Decision Science Letters 12 (2023) 279–290

Contents lists available at GrowingScience

Decision Science Letters

homepage: www.GrowingScience.com/dsl

Exploring the quality of the higher educational institution website using data mining techniques

Mohammed Hameed Afif[a*]

[a]*Department of Management Information Systems, College of Business Administration, Prince Sattam bin Abdulaziz University, Saudi Arabia*

C H R O N I C L E

Article history:
Received: November 26, 2022
Received in revised format: December 20, 2022
Accepted: January 13, 2023
Available online: January 13, 2023

Keywords:
Website Quality
Data mining
Usability quality
Information Quality
Higher Education

A B S T R A C T

The website of higher educational institutes is considered a vital communication channel to provide main resources to their stakeholders. It plays an important role in disseminating information about an institute to a variety of visitors at a time. Thus, the quality of an academic website requires special attention to respond to the users' demands. This study aims to explore the quality of the PSAU website based on data mining techniques. The first step: was collecting opinions about the PSAU website using a survey. After that, data mining processes were used as descriptive and predictive models. The descriptive model was applied to describe and extract the major indicators of website quality. Besides, the predictive model was applied to create models for predicting the website quality level. More than one classification algorithm was used. Naive Bayes and Support Vector Machine have given the best results in all performance indicators, and the achieved accuracy rate for both algorithms was 86% and 84% respectively. The results revealed that the overall quality level of the PSAU website is very good. The usability quality and content quality were very good. The service quality needs more attention. which indicated that the service level is inadequate and needs to be further enhanced. The results of the study should be useful to the deanship of Information Technology at PSAU, and website developers, in redesigning with high quality in terms of its usability, content, and service.

1. Introduction

Nowadays, the website of universities is considered as an essential communication channel. It uses to publish all information about academic programs and services to all stakeholders (Andalib & Danaee, 2013; Peker et al., 2016; Barraood, 2016; Budiman et al., 2020). The universities' websites are a good source of information for students, faculty members, employees, and visitors, and can be used to empower users to learn about the universities easily, and provide different services to all stakeholders (Peker et al., 2016; Andalib & Danaee, 2013; Barraood, 2016; Ismail & Kuppusamy, 2018; Rashida et al., 2021; Karani et al., 2021; Semerádová & Weinlich, 2020). Furthermore, Websites are considered as another appearance for universities to shape their image, as website visitors take an impression of these universities through their websites. Besides that, websites can be used to create a globally competitive advantage to attract visitors, students, scholars, and funding from other places, spreading the prestige of universities over the world (Barraood, 2016; Ismail & Kuppusamy, 2018; Rashida, et al., 2021; Karani et al., 2021). Therefore, the website for the higher education institute is a critical virtual gateway, due to its importance as a communication channel with all stakeholders, and a gateway for performing their activities. Thus, developing a high-quality website for higher educational institutes is a critical goal. Moreover, the University website needs special attention to be made effective and attractive.

Quality evaluation for university website gets the attention of many researchers (e.g. Ahmad & Khan, 2017; Andalib & Danaee, 2013; Barraood, 2016; Budiman et al., 2020; Rante et al., 2020; Rashida, et al., 2021; Sutanto et al., 2021), where

* Corresponding author.
E-mail address: mohaafif@gmail.com (M. H. Afif)

doi: 10.5267/dsl.2023.1.007

several models and methods were employed. The outcomes of these studies highlighted the major issues that were related to the university website quality. Besides that, providing feedback which helps in empowering the website's functionality.

These reasons motivate us to focus on exploring the quality of the higher education institutes' websites to highlight their quality level and discover the issues that face visitors.

The major goal of the current study is to explore the quality level of the Prince Sattam bin Abdelaziz university (PSAU)'s website by using knowledge discovery techniques to determine relationships among different components and discover problems. In addition to predicting the quality level of the PSAU website.

The current study attempts to bridge the gap in the literature using data mining techniques in discovering the knowledge in the area of website quality evaluation as well as explore the quality level of higher educational institutes.

2. Related work

However, to the best of our knowledge, there are few approaches that apply data mining techniques to a dataset containing evaluation Web sites quality. So, this part summarized the related studies that use data mining to discover hidden knowledge from collected data related to users interacting with websites.

Boza et al. (2014) applied knowledge discovery techniques to extract relationships among Nielsen usability components. The study aimed to extract hidden relationships among attributes, and components and discover problems of website usability. Experiments were applied by association rules and decision trees on dataset containing evaluation reports of different Web sites. The produced outcomes indicate that the suggested method was promising to discover interesting relations from this type of data. The discovered patterns and relations are useful for Web site designers and give them insights related to usability issues that must take into account. Al-Omar (2018) applied data mining techniques to discover usability issues faced by users of the learning management system (LMS) at King Abdulaziz University. The major objective of the study was to consolidate similar usability problems into one group. This will help universities to understand faculty members' and students' behaviours, and then find out the most common problems for each group and customize the LMS for each group accordingly. A novel method for usability website evaluation was suggested by Sagar and Saha (2016) and Sagar and Saha (2017); using data mining techniques in combination with traditional usability testing approaches. The objective of the study is to discover common usability problematic patterns that belong to top-50 academic websites. The study uses the ISO9241-151 guidelines under 16 categories to collect data from hundreds of participants. Association rule and decision tree data mining methods are applied to extract hidden relationships and patterns. The findings indicated that most of the issues were found in Search and Social Media categories. Furthermore, users easily locate 50.53% of guidelines on websites as fully functional whereas, 49.46% of characteristics are considered problematic usability features that are not functional on the academic website as a whole. A systematic review of predictive data mining techniques in exploring the factors that influence students' performance in higher education was presented by Abu Saa et al. (2019). The major aim of the study was to identify the most common factors that influence the student performance studied by researchers, as well as determine which data mining methods were applied to identify these factors. Authors focus on only 36 research articles out of a total of 420 from 2009 to 2018. The outcomes indicated that the common factors were classified into four categories, namely students' previous grades and class performance, students' e-Learning activity, students' demographics, and students' social information. Furthermore, the results also showed that the most common data mining methods used to predict and classify students' factors are decision trees, Naïve Bayes classifiers, and artificial neural networks. El-Halees, (2014) applied an opinion-mining technique as an automatic technique to measure subjective usability. The method develops a model for extracting knowledge from the opinion to enhance subjective software usability. The proposed approach focused on three usability quality factors: effectiveness, efficiency, and satisfaction. Besides, four software were used to evaluate the proposed model. A set of experiments were conducted on a dataset containing 565 reviews divided into 345 positive and the remaining negative. The obtained result is 85% of accuracy. Authors Lin et al. (2022)presented a systematic literature review on opinion mining for software development activities. The review included more than 180 articles. One of the main points for the review related to the usage of opinion mining for assessing the quality of software products. Furthermore, the authors explored articles that applied opinion mining in software engineering in collecting informative app reviews to understand how developers can improve their products and revise their release plans. Besides opinion mining has been used to assess the quality of software products.

