Int. J. Advance Soft Compu. Appl, Vol. 13, No. 2, July 2021
Print ISSN: 2710-1274, Online ISSN: 2074-8523

Automating SWOT Analysis Using Machine Learning Methods

Ahmad Abu-Alaish[1], Ghaith Jaradat[2], Mahmoud Al-Shugran[3], Iyas Alodat[4]

[1,3,4]Department of Computer Science, Faculty of Computer Science & Information Technology, Jerash University, 26150-311, Jerash, Jordan.
e-mail: ahmad.abualaish@jpu.edu.jo, m.shugran@jpu.edu.jo, eyas.odat@jpu.edu.jo
[2]Department of Computer Science, Faculty of Computer Science & Informatics, Amman Arab University, 2234-11953, Mubis-Amman, Jordan.
e-mail: g.jaradat@aau.edu.jo

Abstract

Quality assurance is one major concern for the faculty of Computer Science and Information Technology (FCSIT) at Jerash University. It involves eight standards including the strategic planning. SWOT analysis is a method meant for assisting the formulation of strategy and planning. An application to strategic planning process formulation for the FCSIT is described. This research studies the SWOT analysis with a major concern of drawing more conclusions using machine learning methods. Data mining is a subfield of machine learning, which focuses on exploratory data analysis using supervised or unsupervised learning. Data mining techniques help fetching required knowledge from raw data to make decisions more confidently interpreted and automated. In this study, regression, classification, clustering, association rules, attributes selection techniques are used to mine data from the SWOT analysis. Using Weka workbench, results of each technique is obtained and interpreted with the favor of the factors that have impact on the success of the strategic plan. The outcome presents a high level of satisfaction regarding employee, and a vibrant level of satisfaction regarding students. Therefore, the developed quality assurance framework is stable but needs more improvements to overcome the dissatisfaction of many students regarding services, supervision, awards and activities.

Keywords: *Machine learning methods; data mining techniques; SWOT analysis; FCSIT, Weka.*

1 Introduction

In order to meet the standards of quality assurance, Jerash University for the last three decades - and recently the faculty of Computer Science and Information Technology (FCSIT) - involves a continuous strategic development which includes three components of: (i) an annual development of a corporate plan for submission to the Jordanian Higher Education Funding Council; (ii) an annual five-year planning process undertaken by the strategy committee; and (iii) the formulation and sometimes adoption of strategic initiatives throughout the year.

The Steering Committee in the university considered that a whole new strategic plan is needed to overcome current problems and to seek effective development. These problems are recently occurred due to the rapidly changing situation of the decisions of the higher education's ministry. It was agreed that the committee would have a strategic planning for three academic years which would aim to produce recommendations for future consideration. It was agreed that a SWOT (Strength, Weakness, Opportunities, Threats) analysis would form the core of the strategy, which would be facilitated by the author who was the representative member of the steering committee for the FCSIT at the university.

This paper first summarizes the executive plan for the FCSIT with a brief illustration of the strategic planning including SWOT analysis. Then, it illustrates the methodology used for concluding the SWOT analysis more effectively using machine learning methods. Then, it discusses the obtained results and lastly it concludes the study in the favor of the factors that have impact on the success of the strategic plan.

2 Related Work

The FCSIT seeks locally and regionally to be an advanced faculty with distinguished programs, productive academic staffs, and qualified alumni. That is to achieve a superior regional educational system. Therefore, a strategic plan is prepared to consider the vision including the management and staffs for what will be accomplished in the years 2019-2022.

The biggest effort in preparing the strategic plan is determining pivots, goals, and strategies. There are 8 pivots that are considered in the FCSIT for this strategic plan including: governance; accreditation; scientific research, development, postgraduate studies; community service; funding; infrastructure; university environment; support and development of human resources. These pivots identify the main elements of the strategic plan, such as: executive plan; priorities; performance indicators; executive responsibility; timeframe; required materials.

It also monitored the changes and possible achievements in the numbers of students, academic programs, faculty members, facilities and equipment. In addition, it monitored the achievements of the academic staffs in scientific research and publication, authoring, translation, preparation of electronic materials, and community services (e.g., seminars, workshops, and consultations).

