Optimized Composition of Business Process Web Services via QoS-Based Categorization Using Decision Tree Classifier and Knowledge-Based Decision Support

Larisa Ivascu[1,*]

[1]Faculty of Management in Production and Transportation, Politehnica University of Timisoara, Romania

Email: larisa.ivascu@upt.ro

Abstract

Determining web services according to Quality of Service (QoS) restrictions is the topic of discussion in this section. Decision tree classifiers are used to accomplish this classification. Because of the ever-changing and expanding nature of online services, it is necessary to accurately categorize them in order to make choosing them more efficient for consumers. It makes use of decision tree techniques, more especially the C5.0 classifier, this is an advancement over older approaches such as the C4.5 classifiers. It incorporates characteristics like as noisy handling, incomplete information administration, and improved decision-making correctness. Web services are classified into four distinct groups: Outstanding, Good, Average, and Poor. These classifications are determined by QoS metrics that include time to response, accessibility, performance, dependability, and success rate. The choice of features is accomplished utilizing an evolutionary algorithm with a wrapper technique with the goal to maximize the effectiveness of this category. This method minimizes the number of repetitive features and improves the method of classification for the purpose of optimization. The resilience and predicted reliability of the algorithm are ensured by additional approaches like as cross-validation and error reduction. These approaches also address difficulties such as overfitting and redundant characteristics. The construction of integrated web services for complicated corporate operations is a particularly valuable use of this technology, which also considerably improves the procedure for making choices for identifying services and consumption. Service 7 stands out with an impressive 98% performance, while Service 6 and Service 3 are also among the top-performing services. Compared to the others, Service 1, Service 2, Service 5, and Service 4 all exhibit comparatively poor results.

Keywords: VoIP; QoS; C5.0 classifier; Decision tree; SWAM

1. Introduction

It is now possible for varied software programs to communicate and interact with one another in a seamless manner across a variety of locations thanks to the proliferation of web-based services, which have become a vital part of the digital community. A considerable transformation has taken place in the surroundings of distributed computation as a result of their capacity to let diverse systems to interact with one another [1]. These systems are often designed using distinct languages of programming and are operating on various systems. The areas of e-commerce, cloud computing, and numerous other internet-driven fields have been significantly transformed as a result of this change. A significant factor that has contributed to the fast growth of web services is their adaptability, which allows them to provide solutions that are platform-independent and reusable for diverse operational needs. Due to the growing constantly number of services that are now accessible, both businesses and consumers are confronted with the issue of determining which service is the most appropriate for meeting their particular requirements. It is not an easy job to choose the appropriate online service since there is such a wide variety of possibilities available. These options often share capabilities that are similar to one another, but they exhibit significant differences with regard to of quality and reliability. When deciding whether or not a web service is suitable for a given purpose, QoS characteristics, which include response time, availability, throughput, success rate, and dependability, play an essential role [2-3]. The non-functional characteristics of a service that are essential

DOI: https://doi.org/10.54216/AJBOR.120201
Received: June 14, 2024 Revised: September 25, 2024 Accepted: December 13, 2024

for satisfying both the demands of users and the needs of operations are included by these specifications. As a consequence of this, there is an urgent want for computerized programs that can classify, rank, and propose online services based on QoS characteristics. As a viable approach to addressing this difficulty, the topic of web service categorization has attracted a lot of interest in recent years. Within the scope of this discussion, classification refers to the process of categorizing online services into various groups according to predetermined criteria, most often their QoS characteristics.

Furthermore, does this classification method increase the speed of service discovery, but it further enhances its efficacy of discovering services that are in alignment with user needs. Decision- tree classifiers, which are an instance of artificial intelligence approaches, have come to prominence as a potentially useful tool for resolving the complexity that are connected with online service categorization [4-5]. The categorization of things into distinct types is made possible via the use of decision trees, which are complex structures that split data depending on attribute values at each level. Due to the fact that they provide answers for inductive deductions that are both readily apparent and computationally fast, they are especially useful in situations that include big datasets. As a result of its higher effectiveness in terms of correctness and flexibility, the C5.0 decision tree algorithm is an excellent option for web service categorization jobs. This method is an enhanced iteration of previous approaches such as C4.5 and ID3, and it has exhibited outstanding efficiency. Noise tolerance, management of missing data, and error trimming are some of the elements that are included into the C5.0 method. These properties, when taken together, contribute to the algorithm's increased resilience and flexibility.

Using the QoS characteristics of online services, the algorithm divides services into classifications such as Excellent, Good, Average, and Poor [6]. Each of these groups are determined by the algorithm. The first step in the process of categorizing is the development of a decision tree using data for training that reflects the QoS characteristics of a number of different web services' qualities. There are a number of choice criteria represented by the nodes in the tree, and the branches of the tree indicate all of the results that may occur depending on those criteria. Consequently, the leaf nodes correspond to the final categorization categories that have been determined [7]. Because of this organized method, the methodology is able to categorize online services that have not yet been viewed with a high degree of accuracy, thereby making it easier to identify services that are effective.

In order to achieve correct categorization, choosing features is an essential component of the process of classification. This process entails determining which quality of service characteristics are the most relevant [8]. Not only does feature selection lower the overall size of the dataset, but it also gets rid of characteristics that are redundant or useless, which might potentially impair the correctness of the model. Genetics algorithms, which are based on the fundamentals of natural choosing, have been used as an efficient way for selecting features in the approach that has been suggested. These techniques refine an ensemble of potential options in an iterative manner, finally convergence on a subset of characteristics that is optimum. The classifier method's overall effectiveness is improved as a result of the use of evolutionary algorithms and decision tree classifiers. Cross-validating is an empirical method that is used to verify the dependability and generalization of the system for classification. It is an additional essential component of the process [9-10]. Through the process of cross-validating, the dataset is partitioned into various subsets for the purposes of training as well as testing. This guarantees that the efficiency of the model is not influenced by particular data partitions. In addition, error pruning methods are used in order to make easier the decision tree by deleting branches that contribute only a minor amount to the reliability of categorization. The complexities of the model is reduced as a result of this, and the danger of overfitting is reduced as well. Overfitting occurs when the models operate well on data that it has been trained on but badly on data that it has not seen before. The approach that has been developed tackles a number of problems that are associated with traditional categorizing methods [11]. These limitations include managing an abundance of noisy data, the calculation of the appropriate tree depth, and the treatment of missing values for attribute values. A substantial accuracy in classification of 97% is achieved by the technique by the use of the capabilities of the C5.0 algorithm, genetic codes, and error trimming. With this degree of accuracy, the dependability of online provider selection is considerably improved, and consumers are given the chance to make well-informed judgments based on the QoS characteristics that are most important to them. The organization and categorizing of web-based services according to QoS characteristics has far-reaching ramifications for a variety of apps that use them. Enterprises may use the categorization model to pick online services that fit with their operational requirements, such as lowering response times or optimizing dependability, particularly in relation to e-business [12]. For example, enterprises can use that framework in choosing web services.

