to a random position beyond its visual range to move away from the threat.

Figure 3.8 gives the pseudo code of Wolf Search Algorithm.

The function "*Generate_new_location()"* gives a position within its visual range that is not visited by the wolf.

Merging with Other Wolves

In WSA the fitness of the objective function determines the quality of wolf's current position. A wolf always wants to be in a position where there is greater probability of finding food and lower probability of being hunted and it will rarely move into territory occupied by another wolf if that territory is better.

It works as follows:

Initially, each wolf locates other wolves within its visual range and evaluates the quality of position of each of its companions. The best location amongst all is compared with the wolf's position. If there is gain in adopting that best location, the wolf changes its location to the best location and will prey there. Otherwise, the wolf searches in a Brownian Motion with an incremental step size. Equation (2) is used for the implementation of the movement:

$$x(i) = x(i) + \beta_o e^{-r^2}\big(x(j) - x(i)\big) + escape(\quad) \qquad (3.2)$$

where

$escape()$ generates a random position that enables the wolf to hop.

$x(i)$is the location of the wolf

$x(j)$is the neighbour that is in a better position

$\beta_o e^{-r^2}$ is the incentive formula which represents the betterment (gain) achieved by wolf by moving to a new position, where β_o is the origin of food, r is the distance between the wolf and the new position

If there are no better terrains occupied by wolf's peers and the wolf is only in the best position, the other wolves will ultimately crowd to the wolf's current position.

Objective function $f(x),\ x = (x_1, x_2, x_3, \dots \dots, x_d)^T$

Initialize the population of wolves, $x_i\ (i = 1, 2, 3, \dots \dots, W)$

Define and initialize parameters:

r = radius of the visual range

s = step size by which a wolf moves at a time

α = velocity factor of wolf

p_α a user-defined threshold [0 .... 1], determines how frequently an enemy appears

While ($t < generations$ && stopping criteria not met)

For I =1: W // for each wolf

Prey_new_food_initiatively ();

Generate_new_location ();

// check whether the next location suggested by the random number generator is new. If not, repeat generating random location

If $dist(x_i, x_j) < r$ && x_j is better as $f(x_i) < f(x_j)$

x_i moves towards x_j // x_j is a better than x_i

Else-if

x_i = *Prey_new_food_passively ();*

End-if

Generate_new_location ();

If $rand\ () > p_\alpha$

$x_i = x_i + rand() + v_i$ // escape to a new position

End-if

End-for

End-while

Fig. 3.8. Pseudo code of Wolf Search Algorithm

Preying

In nature a wolf looks out a region completely to search for food in a pattern of Brownian motion. WSA structures three types of preying behaviour that take place in that order which are described as follows:

1. Preying initiatively: The objective of the optimization function is represented as food. In this step each wolf checks its visual range to detect prey. The wolf will move step by step in the direction of the prey detected with highest fitness.

2. Prey passively: In case the wolf is not able to find a better position occupied by a peer in the preceding step, or it is not able to find food, it will prey passively by staying alert for incoming threats and also it will check the position of its peers in an attempt to improve its current position.

3. Escape: the wolf escapes quickly when a threat is detected. It relocates itself to a new location in random direction with an "escape distance". Escape prevents all the wolves from getting stuck at a local optimum.

The aforementioned preying steps can be defined mathematically by (3):

$$if\ moving = \begin{cases} x(i) = x(i) + \alpha.r.rand() & //Prey \\ x(i) = x(i) + \alpha.s.escape() & //Escape \end{cases} \tag{3.3}$$

where $x(i)$is the wolf's position

α is the velocity

$rand()$ is a random function with mean value in [-1,1]

v is the visual distance

s is the step size.

$escape()$ is a custom function that generates a position in a random manner which is greater than *v* and less than half of the solution boundary

The upper part of equation is used by Preying initiatively and Preying passively for movement whereas the Escape step uses the lower part of formula.

3.2.2. Classification

Six baseline classifiers have been implemented to measure the effectiveness of feature selection. These classifiers are separately trained with the reduced feature set obtained at the end of previous step. Comparative performance analysis is done to measure the accuracy gain of classifiers.

3.2.3. Implementation

In this work, WEKA is used for both the tasks of feature selection using metaheuristics and classification using supervised Machine Learning. Feature selection is divided into two sub parts: feature evaluator (attribute evaluator) and search methods to find the appropriate features in the search space. The attribute evaluator and search methods used in this study are given in the following Table 3.2 with the parameter values taken.

Table 3.2: Hyper-Parameter Values

Attribute evaluator	***ClassifierSubsetEval***
Search methods	PSO, ESA, WSA
inertiaWeight	0.33
individualWeight	0.34
socialWeight	0.33
Chaotic coefficient	4
Population size	20
Iterations	20

3.3. QualScore$_{site}$ Model

Figure 3.9 illustrates the FIS for website quality prediction based on the fuzzy inference system.

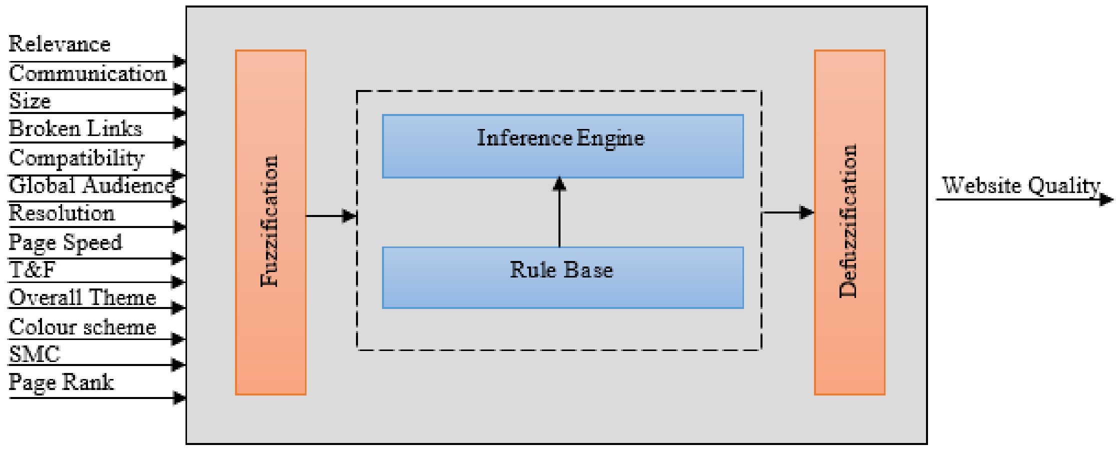


Fig. 3.9. Fuzzy Inference System for Website Quality Quantification

The 13 features of a website acts as input to the fuzzy inference system. Each feature is fuzzified to generate the membership function values. The fuzzified feature values go as input to the inference engine which combine with the rule base to generate an output. The output is defuzzied to generate a crisp score which determines the website quality.

In this study Sugeno type inference model of MATLAB has been used to evaluate the score of websites.

A typical rule in a Sugeno fuzzy model has the following form:

$$\textit{If Input 1} = \mathrm{x} \textit{ and Input 2} = \mathrm{y}, \textit{then Output is } \mathrm{z} = \mathrm{ax} + \mathrm{by} + \mathrm{c} \tag{3.4}$$

For a zero-order Sugeno model, the output level z is a constant $(a=b=0)$

Each rule has a firing strength of the rule represented by w_i which is used to weight the output level z_i.

For example, for an AND rule with Input 1 = x and Input 2 = y, the firing strength is

$$w_i = AndMethod(F_1(x), F_2(y)) \tag{3.5}$$

where $F_1(x)$ *and* $F_2(y)$ are the membership functions for Inputs 1 and 2 respectively. The MIN operator is a natural choice for the logical AND.

The final output of the system is the weighted average of all rule outputs, computed as follow:

$$final\ output = \frac{\sum_{i=1}^{N} w_i z_i}{\sum_{i=1}^{N} w_i} \tag{3.6}$$

where N is the number of rules.

A Sugeno rule operates as shown in the following figure 3.10.

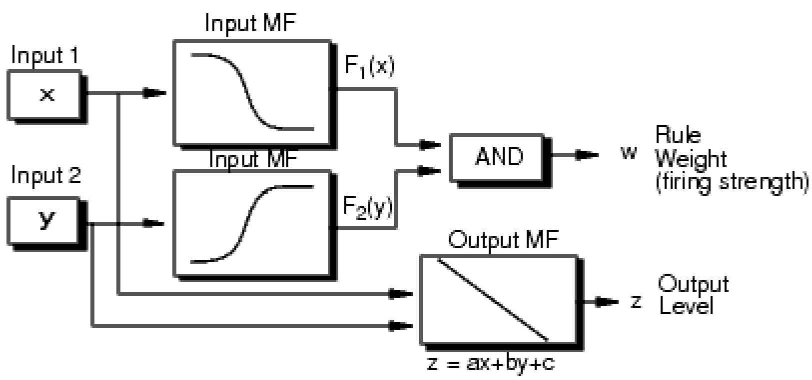


Fig. 3.10. Operation of a Sugeno Rule

The fuzzy logic inference system consists of four components: fuzzifier, inference engine, rule base and defuzzifier [42].