Based on the national strategy of the higher education and scientific research, it is dedicated for graduating students (with their various levels) to meet the needs of the industrial and production sector. With respect to the national priorities, it was an urgent need to provide strategic plans for the FCSIT, that aim to organize and utilize the available human resources and capabilities to achieve the goals that were translated from the objectives for an optimal usage of cost and time with highest levels of quality. Therefore, the FCSIT has responded to the university's recommendations to develop a strategic plan for the coming years 2019-2022.

2.1 Procedures of Implementing the Strategic Planning

The strategic planning is implemented in the following procedures:

2.1.1 SWOT Analysis: Strengths, Weaknesses, Opportunities, Threats

For start, the development of the current strategy was examining potential strategies derived from SWOT analysis, identifying the common elements of these different elements, and the information collected as a result of the analysis. That is to strengthen and develop the faculty's strength and increase its opportunities.

To reduce or eliminate the weaknesses or the threats, there is better cooperation with external and internal beneficiaries, both in the Ministry of Education, or governmental and non-governmental organizations that benefit from the services provided by the FCSIT, other faculties in the university in addition to the corresponding faculties in Jordan.

Internal environment is meant to be the situations, variables, and available resources that have a direct impact on the performance of the faculty which could be controlled. On the other hand, external environment is meant to be a group of variables that has a direct or indirect impact on the activities or decisions which could be out of control. Similar studies were implemented in different fields including competitiveness, profession, and SWOT's conceptual framework such as: [1] [2] [3] [4].

SWOT analysis aims at identifying factors regarding strengths and weaknesses of the FCSIT, as well as opportunities and threats in the environment. Our study considers critical factors presented in table 1. These factors are developed to build on strengths, eliminate weaknesses, and exploit opportunities or counter threats. Strengths and weaknesses are identified by an internal environment of the FCSIT while the opportunities and threats are identified by an external environment. The internal environment examines all aspects of the FCSIT covering, e.g., personnel, facilities, location and services. The external environment scans the economic, social, technological and competitive environment. Based on the work of [5], a variation of SWOT analysis (e.g., the TOWS matrix) is also introduced. In the TOWS matrix various factors are identified and then paired. For example, an opportunity with strength, with the intention of stimulating a new strategic initiative (see Table 2). The TOWS matrix identifies the relationship between factors in four perspectives to come out with the right decisions for a soundful environmental examination of all FCSIT aspects.

Table 1. SWOT factors and strategies

Factor	Strategies
Opportunities O1-O8	• develop the science strategy • develop social sciences • develop relationships • improve undergraduate experience • continuing professional development • fundraising • expansion of postgraduate programs • widening access
Threats T1-T3	• develop the science strategy • fundraising • human resources policy
Strengths S1-S7	• develop the science strategy • develop social sciences • develop relationships • continuing professional development • fundraising • human resources policy • expansion of postgraduate programs
Weaknesses W1-W3	• develop the science strategy • Fundraising • human resources policy

Table 2. TOWS matrix for the FCSIT

Factor	Strengths:	Weaknesses:
Opportunities:	Utilized by strengths. SO strategies: S1+O3, S1+O4 S3+O6, S5+O6 S2+O2 S4+O3 S5+O1 S6+O3, S6+O4, S6+O7 S7+O3, S7+O7	To avoid weaknesses. WO strategies: W1+O2, W1+O3, W1+O4 W2+O5 W3+O3, W3+O5
Threats:	Avoided by strengths. ST strategies: S2+T1, S3+T1, S6+T1 S1+T3, S4+T3, S5+T3	To minimized weaknesses. WT strategies: W1+T1 W2+T3, W3+T3

