مجلـــة العلوم الهندسية والتقنية
مجلة علمية محكمة - تصدر عن كلية الهندسة وكلية الحاسبات والمعلوماتية – جامعة ذمار

A Web Mining Approach for Evaluation of Quality Assurance at University of Science and Technology

Ahmed Sultan Al-Hegami[1], Nathem Saeed Ahmed Al-Hddad[2], Hussein Alkhader Alsaeedi[3]

[1]Faculty of Computers and Information Technology, University of Sana'a, Sana'a, Yemen

[2]Department of Computer Science, University of Science and Technology, Aden, Yemen

[3]Department of Computer Science, University of Science and Technology, Sana'a, Yemen

alhegami@su.edu.ye n.alhaddad@aden.ust.edu.ye alhussein1977@gmail.com

ABSTRACT

In today's digital landscape, there is an abundance of websites containing vast amounts of information. However, the challenge lies in extracting valuable knowledge from this sea of data. Knowledge serves as the foundation for making informed decisions, making it essential to develop efficient mining approaches. This research aims to address this problem by proposing a novel framework. The framework will focus on analyzing diverse datasets, particularly those related to educational, administrative, and student activities at the University of Science and Technology. Through an in-depth analysis of this data, the quality of services provided by the university will be evaluated. The ultimate goal is to make optimal decisions that will enhance the university's overall performance. This includes selecting the best teaching staff and improving various other services to ensure a higher level of efficiency and effectiveness.

Keywords: Data mining, Rules Discovery, Quality Assurance, Web mining, Web Content, Web Usage, KDD.

1. INTRODUCTION

Computer science evolved during the last years of the last century, dramatically and rapidly, besides the emergence of many technologies which served humans such as: database, data warehouse, and data integrated, also environment communication, networks, Internet, etc., that contribute to the rise of information.

There are many websites that involve a lot of data but the problem lies in the fact that taking a decision in getting the information and finding out knowledge is quite difficult. For this reason, the mining approach is required to find a proper solution to this problem.

In this study, the content of the various websites of University of Science and Technology will be analyzed in an attempt to find a solution for the problem of the research which is the evaluation of services quality offered by the university.

Accordingly, the available data will be analyzed to evaluate the educational, administrative and students' activities and services because they are the primary services for any educational institution. Algorithms data mining are used for that to create a general role for prediction process with any new case.

After that, we are going to depend on the role that will occur during the contact with a new process so that we realize from the beginning what is good or not. By this, we will predict the quality of primary services in the university. A new framework will be to extract the patterns from the historical data. The resulted patterns will be the basis to create a general role to data with any new case and predict its result.

2. PROBLEM STATEMENT

The study deals with the following issues:

Discovering knowledge from the large volume of data is difficult and not effectively utilized.

The discovery of uninteresting patterns included repeated, conformed, generalized, and specialized knowledge from data mining process.

3. RELATED WORK

In [3], the authors introduce novel possibilities for leveraging robust graph representations in conjunction with popular machine learning algorithms. They demonstrate how graph distance can be used to assess graph similarity and adapt well-known algorithms like k-means clustering and k-nearest neighbors' classification to operate on graphs. The effectiveness of these techniques is evaluated in the field of web content mining, where multiple approaches for representing web document content using graphs are introduced.

In (Ahmed, 2025) [4], the authors compare different web page representations and highlight the impact of data collection and classification processes on the results. They find that a binary representation yields the best performance in binary classification scenarios, but note that other representations may be more effective as the number of classes increases.

In [5], the authors address the challenges of accessing consumer opinions prior to the internet era. They focus on customer reviews as a valuable source of information and propose a visual analysis system for comparing consumer opinions across multiple products. They introduce a supervised pattern discovery method to automate the identification of product features from reviews.

In [7], the authors present a system that allows analysts to interactively correct errors of an automatic system, making the process more efficient than manual tagging. Future work includes improving automatic techniques, studying the strength of opinions, and extracting useful information from other opinion sources.

In [10], the authors propose a solution for concise and conceptual audience metrics in web analytics. They discuss techniques for mining web pages generated by web servers and introduce term-based metrics. They leverage term grouping algorithms and concept hierarchies from ontologies to aggregate the metrics into concept-based metrics using OLAP tools.