Additionally, in the realm of cloud computing, vendors of services have the ability to use the categorization structure in order to give consumers with suggestions that are specifically suited to meet their requirements for speed. In addition, the technique allows for the building of combined web services, which are constructed by combining different services in order to satisfy the needs of complicated individual users [13]. By correctly classifying specific services, the system guarantees that the composites service will have just those elements that

DOI: https://doi.org/10.54216/AJBOR.120201
Received: June 14, 2024 Revised: September 25, 2024 Accepted: December 13, 2024

are particularly appropriate for it. This results in an increase in both the general effectiveness and the level of pleasure experienced by subscribers. In order to provide an illustration of the approach that has been provided, a block diagram that outlines the primary steps of the categorization process has been shown. Obtaining a dataset that contains QoS characteristics for a number of different online services is the first step in the process. A pre-processing step is performed on this dataset in order to deal with missing values and get rid of noise. This step ensures that the input data is appropriate for categorization.

The subsequent step entails the development of the decision tree via the use of the C5.0 algorithm, while the identification of features is carried out simultaneously through the utilization of human genetic approaches [14-15]. In order to guarantee the dependability and precision of the decision tree that was produced, it is then verified via the process of cross-validation. After that, the tree undergoes pruning in order to lessen the amount of detail and improve the generalization, which ultimately results in the categorization of web-based services into subcategories that have been set. A robust and effective structure for website service categorization is produced as a consequence of the integration of those elements. This structure is able to solve the issues that are given by the dynamic and varied nature of online services. Not only does this technique make the entire process for service discovering more efficient, but it also gives consumers the ability to make judgments based on facts when picking online services that are the most suitable for their individualized requirements.

By automated the categorization process, the structure cuts down on the amount of time and effort that is necessary for services selection [16]. This enables users to concentrate on using the services that have been picked in order to accomplish their goals. It can be concluded that the categorization of online services according to QoS characteristics is an essential factor that enables efficient and successful consumption of services in the digital era. A substantial step forward in this field is represented by the technique that has been suggested, which places a focus on decision- tree classifiers, algorithmic genetics, and errors pruning. Because of its capacity to properly classify services according to the functionality that they possess, it guarantees that users are able to make choices that are well-informed, which in turn maximizes the value that can be generated from web services [17]. The necessity of categorization methods that are both robust and extensible will only increase as the online ecosystem keeps on undergoing further development, which highlights the significance of the technique that is provided in this section by highlighting its relevance.

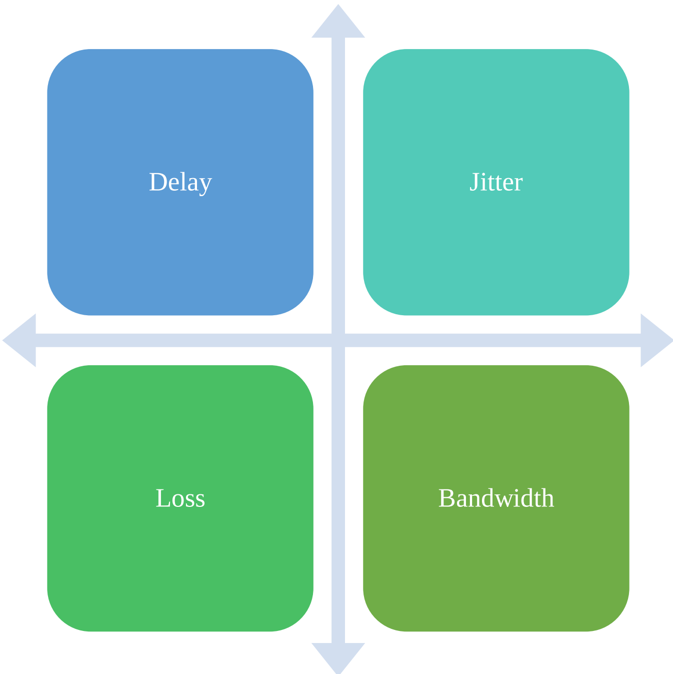

Figure 1. Features of Innovation

This image illustrates the four basic forms of traffic that are measured by QoS in networks. These categories of traffic include bandwidth, delay, loss, and jitter. For the purpose of determining the capability of a network, bandwidth is defined as the quantity information that can be transferred across a system during an interval of time [18]. Delay-sensitive applications, such as video conferences or voice over internet protocol (VoIP), need a delay that monitors the amount of time needed for information to make it from the sender's end to the receivers. As a crucial indicator for measuring the dependability of a system, network loss is the number of data packets that are

DOI: https://doi.org/10.54216/AJBOR.120201
Received: June 14, 2024 Revised: September 25, 2024 Accepted: December 13, 2024

unable to arrive at their intended location. Jitter, which monitors the variance in packet delivery timings, is ultimately responsible for causing interruptions in real-time services like as playing games online or streaming videos. These interruptions have the potential to negatively impact the performance and overall performance of the service. For apps that need high integrity of data and low delay, these criteria are extremely important when it comes to analysing and enhancing the efficiency of the network.

2. Existing Work

Web service categorization using quality of service metrics was investigated by researcher using the J48 method. His research shed light on how to rank online services based on their performance features using confusion matrix evaluation, which he used to evaluate the precision of service selection [19-20]. An easy way to categorize services according to reaction time and dependability is provided by the author technique, which stands out for its efficacy in big datasets.

