Computation of Accessibility Score of Educational Institute Webpages using Machine Learning Approaches

Jinat Ara, and Cecilia Sik-Lanyi

***Abstract*—The availability of digital platforms by ensuring accessibility and usability is considered a virtual gateway to provide a wide array of information to its stakeholders. An accessible web platform can disseminate information among a variety of target audiences. Thereby accessibility of academic web pages requires special attention. Herein we proposed an accessibility computation approach for higher education institute webpage (Homepage) in the context of universities in Hungary. The proposed approach incorporated two machine learning (ML) classifiers: Random Forest (RF), and Decision Tree (DT) to experiment on our custom dataset to compute the overall accessibility score. Performance of ML methods validated through confusion matrix and classification report result. The empirical results of ML methods and statistical evaluation showed poor accessibility scores which depicts that none of the selected web pages are free from accessibility issues associated with disabilities. As such, accessibility is a crucial aspect that needs further concern as most of the considered academic webpages have experienced accessibility issues and showed improvement demands.**

***Index Terms*—Accessibility validation; machine learning methods; questionnaire analysis; academic institute.**

I. Introduction

The growth of web technologies brought immense progress in different spheres of life, especially in accessing digital information [1][2] from various platforms, such as the healthcare sector, banking sector, education sector, and others [3]. In this modern era, focusing on any sector, having a digital platform (e.g., webpages) for providing information is inevitable. In many cases, web pages act as the primary resources, especially for communicating with various stakeholders. Therefore, different stakeholders access web pages frequently for their required information. For example, from the perspective of the university, academic and prospective students might need to download their course curriculum, class schedule, campus news, admission requirements, and other information. Thus, digital platforms or webpages should be accessible to serve their stakeholders to compete globally. However, the emerging concern is that most web pages are not developed concerning the accessibility perspective [4].

The authors are with the Department of Electrical Engineering and Information System, University of Pannonia, Veszprem, Hungary.
E-mail: {jinat.ara@mik.uni-pannon.hu, lanyi.cecilia@mik.uni-pannon.hu).

These developments limit the universal inclusion of potential users. Accessible web generally refers to the design and development of websites (multiple web pages) in a manner that is effective for people with disabilities and without disabilities [5]. Nowadays, web designers and developers are trying to incorporate several complex functionalities (e.g., dynamic, drop-down menu) and components (e.g., images, videos) into their web pages to make them more interactive. Though these interactive functionalities are prominent to attract more people, they limit the accessibility concept for users with disabilities [6] [7]. Therefore, it is increasingly important to design and develop web pages considering accessibility manners and follow accessibility practices to serve equal access to resources. Concerning educational institute webpages, the introduction of accessible university webpages is not only beneficial for students with impairments but also for the university authorities and other associate practitioners for their academic progress. From the experience of the COVID-19 pandemic, unlike in other countries, the importance of improving the quality of life for students with disabilities at higher education levels has increased in Hungary. Despite addressing numerous challenges faced by students with disabilities in past studies, the accessibility aspects for higher education students have been overlooked from the beginning. Thus, the consequence of the current accessibility limitation of higher education webpages has grown dramatically, particularly university webpages.

Over the years, researchers from many countries have conducted accessibility evaluations of web pages concerning issues and benefits of people with disabilities at different levels. The record of past literature depicts that the majority of the proposed approach evaluated the effectiveness of the webpages concerning their quality (e.g., broken link, interactivity) and usability (e.g., HTML page, aesthetic, design, page size) [8-10]. Concerning the increasing number of people with disabilities, researchers from different backgrounds (e.g., technology, education, neurodevelopment) have sought to evaluate the accessibility of online platforms [11]. Some researcher contributed their effort to develop an effective approach to validate the web, identify accessibility issues, and compute accessibility barrier scores. Among several approaches, fuzzy inference-based evaluation, regression model development, and variable magnitude approaches are prominent. Besides, nowadays several automated web accessibility testing tools have been developed that provide interactive accessibility

DOI: 10.36244/ICJ.2024.5.6

reports about the tested website. However, these approaches and existing automated tools incorporate some specific attributes of websites according to the Web Content Accessibility Guidelines (WCAG) to evaluate their accessibility status. Though the latest version of WCAG 2.2 [12] is a complete guideline for accessibility features, it has limited consideration about some issues with people with disabilities such as whether the website is active or deactivated, website has a manual text size adjustment option, manual font family adjustment option, manual color adjustment option, user information requirement, CAPTCHA issues, usefulness of internal/external links, used images, inserted video and audio content. As these features are not directly possible to evaluate in an automatic manner, WCAG does not provide a clear indication about these aspects though without considering these aspects, it is not possible to ensure complete accessibility of the developed websites.

In that manner, our prime focus is to evaluate websites considering the above-mentioned aspects and compute the overall accessibility score based on these aspects. The prime challenges of this work related to the prepared dataset as there is no dataset has been found that evaluated websites according to these ten aspects. Thus, to conduct this work, we used our custom dataset that we prepared according to the ten aspects mentioned previously and incorporated two machine learning approaches (Random Forest (RF), and Decision tree (DT)) that might be effective to validate the accessibility of the web platform. Besides, none of the work showed their contribution incorporating machine learning (ML) classifiers to add the benefits in this particular domain, such as accessibility of the digital platform. The classification result of ML methods has been evaluated through classification accuracy, precision, sensitivity, specificity, and F1 score. Overall, this paper aims to contribute to web accessibility research by proposing an accessibility score computation system to validate the accessibility of university web pages. The proposed system is platform-independent and dynamic, thus can evaluate and perform a comparative analysis of any academic webpages of a different country.