Most of the previous studies depend on models for evaluating website quality by collecting the data by using surveys and then performing different statistical analyses for studying the website quality. The current study collects the data using a survey and then the data mining processes were applied to explore the hidden relationships among different components, among attributes, and highlight the issues that are discovered.

3. Methodology

The methodology for consists of many steps that begin with collecting data and then pre-processing phase after that data mining algorithm and last one is the knowledge discovery phase. Fig 1 shows the methodology steps before a brief introduction about the website quality and its measure is presented as a background about the website quality measures and models.

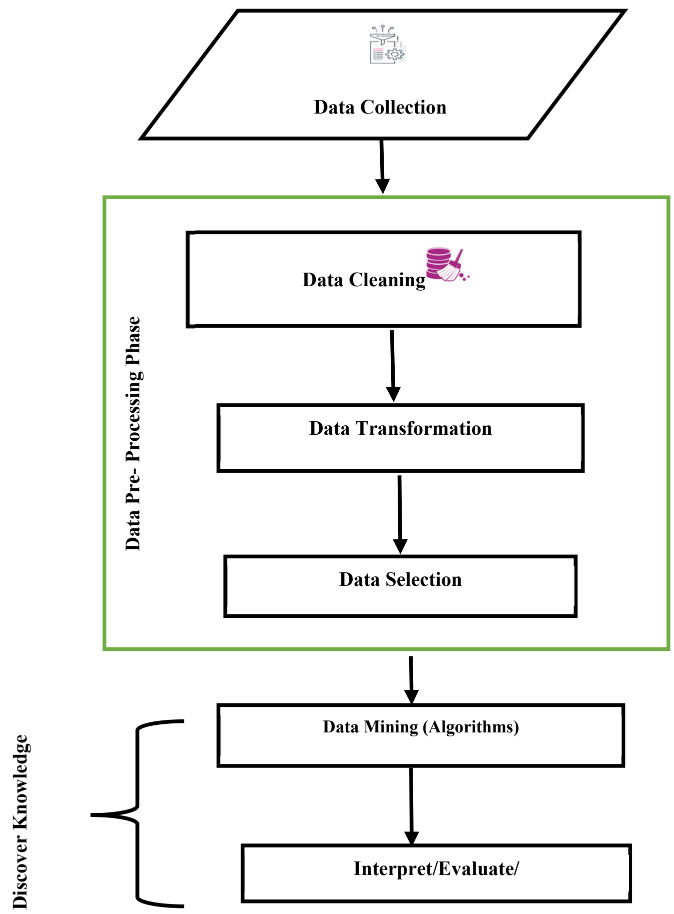


Fig. 1. Methodology Steps

3.1 Website quality and its measures

The term quality is defined as the measure of the ability of a product or service to meet the needs and expectations of the consumer (Semerádová & Weinlich, 2020; Budiman et al., 2020; Andalib & Danaee, 2013). So, website quality is described as the ability of a website to meet user requirements and expectations (Andalib & Danaee, 2013; Barraood, 2016; Rante et al., 2020; Semerádová & Weinlich, 2020).

The major aim of website evaluation is to gain feedback from their users, then analyse these data to identify the issues that face users. Moreover, ensuring that the website is performing its functionality well, and achieving business goals. Therefore, the assessment of website quality is a complex process consisting of measurable, representative items which help to determine the components which require improvement. Website Quality (or Quality of Websites) could be measured from two perspectives: Programmers, and End-users (Andalib & Danaee, 2013). The aspects of website quality from programmers focus on the degree of Maintainability, Security, Functionality, etc. Whilst the end-users are paying more attention to Usability, Efficiency, Creditability, etc. (Andalib & Danaee, 2013; Barnes & Vidgen, 2000; Barnes & Vidgen, 2002). Many studies have been proposed to evaluate website quality using different models and methods (e.g. Semerádová & Weinlich, 2020; Rashida et al., 2021; Budiman et al., 2020; Frisdiantara et al., 2020; Rante et al., 2020; Rashida, et al., 2021; Sutanto et al., 2021; Ahmad & Khan, 2017; Barnes & Vidgen, 2002; Barnes & Vidgen, 2000). The WEBQUAL is one of the popular models which has been used for assessing the quality of websites. It is suggested by Barnes and Vidgen (2002), and widely used by many researchers as listed by Ahmad and Khan (2017). Moreover, the WEBQUAL framework represents a viable quality measurement tool for website quality, and it comprises three main dimensions: Usability, Information Quality, and Service Interaction Quality. In the current study, two dimensions of the WEBQUAL model are used for measuring the website quality, they are usability quality and information quality. The service quality dimension was measured based on ISO 9126 quality standard metrics which are used by researchers in (Barraood, 2016). Table 1 lists all questions which are used for evaluating the website quality for higher educational institutes based on the quality factors used by studies in the literature (Semerádová & Weinlich, 2020; Budiman et al., 2020; Ahmad & Khan, 2017; Barnes & Vidgen, 2002; Barraood, 2016; Ahmad & Khan, 2017; Budiman et al., 2020; Frisdiantara et al., 2020; Rante et al., 2020; Sutanto et al., 2021). The steps of the methodology begin with collecting data and then the pre-processing phase after that data mining algorithm and last one is the knowledge discovery phase. Fig 1 shows these steps.

Table 1
Questions for Dimensions of website quality

No	Usability Quality	No	Information Quality (Content quality)
UQ1	The site easy to learn to operate.	IQ1	The site provides accurate information
UQ2	User interaction of the site is clear and understandable	IQ2	Provides believable information or reliable information
UQ3	The site easy to navigate	IQ3	Provides timely information.
UQ4	The site easy to use.	IQ4	Provides relevant information
UQ5	The site has an attractive appearance.	IQ5	Provides easy to understand information
UQ6	The design is appropriate to the type of site.	IQ6	Provides information at the right level of detail.
UQ7	The website is appropriate in arranging layout information.	IQ7	Presents the information in appropriate format.
Services Quality (Responsiveness and Reliability)			
SQ1	The PSAU website makes it easy to communicate with university		
SQ2	The PSAU website responds to my actions as expected		
SQ3	The PSAU website has improved the engagement with students		

3.2 Data collection

To collect data a questionnaire was designed based on questions in Table 1, which were extracted from many references in the literature as mentioned in the previous section. The questionnaire was divided into two main sections, namely, section A and section B. In specific, section A comprised statements designed to collect data about the respondents like gender, Employment, educational level, and other three questions about the internet usage, website frequency use and type of machine that is used to access the website. Section B ascertains the views of the College of Business Administration faculty members and students on the website quality of Prince Sattam bin Abdulaziz University. A five-point Likert scale was used. The questionnaire was published via google forms. The dataset includes 212 records with 23 features.

3.3 Data Pre-processing phase

This phase is data mining steps which are necessary for preparing data for the mining process. During this phase the collected data was prepared in proper format by encoding some features, replacing the words with numeric value according to Likert scale. After that, the data is cleaned by removing incomplete data, missing value, and unsuitable data. After that, the normalization process was performed by transforming the Likert scale value to range between (0-1). Then, the value of WEBUSE is calculated based on (Chiew & Salim, 2003; Barraood, 2016; Andalib & Danaee, 2013; Ahmad & Khan, 2017). The attribute quality level is developed and assigned its value by calculating the WEBUSE scale. Finally, the feature selection was implemented to select the relevant features to the data mining process, where some attributes were ignored during the mining process.