In [6], the authors develop indicators of the quality of service provided by Egyptian universities through electronic means. They aim to assure students that the university provides high-quality decisions through the internet as part of the educational experience.

In [8], the researchers divide evaluation criteria for educational websites into content and technical aspects. They propose nine criteria to help teachers evaluate websites for personal resource use and classroom implementation, distinguishing between informative and well-designed websites and those that are uninformative and poorly designed.

In [10], the authors discuss the usage of the internet as a source of information and emphasize the importance of evaluating information skills that will be utilized throughout one's lifetime.

4. The NWMA Approach

4.1. The NWMA Approach Framework

The main core of this research is building a good framework to discover the patterns which will lead to evaluate the university services.

This evolution process will support the university to improve its performance. Figure (4.1) shows the framework of the proposed approach. The proposed approach consists of five main steps as follows:

1. Settings the parameters and creating of nodes.

2. Classification of parameters.
3. Computing of Attributes.
4. Generate partial rule.
5. Produce of final results.

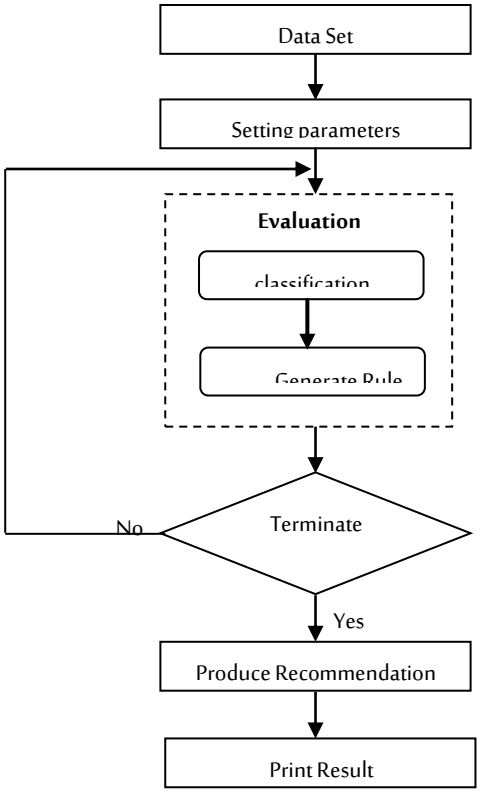


Figure (2): Framework.

4.1.1 Setting parameters

There are many factors on which I will depend in evaluation of the educational services in university of science and technology.

Such factors vary with each other in the degrees of importance.

There are primary and secondary standards. So, the degree of significance has been recorded for every standard.

The objective of this is:

- Achieving flexibility:

To be able to decrease and increase the standards of evaluation and this facilitates the decrease and increase the percentage significance so that it does not affect the total percentage.

- Obtaining accuracy:

All the standards do not have equal values in measuring performance.

However, there is a great significance for some standards of the decision makers.

The following table shows every group with its significance so the total degree is the full percentage %100.

The parameters are classified into two classifications:

4.1.2 Classification of Parameters:

We will classify parameters into two classes:

- **Main Attributes:**

In table (1) shows that:

Student Activities	Administrative services	Educational services
%25	%25	%50

Table (1): Main Attributes.

- **Secondary Attributes**

In table (2) shows that:

Educational services							Administrative services				Student Activities			
The Experience	The Age	Dealing with Students	Using Technology	Library Hours	Using LMS	Official Schedule	Performance of Procedures	Application of Regulations	Dealing with Customer	Quality Systems	Utilizing The Activity	Carrying Out The Activities	Student's interaction with The	Achieving Objective
%25							%25				%50			

Table (2): Secondary Attributes.