- Inputs with their crisp values are fed into the fuzzifier to fuzzify into linguistic values.
- The rule base contains fuzzy IF-THEN rules which is referred by the inference engine to derive the linguistic values for the output linguistic variables.
- The defuzzifier produces the final crisp values from the output linguistic values.

3.3.1. Fuzzification of Crisp Input Values

The website quality prediction task begins with entering of 13 quality attributes to the fuzzifier to fuzzify the crisp input values and determining the degree of membership for each input values. Each of the input linguistic variable is divided into linguistic terms (fuzzy sets) as given in Table 3.3 as per their value.

Membership functions are used to quantify linguistic term and represent a fuzzy set graphically. Figures 3.11 to 3.23 represent the fuzzy sets on the x axis and their corresponding degree of membership values on the y axis for input variables

Table 3.3: Fuzzy Sets for Each Input

S. No.	**Input Variable**	**Fuzzy Sets**			
1	Relevance	Low	Medium		High
2	Communication	MoreOrLess Communicative	Satisfying	Good	Excellent
3	Size	Small	Medium		Large
4	Broken Links	Too Many	Many	Less	Minimal
5	Compatibility	Low Compatibility	Medium Compatibility		Highly Compatible
6	Global Audience	Not Global		Global	
7	Resolution	Low	Medium		High
8	Page Speed	Slow	Medium		Fast
9	Typography & Font	Poor	Average		Good
10	Overall Theme	Poor	Average		Good
11	Color Scheme	Poor	Average		Good
12	Social Media Connectivity	MoreOrLess connective	Average	High	Excellent Connectivity
13	Page Rank	Bottom	Medium		Top