3.4 Data Mining Methods

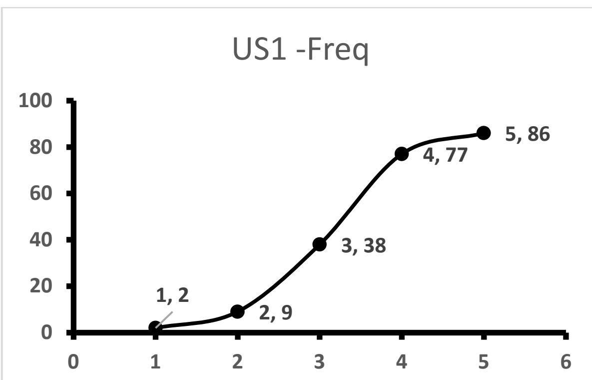
Data mining has two main purposes. First, applied as a descriptive model which is used to determine patterns or relationships in data and serves as a manner to explore the characteristics of the data examined. The second one is applied as a predictive model which is used to create a model for predicting the datasets whose results are unknown (Han, Pei, & Kamber, 2022), (Gupta & Chandra, 2020; Kantardzic, 2011; Abu Saa et al., 2019; Al-Hagery et al., 2020; Yağcı, 2022; LIN, et al., 2022). In this study, the descriptive and predictive models were applied to discover correlations, patterns, relationships and trends by searching through a large amount of data. The major aim of this step is to discover useful information or unknown knowledge from the collected data. Different data mining techniques i.e. classification and filtered attribute evaluator with ranker algorithms were applied to stored data. Generally, the classification technique is applied to assist in predicting website quality. Moreover, classification methods assist in predicting the behaviour of the context of the use of academic websites under evaluation. As an output, a list of rules and knowledge are generated that should be analysed by the researcher.

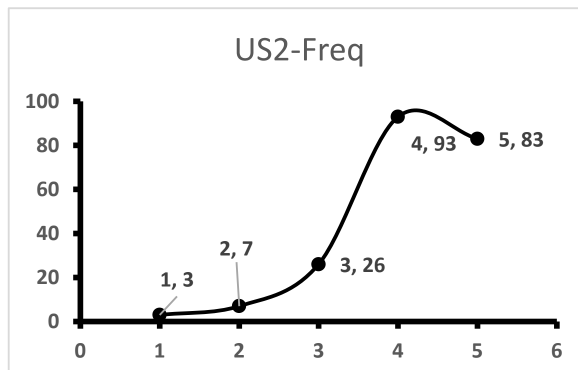
3.5 Data Explanatory

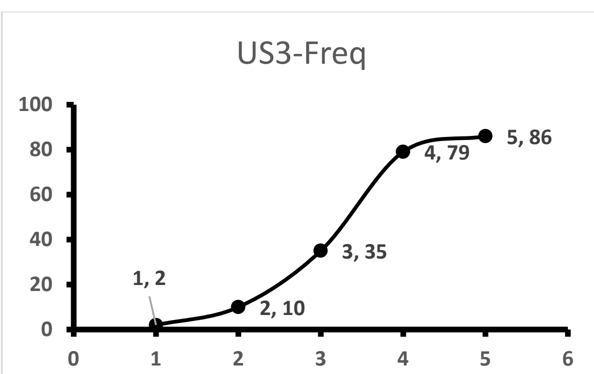
The first step before performing the analysis process is data exploration. The importance of this process is to explore the data and its characteristics through summaries and charts. In addition, exploring the data is helping in understanding and acquiring information about data in more detail. The collected data is divided into four parts. The first one is demographics attributes, and the other parts were regarding usability dimension, information dimension, and service dimension. Table 2 displays a summary about the data including feature name, description, mean, median, dispersion, min, max, and frequency for each feature with its ratio. No missing values were detected. The main indicators that were highlighted first regarding gender 66% were males and 34% were females; while the second was employment where 92% from employment were students and 6% were faculty. The third indicator was related to the educational level 61% have a Bachelor and the next was 33% for secondary. The next pointer was regarding the usage of website frequency per day. The extracted data indicates that 76% use websites less than an hour per day. Regarding the type of device, which was used to surf the website, 63% of respondents use both personal computers and smartphones, while 32% use a smartphone. The quality level attribute represents the level of quality of the website based on WEBUSE scale. The highest value was for Excellent and good levels with a ratio of 76%, and 66% respectively, while acceptable and non-acceptable level gets the same ratio which is 17%.

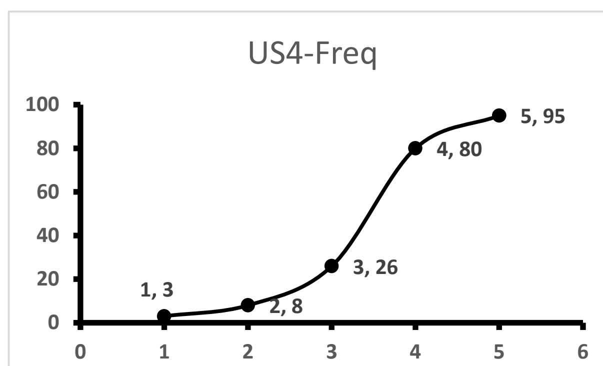
Table 2
Description of demographic features

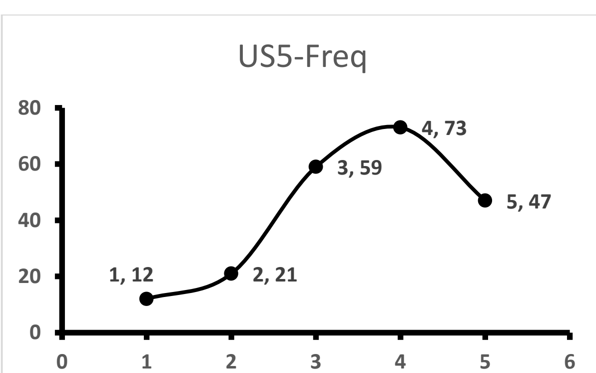
Name	Description	Mean	Median	Dispersion	Min	Max	Frequency		
							Code	Freq	Ratio
F1	This feature represents a gender, it is encoded by 1 for males and 0 for females.	0.655	1	0.72	0.00	1	0 1	73 **139**	34% 66%
F2	This feature refers to employment and is encoded as 1 for students, 2 for faculty, and 3 for employees.	1.10	1	0.32	1	3	1 2 3	**195** 13 4	**92**% 6% 2%
F3	This feature represents the educational level, which is encoded with 1 ,2,3, and 4. (1=secondary,2=Bachelor, 3=Master, 4=Phd)	1,78	2	0.40	1	4	1 2 3 4	70 **129** 2 11	33% 61% 1% 5%
F4	This attribute for frequency of usage for website per day and encoded as, 1=never,2=less than one hour, 3=Between 3 to 6, 4=greater than 6 hours per day	2.20	2	0.24	1	4	1 2 3 4	7 **161** 38 6	3% **76**% 18% 3%
F5	Device type for visiting PSUA website, 1==PC Laptop, 2= Smart Phone, 3= Both	2.58	3	0.22	1	3	1 2 3	10 **68** **134**	5% 32% 63%
Quality Level	The output value or level of quality – which developed based on calculate the WEBUSE value.	-	-	-	-	-	Excellent Good Acceptable NotAccept	76 66 35 35	36% 31% 17% 17%