5 Computing of Attributes:

The following notations are used in the algorithms:

MC=Main Class

SA=Secondary Attributes

CP=Class Degree

SP=Secondary Attributes Degree

WC=Class Weight

WS=Secondary Attributes Weight

1.The following law is used to compute the main class degree:

$$CP = \sum_{i=1}^{n} (\mathrm{SA}) * \boldsymbol{100/Wc}$$

where,

CP=Class Degree

WC=Class Weight

SA=Secondary Attributes

n=the number of Secondary Attributes

2.The following law is used to compute the secondary attributes value:

$$SA = \sum_{i=1}^{n} (\mathrm{SP}) * \boldsymbol{Wc/Ws}$$

where,

SA=Secondary Attributes

SP=Secondary Attributes Degree

WC=Class Weight

WS=Secondary Attributes Weight

n=the number of secondary attributes degree

5.1. Implementation and Experimentation

The Decision Rules Generated from the NWMA approach:

5.1.1 First: Educational Services

1.If teacher age ="$\leq$30" then bad performance

2.If teacher age = "30 .. 40" then good performance

3.If teacher age ="> 40" and teacher expertise ="small then bad performance

4.If teacher age ="> 40" and teacher expertise ="medium" then good performance

5.If teacher age ="> 40" and teacher expertise ="large" then good performance

We used the above rules to building the model that shown in the next figure:

6 Second: The Administrative Services

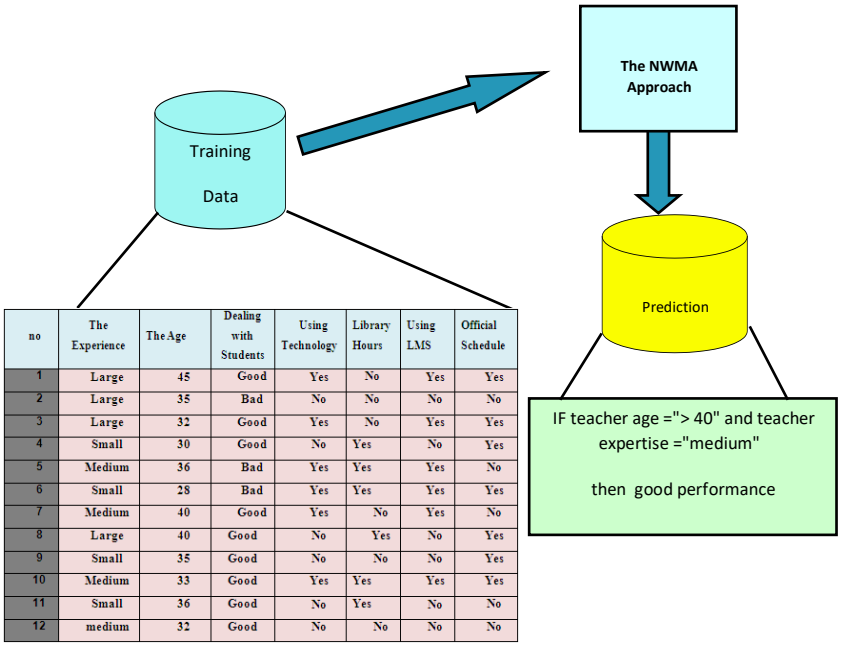
1.If Dealing with Students ="low" then bad performance

2.If Dealing with Students ="medium" then good performance

3.If Dealing with Students="high" and applied of regulations="low" then bad performance

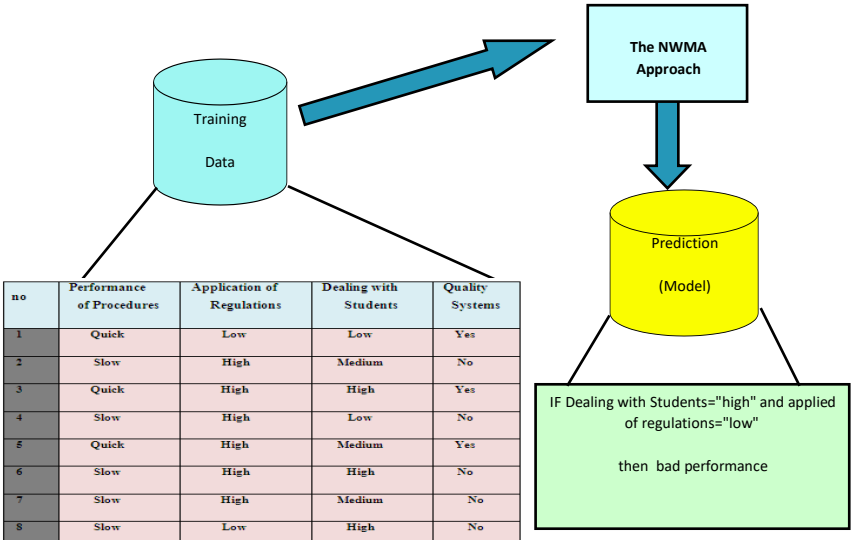
4.If Dealing with Students="high" and applied of regulations="high" then good performance