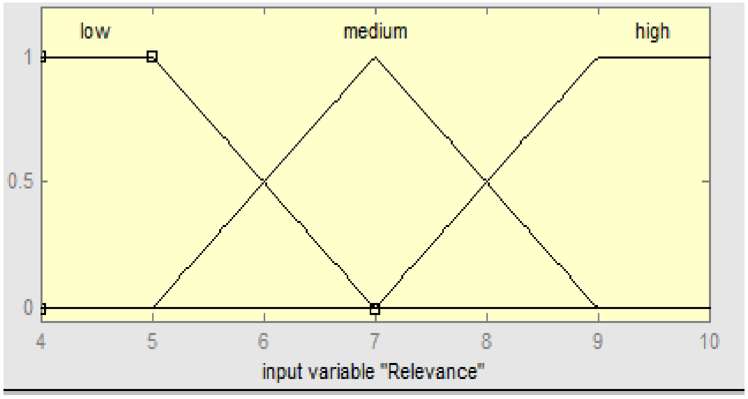


Fig. 3.11. Membership Function for Relevance

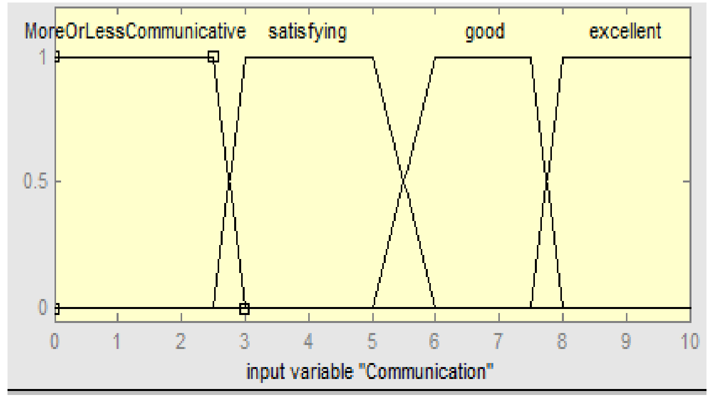


Fig. 3.12. Membership Function for Communication

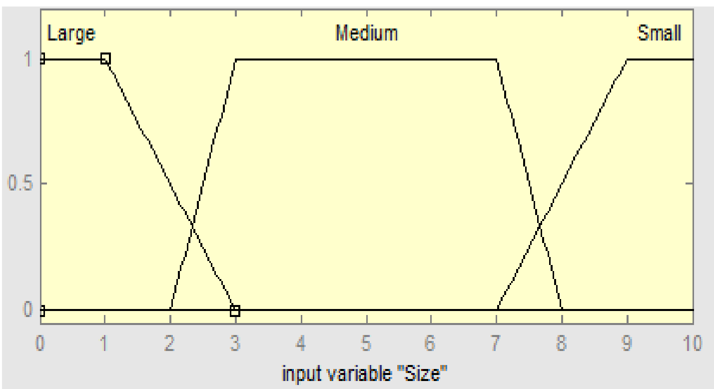


Fig. 3.13. Membership Function for Size

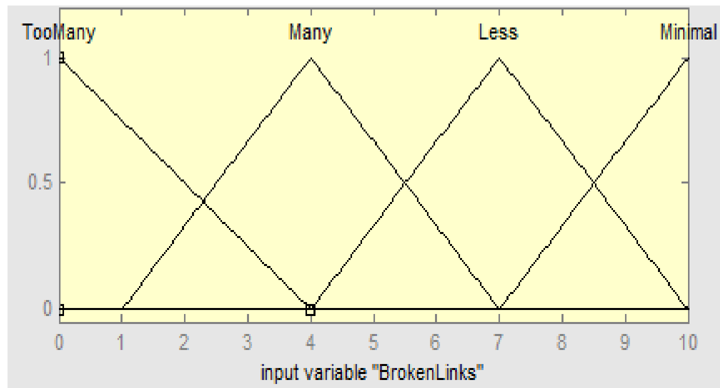


Fig. 3.14. Membership Function for Broken Links

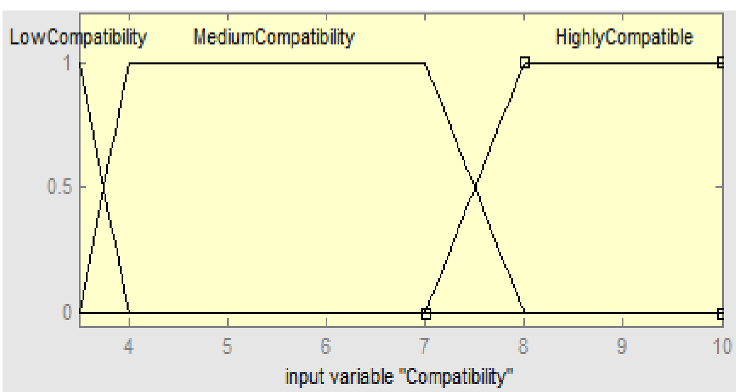


Fig. 3.15. Membership Function for Compatibility

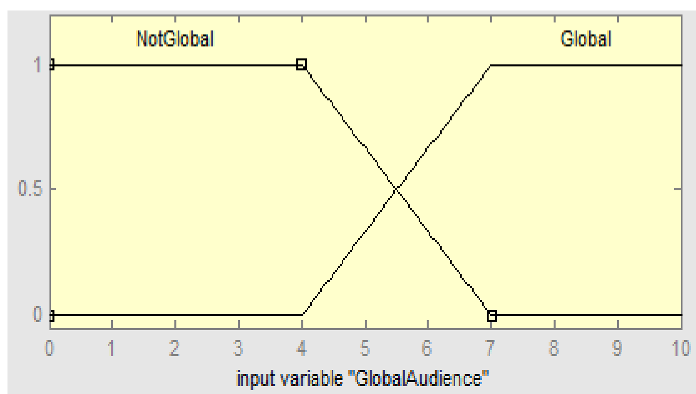


Fig. 3.16. Membership Function for Global Audience

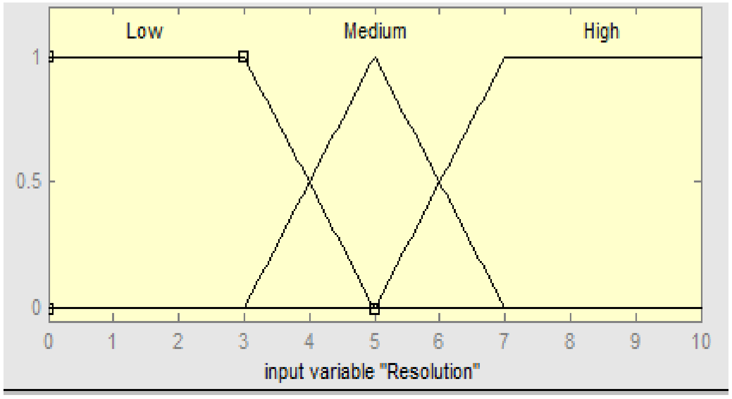


Fig. 3.17. Membership Function for Resolution

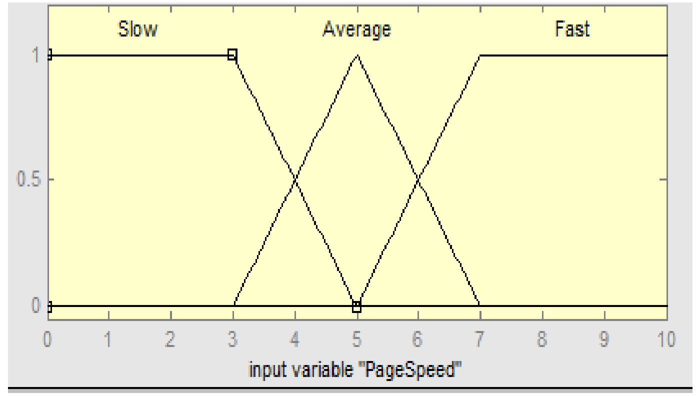


Fig. 3.18. Membership Function for Page Speed

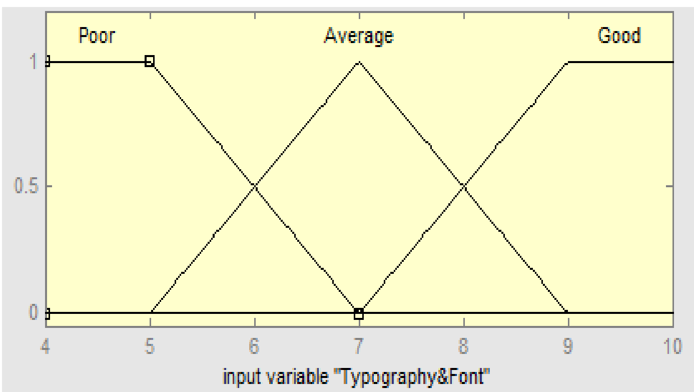


Fig. 3.19. Membership Function for Typography & Font

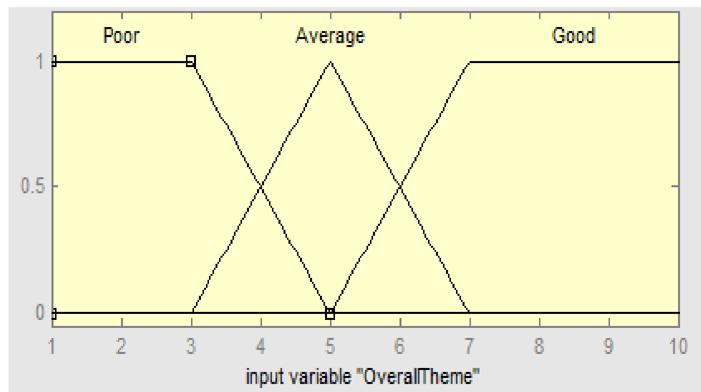


Fig. 3.20. Membership Function for Overall Theme

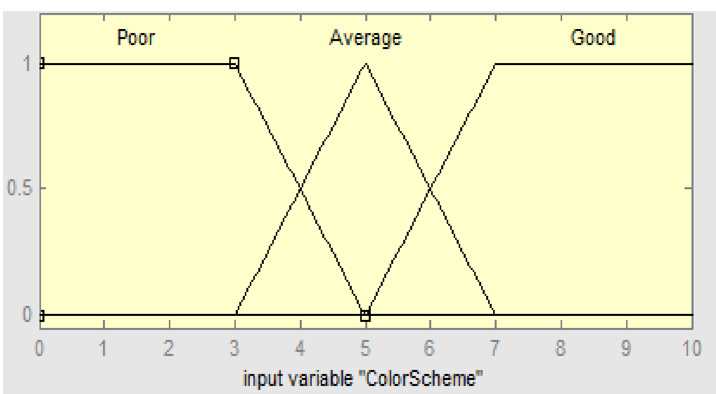


Fig. 3.21. Membership Function for Color Scheme

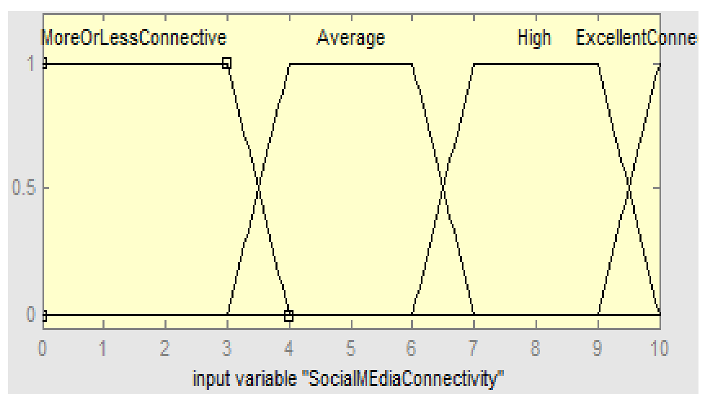


Fig. 3.22. Membership Function for Social Media Connectivity

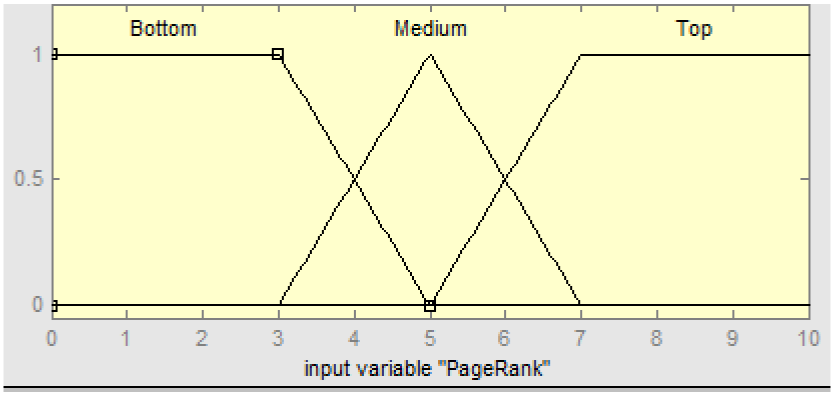


Fig. 3.23. Membership Function for Page Rank

There are 3 outputs linguistic terms corresponding to the output variable *'Quality'* namely: Good, Average, and Poor each with a membership value of 1. The following figure 2.24 depicts the 3 output linguistic terms associated with the output variable of the inference system.

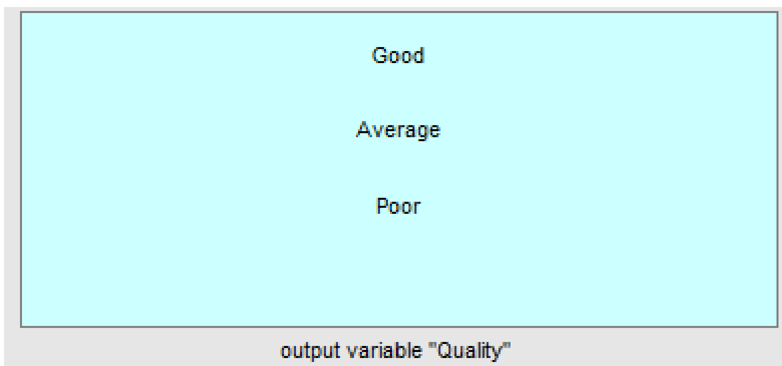


Fig. 3.23. Output Variable 'Quality'

The parameter values associated with these terms are

Good-3

Average-2

Poor-1

The fuzzy inference system outputs the quality score value between 1 and 3. A high quality score near 3 implies a good quality website, whreas a low quality near 1 represents a poor quality website.

3.3.2. Rule Base

A rule base is then constructed based the input parameters which are applicable. These rules result in an aggregate fuzzy set that represents a particular decision regarding the quality of website. In this work there are 13 inputs and some inputs are divided into 3 fuzzy sets while some have been divided into 4 fuzzy sets corresponding to a total of 2,519,424 rules if these sets are combined individually. But implementation of such a large number of rules is impractical so, a rule base is created with 64 rules which has been possible due to the NOT (~) operator. Table I shows a sample rule base for the system under consideration.

1. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Minimal) and (OverallTheme is not Good) and (Communication is not MoreOrLessCommunicative) and (PageSpeed is Fast) and (SocialMediaConnectivity is ExcellentConnectivity) and (Compatibility is HighlyCompatible) then (Quality is Good)

2. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Minimal) and (OverallTheme is not Good) and (Communication is not MoreOrLessCommunicative) and (PageSpeed is Fast) and (SocialMediaConnectivity is ExcellentConnectivity) and (Compatibility is not HighlyCompatible) then (Quality is Average)

3. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Minimal) and (OverallTheme is not Good) and (Communication is not MoreOrLessCommunicative) and (PageSpeed is Fast) and (SocialMediaConnectivity is High) and (Compatibility is HighlyCompatible) then (Quality is Good)

4. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Minimal) and (OverallTheme is not Good) and (Communication is not MoreOrLessCommunicative) and (PageSpeed is Fast) and (SocialMediaConnectivity is High) and (Compatibility is not HighlyCompatible) then (Quality is Average)

5. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Minimal) and (OverallTheme is not Good) and (Communication is not MoreOrLessCommunicative) and (PageSpeed is Fast) and (SocialMediaConnectivity is Average) then (Quality is Average)

6. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Minimal) and (OverallTheme is not Good) and (Communication is not MoreOrLessCommunicative) and (PageSpeed is Fast) and (SocialMediaConnectivity is MoreOrLessConnective) then (Quality is Average)

7. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Minimal) and (OverallTheme is not Good) and (Communication is not MoreOrLessCommunicative) and (PageSpeed is not Fast) and (SocialMediaConnectivity is ExcellentConnectivity) then (Quality is Average)

8. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Minimal) and (OverallTheme is not Good) and (Communication is not MoreOrLessCommunicative) and (PageSpeed is not Fast) and (SocialMediaConnectivity is High) then (Quality is Average)

9. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Minimal) and (OverallTheme is not Good) and (Communication is not MoreOrLessCommunicative) and (PageSpeed is not Fast) and (SocialMediaConnectivity is Average) and (Compatibility is HighlyCompatible) then (Quality is Average)

10. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Minimal) and (OverallTheme is not Good) and (Communication is not MoreOrLessCommunicative) and (PageSpeed is not Fast) and (SocialMediaConnectivity is Average) and (Compatibility is not HighlyCompatible) and then (Quality is Poor)

11. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Minimal) and (OverallTheme is not Good) and (Communication is not MoreOrLessCommunicative) and (PageSpeed is not Fast) and (SocialMediaConnectivity is MoreOrLessConnective) and (Compatibility is HighlyCompatible) then (Quality is Average)

12. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Minimal) and (OverallTheme is not Good) and (Communication is not MoreOrLessCommunicative) and (PageSpeed is not Fast) and (SocialMediaConnectivity is MoreOrLessConnective) and (Compatibility is not HighlyCompatible) then (Quality is Poor)

13. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Minimal) and (OverallTheme is not Good) and (Communication is MoreOrLessCommunicative) and (PageSpeed is Fast) and (SocialMediaConnectivity is ExcellentConnectivity) then (Quality is Average)

14. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Minimal) and (OverallTheme is not Good) and (Communication is MoreOrLessCommunicative) and (PageSpeed is Fast) and (SocialMediaConnectivity is High) then (Quality is Average)

15. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Minimal) and (OverallTheme is not Good) and (Communication is MoreOrLessCommunicative) e) and (PageSpeed is Fast) and (SocialMediaConnectivity is Average) and (Compatibility is HighlyCompatible) then (Quality is Average)

16. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Minimal) and (OverallTheme is not Good) and (Communication is MoreOrLessCommunicative) and (PageSpeed is Fast) and (SocialMediaConnectivity is Average) and (Compatibility is not HighlyCompatible) then (Quality is Poor)

17. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Minimal) and (OverallTheme is not Good) and (Communication is MoreOrLessCommunicative) and (PageSpeed is Fast) and (SocialMediaConnectivity is MoreOrLessConnective) and (Compatibility is HighlyCompatible) then (Quality is Average)

18. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Minimal) and (OverallTheme is not Good) and (Communication is MoreOrLessCommunicative) and (PageSpeed is Fast) and (SocialMediaConnectivity is MoreOrLessConnective) and (Compatibility is not HighlyCompatible) then (Quality is Poor)

19. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Minimal) and (OverallTheme is not Good) and (Communication is MoreOrLessCommunicative) and (PageSpeed is not Fast) and (SocialMediaConnectivity is ExcellentConnectivity) and (Compatibility is HighlyCompatible) then (Quality is Average)

20. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Minimal) and (OverallTheme is not Good) and (Communication is MoreOrLessCommunicative) and (PageSpeed is not Fast) and (SocialMediaConnectivity is ExcellentConnectivity) and (Compatibility is not HighlyCompatible) then (Quality is Poor)

21. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Minimal) and (OverallTheme is not Good) and (Communication is MoreOrLessCommunicative) and (PageSpeed is not Fast) and (SocialMediaConnectivity is High) and (Compatibility is HighlyCompatible) then (Quality is Average)

22. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Minimal) and (OverallTheme is not Good) and (Communication is MoreOrLessCommunicative) and (PageSpeed is not Fast) and (SocialMediaConnectivity is High) and (Compatibility is not HighlyCompatible) then (Quality is Poor)

23. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Minimal) and (OverallTheme is not Good) and (Communication is MoreOrLessCommunicative) and (PageSpeed is not Fast) and (SocialMediaConnectivity is Average) then (Quality is Poor)

24. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Minimal) and (OverallTheme is not Good) and (Communication is MoreOrLessCommunicative) and (PageSpeed is not Fast) and (SocialMediaConnectivity is MoreOrLessConnective) then (Quality is Poor)

25. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Less) and (OverallTheme is not Good) and (Communication is not MoreOrLessCommunicative) and (PageSpeed is Fast) and (SocialMediaConnectivity is ExcellentConnectivity) then (Quality is Average)

26. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Less) and (OverallTheme is not Good) and (Communication is not MoreOrLessCommunicative) and (PageSpeed is Fast) and (SocialMediaConnectivity is High) then (Quality is Average)

27. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Less) and (OverallTheme is not Good) and (Communication is not MoreOrLessCommunicative) and (PageSpeed is Fast) and (SocialMediaConnectivity is Average) and (Compatibility is HighlyCompatible) then (Quality is Average)

28. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Less) and (OverallTheme is not Good) and (Communication is not MoreOrLessCommunicative) and (PageSpeed is Fast) and (SocialMediaConnectivity is Average) and (Compatibility is not HighlyCompatible) then (Quality is Poor)

29. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Less) and (OverallTheme is not Good) and (Communication is not MoreOrLessCommunicative) and (PageSpeed is Fast) and (SocialMediaConnectivity is MoreOrLessConnective) then and (Compatibility is HighlyCompatible) (Quality is Average) (1)

30. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Less) and (OverallTheme is not Good) and (Communication is not MoreOrLessCommunicative) and (PageSpeed is Fast) and (SocialMediaConnectivity is MoreOrLessConnective) and (Compatibility is not HighlyCompatible) then (Quality is Poor)

31. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Less) and (OverallTheme is not Good) and (Communication is not MoreOrLessCommunicative) and (PageSpeed is not Fast) and (SocialMediaConnectivity is ExcellentConnectivity) and (Compatibility is HighlyCompatible) then (Quality is Average)

32. If (Relevance is High) and (PageRank is Top) and and (BrokenLinks is Less) and (OverallTheme is not Good) and (Communication is not MoreOrLessCommunicative) and (PageSpeed is not Fast) and (SocialMediaConnectivity is ExcellentConnectivity) and (Compatibility is not HighlyCompatible) then (Quality is Poor)

33. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Less) and (OverallTheme is not Good) and (Communication is not MoreOrLessCommunicative) and (PageSpeed is not Fast) and (SocialMediaConnectivity is High) and (Compatibility is HighlyCompatible) then (Quality is Average)

34. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Less) and (OverallTheme is not Good) and (Communication is not MoreOrLessCommunicative) and (PageSpeed is not Fast) and (SocialMediaConnectivity is High) and (Compatibility is not HighlyCompatible) then (Quality is Poor)

35. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Less) and (OverallTheme is not Good) and (Communication is not MoreOrLessCommunicative) and (PageSpeed is not Fast) and (SocialMediaConnectivity is Average) then (Quality is Poor)

36. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Less) and (OverallTheme is not Good) and (Communication is not MoreOrLessCommunicative) and (PageSpeed is not Fast) and (SocialMediaConnectivity is MoreOrLessConnective) then (Quality is Poor)

37. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Less) and (OverallTheme is not Good) and (Communication is MoreOrLessCommunicative) and (PageSpeed is Fast) and (SocialMediaConnectivity is ExcellentConnectivity) then and (Compatibility is HighlyCompatible) (Quality is Average)

38. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Less) and (OverallTheme is not Good) and (Communication is MoreOrLessCommunicative) and (PageSpeed is Fast) and (SocialMediaConnectivity is ExcellentConnectivity) and (Compatibility is not HighlyCompatible) then (Quality is Poor)

39. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Less) and (OverallTheme is not Good) and (Communication is MoreOrLessCommunicative) and (PageSpeed is Fast) and (SocialMediaConnectivity is High) then and (Compatibility is HighlyCompatible) (Quality is Average)

40. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Less) and (OverallTheme is not Good) and (Communication is MoreOrLessCommunicative) and (PageSpeed is Fast) and (SocialMediaConnectivity is High) and (Compatibility is not HighlyCompatible) then (Quality is Poor)

41. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Less) and (OverallTheme is not Good) and (Communication is MoreOrLessCommunicative) and (PageSpeed is Fast) and (SocialMediaConnectivity is Average) then (Quality is Poor)

42. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Less) and (OverallTheme is not Good) and (Communication is MoreOrLessCommunicative) and (PageSpeed is Fast) and (SocialMediaConnectivity is MoreOrLessConnective) then (Quality is Poor)

43. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Less) and (OverallTheme is not Good) and (Communication is MoreOrLessCommunicative) and (PageSpeed is not Fast) then (Quality is Poor)

44. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Many) then (Quality is Poor) (1)

45. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is TooMany) then (Quality is Poor)

46. If (Relevance is High) and (PageRank is not Top) then (Quality is Poor)

47. If (Relevance is not High) then (Quality is Poor)

48. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Minimal) and (OverallTheme is Good) and (GlobalAudience is Global) and (Size is not Large) and (Resolution is not Low) and (Typography&Font is not Poor) and (ColorScheme is not Poor) then (Quality is Good)

49. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Minimal) and (OverallTheme is Good) and (GlobalAudience is Global) and (Size is not Large) and (Resolution is not Low) and (Typography&Font is not Poor) and (ColorScheme is Poor) then (Quality is Average)

50. If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Minimal) and (OverallTheme is Good) and (GlobalAudience is Global) and (Size is not Large) and (Resolution is not Low) and (Typography&Font is Poor) and (ColorScheme is not Poor) then (Quality is Average)

3.3.3. Inference Engine

Once all crisp input values have been fuzzified into their respective linguistic values, the inference engine will refer the fuzzy rule base to derive linguistic values for the output fuzzy variables. Aggregation and composition are the two main steps in the inference process [42]. Aggregation includes computation of the values of the IF (antecedent) part of the rules while composition includes calculating the values of the THEN (conclusion) part of the rules.

Each condition in the IF part of a rule is assigned a degree of truth based on the degree of membership of the corresponding linguistic term further computing the minimum (MIN) of the degrees of truth of the conditions which becomes the degree of truth of the IF part. This is assigned as the degree of truth of the THEN part. This is done during aggregation.

3.3.4. Defuzzification

The last phase in the fuzzy expert system is the defuzzification of the linguistic values of the output linguistic variables into crisp values. The Sugeno type of inference system uses two methods of defuzzification [31]. These are wtsum and wtaver. The wtsum determines the degree of membership of the output fuzzy set as the sum of output values generated from each rule. The wtaver calculates the crisp value of the output as the weighted average of the output from each rule. In this work, wtaver has been used for defuzzification to obtain the crisp output.

CHAPTER 4
RESULTS AND ANALYSIS

This chapter describes the experimental results and the analysis to account for the tests performed.

4.1. Experimental Results of Optimized Website Quality Analytics

This research comprises of various soft computing techniques for website quality prediction. It includes a comparison of three metaheuristic optimization algorithms for website quality analytics two of which are inspired by animal behavior. PSO, ESA and WSA have been applied for feature selection followed by application of various classification algorithms and the improvement in classification accuracy has been observed. Accuracy has been considered to measure the effectiveness and efficiency of a classifier.

4.1.1. Feature Selection using PSO

In this work, 13 quality attributes have been evaluated and PSO is used to select the best subset which optimizes the classifier's performance. Table 4.1 illustrates the features selected by PSO algorithm with different classifiers and the length of feature subset.

Table 4.1: Features Selected by PSO and Features Subset Length.

Classifier	Feature Subset	Length of feature subset	Reduced features	Percentage reduction (%)
SVM	1. relevance 2. communication 3. size 4. broken links	12	1	7.69

	5. global audience 6. resolution 7. loading time (page speed) 8. typography and font 9. overall theme 10. colour scheme 11. social media connectivity 12. key word (page rank)			
RF	1. global audience 2. resolution 3. loading time (page speed) 4. typography and font 5. colour scheme 6. social media connectivity 7. keyword (page rank)	7	6	46.15
K-NN	1. communication 2. broken links 3. resolution	3	10	76.92
DT	1. communication 2. size 3. broken links 4. compatibility 5. global audience 6. resolution 7. loading time (page speed) 8. typography and font 9. colour scheme 10. social media connectivity 11. keyword (page rank)	11	2	15.38
MLP	1. relevance 2. communication 3. size 4. broken links 5. compatibility	12	1	7.69

	6. global audience 7. resolution 8. loading time (page speed) 9. overall theme 10. colour scheme 11. social media connectivity 12. keyword (page rank)			
NB	1. communication 2. broken links 3. global audience 4. resolution 5. loading time (page speed) 6. typography and font 7. social media connectivity	7	6	46.15

An average of 33% reduction of features is observed with the highest reduction of 76.92% with K-NN. The following Figure 4.1 represents the pie chart of the PSO feature selection.

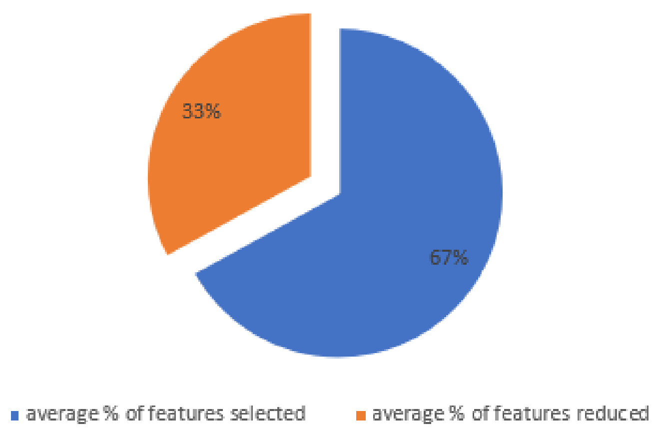


Fig. 4.1. Feature Selection using PSO

4.1.2. Feature selection using ESA

ESA is used to select the best subset out of 13 quality attributes which optimizes the classifier's performance. Table 4.2 represents the features selected by ESA algorithm when implemented with the classifiers and the feature subset length for each classifier.

Table 4.2: Features Selected by ESA and Feature Subset Length

Classifier	Feature Subset	Length of feature subset	Reduced features	Percentage reduction (%)
NB	1. communication 2. broken links 3. global audience 4. resolution 5. loading time (page speed) 6. typography and font 7. social media connectivity	7	6	46.15
SVM	1. relevance 2. communication 3. size 4. broken links 5. global audience 6. resolution 7. loading time (page speed) 8. typography and font 9. overall theme 10. colour scheme 11. social media connectivity 12. key word (page rank)	12	1	7.69

MLP	1. relevance 2. communication 3. size 4. broken links 5. compatibility 6. global audience 7. resolution 8. loading time (page speed) 9. overall theme 10. colour scheme 11. social media connectivity 12. keyword (page rank)	12	1	7.69
KNN	1. communication 2. size 3. broken links 4. resolution 5. loading time (page speed) 6. colour scheme	6	7	53.84
DT	1. relevance 2. communication 3. size 4. broken links 5. compatibility 6. global audience 7. resolution 8. typography and font 9. colour scheme 10. social media connectivity 11. keyword (page rank)	11	2	15.38

RF	1. communication 2. broken links 3. global audience 4. loading time (page speed) 5. overall theme 6. colour scheme 7. social media connectivity	7	6	46.15

An average of 29% reduction of features is observed with the highest reduction of 53.84%. The following figure 4.2 represents the pie chart of the ESA feature selection.

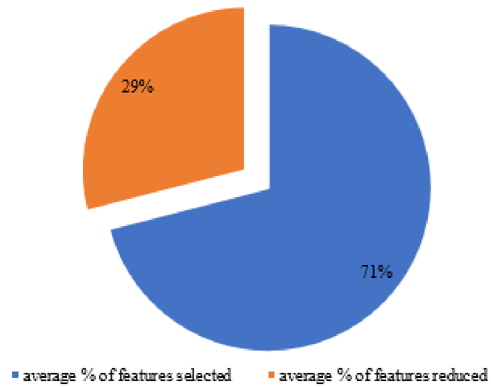


Fig. 4.2. Features Selected by ESA

4.1.3. Feature selection using WSA

Best subset out of 13 quality attributes is selected using WSA for optimization of results. Table 4.3 illustrates the features selected and the length of feature subset for different classifiers when WSA is used for feature selection.