The remainder of this paper is structured as follows. Section 2 presents a brief discussion of the background and related work about web accessibility. Section 3 presents the methods and materials of the proposed framework, including dataset preparation and description, and system architecture and design. Section 4 presents the experimental analysis including classifiers and website performance analysis. Section 5 presents a detailed discussion. Finally, the paper is concluded through conclusion and future work.

II. Related Studies

Higher education institutions (e.g., universities) play a vital role in developing society and in the scientific community by educating the young mind. For example, university websites (multiple web pages) are responsible for providing information to the community. As such evaluating webpages, researchers considered several methods to identify the quality of the webpage to ensure the inclusion of stakeholders with disabilities. Numerous past studies focused on the education domain to evaluate their accessibility. For example, Chopra et al. [13] enhanced the importance of e-learning at the higher education level. They conducted a questionnaire-based statistical evaluation of university websites. They added that website quality, service quality, and information quality are leading factors to influence user satisfaction and net benefits, which is crucial to consider during web development. Mittal et al. [14] proposed a website quality evaluation process using fuzzy logic/technique. They incorporated fuzzy logic to assess websites in terms of several metrics such as loading time, response time, mark-up validation, broken link, accessibility error, size, page rank, frequency of update, traffic, and design optimization. In another work, Malhotra et al. [15] focused on website quality prediction through an automated Web Metrics Analyzer tool called Neuro-fuzzy inference models. This study confirms that a fuzzy logic-based website analyzer is feasible for predicting the quality of the website.

Further, few studies proposed several frameworks to evaluate academic or higher institute websites. As such, Rashida et al. [16] developed an automated web-based tool to investigate university websites following the content of information, loading time, and overall performance. Their result shows that most university websites did not meet users' satisfaction. Olaleye et al. [17] proposed a framework called WebFUQII based on the web analytical tools WebQual and SITEQUAL, considering ease of use, processing speed, aesthetic design, interactive responsiveness, entertainment, and trust and usefulness.

Concerning accessibility, Alahmadi [18] proposed a multi-model accessibility evaluation framework for university websites for deaf, visually impaired, and Deaf-blindness students. They incorporated automated tools to generate accessibility reports, source code mining to evaluate media content accessibility errors, and human evaluation to validate the assessment result. Focusing on particular disability types, such as people with vision impairment, Hassouna et al. [19] proposed a framework incorporating manual assessment (user and expert testing) and automatic assessment (Cynthiasays) to evaluate the accessibility of university websites.

Following this, few studies considered only automatic accessibility testing tools to investigate the accessibility of higher institute websites. For example, Verkijika et al. [20] evaluated 26 South African university websites through two automated validators (e.g., AChecker and TAW). This investigation reveals that the appearance of broken links and failing Google Mobile-friendly Test is the frequent issue that leads to accessibility issues. AlMeraj et al. [21] evaluated the accessibility of 41 higher education institution websites in Kuwait. They validated considering several automatic accessibility testing tools, specifically AChecker, Total Validator, MAUVE++, WAVE, and HTML/CSS/ARIA. This evaluation concludes that during the website design and development, accessibility is not accounted which leads to an increased number of inaccessible websites.

Nowadays, web researchers have shown their active participation in considering machine learning methods to evaluate the quality and usability of websites. For example, Dhiman et al. [22] proposed the most recent work that focused on machine learning methods to evaluate the performance of tools to identify website quality. They have implemented logistic regression and six machine learning methods, such as Random Forest, Adaboost, Bagging, Multilayer Perceptron, and Bayes Net. They depict that the machine learning model is more effective in evaluating website quality than other approaches. Addressing this research work, we aim to extend our evaluation by focusing on accessibility validation of the higher institute websites using ML techniques focusing on the eight aspects related to the website features. To evaluate the accessibility, we considered top university web pages in Hungary.

III. Material and Methods

This work has focused on webpage accessibility evaluation incorporating two machine learning classifiers. The prime objective is to observe the performance of the selected ML classifier to identify its effectiveness and then evaluate website accessibility according to the classification result. Figure 1 shows the working diagram that illustrates the materials and methods used in this study.

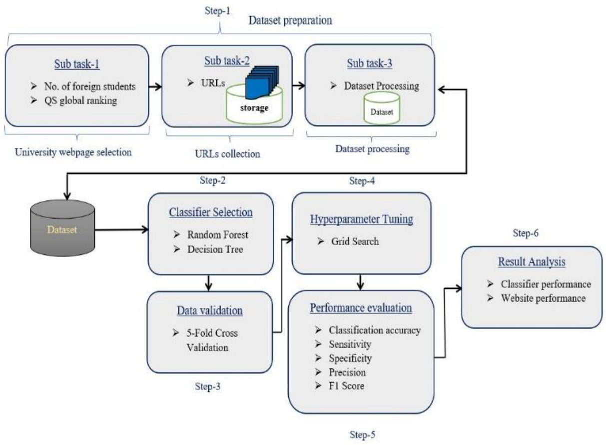

Fig.1. The System Architecture of the proposed model

A. Dataset Preparation

Dataset preparation is represented through multiple sub-tasks to get university webpage information to evaluate their quality and accessibility. University webpage selection, URL collection, and Dataset processing are the three sub-tasks of dataset preparation, as shown in Figure 1.

a) University Webpage Selection: The University webpage is the primary resource for a wide array of information such as departments, subjects, tuition fees, faculty, research groups, scholarships, etc., that help prospective students in their university selection. To investigate the quality and accessibility, initially, university selection is an important and complex task. To conduct this task, we considered Webometrics university ranking [23] for selecting the top five universities in Hungary. Then, we ranked the top five universities according to their number of international students and QS world ranking, as shown in Table 1.