Table 1: Overview of related work

Method	Advantage	Strength of the Work	Research Gap
J48 Algorithm [21]	Sorts services according to their quality of service using a decision tree and a confusion matrix.	Effective categorization for massive datasets, simple to understand.	Problems with scalability, controlling noise, and maintaining data integrity.
C5 Classifier [22]	Revised and enhanced version of C4.5; now handles missing data and the overfitting more effectively.	Resolves missing values and huge datasets with ease; dependable.	Deals poorly with situations when the service quality of service is changing quickly.
Ranking Model with PCA [23-24]	Creates subsets of the dataset that are smaller to facilitate categorization.	Accelerates the process of finding and classifying services.	Potentially need sophisticated trimming methods due to sensitivity to superfluous characteristics.
Naive Bayesian Network (SWAM) [25]	Sorts features according to their categorization weights using Principal Component Analysis (PCA).	Service rankings using quality of service metrics are quite accurate.	Too computationally heavy for use in real-time applications; performs poorly in fluid settings.
Genetic Algorithm	Quick and easy categorization for simpler datasets.	Simple to execute and uses little computing power.	A decline in performance while dealing with massive amounts of multi-dimensional quality-of-service data.
Cross-Validation	Reducing unnecessary features is the goal of feature selection using the wrapper technique.	By picking the most relevant quality of service criteria, the model's accuracy is improved.	Relies on static datasets alone; more study is needed to include dynamic QoS changes.

DOI: https://doi.org/10.54216/AJBOR.120201
Received: June 14, 2024 Revised: September 25, 2024 Accepted: December 13, 2024

Error Pruning	Contributes to the model's validation by allowing for the testing of subsets of data.	Strengthens the categorization model and makes it more applicable to different scenarios.	Optimizing is necessary for complicated multidimensional elements and large-scale datasets.
Artificial Neural Networks (ANN)	Trims the decision tree of branching that don't add accuracy.	Makes the model more efficient and aids in avoiding overfitting.	When trimming is very severe, it might cause under fitting.
Support Vector Machines (SVM)	Adapts weights to ensure precise categorization based on learned data.	Extremely flexible, particularly when dealing with complicated, nonlinear data.	Big datasets need a lot more processing power and a lot of time for training.
K-Nearest Neighbors (KNN)	Uses hyperplanes for data classification with the goal of maximizing margin among classes.	Impressive generalizability; works well with high-dimensional feature spaces.	Complex; requires improved management of non-linear correlations in quality-of-service data; energy demanding to compute.
Random Forest Classifier	Closest-point-based classification is a non-parametric technique.	Easy to understand and use; suitable for datasets of a modest to medium size.	Struggles with high-dimensional data; slower with larger datasets.
Fuzzy Logic Systems	A technique that uses an ensemble of decision trees to improve accuracy.	Constructive, resistant to overfitting, and adept in dealing with missing data.	Potentially inefficient when dealing with high-dimensional datasets on a big scale.
Clustering Methods (K-Means, DBSCAN)	Deals with lack of clarity and precision while classifying data.	The scenario is effective when the QoS criteria are unclear or both.	Requires fine-tuning to account for QoS variations; limited by the number of clusters chosen.
Markov Decision Processes (MDP)	Organizes services into groups according to their shared quality of service characteristics.	Assists in categorizing services based on shared performance measures.	Mathematically difficult; difficult to execute in large-scale real-time scenarios.
Principal Component Regression (PCR)	Models the process of classifying services as a series of decisions.	Beneficial for ever-changing situations with changing quality of service circumstances.	Dimension reduction may lead to data loss, which can impact how complicated datasets are processed.

DOI: https://doi.org/10.54216/AJBOR.120201
Received: June 14, 2024 Revised: September 25, 2024 Accepted: December 13, 2024

Linear Discriminant Analysis (LDA)	Directs attention to primary components, which decreases dimensionality and boosts accuracy in classification.	Facilitates the preservation of critical information while minimizing feature space.	Unsuitable for QoS data with complicated, nonlinear connections.
Evolutionary Algorithms (EA)	Determines the most effective linear combination of attributes for classifying data.	Useful when classes can be easily separated along a linear path.	Needs a lot of processing power and tweaking to work under changing quality of service circumstances.

The approach may encounter scaling issues in dynamic settings where QoS measures are subject to frequent changes, and the study has shown that it has trouble dealing with datasets that include noise. C5 classifier, developed by the researcher is an upgrade to C4.5 that enhanced classification accuracy and dealt with missing data better. C5.0 is well-suited for bigger datasets since it has improvements in memory economy and overfitting avoidance. However, researcher did note that these models aren't very flexible, therefore they can't keep up with the dynamic QoS requirements of online services.

3. Objective of the research work

Presenting research in this chapter has as its goal the development of a reliable system for categorizing web services according to their QoS characteristics. Classifying services as Excellent, Good, Average, or Poor according to critical quality of service metrics including availability, performance, reaction time, and dependability is the aim to enhance the website's service choice process. An effective model for web service classification may be constructed using this approach by using decision tree classifiers, most especially the C5.0 algorithm. Feature selection using neural networks is another area of study; this technique aids in the optimization of the categorizing procedure by removing superfluous or unnecessary features, thereby improving the reliability of the model. This study further improves the categorizing model's generalizability and durability by using cross-validation and error reduction. The end goal is to create a trustworthy, accurate, and scalable system for finding and choosing online services so that organizations and individuals may make better decisions in complex, ever-changing settings.

4. Motivation for the research work

An ever-increasing variety and number of online services is driving the research discussed in this chapter. With more and more online services being relied upon by both organizations and consumers, the difficulty of choosing the best service according to non-functional criteria like availability, dependability, and reaction time becomes paramount. Manual selection is time-consuming and prone to mistakes since there are so many providers out there with comparable features but different performance. Because of this, there is a need for a scalable, effective, and automatic way of categorizing and sorting web services according to their QoS qualities.