Table 4.3: Features Selected by WSA and Feature Subset Length

Classifier	Feature Subset	Length of feature subset	Reduced features	Percentage reduction (%)
NB	1. communication 2. size 3. broken links 4. compatibility 5. global audience 6. resolution 7. loading time (page speed) 8. typography and font 9. overall theme 10. social media connectivity 11. keyword (page rank)	11	2	15.38
SVM	1. relevance 2. communication 3. size 4. broken links 5. compatibility 6. global audience 7. resolution 8. typography and font 9. overall theme 10. colour scheme 11. social media connectivity 12. key word (page rank)	12	1	7.69

MLP	1. relevance 2. communication 3. size 4. broken links 5. compatibility 6. global audience 7. resolution 8. loading time (page speed) 9. typography and font 10. overall theme 11. social media connectivity 12. keyword (page rank)	12	1	7.69
KNN	1. communication 2. size 3. loading time (page speed) 4. typography and font 5. colour scheme 6. social media connectivity 7. keyword (page rank)	7	6	46.15
DT	1. relevance 2. communication 3. size 4. broken links 5. compatibility 6. global audience 7. resolution 8. typography and font 9. colour scheme	11	2	15.38

	10. social media connectivity 11. keyword (page rank)			
RF	1. relevance 2. broken links 3. loading time (page speed) 4. typography and font 5. colour scheme 6. social media connectivity 7. keyword (page rank)	7	6	46.15

An average of 23% reduction of features is observed with the highest reduction of 46.15% given by KNN and RF. The following Figure 4.3 represents the pie chart of the WSA feature selection.

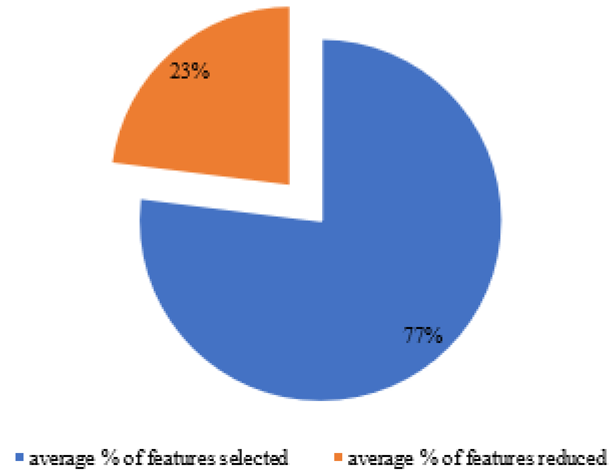


Fig. 4.3. Features Selected by WSA

4.1.4. Comparative Analysis of Performance

In order to gain insight into how PSO, ESA and WSA feature selection for classification of website quality works, we carried out some experiments using the selected significant features. The feature subset selected by PSO, ESA and WSA are further used to train various classifiers to evaluate the performance of classification. The following Table 4.4 gives the summarised comparative results of various classifiers and their performance with PSO, ESA and WSA.

Table 4.4: Comparative Results of Classification without Optimization and with Optimization.

Classifier	Without Feature Selection (%)	PSO (%)	Improvement with PSO (%)	ESA (%)	Improvement with ESA (%)	WSA (%)	Improvement with WSA (%)
NB	74.57	90.66	15.50	87.15	12.58	80.10	5.53
SVM	90.03	93.51	0.63	93.53	3.50	94.88	4.85
MLP	88.64	94.30	2.69	92.80	4.16	94.99	6.35
KNN	78.81	93.35	12.97	80.55	1.74	82.82	4.01
DT	66.27	99.14	27.57	74.66	8.39	75.46	9.19
RF	81.37	99.36	17.08	84.38	3.01	86.10	4.73

Results clearly convey that the classification performance is improved with the feature subset as selected by PSO, ESA and WSA. The average improvement of 12.74% was observed with the implementation of PSO while the average improvement of 5.56% and 5.77% have been observed after using ESA and WSA respectively.

Figure 4.4 gives the graphical illustration of the comparative accuracy of classifiers without optimization and after the feature subset selection.

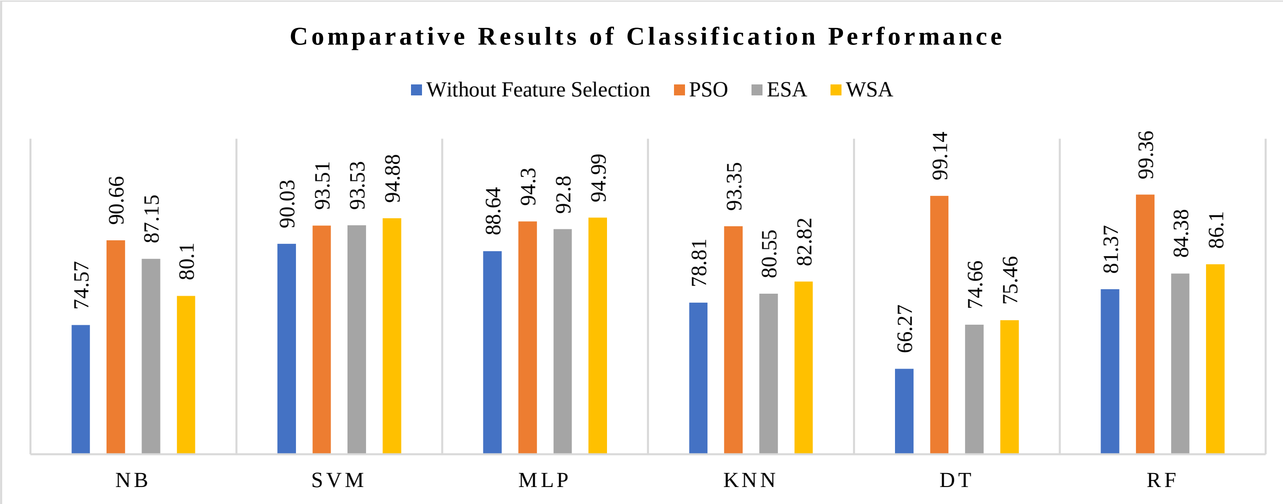


Fig. 4.4. Comparative Performance Results.

With the implementation of NB, KNN, DT, and RF, PSO feature selection outperforms ESA and WSA while with the implementation of SVM and MLP, WSA feature selection gives the best results. Highest accuracy is given by PSO+RF i.e. 99.36%.

Since the maximum reduction in features was observed to be using the PSO Search for feature selection with the value of 33% and most of the classifiers perform best with the features selected by PSO giving an average improvement of 12.74%, the study concludes that PSO for feature selection outperforms ESA and WSA for the website quality prediction task.

4.2. Experimental Results of $\text{QualScore}_{\text{site}}$ Model

The fuzzy logic-based Website quality quantifier has been implemented in MATLAB R2016b. The 13 inputs of a website have been fed into the system with their crisp values. An illustrative of how the website quality score has been generated as an output of fuzzy inference system has been given. The performance assessment of $\text{QualScore}_{\text{site}}$ model has been done using four performance evaluators.

4.2.1. Algorithm with an Illustrative Example

The example of quality quantification has been illustrated below for a sample website.

i) Input: The input values for a sample website has been given in Table 4.5.

Table 4.5: Sample Inputs

Input Variable	Value
Relevance	10
Communication	10
Size	10
Broken Links	8
Compatibility	10
Global Audience	10
Resolution	7
Page Speed	10
Typography & Font	10
Overall Theme	10
Color Scheme	9
Social Media Connectivity	7
Page Rank	6

ii) Fuzzification of crisp values of the inputs: Through the use of membership functions defined for each fuzzy set for each linguistic variable as shown in the figure 3 to figure 15 the degree of membership of a crisp value in each fuzzy set is determined.

iii) Fire the rule bases that correspond to these inputs. Based on the value of the fuzzy membership function values for the example under consideration, the following rules apply:

- *Rule 46:* If (Relevance is High) and (PageRank is not Top) then (Quality is Poor)
- *Rule 48:* If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Minimal) and (OverallTheme is Good) and (GlobalAudience is Global) and (Size is not Large) and (Resolution is not Low) and (Typography&Font is not Poor) and (ColorScheme is not Poor) then (Quality is Good)
- *Rule 58*: If (Relevance is High) and (PageRank is Top) and (BrokenLinks is Less) and (OverallTheme is Good) and (GlobalAudience is Global) and (Size is not Large) and (Resolution is not Low) and (Typography&Font is not Poor) and (ColorScheme is not Poor) then (Quality is Average)

iv) Execute the inference engine:

Rule 46: Relevance is High with degree of membership value 1 and PageRank is not Top with membership value of 0.5. So, $w_1 = AndMethod(1{,}0.5)$. As the *AndMethod* always returns minimum, therefore w_1 obtained from this rule has the value 0.5. Also, as per this rule the quality is Poor, hence, $z_1 = 1$.
According to rule 48, all the fuzzy sets have membership value of 1 except of Top (PageRank) and Minimal (BrokenLinks). The PageRank is Top with membership value 0.5 and BrokenLinks is Minimal with the strength of 0.2. Hence, $w_2 = 0.2$. the quality stated by this rule is Good. So, the value of z_2 is 3. Similarly, as per rule 58, $w_3 = 0.5$ and $z_3 = 2$.