TABLE I
DEMOGRAPHIC INFORMATION OF THE SELECTED WEBSITES

Websites	URLs	No. of foreign students	QS ranking
Web-1	https://u-szeged.hu/english	5000	551
Web-2	https://unideb.hu/en	4000	600
Web-3	https://www.elte.hu/en/	3000	700
Web-4	http://www.bme.hu/?language=en	1900	801
Web-5	https://www.ceu.edu/	962	124 (by subject)

b) Collection of URLs: The Uniform Resource Locator (URL) is a valuable resource for information extraction from the university website. For the selected top five universities, we stored their homepage URLs. Table 1 shows the university list with its URLs.

c) Dataset Processing: For dataset processing, we conducted a preliminary survey to understand the importance of our considered ten (10) aspects (availability, manual text size adjustment option, manual font family adjustment option, manual color adjustment option, user information requirement, CAPTCHA, usefulness of internal/external links, images, inserted video, and audio content) in terms of their effectiveness to the people with disability to represent the accessibility of the websites. Six users participated in this survey including vision problems (4), and cognitive problems (2) as the selected eight features are more likely to cause difficulty to these groups of people with disabilities. All the participants are active on the Internet platform for their professional work and daily activities and are between 25-50 years old. In the online survey, we asked participants to provide their opinions about whether these ten aspects are useful for them to understand the website content effectively. All the users expressed their positive opinion that these aspects are useful for understanding the web content and can improve their experience in internet platform browsing.

According to the positive feedback from the preliminary survey, we prepared our dataset incorporating human observation where 23 participants observed the selected five websites according to the selected ten aspects. All participants were university students from the Department of Electrical Engineering and Information Systems, at the University of Pannonia, Hungary. All the participants had adequate knowledge of the field of interactive design and web development. To evaluate the selected websites, we arranged online participation where we shared our prepared 10 questions related to the considered ten aspects (shown in Table 2) and website resources that need to be observed. All the participants observed the websites and answered each of the questions according to their understanding. After obtaining the responses from users for five selected websites, we labeled the responses in terms of ‘Accessible’, ‘Partially Accessible’, and ‘Not Accessible’ metrics where for all positive responses or ten positive responses, we labeled them as “Accessible”; for having >6 negative response (out of 10), we labeled it as “Not accessible” and rest of the responses having =<6 negative response (out of 10) were labeled as “Partially Accessible”. Figure 2 shows the dataset preparation flowchart to represent the entire process in detail. In total 23 responses were recorded to each dataset related to eight features of the website and classified into three levels of accessibility status. However, the five tested websites' observation results have been incorporated as labeled datasets in our system to conduct the experimental analysis. As our data are categorical, we followed LabelEncoding to encode the data (label/categories). LabelEncoding is a popular categorical data encoding process. We used LabelEncoding through the sklearn LabelEncoder () function.

TABLE II
SELECTED QUESTIONS FOR THE PRELIMINARY SURVEY

Questions	Response
Question-1: Is the webpage available?	Yes/No
Question 2: Does the webpage have a manual text size adjustment option?	Yes/No
Question 3: Does the webpage have a manual font family adjustment option?	Yes/No
Question 4: Does the webpage have a manual color adjustment option?	Yes/No
Question 5: Does the webpage require user information?	Yes/No
Question 6: Does the webpage require CAPTCHA?	Yes/No
Question 7: Do the internal/external links are useful?	Yes/No
Question 8: Does the webpage images are useful?	Yes/No
Question 9: Is the webpage video content useful?	Yes/No
Question 10: Is the webpage audio content useful?	Yes/No

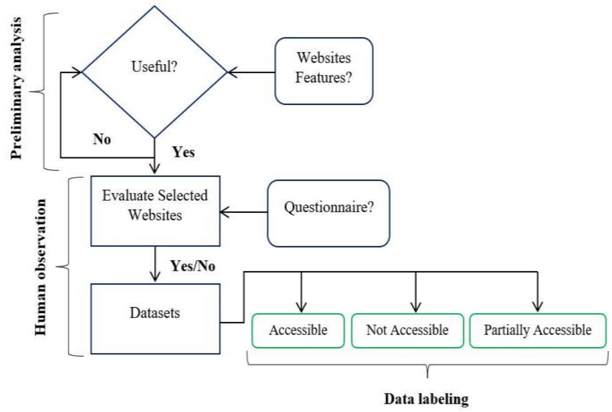

Fig. 2. Flowchart of Dataset preparation

B. *System Architecture and Design*

The system architecture and design have been described through step-2, step-3, step-4, and step-5 (according to Figure 1). These four steps are described in the following subsections:

a) ML Classification Algorithms: Machine learning is a branch of artificial intelligence that employs statistics, probabilities, absolute conditionality, boolean logic, and unconventional optimization strategies to learn and classify patterns through predictive models [24]. ML has both supervised and unsupervised models for classification and regression problems. Thus, we used two most commonly used supervised ML classifiers or algorithms: Random Forest (RF), and Decision Tree (DT).