5. The Projected method

One of the many methods that are used in the method of classification is the C5.0 algorithm, which is an improvement on the C4.5 algorithm. When applied to the huge data collection, the classification method known as C5.0 is used. GALA12 is the acronym. The performance and memory of C5.0 are superior to those of C4.5 thanks to the improvements. Partitioning the sample according to the area that offers the most information gain is how the C5.0 model does its task. Through the use of the C5.0 model, samples may be divided according to the information gain field that is the most significant.

Among the many algorithms that may be used to form the decision tree, some of the more prominent ones are ID3, C4.5, and C5.Quinlan is the one who first presented the ID3 categorization technique, which is used to generate decision trees based on the data and constraints that are provided. There are three different kinds of data sets that are used. The layout of the file is determined by the non-categorical properties, which are the first kind. Both the initial training set and the test data set are included in both of the subsequent sets, respectively. When categories data is considered, it is possible to discover categorical values, which often include values such as "good," "average," and "satisfactory". There is a lot of difficulty in dividing the continuous spectrum of data using the ID3 method. With the implementation of the c4.5 algorithm, which adheres to the same criteria as ID3, this issue has been addressed.

DOI: https://doi.org/10.54216/AJBOR.120201
Received: June 14, 2024 Revised: September 25, 2024 Accepted: December 13, 2024

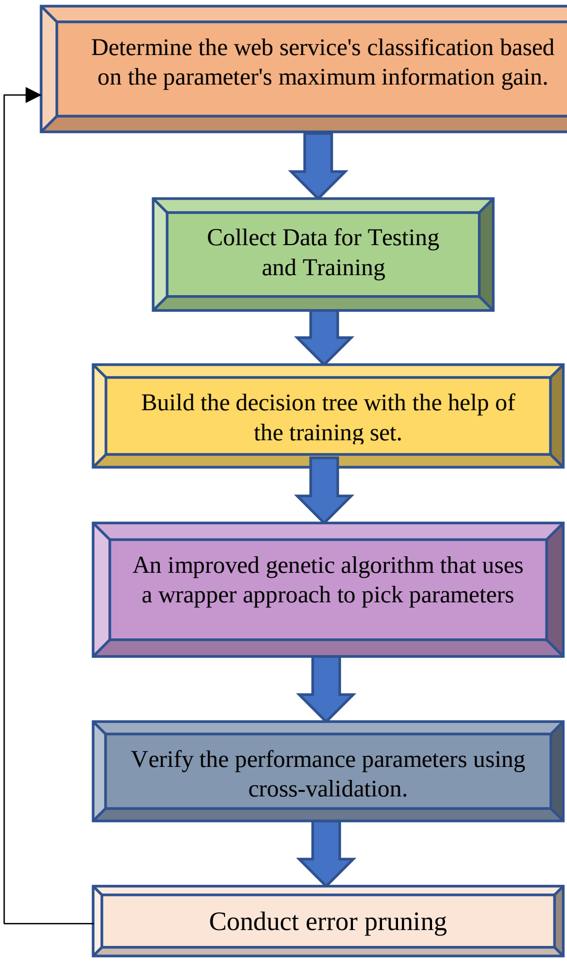

Figure 2. The suggested procedure Web service categorization graphic

The C5 classifiers is a better version of the C4.5 algorithm, and it adheres to the principles that were used by the C4.5 method. In addition to that, it comes with a few extra features already included:
1. the extensive decision tree may be seen as a collection of guidelines.
2. Provides an acknowledgement of the noise & the data that is missing
3. A solution has been found for the issue of overfitting and error pruning
4. It is feasible to anticipate aspects that are relevant to the situation.

As a result of these extra attributes, this technique has been improved and used to the categorization of web-based services in order to get a compositional structure that is effective. It is possible for decision trees to have a number of shortcomings, including but not limited to the following: irrelevant attributes, decision making borders, duplication of sub-trees, continuing category attribute, emphasis on important characteristics, and lacking values of attributes. These issues are addressed by the technique that has been presented, which makes use of picking features, error pruning, validation across features, and validation of model's functions. In addition, it establishes the level of complexity of the decision tree.

$$J(R) = -\sum_{j=1}^{N} q_j log_2(q_j) \quad (1)$$

The variability or impurity of a dataset is measured by its entropy. The percentage of samples in class j is denoted as q_j , and there is a total of m classes.

$$F(B) = -\sum_{i=1}^{m} q_{m,n}\, log_2(q_{m,n}) \quad (2)$$

All n potential splits (such as Excellent, Good, Average, and Poor) for an attribute A are accounted for by entropy. The likelihood of a sample being a member of class i in subset j is denoted as $q_{m,n}$.

$$G(B) = J(R) - F(B) \quad (3)$$

DOI: https://doi.org/10.54216/AJBOR.120201
Received: June 14, 2024 Revised: September 25, 2024 Accepted: December 13, 2024

G(B): Decreasing entropy when dataset is partitioned according to attribute B. The entropy of the whole dataset prior to splitting is denoted by $J(R)$. Weighted entropy after splitting by B is denoted as $F(B)$.

$$GR(B) = \frac{G(B)}{SI(B)} \quad (4)$$

$$SI(B) = -\sum_{j=1}^{N} q_j log_2(q_j) \quad (5)$$

Where G(B): Decreasing entropy when dataset is partitioned according to attribute B. The entropy of the whole dataset prior to splitting is denoted by I(S). Weighted entropy after splitting by B is denoted as E(A). The intrinsic knowledge acquired by dividing data by B is measured by $SI(B)$. Post-splitting percentage of samples in subset j is denoted as q_j .

$$RT = v_E - v_S \quad (6)$$

When the service finishes responding, the time is represented by v_E. When the service that received the request is started, the time is represented by v_S.

$$ND = E + M + C \quad (7)$$

Decision tree depth (E), node count (M), and branch count (C) added together.

$$E(y) = \frac{1}{1+F(y)} \quad (8)$$

The fitness coefficient $E(y)$ determines how good a solution y is, whereby the error in misclassification of the parameters that were chosen is denoted as $F(y)$.

$$D = \propto Q_1 + (1-\propto)Q_2 \quad (9)$$

Creates a new potential candidate D by merging parental solutions Q_1 and Q_2, with the blend factor α controlling the process.

$$N_j = Q_j + \beta\Delta \quad (10)$$

Modifies each unique Q_j by a little amount (βΔ), which introduces unpredictability.

$$F_q = \sum_{j=1}^{m} z_j . f_j \quad (11)$$

An error f_j at each node in the subtree, with a weight of z_j .