So, the final values obtained are as the following Table 4.6.

Table 4.6: Output of Each Rule

Variable	**Value**
w_1	0.5
w_2	0.2
w_3	0.5
z_1	1
z_2	3
z_3	2

v) Defuzzification: We use wtaver for defuzzification.

$$final\ output = \frac{\sum_{i=1}^{N} w_i z_i}{\sum_{i=1}^{N} w_i} \quad (4.1)$$

Final output= 1.750

Hence, the website quality quantification score generated by the system for the sample website is 1.750

4.2.2. Performance Assessment of QualScore$_{site}$ Model

The performance of *QualScore$_{site}$* model is assessed based on four evaluators namely: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), MAE/Mean Ratio, and Accuracy.

The MAE

The MAE measures the size of the error in units. It is calculated as the average of the unsigned errors as follows [43]:

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \quad (4.2)$$

where, y_i is the predicted value of instance i,

x_i is the labelled (target) value of instance i,

n is the number of instances.

The MAE value calculated for QualScore$_{site}$ Model is *0.563065.*

The MAPE

The MAPE measures the size of the error in percentage terms. It is used as a loss function in prediction models [44]. It is calculated as the average of the unsigned percentage error, as given below:

$$\left(\frac{1}{n} \sum_{i=1}^{n} |y_i - x_i|\right) * 100\ \% \quad (4.3)$$

where, y_i is the predicted value of instance i,

x_i is the labelled (target) value of instance i,

n is the number of instances.

The MAPE value calculated for $\text{QualScore}_{\text{site}}$ model is *38.42338%*.

The MAE/Mean Ratio

The MAD/Mean ratio is an alternative to the MAPE [44]. The statistic is calculated exactly as the name suggests—it is simply the MAD divided by the Mean as given below:

$$MAE/Mean\ Ratio = \frac{MAE}{\frac{\sum_{i=1}^{n} x_i}{n}} \tag{4.4}$$

where, MAE is the Mean Absolute Error of the system as calculated in equation (4.2),

x_i is the labelled (target) value of instance i,

n is the number of instances.

The MAE/Mean ratio of the proposed model is calculated as *0.23708*

The Accuracy

The Accuracy is defined as proximity of a measurement to its true value [45]. The accuracy is calculated as (1-MAE) *100%. The following formula represents the accuracy.

$$Accuracy = 1 - \left(\frac{\frac{\sum_{i=1}^{n} |y_i - x_i|}{n}}{\frac{\sum_{i=1}^{n} x_i}{n}} \right) * 100\% \tag{4.5}$$

The proposed model gives the accuracy as *76.292%*.

Table 4.7: Performance of $QualScore_{site}$ Model

Performance Evaluator	Value
MAE	0.563
MAPE	38.42%
MAE/Mean Ratio	0.237
Accuracy	76.29%.

The figure 4.5 graphically illustrates the performance of the model when all the performance evaluator values are converted into percentages.

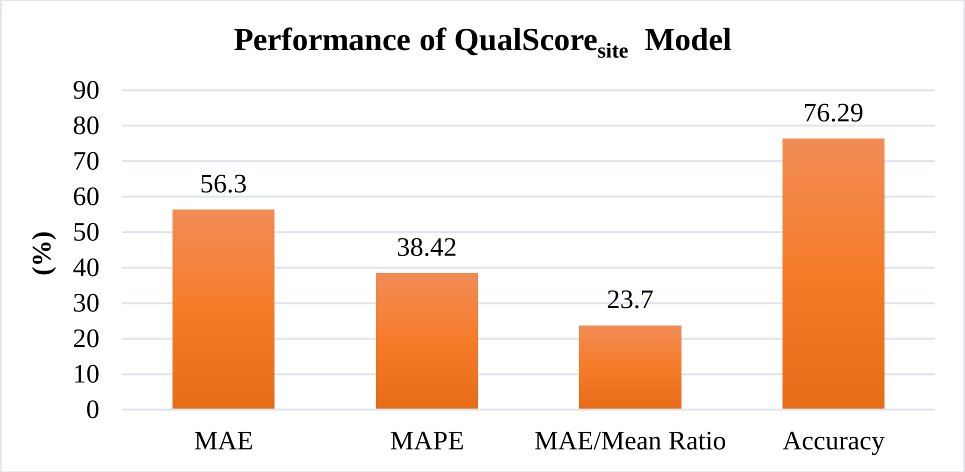


Fig. 4.5. Performance of $QualScore_{site}$ Model

CHAPTER 5
CONCLUSION AND FUTURE SCOPE

5.1. Conclusion of Research

Web quality analytics is the process of examining websites to uncover patterns, correlations, trends, insights and other useful information which can be utilized to optimize web usage and to improve the quality of website. This research proposed an analytical model to classify websites into good, average and poor quality based on attributes related to user satisfaction and acceptance. An empirical study to classify websites from 7 top-level domains using soft computing techniques, including classifiers (Naïve Bayesian, Support Vector Machines, Multi-Layer Perceptron, K-Nearest Neighbor, Decision Tree, and Random Forest) as well as fuzzy logic is conducted. The baseline classifiers are evaluated using an optimal feature selection metaheuristic algorithms Elephant Search Algorithm and Wolf Search Algorithm. The study is performed with 700 websites which were analysed based on 13 attributes (features). The optimized website quality analytics model using metaheuristic-based feature selection gave a better performance than fuzzy inference-based website quality quantification. The accuracy of $\mathit{QualScore}_{site}$ model was observed to be 76.29% whereas the supervised machine learning methods with feature selection gave an extremely better performance.

An average 12.74% improvement in accuracy was observed using PSO feature selection method, 5.56% using the ESA feature selection and 5.77% was observed using WSA for feature selection. Out of various classifiers studied PSO+RF gave the best performance with 99.36% of accuracy. K-NN gave the highest percentage of feature reduction (76.92%) when implemented with PSO. The study confirms that PSO for feature selection in website quality analysis task outperforms ESA and WSA feature selection.

The key contributions of this research are as follows:

- Preparation of website quality dataset based on 13 key quality attributes.

- Optimised Model to predict the quality of website and classify it into good, average and poor based on metaheuristic-based feature selection and machine learning.

- Automatic model for quality quantification based on fuzzy inference system which generates a website quality score.

5.2. Future Research Directions

As a future direction of work, other quality attributes of a website such as security, accuracy of information, last update date of a website may further assist in website quality prediction. As another promising direction of future research, deep learning techniques can be studied and validated for website quality prediction. Moreover, other soft computing techniques such as Neural Networks and Probabilistic Regression can be implemented to determine the quality of websites.

References

[1] Bhatia, MPS.; Kumar, A. Information Retrieval & Machine Learning: Supporting Technologies for Web Mining Research & Practice. *Webology*, 2008, *5*(2), 5.

[2] Kumar, A.; Gupta, D. Paradigm shift from conventional software quality models to web based quality models. *Int. J. Hybrid Intell. Syst*, 2018, *14*(3), 167–179.

[3] Hartmann, J., De Angeli, A., Sutcliffe, A.: 'Framing the user experience: information biases on website quality judgement'.Proc. SIGCHI Conf. Human Factors in Computing Systems, April 2008, pp. 855-864, ACM.

[4] Loiacono, E.T., Watson, R.T., Goodhue, D.L.: 'WebQual: A measure of website quality'. Marketing theory and applications., 2002, 13, (3), pp.432-438.

[5] Sørum, H., Medaglia, R., Andersen, K.N.: 'Assessment of website quality: Scandinavian web awards right on track?., Int. Conf. Electronic Government, Berlin, Heidelberg, August 2009, pp. 198-209, Springer,.

[6] Sobecki, J. and Żatuchin, D.: 'Knowledge and data processing in a process of website quality evaluation'., New challenges in computational collective intelligence, Berlin, Heidelberg, 2009, pp. 51-61. Springer. the measurement of information systems service quality'. MIS Quarterly, 21, (2), 1997, pp 223-240.

[7] Kettinger, W.J., Lee, C.C.: 'Pragmatic perspectives on the measurement of information systems service quality'. MIS Quarterly, 21, (2), 1997, pp 223-240.

[8] Parasuraman, A., Zeithaml, V.A., Berry, L.L.: 'SERVQUAL: a multiple-item scale for measuring consumer perceptions of service quality'. Journal of Retailing, 1998, 64, (1), p.12. 1988, pp 12-40.

[9] Kaynama, S., Black, C. I.:'"A proposal to assess the service quality of online travel agencies: an exploratory study.' Journal of Professional Services Marketing, 2000, 21, (1), pp. 63-88. 2000, pp 63-88.