1) *Random Forest Classifier*
Random Forest (RF) is a popular and most effective supervised machine learning classifier in classification and regression problems. The prime objective of RF is to reduce classification errors. It performs by building a decision tree by taking samples randomly. To classify the output, it takes the majority of the voted result. It aggregates the decisions by taking the average results of all the trees to improve the predictive accuracy and control the over-fitting problem. The advantage of the random forest classifier is to handle the data set containing continuous variables for regression and categorical variables in classification [25]. However, it provides the best performance for categorical variables in classification problems.

2) *Decision Tree Classifier*
Decision Tree (DT) is a supervised machine-learning algorithm for classification problems. It is the most popular and widely used ML algorithm. The prime goal of this algorithm is to predict the output value considering the target value and represent the solved problem in the way of tree representation called a decision tree with a leaf node or decision node [26]. A decision tree has internal and external nodes where the internal node takes part in the decision-making to make decision. In the decision tree, leaf nodes represent the class label, and the internal nodes represent the attributes. The main objective of using the Decision Tree in this work is to predict the target class instances using the decision rule learned from the prior data. In the build decision tree, root nodes classify the instances with different features where root nodes have multiple branches and the leaf nodes represent the classification result. The Decision tree chooses a node according to the highest information gain among all the attributes. The best way to information gain is to calculate entropy. Entropy is the quantified measurement of the amount of uncertainty of random instances, as shown in equation 1, where x= random instance, xi = possible outcomes, and P(xi) = probability of possible outcomes. By the value of entropy for any random instances, we can calculate the information gain through equation 2.

$$Entropy, H(X) = -\sum_{i=1}^{n} P(x|i) log P(x|i) \qquad (1)$$

$$InformationGain = 1 - Entropy \qquad (2)$$

b) Hyperparameter Tunning: It is noteworthy that there are several machine learning classifiers, and every ML classifier requires different constraints, weights, or learning rates to generalize the data patterns. Failure of the appropriate parameter selection might lead to differences in the final result. For example, every iteration of the classifier resulted in a different accuracy. Therefore, to optimally solve the machine learning problem and improve classification accuracy, it is crucial to appropriate parameter selections. The best way to appropriate parameter setting is hyperparameter tuning or hyperparameter optimization. Hyperparameter tuning is the process of optimal parameter setting process for a machine learning model. It is a crucial task to implement any ML model as it directly optimizes the performance of ML classification. It allows for defining all possible parameters for testing all the combinations to maximize the classification results [27]. This optimization process also allows the use of Cross-Validation (CV) to estimate the generalization performance. A crucial aspect that needs to be mentioned is that every ML classifier has different default parameter requirements. Following the default parameter, it allows the setting of the optimization parameter. Several hyperparameter optimization methods are available to select the best parameter. Herein we have used a grid search approach to evaluate classification accuracy with different combinations of parameters using a 5-fold CV. Grid search is a popular method for parameter fitting [28] that is implemented with the Scikit-learn library. For details on how parameters influence the decision of machine learning models, we suggested the literature proposed by Wu et al. [29]. In this study, we employed Grid search as its most traditional hyperparameter optimization technique, also known as parameter sweep. It is the most effective and time-constrained procedure that takes a longer time than other optimization techniques and returns better optimization results. To limit the grid search complexity, seven relevant parameters for RF and DT have been selected. The optimal values are shown in Table 3, and Table 4 which list the selected parameters of RF, and DT with definition, default values, grid values, and the optimal values for the optimization process.

TABLE III
SELECTED HYPERPARAMETER OF RF

Random Forest Parameters				
Parameter	**Definition**	**Default**	**Grid Values**	**Optimal**
criterion	The quality of a split	Gini	Gini, entropy	Gini
max_depth	Depth of a tree	None	None	None
max_features	Maximum feature of a tree	auto	auto, sqrt, log2	auto
min_samples_leaf	Minimum sample in a leaf node of a tree	1	1, 5, 8, 10, 15, 20	1
min_samples_split	Minimum samples to be split	2	2, 5, 10, 15, 20, 25	2
n_estimators	The number of estimators	100	100, 110, 120, 130, 140, 150	130
random_state	The random state of a node	None	None	None

TABLE IV
SELECTED HYPERPARAMETER OF DT

Decision Tree Parameters				
Parameter	**Definition**	**Default**	**Grid Values**	**Optimal**
criterion	The quality of a split	Gini	Gini, entropy	Gini
max_depth	Depth of a tree	None	None	None
max_features	Maximum feature of a tree	None	None	None
max_leaf_nodes	The maximum leaf node of a tree	None	None	None
min_samples_leaf	Minimum sample in a leaf node of a tree	1	1,2,3,4,8,12, 18,20	2
min_samples_split	Minimum samples to be split	2	2,3,4,5,8,10, 15	2
random_state	The random state of a node	None	None	None

c) Cross-Validation (CV): The most effective way to validate the performance of an ML model is to train a model with available data and test its classification performance using a newly separated dataset. Another popular technique is the Train-Test Split. It is the process of data splitting before model development and using the separated data for performance validation. However, these processes require a substantial amount of data for validation. Generally, CV is used to evaluate the performance of learning algorithms or models by partitioning data into a training set for pattern learning and a testing set for model evaluation [26]. The prime idea is to split the dataset into training and test sets according to the user-defined number of partitions such as K-fold. First, the dataset is divided into k folds where k-1 fold is used for training and the remaining fold for testing. In our experiment, we split the entire dataset into 70:30, where 70% of the data was used for training and 30% for testing purposes. We trained our model on the training set considering a five-fold partitioning setting and evaluated the model through a testing set for performance measurement. For hyperparameter tuning, we also applied five-fold partitioning on each fold of the training set.