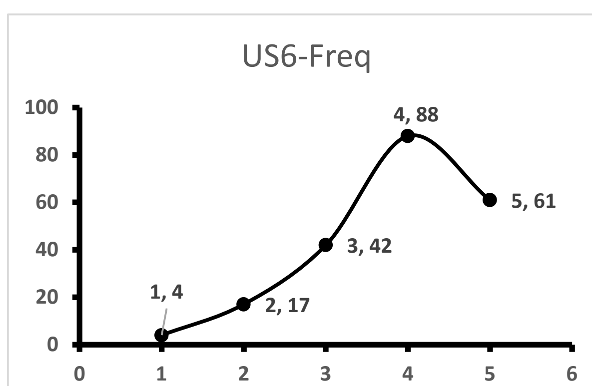


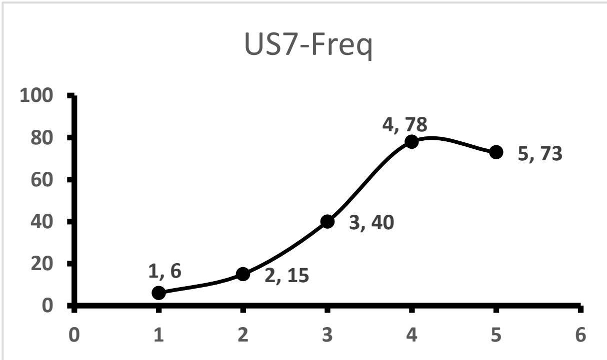












Figs. [1-7]. Usability attributes distributions

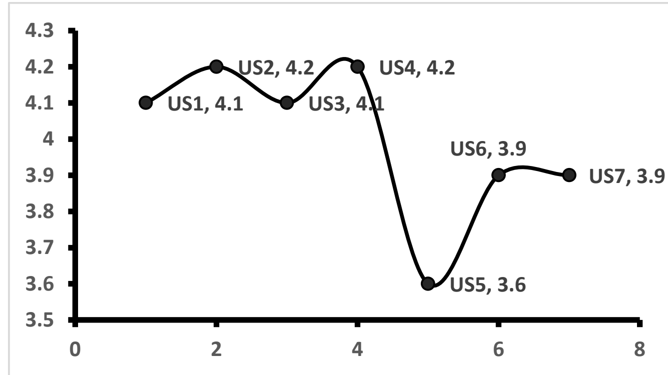


Fig. 8. Usability Quality Attributes indicators

The next part of dataset is related to usability dimension which contains seven attributes (US1 - US7). Graphs from 1 to 7 show the distribution for each feature including the scale and the frequency for each one. Features US1, US2, US3, US4, and US7 have got the highest rating, while the remainder features US5 and US6 given good ratings. Table 3 reports a descriptive statistic for usability attributes which include attribute Id, description, mean, median, dispersion, min, and max. The mean was in the range of 4.22 and 3.58. The standard deviation which reflects the dispersion of the dataset was in the range (of 0.21 and 0.31). The major indicators extracted are summarized as the following attributes US1, US2, US3, US4, and US7 were excellent, and most website visitors express their satisfaction with the visibility of these features. Furthermore, features US5 and US6 need more attention to understand the challenges and improve the quality of those features as highlighted in Fig. 8. In summary, the usability quality for PSAU website is good, and has high level of satisfaction as illustrated in figures.

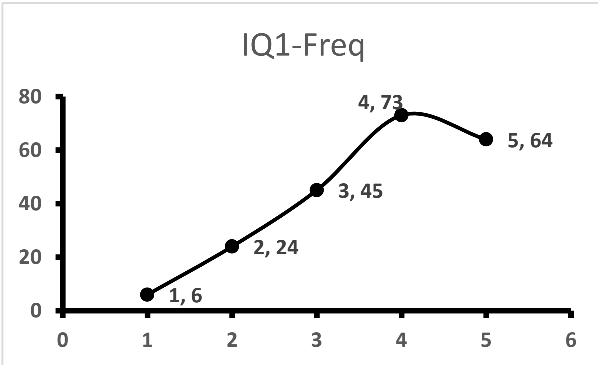
Table 3
Details about dataset usability features

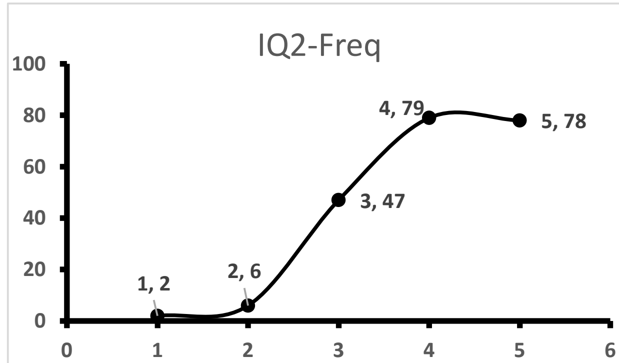
Id	Description	Mean	Median	Dispersion	Min	Max
US1	This feature is referred to Q1 in table 1 – The Likert scale was used for encoding the dataset	4.11	4	0.22	1	5
US2	This feature is referred to Q2 in table 1 – The Likert scale was used for encoding the dataset	4.16	4	0.21	1	5
US3	This feature is referred to Q33 in table 1 – The Likert scale was used for encoding the dataset	4.12	4	0.22	1	5
US4	This feature is referred to Q4 in table 1 – Likert scale was used for encoding dataset	4.21	4	0.21	1	5
US5	This feature is referred to Q5 in table 1 – Likert scale was used for encoding dataset	3.58	4	0.31	1	5
US6	This feature is referred to Q6 in table 1 – Likert scale was used for encoding dataset	3.87	4	0.25	1	5
US7	This feature is referred to Q7 in table 1 – Likert scale was used for encoding dataset	3.93	4	0.26	1	5
Usability Quality	Represent average for all features of usability dimension.	4	4	0.191	1.57	5

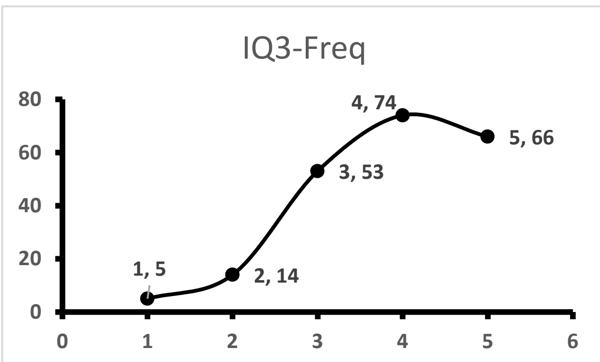
The content quality dimension is the third part of the data which includes seven attributes from IQ1 to IQ7. Table 4 reports summary for information quality data which include attribute Id, description, mean, median, dispersion, min, and max. Charts from 9 to 16 show the distribution for each feature including the scale and the frequency for each choice. Attributes IQ2, IQ4, IQ5, IQ6, and IQ7 have got the highest rating. The remainder features IQ1 and IQ3 have accepted ratings as listed in figure 17. In general, the content quality for the PSAU website is good. Moreover, features IQ1 and IQ3 that are related to information accuracy and content update require more attention to empower their satisfaction.