We used the above rules to building the model that shown in the next figure:



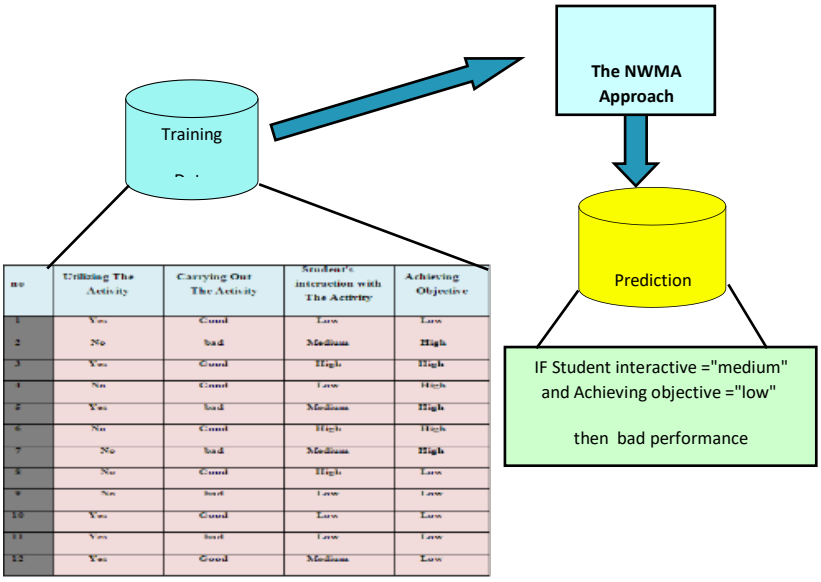
no	The Experience	The Age	Dealing with Students	Using Technology	Library Hours	Using LMS	Official Schedule
1	Large	45	Good	Yes	No	Yes	Yes
2	Large	35	Bad	No	No	No	No
3	Large	32	Good	Yes	No	Yes	Yes
4	Small	30	Good	No	Yes	No	Yes
5	Medium	36	Bad	Yes	Yes	Yes	No
6	Small	28	Bad	Yes	Yes	Yes	Yes
7	Medium	40	Good	Yes	No	Yes	No
8	Large	40	Good	No	Yes	No	Yes
9	Small	35	Good	No	No	No	Yes
10	Medium	33	Good	Yes	Yes	Yes	Yes
11	Small	36	Good	No	Yes	No	No
12	medium	32	Good	No	No	No	No

Figure (3): Module building for The Administrative Services



no	Performance of Procedures	Application of Regulations	Dealing with Students	Quality Systems
1	Quick	Low	Low	Yes
2	Slow	High	Medium	No
3	Quick	High	High	Yes
4	Slow	High	Low	No
5	Quick	High	Medium	Yes
6	Slow	High	High	No
7	Slow	High	Medium	No
8	Slow	Low	High	No
9	Quick	Low	Low	No
10	Slow	Low	Low	Yes
11	Quick	Low	Low	Yes
12	Quick	Low	Medium	Yes

Figure (4): module building for The Educational Services.



no	Utilizing The Activity	Carrying Out The Activity	Student's interaction with The Activity	Achieving Objective
1	Yes	Good	Low	Low
2	No	bad	Medium	High
3	Yes	Good	High	High
4	No	Good	Low	High
5	Yes	bad	Medium	High
6	No	Good	High	High
7	No	bad	Medium	High
8	No	Good	High	Low
9	No	bad	Low	Low
10	Yes	Good	Low	Low
11	Yes	bad	Low	Low
12	Yes	Good	Medium	Low

Figure (5): Module building for Students Activities Services

7 Third: Students the Activities Services

1.If Student interactive="low" then bad performance

2.If Student interactive ="medium" and Achieving objective ="low" then bad performance