d) Performance Evaluation Metrics: The evaluation metrics for measuring the classifier performance are derived from the binary confusion matrix, as represented in Figure 3. We employed True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) counts for calculating the classification accuracy, precision, recall, and f1 score of the classifier. TP, FP, FN, and TN have been explained in following through equations 3-7.

True Positive (TP): Number of predicted instances as positive which are originally positive.

False Positive (FP): Number of predicted instances as positive which are originally negative.

True Negative (TN): Number of predicted instances as negative which are originally negative.

False Negative (FN): Number of predicted instances as negative which are originally positive.

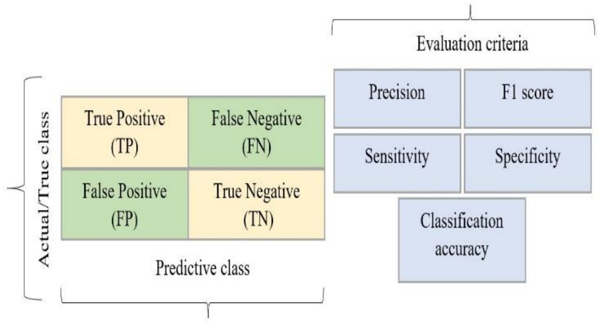

Fig. 3. Confusion Matrix with several evaluation metrics

Classification accuracy: Classification accuracy is the proportion of the number of correctly classified samples (Equation 3).

$$Classification\ Accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (3)$$

Precision: Precision is the proportion of the samples that are actually true (Equation 4).

$$Precision = TP/(TP + FP) \quad (4)$$

Sensitivity: Sensitivity is the proportion of total correctly predicted samples by the learning algorithm (Equation 5).

$$Sensitivity = TP/(TP + FN) \quad (5)$$

Specificity: Specificity is the proportion of the correctly predicted negative sample with all negative samples (Equation 6).

$$Specificity = TN/(TN + FP) \quad (6)$$

F1 score: F1 score is the harmonic mean of precision and recall. To achieve the best performance, F1 score should be one, and for the lowest performance, it's usually zero (Equation 7).

$$F1\ Score = 2 * (Precision * Sensitivity)/(Precision + Sensitivity) \quad (7)$$

IV. EXPERIMENTAL ANALYSIS

This section discusses the experimental setting, evaluation of classifier performance, and evaluation of website performance. The primary aim of this section is to investigate the effects of five datasets with different ratios on selected ML classifiers to verify their performance. This task has accomplished thorough experiments on multiple datasets using selected ML classifiers. Then ML classifier is applied to identify the effects of the dataset and derive the performance of the applied classifier considering various evaluation metrics.

A. Experimental Setting

The experiments with different datasets were performed considering Python programming language in a jupyter notebook environment. The experiments were run on a computer with an integrated 2.5 GHz processor and 8 GB RAM. We employed two ML classifiers: Random Forest and Decision Tree for their efficiency and reliability [30]. Every experiment was performed by taking datasets in CSV format as input to the ML classifier and splitting the dataset into a 70:30 ratio.

B. Performance Evaluation and Accessibility Score Computation

The confusion matrix of experimented datasets is tabulated in Table 5 for two selected classifiers: Random Forest (RF), and Decision Tree (DT), respectively. Table 5 also elucidates different measurements for classification result assessment, such as precision, sensitivity, specificity, F1 score, and overall accuracy to illustrate the effectiveness of the selected machine learning classifier. This table illustrates that the Random Forest (RF) classifier performs well for all five tested datasets.