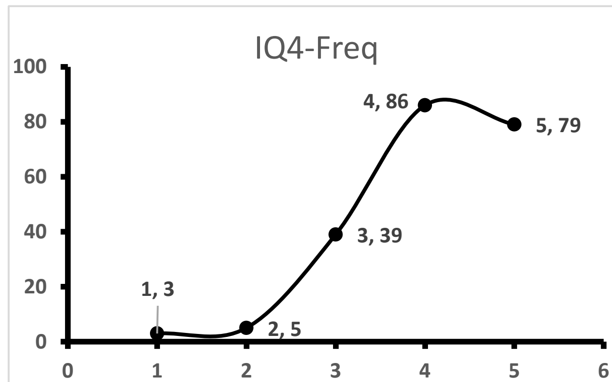
Table 4
Details about dataset information Quality (content quality)

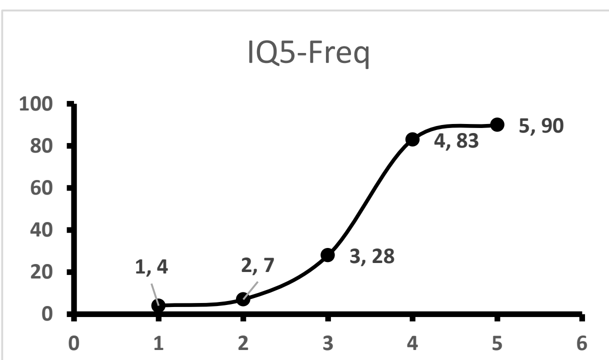
Id	Description	Mean	Median	Dispersion	Min	Max
IQ1	This feature is referred to in Q1 in table 1 – The Likert scale was used for encoding the dataset	3.78	4	0.29	1	5
IQ2	This feature is referred to in Q2 in table 1 – The Likert scale was used for encoding the dataset	4.06	4	0.22	1	5
IQ3	This feature is referred to in Q3 in table 1 – The Likert scale was used for encoding the dataset	3.86	4	0.26	1	5
IQ4	This feature is referred to in Q4 in table 1 – The Likert scale was used for encoding the dataset	4.10	4	021	1	5
IQ5	This feature is referred to in Q5 in table 1 – The Likert scale was used for encoding the dataset	4.17	4	0.22	1	5
IQ6	This feature is referred to in Q6 in table 1 – The Likert scale was used for encoding the dataset	4.01	4	0.23	1	5
IQ7	This feature is referred to in Q7 in table 1 – The Likert scale was used for encoding the dataset	4.01	4	0.22	1	5
Information Quality	Represent average for all features of information quality dimension.	3.99	4	0.22	1	5

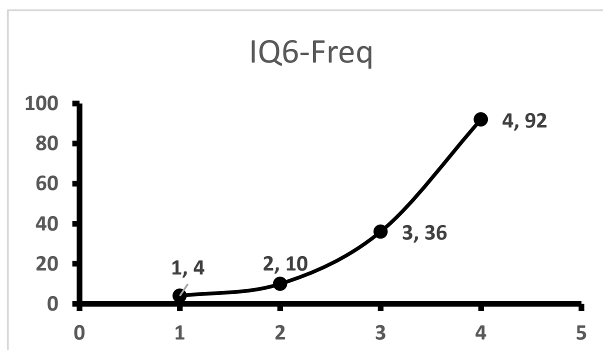


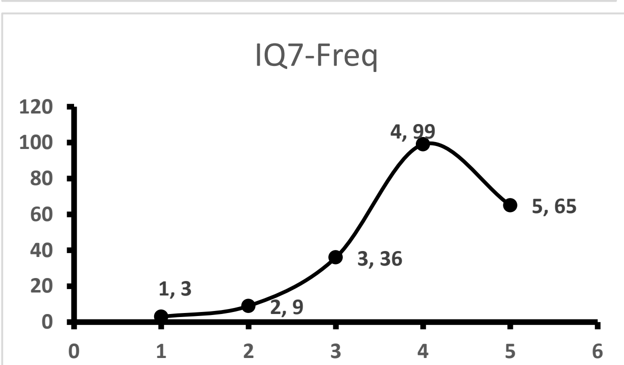












Figs. [9-16]-Information Quality attributes distributions

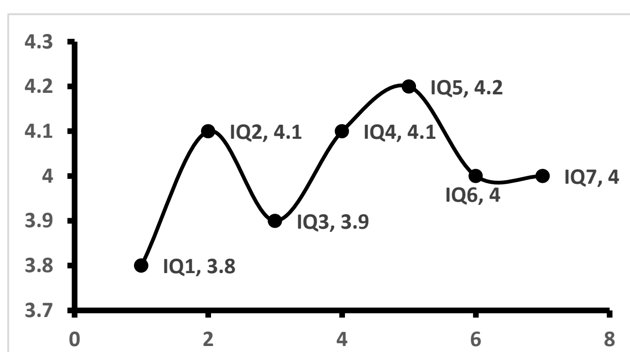
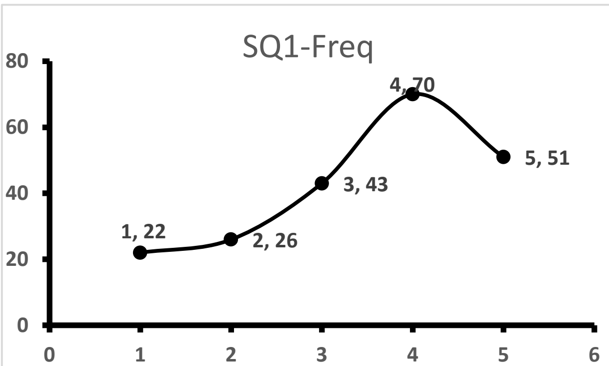
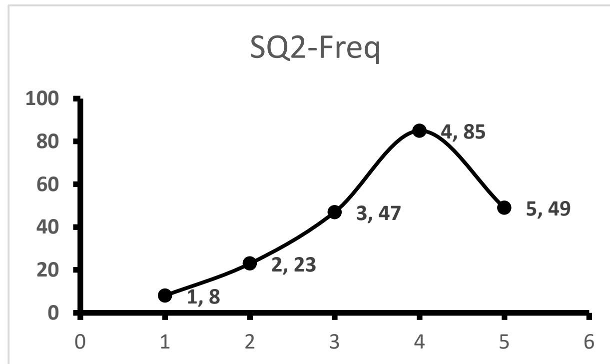