3.If Student interactive ="medium" and Achieving objective ="high" then good performance

4.If Student interactive ="high" and Achieving objective ="low" then good performance

5.If Student interactive ="high" and Achieving objective ="high" then good performance

We used the above rules to building the model that shown in the next figure:

6.2. Prediction

- **Prediction for The Educational Services**

Now, we will use the new record as attesting data form prediction. So we want to predict the performance of the following the teacher by using the next:

The Experience = " Medium "

The Age = "43"

Dealing with Students ="Good"

Using Technology = "No"

Library Hours = "Yes"

Using LMS =" No"

Official Schedule= "No"

Performance ="?"

We cannot use the rule no. 1

And We cannot use the rule no. 2

And We cannot use the rule no. 3

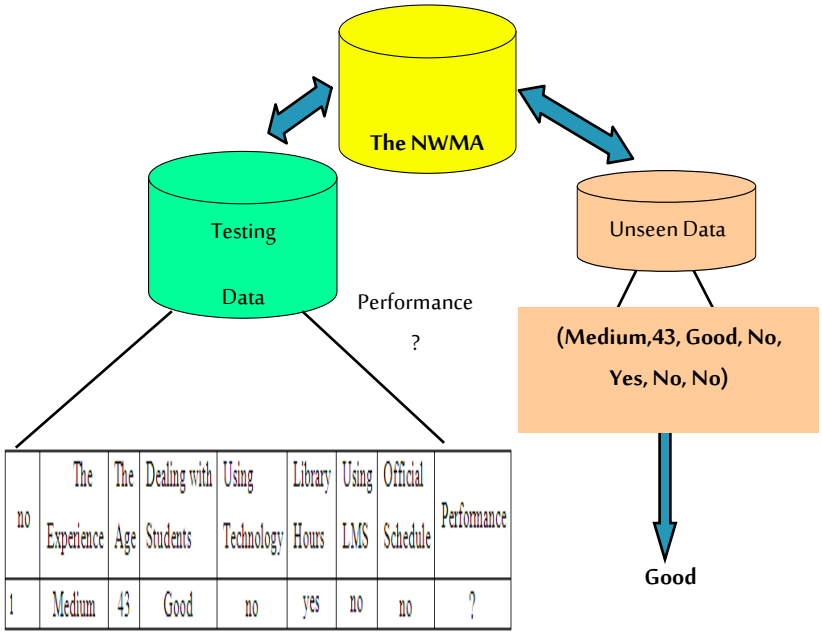
But we can use the rule no.4

IF teacher age ="> 40" and teacher expertise ="medium"

then good performance

If 43>40 (TRUE) and "medium" ="medium"(TRUE) then good performance

The use of the model in prediction is shown in the in the following figure:



no	The Experience	The Age	Dealing with Students	Using Technology	Library Hours	Using LMS	Official Schedule	Performance
1	Medium	43	Good	no	yes	no	no	?

Figure (6): Using Model in Prediction for the Educational Services

- **Prediction for the Administrative Services**

Now we will use the new record as attesting data form prediction. So we want to predict the performance of new management by using the next:

Performance of Procedures ="Quick"

Application of Regulations ="low"

Dealing with Students ="high"

Quality Systems ="yes"

Performance ="?"

We cannot use the rule no. 1

And We cannot use the rule no. 2

But we can use the rule no.3

IF Dealing with Students="high" and applied of regulations="low"

then bad performance

If high=high (TRUE) and "low" ="low"(TRUE) then bad performance

The use of the model in prediction is shown in the in the following figure:

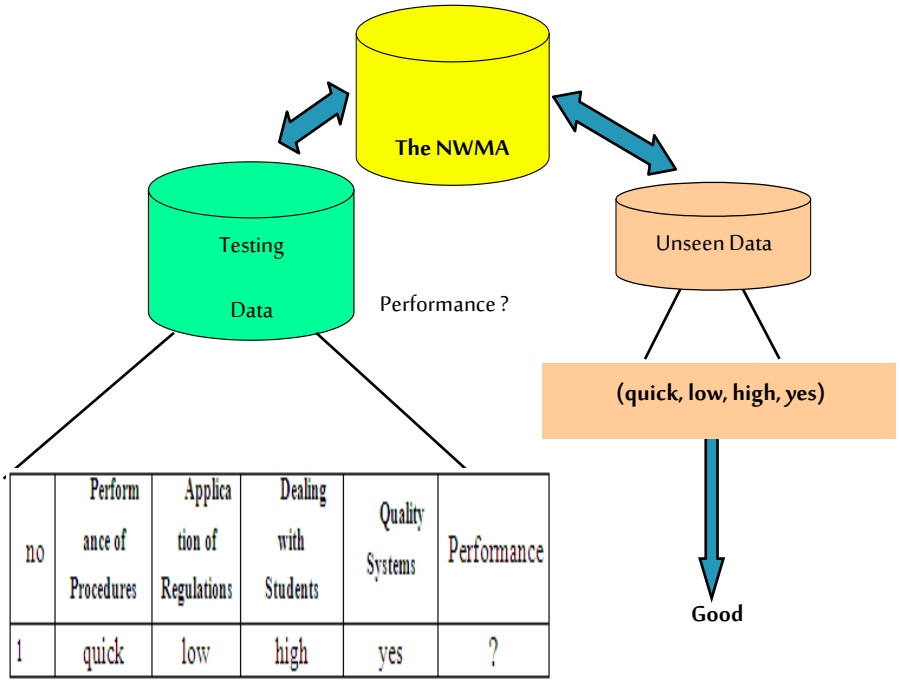


Figure (7): The Use of Model in Prediction for The Administrative Services

- **Prediction for Students the Activities Services**

Now we will use the new record as attesting data form prediction. So we want to predict the performance of new activity by using the next:

Utilizing The Activity = "No"

Carrying Out The Activity ="bad"

Student's interaction with The Activity= " Medium"

Achieving Objective = "low"

Performance ="?"

We cannot use the rule no. 1

But we can use the rule no.2

IF Student interactive ="medium" and Achieving objective ="low" then bad performance

If medium=medium (TRUE) and low="low"(TRUE) then bad performance

The use of the model in prediction is shown in the in the following figure:

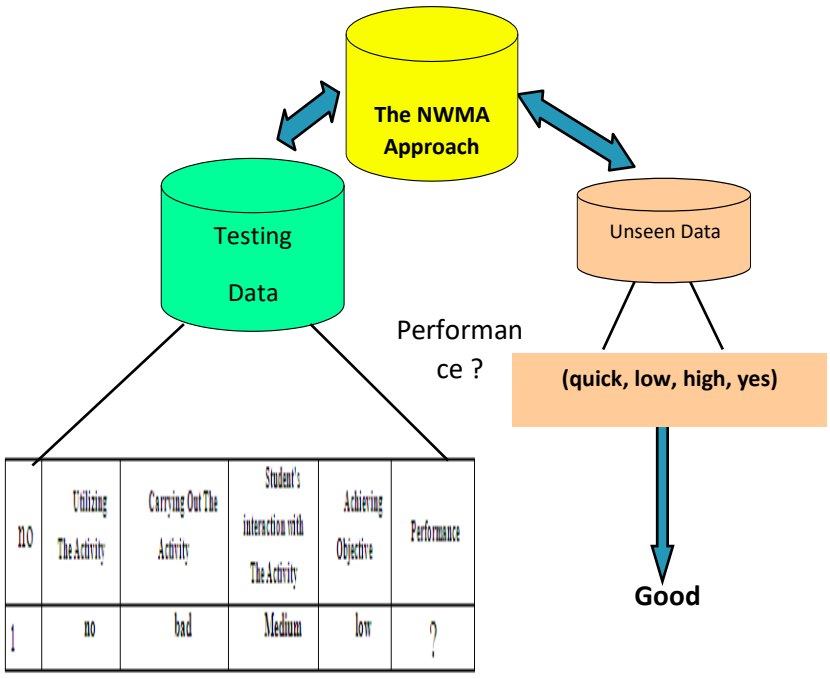


Figure (7): The Use of Model in Prediction for Students' Activities Services

7.1. Conclusions:

In this research, we proposed a web mining approach for evaluation of quality assurance at University of Science and Technology.