TABLE V
CLASSIFICATION RESULTS OF FIVE DATASETS USING RF AND DT

Random Forest (tested data)									
		Accessible (0)	**Not Accessible (1)**	**Partially Accessible (2)**	**Classification Metrics**				
Datasets	**Actual class**	**Predictive class**			**Precision**	**Sensitivity**	**Specificity**	**F1 score**	**Accuracy**
Dataset-1	Accessible (0)	**3**	0	0	1.0	0.95	0.99	1.0	0.97
	Not Accessible (1)	0	**9**	0					
	Partially Accessible (2)	0	2	**9**					
Dataset-2	Accessible (0)	**3**	0	0	1. 0	0.86	0.99	1.0	0.99
	Not Accessible (1)	0	**8**	0					
	Partially Accessible (2)	0	0	**12**					
Dataset-3	Accessible (0)	**5**	0	0	1. 0	0.86	0.99	1.0	0.97
	Not Accessible (1)	0	**4**	0					
	Partially Accessible (2)	0	2	**12**					
Dataset-4	Accessible (0)	**1**	0	0	1.0	0.94	0.99	0.97	0.99
	Not Accessible (1)	0	7	0					
	Partially Accessible (2)	0	0	**15**					
Dataset-5	Accessible (0)	**4**	0	0	1.0	1.0	1.0	1.0	1.0
	Not Accessible (1)	0	**9**	0					
	Partially Accessible (2)	0	0	**10**					
Decision Tree (tested data)									
Dataset-1	Accessible (0)	**2**	0	0	0. 97	0.97	0.97	0. 96	0.95
	Not Accessible (1)	0	**8**	2					
	Partially Accessible (2)	0	1	**10**					
Dataset-2	Accessible (0)	**4**	0	0	0.98	0.95	0.93	0.95	0.94
	Not Accessible (1)	0	7	3					
	Partially Accessible (2)	0	0	**9**					
Dataset-3	Accessible (0)	**5**	0	0	1.0	1.0	1.0	1.0	1.0
	Not Accessible (1)	0	**5**	0					
	Partially Accessible (2)	0	0	**13**					
Dataset-4	Accessible (0)	**2**	0	0	0.96	0.94	0.98	0.98	0.94
	Not Accessible (1)	0	**12**	0					
	Partially Accessible (2)	0	1	**8**					
Dataset-5	Accessible (0)	**4**	0	0	0.91	0.92	0.94	0.94	0.92
	Not Accessible (1)	0	**6**	3					
	Partially Accessible (2)	0	0	**10**					

To compute the accessibility score, we quantify the score of each class (0, 1, 2) based on the number of samples of predicted data (Table 5) as shown in Equation 8. We set the severity score based on the importance of three classes as shown in Equation 9. The computed score of each class has been scaled down by multiplying their severity level as shown in Equation 10 and computing the final score through Equation 11.

$$Accessible\ (\alpha) = TP_0,\ Partially\ Accessible\ (\beta) = TP_1,\ Not\ Accessible\ (\Upsilon) = TP_2 \quad (8)$$

$$\epsilon_\alpha = 0.2,\ \epsilon_\beta = 0.1,\ \epsilon_\Upsilon = 0.01 \quad (9)$$

$$Accessible = [\alpha * \epsilon_\alpha],\ Partially\ Accessible = [\beta * \epsilon_\beta],\ Not\ Accessible = [\Upsilon * \epsilon_\Upsilon] \quad (10)$$

$$Total\ accessibility\ score = \frac{\{(Accessible + Partially\ Accessible) - Not\ Accessible\}}{n} * 100 \quad (11)$$

The computed accessibility score for each classifier with their average score and standard deviation (SD) is shown in Table 6. According to the computed accessibility score, it illustrates that dataset 3 has a higher accessibility score than other experimented datasets. Besides, dataset 4 experienced with the lowest accessibility score. In the context of accessibility, the higher the accessibility score the lower the accessibility barrier, and the lower the accessibility score the higher the accessibility barrier. Also, the standard deviation represents the difference between the computed results of two implemented classifiers. The highest difference (in terms of SD) among two classifiers results was observed for dataset 4 and dataset 5.

TABLE VI
ACCESSIBILITY SCORE COUNT OF TESTED DATASET

Dataset	RF	DT	Avg. Accessibility Score	SD
Dataset-1	47%	44%	45.5%	2.12
Dataset-2	57%	53%	55%	2.23
Dataset-3	72%	75%	73.5%	2.12
Dataset-4	25%	36%	30.5%	7.77
Dataset-5	53%	58%	55.5%	3.53

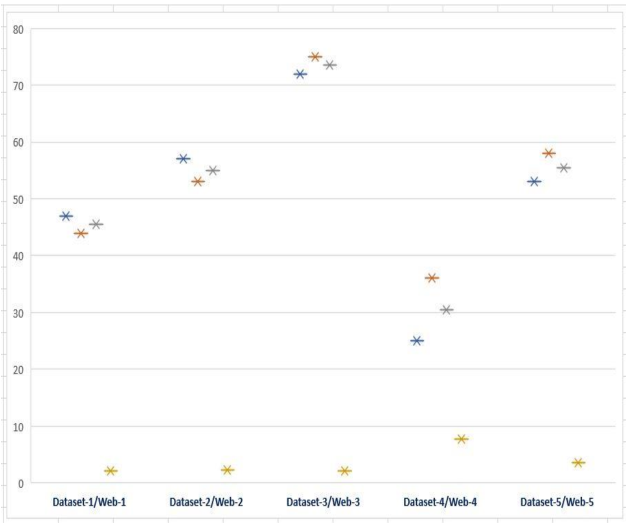

Fig. 4. Graphical representation of the computed accessibility score of five selected university websites.

Figure 4 shows the graphical representation of the computed accessibility score (for both RF and DT), their average accessibility score, and SD values for each dataset where we reference the results of each dataset with their associated dataset/website. It shows that the website of the Budapest University of Technology and Economics (Dataset 4/Web-4) has the lowest accessibility score than other selected university websites. In contrast, the website of Eotvos Lorand University (Dataset 3/Web-3) has the highest accessibility score. From these computed scores, it is noteworthy that none of the selected university websites were found completely accessible or barrier-free for people with several disabilities in accordance with our selected ten features.

V. DISCUSSION

This work presents an extensive study of university website accessibility score computation in Hungary. Generally, university websites are the prime source of information and services for native and international students and stakeholders. Thus, there’s an emerging need to identify the accessibility status of university websites. Besides, there is no previous research work conducted in Hungary to evaluate the university websites or academic websites of this country. Addressing these gaps, we proposed a web accessibility evaluation approach considering machine learning algorithms to compute the accessibility score of the selected university web pages of Hungary.