Fig. 17. Information Quality Attributes indicators

Table 5
Details about dataset Part Service Quality

Id	Description	Mean	Median	Dispersion	Min	Max
SQ1	This feature is referred to in Q1 in table 1 – The Likert scale was used for encoding the dataset	3.48	4	0.36	1	5
SQ2	This feature is referred to in Q2 in table 1 – The Likert scale was used for encoding the dataset	3.68	4	0.29	1	5
SQ3	This feature is referred to in Q3 in table 1 – The Likert scale was used for encoding the dataset	3.90	4	0.27	1	5
Service Quality	Represent average for all features of the Service quality dimension.	3.66	4	0.28	1	5





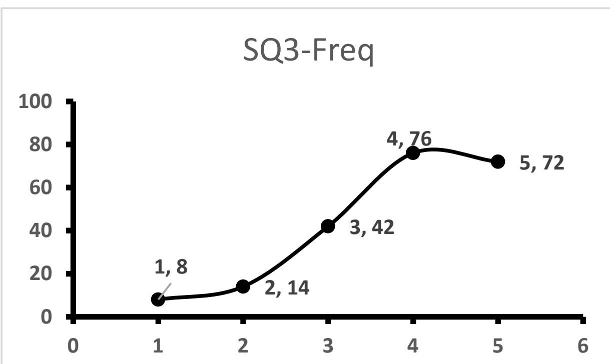


Fig. [18-20]. Service Quality attributes distributions

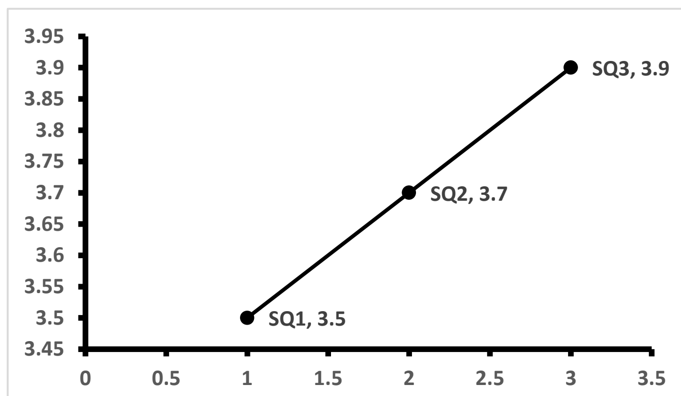


Fig. 21. Service Quality Attributes indicators

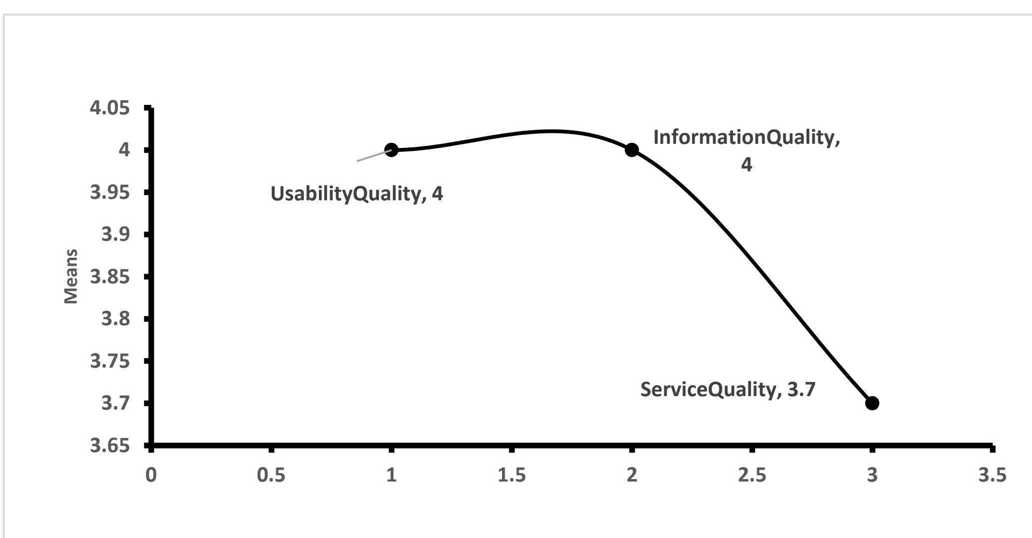


Fig. 22. Website Quality Attributes indicators

The last part of the data is the service quality dimension which contains three attributes from SQ1 to SQ3. Graphs from 18 to 20 show the distribution for each feature. All Attributes have accepted ratings. Table 5 list the major indicators for service quality attributes which include attribute Id, description, mean, median, dispersion, min, and max. The mean was in the range (3.90 – 3.48), while the median was equal for all features. The standard deviation was in a range between (0.27 and 0.36). Figure 21 displays the summary of service quality attributes. Figure 22 illustrates the website quality indicators for all dimensions where the usability and content quality were good. The service quality is acceptable and needs more attention to increase the satisfaction level and improve it . As a result, the quality level of the PSAU website is very good based on the discovered information from the collected dataset**.**

4. Experiments and Results

The experiments are implemented using orange machine learning software and MSExcel 365 on the machine has a core i7 CPU with 16 GB of RAM. The Fig. 23 shows the workflow diagram for the suggested method using orang platform. The workflow demonstrates the steps that are implemented to achieve the study objectives. As displayed on the figure there are more than one classification algorithms used. Furthermore, experiments are performed on each dimension alone to discover the patterns, hidden relationships among features and predicate the quality level. After that, the overall website quality was explored. Regarding the usability, the quality of the website is classified into four classes representing the quality level of usability. These levels are Excellent, Good, acceptable and not acceptable. The obtained outcomes are reported in Table 6. The table contains in the first column algorithm name, then the performance indicators which are Area Under the Curve (AUC), classification accuracy (CA), F1, precision and recall. The AUC is the measure of the ability of a classifier to distinguish between classes and is used as a summary of the ROC curve. The higher value of the AUC, the better performance of the model at distinguishing between the positive and negative classes. As Reported in Table 6 the Naive Bayes gave the highest outcomes for all performance indicators, after that the logistic regression is in the second rank where they give 91%, 75%, 75%,74% and 75%. for AUC rate, 75% for accuracy and 93% and 91%.