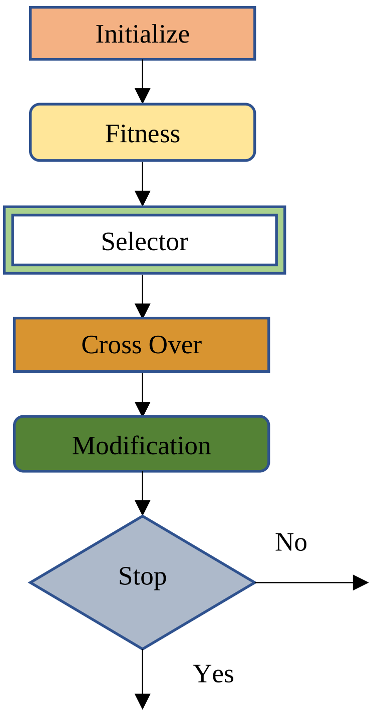

Figure 3. A Genetic Algorithm Using the Wrapped Approach

DOI: https://doi.org/10.54216/AJBOR.120201
Received: June 14, 2024 Revised: September 25, 2024 Accepted: December 13, 2024

Through the process of choosing a feature, a subset of traits is chosen from the initial feature set while any change taking place. This ensures that the material values of the initial features are preserved. In the field of artificial intelligence and data mining, the approach of reducing dimension was used for the choice of features. The quicker model is constructed by the process of selecting characteristics, which helps eliminate features that are unnecessary, redundant, and noisy. This is accomplished by lowering the number of features. In order to get a more accurate estimate, genetic methods are used to populations of people. Heuristics for searching are used in the process of selection by nature. It employs a group of people, which is referred to as the general public, in order to provide a solution to the issue. Genes provide the basis for a person's characteristics. In order to form a chromosome, the genes are linked together. Genes are referred to as factors in this context, whereas chromosomal are denoted as solutions. Ones and zeros are used to form a string that is used to represent persons.

5.1 Fitness: A fitness score is assigned to every single person by the system, which then determines the status of each person in terms of selecting for reproducing.

5.2 Selector: The selection of persons based on their fitness assessments.

5.3 Cross Over: Selected from the random genetic material of the parents who were matted.

5.4 Modification: Preserves the variety of populations in order to forestall the occurrence of early convergence.

The process of analysing and contrasting different learning methods is known as cross-validation. This technique involves separating data into two phases: the first phase is used for learning or train a model, and the second phase is used to verify the model's accuracy. In order to achieve higher levels of accurate classification, the level of detail of the model should be raised. Changing the parameters of the model will result in a boost in its level of complexity. Lessening of Errors In artificial intelligence, pruning is a strategy that minimizes the total number of decision trees by deleting areas of the tree that give little ability to categorize cases. This approach is used to help decrease the complexity of decision trees. Through the decrease of overfitting and the elimination of portions of a classification that might be generated from noisy or incorrect data, pruning serves two purposes: first, it reduces the level of difficulty of the final algorithm and second, it improves the assumed precision of the classification method. The C5.0 model has the capability to divide samples according to the information gain field that is the most significant. Having said that, there are several challenges involved in the process of learning decision trees. Choosing how deeply to build the decision tree is a challenging choice to make because of the complexity involved. To add insult to injury, selecting an effective choice of attribute method and managing training data that contains missing value for attributes are also challenging tasks. All of the problems have been fixed, and the complex nature of the design has also been studied in the method that has been presented.

There are many degrees of service providing attributes that are characterized by the service categorization. When it comes to services, there are four categories: Outstanding, Very good, average, poor. The classification is distinguished based on the overall quality assessment of the chosen parameters of QoS and the average values of the parameters, as was stated in phase I. A classification and selection of the most important amenities inside the tree structure is carried out by the suggested algorithm. To begin, a classifier is trained and put through its paces. After that, the decision tree that was produced is used to categorize data that has not yet been viewed. During choosing attributes, it is just concentrating on the features that are relevant to the application.

$$B_{DU} = \frac{1}{l}\sum_{j=1}^{l} B_j \qquad (12)$$

An B_{DU}: Accuracy in cross-validation. l is the cross-validation fold count. The accuracy for the j-th fold is denoted by B_j.

$$P_{normalize} = \frac{P_j - P_n}{P_m - P_n} \qquad (13)$$

Value of the normalized QoS is P_j. The first quality-of-service value. P_n , P_m : The dataset's minimum and greatest perceived quality of service values.

$$H(B) = 1 - \sum_{j=1}^{n} q_j^2 \qquad (14)$$

B: The quality that is being assessed. m: The total number of classes. q_j^2: the proportion of the samples in class i. Splits are better when the $H(B)$ value is lower.

$$Q = \propto . E + \beta . P \qquad (15)$$

Using the overfitting measure (P) and tree depth (E), it penalizes inappropriate training. α and β: Crime weights.

$$G_o = am_G L(G) \qquad (16)$$

DOI: https://doi.org/10.54216/AJBOR.120201
Received: June 14, 2024 Revised: September 25, 2024 Accepted: December 13, 2024

The loss function $L(G)$ is dependent on the hyperparameter (G).

$$P_a = \sum_{j=1}^{n} z_j . P_j \quad (17)$$

$$\sum_{j=1}^{m} z_j y_j + c = 0 \quad (18)$$

The feature y_j and the weight allocated to it are denoted by z_j, while the bias term is denoted by c. P_j: Quality of Service measure j (for instance, reaction time, availability). There is a total of m quality of service measures, and each metric has a weight of z_j.

6. Results

Evaluation of the efficiency of the suggested approach is carried out using the metrics that have been given as well as its parameters. This section also includes a discussion of the parameters and metrics in question. The QoS characteristics and measures that are used in this suggested system.

6.1 Response Time: The amount of time it took for the web service to reply to the request. The web service is evaluated using it. Web service users demand lower response latency (ms).

6.2 Reliability: A web service's capacity to carry out its specified operations within a certain time frame and set of parameters.