We implemented a framework using C++ programming Language. The framework is experimented and evaluated using real dataset and results have been presented.

The Performance of services in UST depended on educational services.

7.2. Future work:

Future work should consist of more experiments with other dataset, as well as more elaborated experiments to optimize several parameters of the algorithm. In addition, the framework can be implemented and experimented with large dataset by programming it in any programming language.

References:

[1] Julio Ponce And Adem Karahoca, Data Mining and Knowledge Discovery in Real Life Applications.

[2] Olfa Nasraoui, Myra Spiliopoulou, Jaideep Srivastava, Bamshad Mobasher, Brij Masand, Advances in Web Mining and Web Usage Analysis.

[3] Amento, L. Terveen, and W. Hill. "Does Authority Mean Quality? Predicting Expert Quality Ratings of Web Documents. In Proc. of the 23rd ACM SIGIR Conf. on Research and Development in Information Retrieval", 2000, pp. 296–303.

[4] Ayres, J. Gehrke, T. Yiu, and J. Flannick. "Sequential Pattern Mining Using Bitmaps In Proc. of the Eighth ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining (KDD'02)", 2002, pp. 429–435.

[5] Baeza-Yates and B. Ribeiro-Neto. "Modern Information Retrieval. Addison-Wesley", 1999.

[6] Barbar, Y. Li and J. Couto. COOLCAT. "an Entropy-Based Algorithm for Categorical Clustering. In Proc. of the 11th Intl. Conf. on Information and knowledge management (CIKM'02)", 2002, pp. 582–589.

[7] Berendt, B. Mobasher, M. Nakagawa, and M. Spiliopoulou. "The Impact of Site Structure and User Environment on Session Reconstruction in Web Usage Analysis. In Proc. of the KDD'02 WebKDD Workshop", 2002.

[8] Blum, and S. Chawla. "Learning from Labeled and Unlabeled Data Using Graph Mincuts. In Proc. of Intl. Conf. on Machine Learning (ICML'01)", 2001, pp.19–26.

[9] Brunk and M. J. Pazzani. "An Investigation of Noise-Tolerant Relational Concept Learning Algorithms. In Proc. of the 8th Intl. "Workshop on Machine Learning", 1991, pp. 389–393.

[10] Buchner and M. D. Mulvenna. "Discovering Internet Marketing Intelligence through Online Analytical Web Usage Mining. In Proc. of the ACM SIGMOD Intl. Conf. on Management of Data (SIGMOD'99)", 1999, pp. 54–61.

[11] Buehrer, S. Parthasarathy, and A. Ghoting. "Out-of-core frequent pattern mining on a commodity PC. In Proc. of the ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining (KDD'06)", 2006, pp. 86 – 95.

[12] Burges. "A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery, 2(2)", 1998, pp. 955–974.

[13] Cooley, B. Mobasher, and J. Srivastava. "Data Preparation for Mining World Wide Web Browsing Patterns. Knowledge and Information Systems, 1(1)", 1999, pp. 5–32.

[14] Dhillon. "Co-Clustering Documents and Words Using Bipartite Spectral Graph Partitioning. In Proc. of the 7th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining (KDD'01)", 2001, pp. 269–274.

[15] Dhillon, S. Mallela, and D. S. Modha. Information-Theoretic Co-Clustering. In Proc. of The Ninth ACM SIGKDD Intl. Conf. "on Knowledge Discovery and Data Mining (KDD'03)", 2003, pp. 89–98.

[16] Domingos, and M. J. Pazzani. "On the Optimality of the Simple Bayesian Classifier under Zero-One Loss. Machine Learning", 1997,29(2–3), pp. 103–130.

[17] http://www.flazx.com

[18] http://www.ust.edu.ye

[19] http://www.aden.ust.edu.ye

[20] http://www.taiz.ust.edu.ye

[21] http://www.hoda.ust.edu.ye

[22] http://kdd.ics.edu

[23] Ahmed, M. R. A. (2025). Accreditation and Quality Assurance: Exploring Impact and Assessing Institutional Change in the US and Saudi Arabian Higher Education Institutions. *Arts for Linguistic & Literary Studies*, *7*(1), 626–639. https://doi.org/10.53286/arts.v7i1.2419