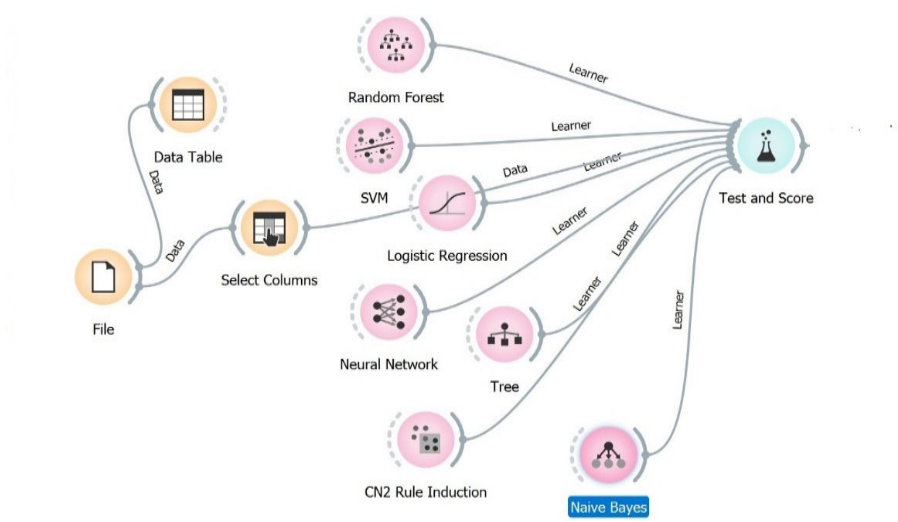


Fig. 23. The workflow diagram for the method

Table 6
Results for usability quality

Model	AUC	CA	F1	Precision	Recall
Tree	0.79	0.64	0.64	0.64	0.64
SVM	0.90	0.73	0.71	0.71	0.73
Random Forest	0.87	0.68	0.68	0.68	0.68
Neural Network	0.90	0.71	0.70	0.69	0.71
Naive Bayes	**0.93**	**0.79**	**0.78**	**0.78**	**0.79**
Logistic Regression	**0.91**	**0.75**	**0.75**	**0.74**	**0.75**
CN2 rule inducer	0.83	0.64	0.65	0.64	0.65

Table 7 illustrates the produced results for information quality dimension. The best AUC achieved was by Naive Bayes with 92%, then the SVM with rating 91%. The highest classification accuracy (CA) was produced by SVM, Naive Bayes and logistic regression with 76%, 72% and 72% respectively. SVM got the best values for other performance indicators F1, precision and recall.

Table 7
Results for Information quality

Model	AUC	CA	F1	Precision	Recall
Tree	0.82	0.64	0.64	0.64	0.64
SVM	0.91	0.76	0.73	0.73	0.75
Random Forest	0.88	0.67	0.66	0.66	0.66
Neural Network	0.90	0.70	0.70	0.70	0.70
Naive Bayes	0.92	0.72	0.71	0.71	0.72
Logistic Regression	0.90	0.72	0.71	0.71	0.72
CN2 rule inducer	0.82	0.65	0.65	0.64	0.65

Regarding the service quality of the website is classified into four classes these represented the quality level of services these levels are Excellent, Good, acceptable and not acceptable. Table 8 displays obtained results for predicting the level of services. As listed in table Naive Bayes given the best results in the most performance metrics, with the following rating: 86%,65%,66%, 68% and 65%.

Table 8
Results for Services quality

Model	AUC	CA	F1	Precision	Recall
Tree	0.75	0.61	0.61	0.61	0.61
SVM	0.81	0.57	0.56	0.58	0.57
Random Forest	0.81	0.62	0.61	0.61	0.62
Neural Network	0.80	0.57	0.56	0.56	0.57
Naive Bayes	0.86	0.65	0.66	0.68	0.65
Logistic Regression	0.83	0.59	0.57	0.58	0.59
CN2 rule inducer	0.73	0.54	0.54	0.54	0.54

As mentioned in previous sections the major objective is to discover the quality level for PSAU website by extracting the hidden knowledge and patterns among different features. The classification algorithms are used to predicate the quality level. Table 9 reported the produced results to predict the quality level for the PSAU website. The Naïve Bayes algorithm given the best outcome for all performance indicators where the values are (97%, 86%, 86%, 86% and 86%); After that, the SVM classifier gives the second highest results in all performance aspects following (97%, 84%, 84%, 84% and 84%). Fig. 24 and Fig. 25 demonstrated these results.

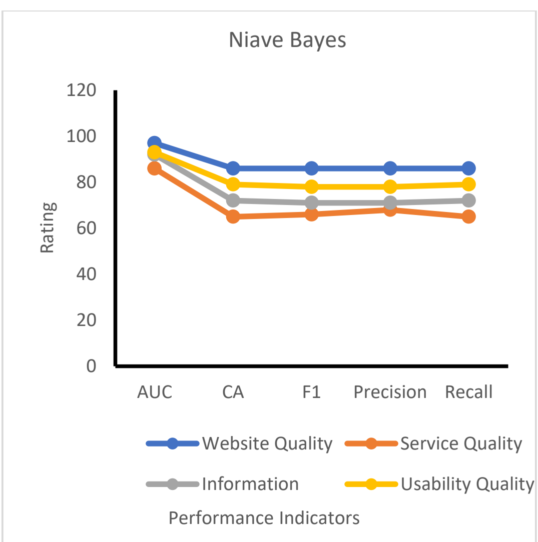


Fig. 24. Results for Naive Bayes algorithm

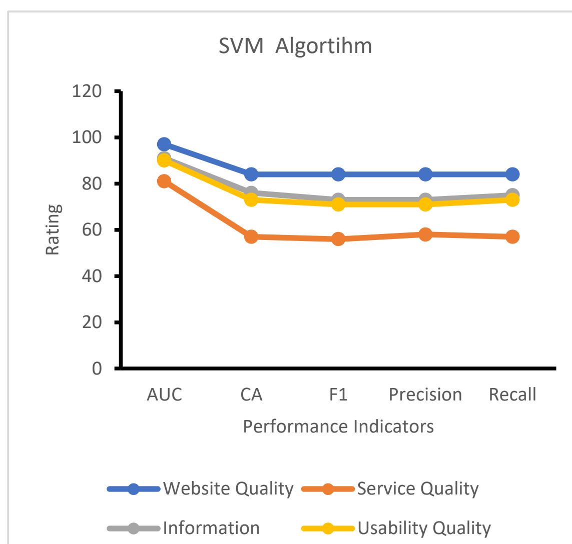


Fig. 25. Results for SVM algorithm

Table 9
Results of Web quality level

Model	**AUC**	**CA**	**F1**	**Precision**	**Recall**
Tree	0.85	0.74	0.74	0.74	0.74
SVM	0.97	0.84	0.84	0.84	0.84
Random Forest	0.95	0.80	0.80	0.81	0.80

Neural Network	0.94	0.80	0.80	0.80	0.80
Niave Bayes	0.97	0.86	0.86	0.86	0.86
Logistic Regression	0.97	0.82	0.82	0.82	0.82
CN2 rule inducer	0.88	0.72	0.71	0.70	0.72

In general, the quality of the PSAU website is very good. The service quality needs more attention to overcome the issues that faced the users of the website in all its features. Furthermore, the usability and information quality dimensions are also very good except some of their features need more attention to improve the overall usability and information quality of the PSAU website and enhanced satisfaction of the user's website.

Therefore, this study contributes to the website quality assessment field in the following ways:

-The study highlights the problems and the website quality level of PSAU university based on three dimensions: usability quality, information quality and service quality.
The study enables designers to avoid mistakes that were discovered and to develop a website with better quality levels.
The results of the study help in identifying the components that require more attention from website designers in creating websites that have high-quality levels of usability, content, and services. Besides, that will help stakeholders accomplish their work efficiently, and easily and enhance the satisfaction level. Furthermore, the study assists the IT team at PSAU university in improving a website to be more effective, attractive, reliable, and easy to use by its users. In addition, it highlights the main issues that face website users and requires more attention to enhance their satisfaction.