3 Methodology and Procedures

This section conducted a detailed description of the procedures that are used in implementing the study, including: identifying the method of the study; description the study population, study sampling, questionnaires preparation, study validation and consistency, study processes, study variables, and the statistical method that are used in processing the obtained results.
To measure the variables of the study and its dimensions, the researchers relied on measuring accuracy and parameters settings of machine learning methods and measuring the features of the FCSIT dataset. First, by presenting questions that reflects the dimensionality of the dataset and the features of the implemented methods. Second, by analyzing the outcomes of those methods to answer those questions.
In this study, an automation of the SWOT analysis via a number of machine learning methods (namely, *WekaDeeplearning4j*, *GenClust++*, and *Apriori*) based on the surveys and/or questionnaires conducted in the FCSIT. The contribution of this paper is outlined as follows:
a. Quality management part:

i. An iterative framework of developing strategic planning for the FCSIT.
ii. Factors that trigger a range of potential strategic initiatives.
iii. Utilizing 3 machine learning methods to predict SWOT analysis.
iv. Obtaining a high level of member's satisfaction.

b. Automation part (using machine learning):

i. The classification, predictive and association models of *GenClust++*, *WekaDeeplearning4j*, and *Apriori*; respectively have an automated phased machine learning process, starting with grouping, training a Convolutional Neural Network (CNN), extracting features, and finally rule set generation by mapping associations between critical factors.
ii. The three models are trained and tested on a relatively large dataset.
iii. The three models have shown their capability to obtain very accurate results across four perspectives.

Similar automation for different fields such as medical implementations can be found in [6] [7] [8] [9].

3.1 Study Questions

To achieve the objectives of the study and answering its research questions, a correlative descriptive survey is used which is considered suitable for this study. These research questions are:

1. *What is the level of the academic quality at the FCSIT from the perspective of faculty members?*

2. *Is there a correlation of a statistical significance at the level between student satisfaction and lecturers' satisfaction from lecturers' perspective?*
3. *Are there differences of a statistical significance at the level ($\alpha \leq 0.05$) between the means of respondents' answers (from the study sample) regarding their satisfaction level from students and lecturers perspective referring to the demographic variables (responder, task, rank)?*

Meanwhile, for the machine learning part, two research questions are addressed:

4. *What features/attributes of the FCSIT dataset that have the greatest impact on the classification and prediction tasks?*
5. *How accurate is the three models?*

This comprises testing a number of configurations for three models and then selecting the best configuration based on their results. We first clarify the FCSIT dataset (see sections 3.2 and 3.3), then the Weka machine learning workbench, then the three techniques *WekaDeeplearning4j*, *GenClust++*, and *Apriori*. Finally, we demonstrate our obtained results.

3.2 Study Population and Sample

The population is comprised of 4200 students, 165 lecturers, and 110 administrators of Jerash University for the academic years 2016-2017, 2017-2018, and 2018-2019. The sample contains students, lecturers, and administrators of the FCSIT. They were randomly selected out of the total study population. Table 3 shows the distribution of the sample members based on the study variables.

3.3 Study Questionnaire

Questionnaires with biographies and Likert-Scales are used for conducting the study, where each question has been developed by reviewing related literature, specifying domains, customizing paragraphs for each domain, conducting a pilot study for properness examination and, validation and verification by a well experienced academic staff. Finally, conducting the final draft containing two sections:

1. Initial data for the study sample which consists of 7 variables including gender, nationality, year of study, graduation year, qualification, department, age, experience, rank, task, and responder.
2. Concerns with the satisfaction of FCSIT members about the provided services, programs, awards, and activities. It is divided into 8 domains to measure the degree of satisfaction for students, lecturers, and administrators. With a total of 111 paragraphs, these domains include satisfaction of academic services, activities evaluation, and privileges.

A number of 327 questionnaires is distributed targeting groups of students (including alumni) and academics (including administrators).

Table 3. The distribution of the sample members based on the study variables

	Categories	Frequency	Percentage
Responder	Student	243	75%
	Employee (Academic, Administrator)	84	25%
	Student		
Gender	Male	185	76%
	Female	58	24%
Nationality	Jordanian	202	83%
	Non-Jordanian	41	17%
Year of study	1st	58	24%
	2nd	82	33%
	3rd	58	24%
	4th	45	19%
Department	Computer Science	129	53%
	Computer Networks	98	40%
	Computer Information Systems	16	7%
	Employee		
Gender	Male	62	74%
	Female	22	26%
Nationality	Jordanian	75	89%
	Non-Jordanian	9	11%
Task	Academic	36	43%
	Administrative academic	48	57%
Rank	Full Professor	2	5%
	Associate Professor	14	40%
	Assistant Professor	20	55%
Experience	Less than 5 years	19	22%
	5-10 years	45	54%
	More than 10 years	20	24%
Age	Less than 30	8	10%
	31-40	40	48%
	41-50	24	28%
	More than 51	12	14%
Qualification	BSc	27	32%
	MSc	25	30%
	PhD	32	38%