6.3 Performance: Performance is a multi-dimensional statistic that measures how well the system classifies online services, how well it adapts to new situations, and how efficiently it uses its resources.

6.4 Availability: When the service is called, this is a signal that the network is ready to go. In order to keep their customers happy, providers of services ought to deliver their online services with a high availability ratio.

Table 2: Investigative DL models in connection to suggested approaches.

Web Services	
Services	Response Time (ms)
Service 1	106.00
Service 2	321.50
Service 3	783.81
Service 4	524.11
Service 5	538.50
Service 6	249.00
Service 7	78.00

The seven online services' response times show how different they are in terms of efficiency. With a response time of only 78.00 ms, Service 7 tops the charts for quickness and effectiveness. In contrast, Service 3's reaction time of 783.81 ms is the longest of the bunch, which makes it the least ideal for uses that need instantaneous replies. Both Service 1 (106.00 ms) and Service 6 (249.00 ms) have comparatively low response times, indicating that they are quicker than the majority of another services. Services 2, 4, and 5 are in the center of the pack, with reaction times of 321.50 ms, 524.11 ms, and 538.50 ms, respectively, indicating more slowly efficiency. From what we can see from these differences, Service 7 is the best option for jobs that need a quick response, although Services 3, 4, and 5 could work better in situations where it is not strictly necessary.

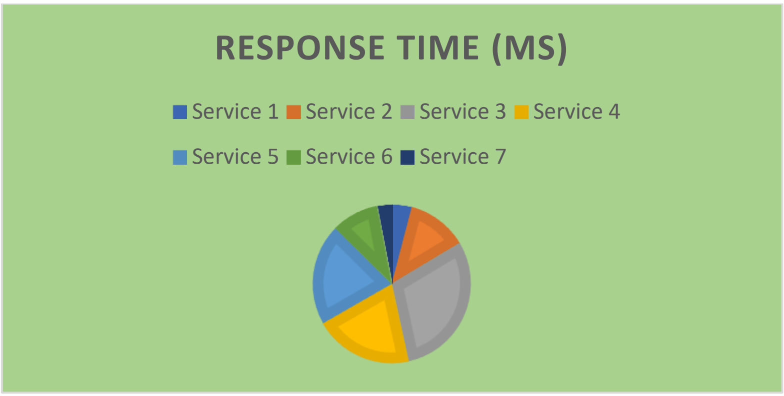

Figure 4. Assessment of ML models in relation to conventional methods.

DOI: https://doi.org/10.54216/AJBOR.120201
Received: June 14, 2024 Revised: September 25, 2024 Accepted: December 13, 2024

Table 3: We compare the proposed technique with current methods.

Web Services	
Services	Reliability (%)
Service 1	63
Service 2	61
Service 3	89
Service 4	76
Service 5	67
Service 6	73
Service 7	92

Service 7's dependability of 92% makes it the most trustworthy alternative for crucial jobs, however the reliability of the seven online services displays a broad range. With a dependability of 89%, Service 3 followed soon after, demonstrating great proficiency in consistently delivering service. The moderate reliability values of 76% for Service 4 and 73% for Service 6 indicate that they are adequate for many applications. While 67% and 63% dependability ratings are significantly under an average, respectively, for non-critical activities, services 5 and 1 are nevertheless usable. Among the services, Service 2 has the least dependability at 61%, suggesting a greater probability of breakdowns or uneven performances. While Service 2 may need some tweaks to be more dependable, Service 7 is clearly the more reliable option overall.

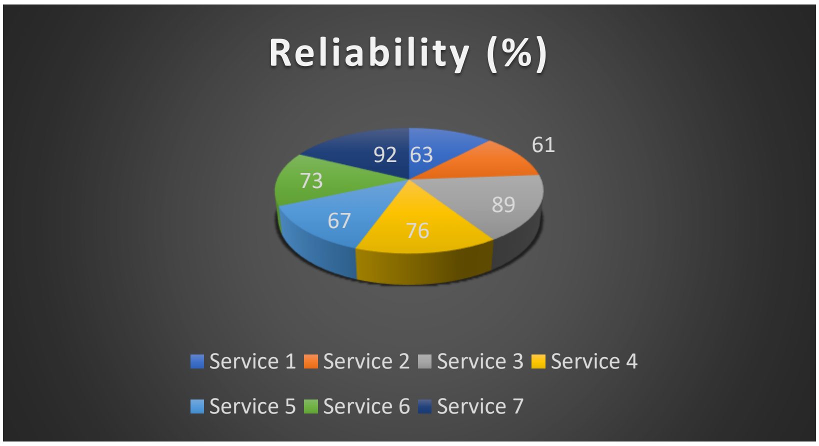

Figure 5. Comparing ML models to more conventional methods for evaluation.

Table 4: Parameters of algorithms sourced from nature

Web Services	
Services	Availability (%)
Service 1	82
Service 2	96
Service 3	94
Service 4	88
Service 5	74
Service 6	71
Service 7	99

DOI: https://doi.org/10.54216/AJBOR.120201
Received: June 14, 2024 Revised: September 25, 2024 Accepted: December 13, 2024

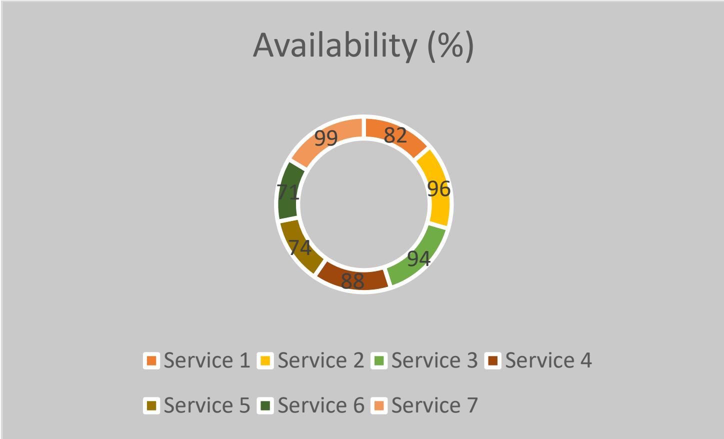

Figure 6. Effectiveness of different system

With an availability rate of 99%, Service 7 is the most trustworthy of the seven online services and is frequently available. Next in line are Service 2 and Service 3, both of which demonstrate excellent uptime performances with availability ratings of 96% and 94%, correspondingly. Application that need constant service access may also take use of Service 4's 88% availability. The availability of Service 1 is moderate at 82%, while that of Services 5 and 6 is the lowest at 74% and 71%, correspondingly. Given these lowered results, it seems they aren't very reliable for mission-critical apps that need a lot of uptime. While Services 5 and 6 could require some serious TLC to live up to the higher availability requirements, Service 7 is clearly the most user-friendly