5. Conclusion and Future works

This study has presented a novel method for exploring the quality level of the website of Prince Sattam bin Abdulaziz University. University websites are developed to provide information and services to its stakeholders. There is almost no research work conducted to evaluate the quality of the PSAU website. To meet this gap, a novel method is proposed that depends on applying data mining processes to explore the quality indicators of the PSAU website and predict its quality level. The proposed method depends upon first: collecting the data about the quality dimensions which include usability quality, information quality, and service quality that shape the components for the quality of the website. After that, the pre-processing phase was applied to preparing data for the mining process. A descriptive model was applied to identify the pattern and relationships between components. A predictive model was used to predict the quality level of the PSAU website using seven common classification algorithms which are: Tree, SVM, Random Forest, Neural Network, Naive Bayes, Logistic Regression, and CN2 rule inducer. Naive Bayes and Support Vector Machine both achieved the best outcomes in all performance indicators. Then the obtained results are interpreted by the researcher. The main contribution of the study is the website quality of the PSAU was good in usability and information quality dimensions. Moreover, service quality required more attention given to empower satisfaction. Most features for service quality must improve to increase the satisfaction of users to these features. The researcher hopes that study will help website designers and the IT team at PSAU university to enhance the quality of the PSAU website. In the future, the author plans to expand this study by using other data mining techniques to evaluate the PSAU university website. In addition, can extend the evaluation of the quality of the website and other users of the website. Furthermore, can develop a subjective evaluation of the PSAU university website by using opinion-mining approaches.

Acknowledgment

The author would like to thank Prince Sattam bin Abdulaziz University. This study is supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2023/R/1444).

References

Abu Saa, A., Al-Emran, M., & Shaalan , K. (2019). Factors affecting students' performance in higher education: a systematic review of predictive data mining techniques. *Technology, Knowledge and Learning*, pp. 567-598.

Ahmad, A., & Khan, M. N. (2017). Developing a website service quality scale: A confirmatory factor analytic approach. *Journal of internet Commerce, 16*(1), 104-126.

Al-Hagery, M. A., Alzaid, M. A., Alharbi , T. S., & Alhanaya, M. A. (2020). Data mining methods for detecting the most significant factors affecting students' performance. *International Journal of Information Technology and Computer Science*, 1--13.

Al-Omar, K. (2018). Evaluating the Usability and Learnability of the" Blackboard" LMS Using SUS and Data Mining. *In 2018 Second International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 386--390). IEEE.

Andalib, Z., & Danaee, H. (2013). A study on measuring the quality of university website. *Management Science Letters, 3*(7), 1955-1960.

Barnes , S., & Vidgen, R. (2000). WebQual: an exploration of website quality. *ECIS 2000 proceedings*, (p. 7).

Barnes, S. J., & Vidgen, R. T. (2002). An integrative approach to the assessment of e-commerce quality. *Journal of Electronic Commerce Research, 3*(3), 114-127.

Barraood, S. O. (2016). *Quality Evaluation Model Of Yemeni Universities Websites From Students Perspectives (Master Thesis).* University Utara Malaysia.

Boza, B. C., Schiaffino, S., Teyseyre, A., & Godoy, D. (2014). An Approach for Knowledge Discovery in a Web Usability Context. *In Proceedings of the 13th Brazilian Symposium on Human Factors in Computing Systems.*

Budiman, E., Puspitasari, N., Taruk, M., & Maria, E. (2020). Webqual 4.0 and ISO/IEC 9126 Method for website quality evaluation of higher education. *The 2nd International Conference on Applied Science and Technology (iCAST)* .

Chiew, T. K., & Salim, S. S. (2003). Webuse: Website usability evaluation tool. *Malaysian Journal of Computer Science*, 47-57.

El-Halees, A. M. (2014). Software Usability Evaluation Using Opinion. *Journal of software, 9*(2).

Frisdiantara, C., Qamar, K., Ardian, Y., & Rahman , E. F. (2020). The effect of website quality using webqual 4.0 method on student's decision in registering at university of Kanjuruhan malang. *In Proceedings of the International Conference on Industrial Engineering and Operations Management. Dubai.* Dubai.

Gupta, M. K., & Chandra, P. (2020). A comprehensive survey of data mining. *International Journal of Information Technology, 4*(12), 243-1257.

Han, J., Pei, J., & Kamber, M. (2022). *Data mining: concepts and techniques.* Morgan Kaufmann.

Ismail , A., & Kuppusamy, K. (2018). Accessibility of Indian universities' homepages: An exploratory study. *Journal of King Saud University-Computer and Information Sciences*, 268-278.

Kantardzic, M. (2011). *Data mining: concepts, models, methods, and algorithms.* John Wiley & Sons.

Karani, A., Thanki, H., & Achuthan, S. (2021). Impact of University website Usability on Satisfaction a Structural Equational Modeling approach. *Management and Labour Studies, 46*(2), 119-138.

LIN, B., Cassee, N., Serebrenik, A., Bavota, G., Novielli, N., & Lanza, M. (2022). Opinion mining for software development: a systematic literature review. *ACM Transactions on Software Engineering and Methodology (TOSEM), 31*(3), 1-41.

Peker, S., Kucukozer-Cavdar, S., & Cagiltay , K. (2016). Exploring the relationship between web presence and web usability for universities: a case study from Turkey. *Program electronic library and information systems*, 157--174.

Rante, R. R., Fauzan, M., & Hermawan, H. (2020). Website Quality Measurement of Higher Education Services Institution Region IV Using Webqual 4.0 Method. *International Journal of Advances in Data and Information Systems, 1*(2), 89-102. doi:https://doi.org/10.25008/ijadis.v1i2.185

Rashida, M., Islam, K., Kayes, A., Hammoudeh, M., Arefin, M. S., & Habib, M. A. (2021). Towards developing a framework to analyze the qualities of the university websites. *Computers*, 1-16.

Sagar, K., & Saha, A. (2016). Enhancing Usability Inspection Through Data-Mining Techniques An Automated Approach for Detecting Usability Problem Patterns of Academic Websites. *In International Conference on Intelligent Human Computer Interaction, IHCI 2016. Lecture Notes in Computer Science. 10127*, pp. 229-247. Springer, Cham. doi:https://doi.org/10.1007/978-3-319-52503-7_19

Sagar, K., & Saha, A. (2017). Qualitative usability feature selection with ranking a novel approach for ranking the identified usability problematic attributes for academic websites using data-mining techniques. *Human-centric Computing and Information Sciences*, 1-24.

Semerádová, T., & Weinlich, P. (2020). *Website quality and shopping behavior: Quantitative and qualitative evidence.* Springer.

Sutanto, N. H., Kusrini, K., & Nasiri, A. (2021). UNIVERSITY SERVICE WEBSITE QUALITY MEASUREMENT WITH WEBQUAL 4, 0 (CASE STUDY: FACULTY OF BUSINESS AND ECONOMICS ISLAMIC UNIVERSITY OF INDONESIA. *Jurnal Riset Informatika, 3*(3), 225-232. doi:https://doi.org/10.34288/jri.v3i3.231

Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, pp. 1-19.