3.4 Similar Approaches (Mining Survey Data for SWOT Analysis)

Some researchers performed a SWOT analysis with the integration of fuzzy analytic hierarchy process such as [10]. This fuzzy approach allows decision-makers to provide fuzzy judgments in pair-wise comparisons instead of exact judgments in order to fully reflect a style of human thinking which in turn produces the sensible quantitative values for the SWOT factors. The main step is calculating the relative importance from fuzzy values identified by decision makers.

The hybrid approach of [11] quantifies SWOT factors based on the preference of multiple decision makers on SWOT factors and groups to provide more versatile information for evaluating the relative importance of SWOT factors.

Hence, the main goal is to make prioritized SWOT factors, where quantitative analysis mainly focuses on determining and computing a relative importance. In some cases, this makes SWOT factors measurable which are produced only based on the university's perspective without considering a lecturer's or student's perspective. Thus, this research paper suggests the use of administrators and lecturer's evaluation, and more importantly the use of students' orientation in the SWOT framework in order to make better use of the analysis. This may also be useful for conducting data mining techniques for evaluation and orientation. This may provide a concrete evaluation to test the validity of the approach used for FCSIT.

3.5 Machine Learning Methods

For further assessment of the strategic planning development framework, a machine learning method is applied afterwards to assist the outcomes of the analysis and derive more conclusions. Machine learning is a subset field of artificial intelligence mainly concerns with designing an intelligent agent that perceives and make decisions to maximize the possibilities of achieving its goal without being explicitly programmed. It is divided into three main categories of data mining techniques: supervised learning (e.g., classification and regression), unsupervised learning (e.g., clustering and reduction), and reinforcement learning (e.g., reward maximization). For more information, refer to [12] [13] [14] [15].

3.5.1 Data Mining Techniques

Data mining is a powerful tool to study patterns and relations in numerous data for different applications. It consists of two main tasks, prediction and description. Prediction extracts patterns based on the current data to predict either unknown or future data, while description extracts patterns for describing common characteristics of the data [16]. In one hand, in the prediction task, principal component analysis, neural networks, fuzzy logic, and other learning tools are used for classification. In the other hand, in the description task, clustering is used to conduct unsupervised learning to help understand and predict values for new

data based on training dataset conducted by using decision trees and association rules. For more details refer to [13] [17] [18] [19] [20].
One of the main tasks of our study is measuring importance as the relative strength of attributes that contributed to the overall satisfaction of administrators, lecturers, and students. This requires methods to discover or extract relationship between an input and a predefined class. Hence, prediction and classification are the most suitable methods to be applied in measuring importance. Consequently, this measures the performance of the SWOT framework.