Table 4: Exploratory DL models in relation to proposed methods

Web Services	
Services	Performance (%)
Service 1	57
Service 2	79
Service 3	81
Service 4	69
Service 5	76
Service 6	85
Service 7	98

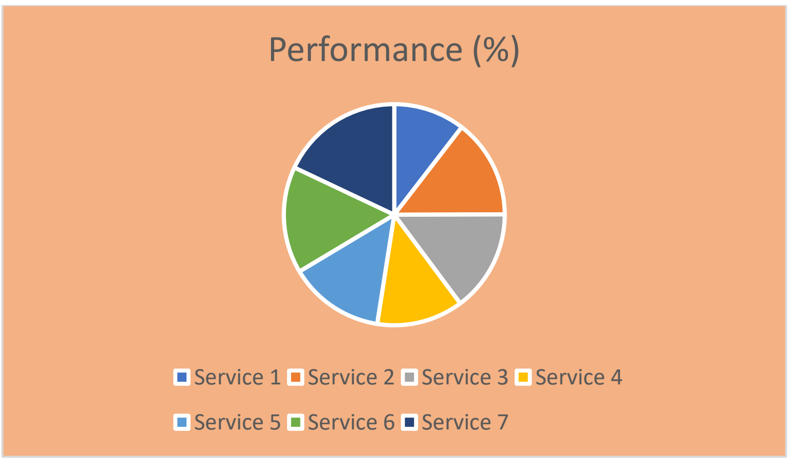

Figure 7. Effectiveness of different models

DOI: https://doi.org/10.54216/AJBOR.120201
Received: June 14, 2024 Revised: September 25, 2024 Accepted: December 13, 2024

After comparing the seven online services' efficiency, we found that Service 7 was the most efficient and reliable, earning a score of 98%. After that, Service 3 and Service 6 both demonstrate significant capabilities that are well-suited to complex applications, with 81% and 85% efficiency scores, correspondingly. With ratings of 79% and 76%, respectively, Service 2 and Service 5 are dependable options for general purposes, falling into a relatively outstanding performance. Category. With 69% efficiency, Service 4 performs around average, while Service 1 comes last with 57%, suggesting it could not be the best choice for jobs that need a lot of power. Services 6 and 3 also achieved outstanding levels of performance, but Service 7 stood out as the highest performer generally.

6. Conclusion

Utilizing decision tree classifiers, in particular the C5.0 algorithm, the chapter concludes by presenting a complete method for categorizing online services according to the QoS characteristic of such services. The technique that has been offered improves the effectiveness of service research and choosing by classifying internet services into categories such as Excellent, Good, Average, and Poor. It is possible to increase the reliability of the decision tree structure while simultaneously reducing its level of detail by including error pruning methods and genetic algorithms for selecting features. Furthermore, the use of cross-validating guarantees the dependability and generality. of a framework that is applied to a wide variety of datasets. Despite these developments, there are still a number of issues and limits that need to be addressed, especially with regard to the management of dynamic and adaptable quality of service indicators. When it comes to real-time services selections in massive networks, the computationally complex nature of rankings algorithms and classifications such as C5.0 eventually results in a bottleneck. Particularly with regard to the management of large amounts of data in real-time settings, the capacity for expansion of the model needs to be the primary emphasis of future study. For the purpose of ensuring the system continues to be both resilient and adaptive, more research is required to modify the categorization models so that they can take into account the quickly changing quality of service features. The investigation of the incorporation of methodologies for machine learning, which include deep learning or reinforcement training, may also result in the development of systems for categorizing that are more efficient and adaptable for software applications. For optimum judgments, decision-makers ought to take seriously various unique elements; future research might apply multi-criteria decision-making procedures to assess new product marketing tactics.

Funding: “This research received no external funding”

Conflicts of Interest: “The authors declare no conflict of interest.”

References

[1] J. Tanha, M. van Someren, and H. Afsarmanesh, ‘‘Semi-supervised self training for decision tree classifiers,’’ Int. J. Mach. Learn. Cybern., vol. 8, no. 1, pp. 355–370, Feb. 2017.

[2] Esraa Mohamed, The Relationship between Artificial Intelligence and Internet of Things: A quick review, Journal of Cybersecurity and Information Management, Vol. 1 , No. 1 , (2020) : 30-34 (Doi : https://doi.org/10.54216/JCIM.010101)

[3] Dr. Ajay B. Gadicha , Dr. Vijay B. Gadicha, Implicit Authentication Approach by Generating Strong Password through Visual Key Cryptography, Journal of Cybersecurity and Information Management, Vol. 1 , No. 1 , (2020) : 5-16 (Doi : https://doi.org/10.54216/JCIM.010102)

[4] B. S. Balaji, S. Balakrishnan, K. Venkatachalam, and V. Jeyakrishnan, ‘‘Retracted Article: Automated query classification based web service similarity technique using machine learning,’’ J. Ambient Intell. Humanized Comput., vol. 12, no. 6, pp. 6169–6180, Jun. 2021.

[5] M. S. Das, A. Govardhan, and D. V. Lakshmi, ‘‘Classification of web services using data mining algorithms and improved learning model,’’ TELKOMNIKA (Telecommun. Comput. Electron. Control), vol. 17, no. 6,p. 3191, Dec. 2019

[6] M.-T. Wu, ‘‘Confusion matrix and minimum cross-entropy metrics based motion recognition system in the classroom,’’ Sci. Rep., vol. 12, no. 1, pp. 1–10, Feb. 2022.