[10] Richins, M.: 'Measuring emotions in the consumption experience'. Journal of Consumer Research, 1997, 24, (2), pp. 127-146.

[11] Hoffman, D. L., Novak, T. P., Chatterjee, P.: 'Commercial scenarios for the Web: opportunities and challenges'. Journal of Computer-Mediated Communication., 1997, 1, (3).

[12] Ghose, S., Dou, W.: 'Interactive functions and their impacts on the appeal of Internet presence sites'. Journal of Advertising Research, 1998, 38, (2), pp 29-43.

[13] Steuer, J.: 'Defining virtual reality: dimensions determining telepresence'. Journal of Communication., 1992, 42, (4), pp 73-93.

[14] Fong, S.; Biuk-Aghai, R.P.; Millham, R.C. Swarm Search Methods in Weka for Data Mining. In: Proceedings of the 10th International Conference on Machine Learning and Computing, February **2018**, pp. 122-127. ACM.

[15] Kumar, A., Sangwan, S.R. and Nayyar, A., 2019. Rumour veracity detection on twitter using particle swarm optimized shallow classifiers. *Multimedia Tools and Applications*, pp.1-19.

[16] Kennedy, l.; Eberhart, R. Particle Swarm Optimization, In: Proceedings of IEEE International Conference on Neural Networks IV. **1995**, pp. l942- 1948.

[17] Deb, S.; Fong, S.; Tian, Z. Elephant search algorithm for optimization problems. In: Digital Information Management (ICDIM), Tenth International Conference on October **2015**, pp. 249-255. IEEE.

[18] Tang, R.; Fong, S.; Yang, X.S.; Deb, S. Wolf search algorithm with ephemeral memory. In: Digital Information Management (ICDIM), Seventh International Conference on August **2012**, pp. 165-172. IEEE.

[19] Xin-She Yang. Firefly algorithms for multimodal optimization, *Stochastic Algorithms: Foundations and Applications, SAGA*, 2009, Lecture Notes in Computer Sciences, **2009**, *5792*, 169-178.

[20] Xin-She Yang. A New Metaheuristic Bat-Inspired Algorithm, *Nature lnspired Cooperative Strategies for Optimization*, N1CSO, **2010**, Volume *284* of the series Studies in Computational lntelligence, 65-74.

[21] Murphy, K.P. Naive bayes classifiers. *University of British Columbia*, *18*, **2006**.

[22] Rish, I. An empirical study of the naive Bayes classifier. In: IJCAI 2001 workshop on empirical methods in artificial intelligence, August **2001**, *3*(22), pp. 41-46, IBM.

[23] Hearst, M.A.; Dumais, S.T.; Osuna, E.; Platt, J.; Scholkopf, B. Support vector machines. *IEEE Intelligent Systems and their applications*, **1998**, *13*(4), 18-28.

[24] Meyer, D.; Wien, F.T. Support vector machines. *R News*, **2001**, *1*(3), 23-26.

[25] Peterson, L.E. K-nearest neighbor. *Scholarpedia*, **2009**, *4*(2), 1883.

[26] Kruse, R.; Borgelt, C.; Klawonn, F.; Moewes, C.; Steinbrecher, M.; Held, P. Multi-layer perceptrons. In*:* Computational Intelligence, **2013**, pp. 47-81. Springer, London.

[27] Quinlan, J.R. Induction of decision trees. *Machine learning*, **1986**, *1*(1), 81-106.

[28] Breiman, L. Random forests. *Machine learning*, **2001**, *45*(1), 5-32.

[29] https://www.sciencedirect.com/topics/engineering/fuzzy-inference-system

[30] Nedeljkovic, I., 2004. Image classification based on fuzzy logic. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *34*(30), pp.3-7.

[31] https://in.mathworks.com/products/fuzzy-logic.html

[32] Schubert, P., Selz, D.: 'Web assessment-measuring the effectiveness of e-commerce sites going beyond traditional marketing paradigms'. Hawaii Int. Conf. Systems Science, Hawaii, January 1999, pp. 10.

[33] Selz, D., Schubert, P.: 'Web assessment-a model for the evaluation and the assessment of successful electronic commerce applications'., 1998., Proc. Thirty-First Hawaii Int. Conf. System Sciences., January 1998, 4, pp. 222-231, IEEE.

[34] Schubert, P., Dettling, W.: 'Extended Web assessment method (EWAM)-evaluation of e-commerce applications from the customer's viewpoint'. Hawaii Int. Conf. Systems Science, Hawaii, 2002, pp. 10-pp, IEEE.

[35] Bernard, M., Mills, M.: 'So, what size and type of font should I use on my website'. Usability news, 2000, 2, (2), pp.1-5.

[36] Cyr, D., Head, M., Larios, H.: 'Colour appeal in website design within and across cultures: A multi-method evaluation'. International journal of human-computer studies, 2010, 68, (1-2), pp.1-21.

[37] Lichtlé, M.C.: 'The effect of an advertisement's colour on emotions evoked by attitude towards the ad: The moderating role of the optimal stimulation level'. International Journal of Advertising, 2007, 26, (1), pp.37-62.

[38] Jacobs, K.W., Hustmyer, G.H.: 'Effects of four psychologically primary colors on GSR, heart rate, and respiration rate, Perceptual and Motor Skills', 1974, 41, pp.763–766.

[39] Bonnardel, N., Piolat, A., Le Bigot, L." 'The impact of colour on Website appeal and users' cognitive processes'. Displays, 2011, 32, (2), pp.69-80.

[40] Wang, X.; Yang, J.; Teng, X.; Xia, W.; Jensen, R. Feature selection based on rough sets and particle swarm optimization'. *Pattern recognition letters*, 2007, *28*(4), 459-47.

[41] Moraglio, A., Di Chio, C. and Poli, R. 2007. Geometric particle swarm optimisation. In *European Conference on Genetic Programming*, 125-136. Springer, Berlin, Heidelberg.

[42] Pathak, J., Vidyarthi, N. and Summers, S.L., 2005. A fuzzy-based algorithm for auditors to detect elements of fraud in settled insurance claims. *Managerial Auditing Journal*, *20*(6), pp.632-644.

[43]https://medium.com/human-in-a-machine-world/mae-and-rmse-which-metric-is-bettere60ac3bde13d

[44] https://www.forecastpro.com/Trends/forecasting101August2011.html

[45] Kumar, A. and Joshi, A., 2017, March. Ontology Driven Sentiment Analysis on Social Web for Government Intelligence. In *Proceedings of the Special Collection on eGovernment Innovations in India* (pp. 134-139). ACM.

Appendix 1

Snippet of Dataset

The following Table A 1.1 represents the snippet od dataset used in this work.

Table A 1.1: Snippet of the dataset used (QA: Quality Attribute)

Domain	Website Name	Quality Attributes												
		QA1	QA2	QA3	QA4	QA5	QA6	QA7	QA8	QA9	QA10	QA11	QA12	QA13
.com	www.ratetag.com	9	7.5	9	2	10	4	8	6	10	9	9	1	4
	www.tradus.com	10	6	10	8	10	9	8	10	10	10	7	5	5
.net	www.mixgame.net	10	0	8	8	10	10	10	6	0	10	7	0	2
	www.overclock.net	7	1	8	9	10	10	6	3	7	7	6	0	7
.org	www.dostor.org	5	2	7	5	7	10	0	4	3	5	4	0	6
	www.vote.org	9	2.5	10	8	10	4	10	10	10	4	9	10	0
.int	www.eac.int	10	7	10	10	7	10	4	5	6	10	9	0	5
	www.arc.int	9	7	10	10	7	10	8	7	3	7	8	4	5
.gov	www.nationalarchives.gov.uk	8	9	8	9	9	7	8	8	7	8	8	9	7
	www.opapp.gov.ph	5	2	4	5	4	6	7	8	3	6	3	2	3
.edu	www.warrington.ufl.edu	8	7	7	8	10	7	8	8	9	10	9	8	4
	www.aimc.edu	5	8	5	5	9	3	4	8	8	5	5	7	3
.mil	www.nationalguard.mil	8	10	7	8	8	7	9	8	7	8	8	9	5
	www.army.mil	8	9	9	3	9	9	9	9	9	9	10	8	8

Appendix 2

List of Publications

(Published Papers)

1. Kumar, A., and Arora, A., 2019. A Filter-Wrapper based Feature Selection for Optimized Website Quality Prediction. In *Amity International Conference on Artificial Intelligence (AICAI),* Dubai, *IEEE*.

(Communicated Papers)

1. Kumar, A., and Arora, A., Website Quality Analytics using Metaheuristic based Optimization.
2. Kumar, A., and Arora, A., Fuzzy Expert System for Website Quality Quantification: $\text{QualScore}_{\text{site}}$.
3. Kumar, A., and Arora, A., An ANFIS Based Compatibility Scorecard for IOT Integration in Websites.