3.5.2 Data Mining Techniques Using Weka

In this study, Weka [1] Machine Learning software is used to measure the importance of the factors presented in the TOWS matrix. Weka is a collection of machine learning algorithms for data mining tasks. It contains tools for data preparation, classification, regression, clustering, association rules mining, and visualization.
For the purpose of assessing the outcomes of the SWOT analysis, this study employs a predictive task using classification and regression algorithm named *WekaDeeplearning4j* (developed by [21]) across 8 datasets of FCSIT. It is a Java-based deep learning implementing a multilayer perceptron classifier, where it has sophisticated neural network architecture. It consists of a Convolutional Layer useful for text embedding; a Dense Layer which connects all of its units to all units of its parent layer; and a Subsampling Layer which subsamples groups of units of the parent layer by different strategies (e.g., mean, maximum). A common batch normalization strategy is applied on the activations of the parent layer. This classifier uses long short-term memory approach and a Global Pooling Layer over time or on sequences. Finally, the Output Layer generates classification or regression outputs.
Then, it employs a descriptive task using clustering algorithm named *GenClust++* and *Apriori*. On one hand, *GenClust++* (developed by [22]) implements clustering based on genetic algorithm for centroid generation. It measures the similarity using the Euclidean Distance. It also uses basic missing value handling from *SimpleKMeans*. If an operation generates a chromosome where all records are assigned to a single cluster, chromosome will be mutated until at least 2 clusters are found. The starting generation for the chromosome selection operation evolves up to 60 generations. On the other hand, association rules algorithm using *Apriori* extracts the most suitable rules for learning from an interesting pattern of the responders. It builds up attribute-value (item) sets that maximize the number of instances that can be explained (coverage of the dataset). The search through item space is very much like the problem of attribute selection and subset search. Selecting attributes is performed using *BestFirst* and a best first search with a greedy hill climbing algorithm (*CorrelationAttributeEval*) for pseudo-randomly

---

[1] Downloaded from the official website of the University of Waikato, New Zealand: https://www.cs.waikato.ac.nz/ml/weka/

choosing attributes. Then, the worth of selected attributes is evaluated considering high correlated features.
These algorithms are selected based on preliminary experiments that are performed the best for the 8 datasets and their attributes. They are implemented instead of the conventional classifiers provided by Weka. The reason behind this, is to guarantee clarity of classes, stability of classification on the same set of training, and flexibility of classification process and its basis. The basis of classification is either simple qualitative (e.g., two classes) or quantitative (numerical) classification. For this study, a simple qualitative classification is conducted.

4 Questionnaire Validity and Consistency

4.1 Validity of the Questionnaire

The contents validity of the questionnaire has been verified by 10 university professors who specialize in educational administration, measurement and evaluation. Their observations and suggestions were studied, as the paragraph that obtained an agreement at 80% was adopted as a measure of their acceptance.

4.2 Validity of the Internal Consistency of the Questionnaire's Paragraphs

To extract the significance of construction validity for the scale, paragraph correlation coefficients for each section was extracted with the total score for the domain to which they belong. This was achieved by conducting a pilot study of 30% of members. Whereas Pearson correlation coefficients represents a significance of validity for each paragraph in the form of a correlation coefficient between each paragraph and overall degree of the section to which it belongs. This was achieved using the statistical package for social science (IBM-SPSS) software.
Paragraphs' correlation coefficients for measuring the degree of FCSIT members' satisfaction with the whole questionnaire is ranged between (-0.417 – 0.702), and with each domain (-0.237 - 0.907). See tables 4 and 5.
It should be noted that all correlation coefficients are of acceptable degrees and statistically significant. Therefore, none of the paragraphs was deleted. It is clear from Tables 4 and 5 that all correlation coefficients for the questionnaire domains and the questionnaire are high. Thus, they are suitable for the purpose of our study. This indicates the strength of the internal cohesion (consistency) of the paragraphs for each domain in the questionnaire.

Table 4. Correlation coefficients between paragraphs, overall degree, and domain to which they belong