Generally, web content accessibility guidelines are widely accepted standards but few issues or aspects associated with people with disabilities are not considered in this standard. For example, if a website does not provide a manual text size adjustment option or manual color adjustment option then the majority of the people with vision disability or color disabilities will face difficulty in navigating the content. Besides, a few websites require user information for browsing some specific content, and few require successful completion of CAPTCHA which is very difficult for people with cognitive disabilities. Some other issues with the use of less-productive/not useful internal/external links, images, and video and audio content also hampered the accessibility aspects for people with special needs. Unfortunately, most of the automated accessibility testing tools also do not consider these accessibility aspects as these issues are very complicated to incorporate into an automated manner [31]. Therefore, accessibility checking considering these aspects might be useful for revealing the true insights of website accessibility.

With this aim, we computed the accessibility score by incorporating the Random Forest (RF) and Decision Tree (DT) techniques and calculating their average score with individual SD calculations. Our experiment result shows that the classification performance of the Random Forest classifier is more significant than the Decision Tree classifier. The average accessibility score shows that Eotvos Lorand University has higher accessibility features (according to the selected features in this research work) than other university websites. However, the computed score of other selected university websites was very poor which represents that most of the selected university web pages are not accessible to people with disabilities in terms of the selected aspects/features. To improve the accessibility in accordance with the selected aspects, tested university web pages need to improve their quality to ensure the complete accessibility objective. In addition, concerning the performance of machine learning classifiers or models, it is interesting to address that machine learning classifiers or models are significant in the evaluation of the accessibility of university websites. However, throughout the experimentation, we have some limitations associated with single-page validation and a small dataset. Therefore, a further investigation is required focusing on the current limitation that will be considered in future work.

VI. Conclusion

The prime objective of this study is to present the accessibility insights of higher institute websites (university websites) by implementing machine learning methods. University websites are an emerging platform to distribute information to students and associated authorities. However, providing an accessible online or accessible website is a relatively challenging task for the web designer and developer. Literature shows that few studies focused on accessibility issues, and there is almost no research work conducted considering the higher institute websites of Hungary. Therefore, we proposed a machine learning approach for computing the accessibility barrier score of the selected university website. We evaluated the result of machine learning methods through several metrics obtained from confusion matrix and classification reports. The computed result predicts that the selected university websites are not free from accessibility issues that reduce their effectiveness. The future work is limited to three focused groups: (i) this work is limited to incorporating two ML models which will be further extended and incorporated with other ML models focusing on the current limitations, (ii) to validate the result, we will incorporate user/ expert testing, and (iii) the entire university websites will be considered for experimenting instead of considering a single webpage/homepage.

References

[1] Allison, R., Hayes, C., McNulty, C. A., & Young, V. (2019). A comprehensive framework to evaluate websites: literature review and development of GoodWeb. *JMIR formative research*, *3*(4), p. e14372.

[2] Longstreet, P., Brooks, S., Featherman, M., & Loiacono, E. (2021). Evaluating website quality: which decision criteria do consumers use to evaluate website quality?. *Information Technology & People*.

[3] Vargas, A. M., Pedraza-Jiménez, R., & Bonilla, L. C. (2020). Website quality: An analysis of scientific production. *El profesional de la información*, *29*(5), p. 22.

[4] Acosta-Vargas, P., Acosta, T., & Lujan-Mora, S. (2018). Challenges to assess accessibility in higher education websites: A comparative study of Latin America universities. *Ieee Access*, *6*, pp. 36 500–36 508.

[5] Dominic, P. D. D., Jati, H., Sellappan, P., & Nee, G. K. (2011). A comparison of Asian e-government websites quality: using a non-parametric test. *International Journal of Business Information Systems, 7*(2), pp. 220–246.

[6] Hackett, S., Parmanto, B., & Zeng, X. (2005). A retrospective look at website accessibility over time. *Behaviour & Information Technology*, *24*(6), pp. 407–417.

[7] Ringlaben, R., Bray, M., & Packard, A. (2014). Accessibility of American university special education departments' web sites. *Universal Access in the Information Society, 13*(2), pp. 249–254.

[8] Devi, K., & Sharma, A. (2016). Framework for evaluation of academic website. International Journal of Computer Techniques, 3(2), pp. 234-239. [9] Montazer, G. A. (2018). University website quality improvement using intuitionistic fuzzy preference ranking model. *Quarterly of Iranian Distance Education Journal, 1*(2), pp. 9–30.

[10] Kaur, S., Kaur, K., & Kaur, P. (2016). An empirical performance evaluation of universities website. *International Journal of Computer Applications, 146*(15), pp. 10–16.

[11] Yerlikaya, Z., & Onay Durdu, P. (2017, July). Evaluation of accessibility of university websites: A case from turkey. In *International Conference on Human-Computer Interaction* (pp. 663–668). Springer, Cham.

[12] Ara, J., & Sik-Lanyi, C. (2023, May). AccGuideLiner: Towards a Modelling Approach of Web Accessibility Requirements following WCAG 2.2. In *2023 IEEE International Conference on Smart Information Systems and Technologies (SIST)* (pp. 10–15). IEEE.