[7] C. Vivek,M. Indu,N. Nandhini, Speech Recognition Using Artificial Neural Network, Journal of Journal of Cognitive Human-Computer Interaction, Vol. 5 , No. 2 , (2023) : 08-14 (Doi : https://doi.org/10.54216/JCHCI.050201)

[8] Vidyul Narayanan,Nithya P.,Sathya M., Effective lung cancer detection using deep learning network, Journal of Journal of Cognitive Human-Computer Interaction, Vol. 5 , No. 2 , (2023) : 15-23 (Doi : https://doi.org/10.54216/JCHCI.050202)

[9] S. Ruuska, W. Hämäläinen, S. Kajava, M. Mughal, P. Matilainen, and J. Mononen, ‘‘Evaluation of the confusion matrix method in the validation of an automated system for measuring feeding behaviour of cattle,’’ Behavioural Processes, vol. 148, pp. 56–62, Mar. 2018.

DOI: https://doi.org/10.54216/AJBOR.120201
Received: June 14, 2024 Revised: September 25, 2024 Accepted: December 13, 2024

[10] N. Agarwal, G. Sikka, and L. K. Awasthi, ‘‘Enhancing web service clustering using length feature weight method for service description document vector space representation,’’ Expert Syst. Appl., vol. 161, Dec. 2020, Art. no. 113682.

[11] M. B. Blake, W. Cheung, M. C. Jaeger, and A. Wombacher, ‘‘WSC-06: The web service challenge,’’ in Proc. 8th IEEE Int. Conf. E-Commerce Technol., 3rd IEEE Int. Conf. Enterprise Comput., E-Commerce, EServices (CEC/EEE), Jun. 2006, p. 62.

[12] Denis A. Pustokhin , Irina V. Pustokhina, FLC-NET: Federated Lightweight Network for Early Discovery of Malware in Resource-constrained IoT, International Journal of Wireless and Ad Hoc Communication, Vol. 6 , No. 2 , (2023) : 43-55 (Doi : https://doi.org/10.54216/IJWAC.060204)

[13] Mohamed Saber , El-Sayed M. El-Kenawy , Abdelhameed Ibrahim , Marwa M. Eid , Abdelaziz A. Abdelhamid, Metaheuristic Optimized Ensemble Model for Classification of SMS Spam in Computer Networks, International Journal of Wireless and Ad Hoc Communication, Vol. 6 , No. 2 , (2023) : 56-64 (Doi : https://doi.org/10.54216/IJWAC.060205)

[14] A. V. Dastjerdi and R. Buyya, ‘‘Compatibility-aware cloud service composition under fuzzy preferences of users,’’ IEEE Trans. Cloud Comput., vol. 2, no. 1, pp. 1–13, Jan. 2014.

[15] M. Razian, M. Fathian, H. Wu, A. Akbari, and R. Buyya, ‘‘SAIoT: Scalable anomaly-aware services composition in CloudIoT environments,’’ IEEE Internet Things J., vol. 8, no. 5, pp. 3665–3677, Mar. 2020.

[16] M. Gao, M. Chen, A. Liu, W. H. Ip, and K. L. Yung, ‘‘Optimization of microservice composition based on artificial immune algorithm considering fuzziness and user preference,’’ IEEE Access, vol. 8, pp. 26385–26404, 2020.

[17] R. Buyya, S. N. Srirama, G. Casale, R. Calheiros, Y. Simmhan, B. Varghese, E. Gelenbe, B. Javadi, L. M. Vaquero, and M. A. Netto, ‘‘A manifesto for future generation cloud computing: Research directions for the next decade,’’ ACM Comput. Surv., vol. 51, no. 5, pp. 1–38, 2018.

[18] Hayder Sabah Salih,Fatema Akbar Mohamed, Fusion-based Diversified Model for Internet of Vehicles: Leveraging Artificial Intelligence in Cloud Computing, Journal of Fusion: Practice and Applications, Vol. 12 , No. 2 , (2023) : 54-69 (Doi : https://doi.org/10.54216/FPA.120205)

[19] Zeena N. Al-kateeb,Dhuha Basheer Abdullah, A Smart Architecture Leveraging Fog Computing Fusion and Ensemble Learning for Prediction of Gestational Diabetes, Fusion: Practice and Applications, Vol. 12 , No. 2 , (2023) : 70-87 (Doi : https://doi.org/10.54216/FPA.120206)

[20] G. Chen, T. Jiang, M. Wang, X. Tang, and W. Ji, ‘‘Modeling and reasoning of IoT architecture in semantic ontology dimension,’’ Comput. Commun., vol. 153, pp. 580–594, Mar. 2020.

[21] W. Chen, B. Liu, H. Huang, S. Guo, and Z. Zheng, ‘‘When UAV swarm meets edge-cloud computing: The QoS perspective,’’ IEEE Netw., vol. 33, no. 2, pp. 36–43, Mar./Apr. 2019.

[22] S. K. Gavvala, C. Jatoth, G. R. Gangadharan, and R. Buyya, ‘‘QoS-aware cloud service composition using eagle strategy,’’ Future Gener. Comput. Syst., vol. 90, pp. 273–290, Jan. 2019.

[23] M. Vučetić, M. Hudec, and B. Božilović, ‘‘Fuzzy functional dependencies and linguistic interpretations employed in knowledge discovery tasks from relational databases,’’ Eng. Appl. Artif. Intell., vol. 88, Feb. 2020, Art. no. 103395.

[24] M. Sözat and A. Yazici, ‘‘A complete axiomatization for fuzzy functional and multivalued dependencies in fuzzy database relations,’’ Fuzzy Sets Syst., vol. 117, no. 2, pp. 161–181, Jan. 2001.

[25] B. Sheng, O. M. Moosman, B. Del Pozo-Cruz, J. Del Pozo-Cruz, R. M. Alfonso-Rosa, and Y. Zhang, ‘‘A comparison of different machine learning algorithms, types and placements of activity monitors for physical activity classification,’’ Measurement, vol. 154, Mar. 2020, Art. no. 107480.

DOI: https://doi.org/10.54216/AJBOR.120201

Received: June 14, 2024 Revised: September 25, 2024 Accepted: December 13, 2024