Parag.	Corr. Coef. vs Domain	Corr. Coef. vs Questionnaire	Parag.	Corr. Coef. vs Domain	Corr. Coef. vs Questionnaire	Parag.	Corr. Coef. vs Domain	Corr. Coef. vs Questionnaire
1	.579**	.138	38	.356**	.204	75	-.237	.022
2	.516**	.032	39	.711**	.487**	76	.648**	.683**
3	.647**	.333*	40	.774**	.473**	77	.813**	.619**
4	.519**	.311*	41	.761**	.511**	78	.817**	.531**
5	.747**	.352*	42	.721**	.358**	79	.907**	.483*
6	.743**	.217	43	.785**	.435**	80	.754**	.310
7	.735**	.375**	44	.725**	.378**	81	.851**	.392
8	.499**	.223	45	.692**	.396**	82	.869**	.391
9	.758**	.430**	46	.591**	.292*	83	.868**	.593**
10	.744**	.291*	47	.713**	.394**	84	.639**	.537**
11	.800**	.347*	48	.725**	.536**	85	.841**	.466*
12	.622**	.385**	49	.702*	.286	86	.611**	.401*
13	.639**	.291*	50	.716**	.112	87	.632**	.218
14	.712**	.249	51	.818**	.183	88	.782**	.302
15	.627**	.273	52	.599*	.125	89	.631**	.166
16	.526**	.276*	53	.727**	.234	90	.826**	.331
17	.689**	.376**	54	.478	.100	91	.812**	.316
18	.606**	.375**	55	.366	.424	92	.504**	.020
19	.211	.065	56	.806**	.505	93	.605**	.235
20	.596**	.297**	57	.871**	.333	94	.618**	.081
21	.387**	.344**	58	.820**	.480	95	.738**	.510**
22	.559**	.420**	59	.564	.233	96	.837**	.514**
23	.533**	.430**	60	.784**	.702*	97	.851**	.449**
24	.541**	.324**	61	.778**	.188	98	.553**	-.129
25	.577**	.372**	62	.768**	.062	99	.807**	.502**
26	.691**	.336**	63	.268	-.047	100	.586**	.036
27	.719**	.369**	64	.550**	-.353	101	.687**	.427*
28	.720**	.425**	65	.393*	-.417*	102	.627**	-.035
29	.654**	.369**	66	.077	.141	103	.703**	.233
30	.664**	.342**	67	.572**	.190	104	.894**	-.115
31	.606**	.381**	68	.526**	.100	105	.369	.081
32	.730**	.406**	69	.450*	.007	106	.862**	-.049
33	.708**	.472**	70	.363*	.100	107	.754**	-.062
34	.798**	.426**	71	.083	.370*	108	.641**	.566**
35	.579**	.283**	72	.247	-.257	109	.907**	.027
36	.567**	.296**	73	.351	.083	110	.708**	.211
37	.775**	.376**	74	-.131	.109	111	.828**	.285

** Correlation is significant at the 0.01 level
* Correlation is significant at the 0.05 level

Table 5. Correlation coefficients between domains and the overall score

Domain	Academic Supervision	Student Services	Website	Alumni Satisfaction	Activity	Lecturer Satisfaction	Job Satisfaction	Awards
Academic Supervision	1							
Student Services	.059	1						
Website	.165	.006	1					
Alumni Satisfaction	.115	.186	.097	1				
Activity	.145	.238	.267	.178	1			
Lecturer Satisfaction	.065	.381	.009	.133	.128	1		
Job Satisfaction	.043	.162	.144	.456	.154	.522	1	
Awards	.310	.217	.294	.206	.071	.360	.643	1

4.3 Domains' Stability and Questionnaire Consistency

The questionnaire is validated for consistency using reliability statistics, the Cronbach's alpha. The Cronbach's alpha for the whole questionnaire has the value of (.845). All consistency coefficients of the questionnaire's domains are high values and suitable for the study. See table 6.

Table 6. Internal consistency coefficient of Cronbach alpha and stability for domains and total score

Domain	*Stability*	*Internal Consistency*
Academic Supervision	.928	.493**
Student Services	.928	.583**
Website	.913	.598**
Alumni Satisfaction	.943	.414
Activity	.116	.040
Lecturer Satisfaction	.963	.542**
Job Satisfaction	.907	.372*
Awards	.915	.172
The whole questionnaire	.845	

** Correlation is significant at the 0.01 level
* Correlation is significant at the 0.05 level

5 Results and Discussion

This section presents results obtained from the responses of the study sample members for paragraphs in each domain. The responses are first processed using statistical methods, and then they are processed using machine learning methods towards analyzing and interpreting the results.

5.1 First Research Question

What is the level of the academic quality at the FCSIT from the perspective of faculty members? To answer this question, means and standard deviations are extracted for the degree of FCSIT members' satisfaction level from the perspective of students and lecturers. This is illustrated in Table 7 for the eight domains.

Table 7 shows that the total score of the mean for all domains is (3.29), and the standard deviation is (1.06) with a medium score. The means for all domains range between (3.07-3.72). The second domain (student services) is ranked first with a mean of 3.72 and a high score, next came the rest of domains with a medium score, starting from the 6^{th} domain as ranked in the second place and ending at the 5^{th} domain in the final rank.