[13] Chopra, G., Madan, P., Jaisingh, P., & Bhaskar, P. (2019). Effectiveness of e-learning portal from students' perspective: A structural equation model (SEM) approach. *Interactive Technology and Smart Education*.

[14] Mittal, H., Sharma, M., & Mittal, J. P. (2012, January). Analysis and modelling of websites quality using fuzzy technique. In *2012 Second International Conference on Advanced Computing & Communication Technologies* (pp. 10–15). IEEE.

[15] Malhotra, R., & Sharma, A. (2013, August). A neuro-fuzzy classifier for website quality prediction. In *2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (pp. 1274–1279). IEEE.

[16] Rashida, M., Islam, K., Kayes, A. S. M., Hammoudeh, M., Arefin, M. S., & Habib, M. A. (2021). Towards developing a framework to analyze the qualities of the university websites. *Computers, 10*(5), pp. 57.

[17] Olaleye, S. A., Sanusi, I. T., Ukpabi, D., & Okunoye, A. (2018). Evaluation of Nigeria universities websites quality: A comparative analysis.

[18] Alahmadi, T. (2017, April). A multi-method evaluation of university website accessibility: Foregrounding user-centred design, mining source code and using a quantitative metric. *In Proceedings of the 14th International Web for All Conference* (pp. 1–2).

[19] Hassouna, M. S., Sahari, N., & Ismail, A. (2017). University website accessibility for totally blind users. *Journal of Information and Communication Technology, 16*(1), pp. 63–80.

[20] Verkijika, S. F., & De Wet, L. (2020). Accessibility of South African university websites. *Universal Access in the Information Society, 19*(1), pp. 201–210.

[21] AlMeraj, Z., Boujarwah, F., Alhuwail, D., & Qadri, R. (2021). Evaluating the accessibility of higher education institution websites in the State of Kuwait: empirical evidence. *Universal Access in the Information Society, 20*(1), pp. 121–138.

[22] Dhiman, P. (2014, September). Empirical validation of website quality using statistical and machine learning methods. In *2014 5th International Conference-Confluence The Next Generation Information Technology Summit (Confluence)* (pp. 286–291). IEEE.

[23] Hungary Ranking Web of Universities-Webometrics: https://www.webometrics.info/en/europe/hungary

[24] Raschka, S. (2018). Model evaluation, model selection, and algorithm selection in machine learning. *arXiv* preprint *ArXiv*:1811.12808.

[25] Izquierdo-Verdiguier, E., & Zurita-Milla, R. (2020). An evaluation of Guided Regularized Random Forest for classification and regression tasks in remote sensing. *International Journal of Applied Earth Observation and Geoinformation, 88*, p. 102051.

[26] Sisodia, D., & Sisodia, D. S. (2018). Prediction of diabetes using classification algorithms. *Procedia computer science, 132*, pp. 1578–1585.

[27] Vabalas, A., Gowen, E., Poliakoff, E., & Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. *PloS one, 14*(11), p. e0224365.

[28] Yang, L., & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing, 415*, pp. 295–316.

[29] Matošević, G., Dobša, J., & Mladenić, D. (2021). Using machine learning for web page classification in search engine optimization. *Future Internet, 13*(1), p. 9.

[30] Wu, J., Chen, X. Y., Zhang, H., Xiong, L. D., Lei, H., & Deng, S. H. (2019). Hyperparameter optimization for machine learning models based on Bayesian optimization. *Journal of Electronic Science and Technology, 17*(1), pp. 26–40.

[31] Ara, J., & Sik-Lanyi, C. (2022). Investigation of COVID-19 Vaccine Information Websites across Europe and Asia Using Automated Accessibility Protocols. *International Journal of Environmental Research and Public Health, 19(5)*, p. 2867.

Jinat Ara graduated from Southeast University, Bangladesh in 2018 with a bachelor's degree in Computer Science and Engineering (CSE). Following that she obtained her master's degree in 2020 from Jahangirnagar University, Bangladesh from the Department of Computer Science (CS). At the same time, she worked as a Software Engineer at Odyssey Apps, Dhaka, Bangladesh. Currently, she is pursuing her Ph.D. degree at the Faculty of Information Technology, University of Pannonia, Hungary. Her research area for her Ph.D. studies on the area of "accessibility investigation in web platforms or environments". Additionally, she has a keen interest in conducting cutting-edge research and currently, she is carrying around six years of research experience in several areas of computer science. Her area of research interest is 'Machine Learning (ML)', 'Virtual and Augmented Reality', 'Assistive Technologies', 'Web Technologies', 'Artificial Intelligent (AI)', 'Natural Language Processing (NLP)' and 'Automated Vehicles (AV)'.

Cecilia Sik-Lanyi is a Professor at University of Pannonia, Veszprem, Hungary. Sik-Lányi studied Mathematics and Computer Science. She obtained the degree of Dr. Univ. in Physical-chemistry (1993), and Ph.D. in Computer Science (2000), and became a Doctor of Hungarian Academy Sciences (D.Sc.) in 2022. Her research area is virtual reality, serious games, and user interface design, focusing on color science and accessibility. She received several awards, the most important ones are: "Diamond-Award from the Association for the Advancement of Assistive Technology in Europe" in 2015, and the "King Salman Award for Disability Research" in 2018. She has supervised more than 220 BSc and MSc theses since 1997. She has published more than 500 referred articles and conference papers and worked as a guest editor for many renowned journals and books.