Table 7. Means and standard deviations of the degree of members' satisfaction level from the perspective of students and lecturers

Rank	*N*	*Domain*	*Mean*	*Std.*	*Degree (score)*
1	2	Student Services	3.72	0.76	High
2	6	Lecturer Satisfaction	3.34	1.35	Medium
3	3	Website	3.31	1.36	Medium
4	1	Academic Supervision	3.23	1.38	Medium
5	8	Awards	3.23	1.13	Medium
6	7	Job Satisfaction	3.16	1.06	Medium
7	4	Alumni Satisfaction	3.15	1.03	Medium
8	5	Activity	3.07	1.03	Medium
The whole questionnaire			3.29	1.06	Medium

5.2 Second Research Question

Is there a correlation of a statistical significance at the level between student satisfaction and lecturers' satisfaction from lecturers' perspective? Means and standard deviations are extracted for the degree of members' satisfaction level (e.g., student services, job satisfaction) from the perspective of lecturers. This is illustrated in Table 8 for the eight domains.

Table 8. Pearson correlation coefficient for the degree of members' satisfaction level from the perspective of lecturers

		Mean of paragraphs of the Job satisfaction domain
Mean of paragraphs of the student services domain	*Pearson Correlation*	.890**
	Sig.	.000

Table 8 shows that there is a positive statistical significance correlation between the job satisfaction and student services from the lecturers' perspective.

5.3 Third Research Question

Are there differences of a statistical significance at the level ($\alpha \leq 0.05$) between the means of respondents' answers (from the study sample) regarding their satisfaction level from students and lecturers perspective referring to the demographic variables (responder, task, rank)? Means and standard deviations are extracted for the degree of members' satisfaction level from the perspective of students and lecturers referring to the demographic variables (responder, task, rank). This is illustrated in Tables 9, 10.

Table 9. Means, standard deviations and t-tests based on RESPONDER variable of the degree of members' satisfaction level

Domain	*Responder*	*N*	*Mean*	*Std.*	*t-test*	*df*	*Sig,*
Academic Supervision	Student	243	3.061	1.028	50.756	326	.000
	Employee	84	3.004	1.094			
Student Services	Student	243	3.248	1.046	53.463	326	.000
	Employee	84	3.174	1.070			
Website	Student	243	2.866	.971	51.674	326	.000
	Employee	84	3.172	1.078			
Alumni Satisfaction	Student	243	3.048	.996	49.408	326	.000
	Employee	84	3.355	1.072			
Activity	Student	243	2.957	.948	51.726	326	.000
	Employee	84	3.264	1.028			
Lecturer Satisfaction	Student	243	3.091	1.062	53.300	326	.000
	Employee	84	3.140	1.227			
Job Satisfaction	Student	243	2.992	.964	50.686	326	.000
	Employee	84	3.214	1.079			
Awards	Student	243	3.385	1.230	50.127	326	.000
	Employee	84	3.263	1.196			
The whole questionnaire	Student	243	3.089	1.041	54.283	326	.000
	Employee	84	3.154	.994			

Table 10. Means, standard deviations and t-tests based on TASK variable of the degree of members' satisfaction level

Domain	*Task*	*N*	*Mean*	*Std.*	*t-test*	*df*	*Sig,*
Academic Supervision	Academic	36	3.699	.633	57.823	83	.000
	Administrator	48	3.854	.639			
Student Services	Academic	36	3.714	.809	48.148	83	.000
	Administrator	48	3.742	.635			
Website	Academic	36	3.422	.768	55.161	83	.000
	Administrator	48	3.270	.682			
Alumni Satisfaction	Academic	36	3.600	.578	69.277	83	.000
	Administrator	48	3.517	.707			
Activity	Academic	36	3.383	.669	59.331	83	.000
	Administrator	48	3.169	.673			
Lecturer Satisfaction	Academic	36	3.083	.764	47.785	83	.000
	Administrator	48	3.990	.795			
Job Satisfaction	Academic	36	3.638	.639	64.568	83	.000
	Administrator	48	3.424	.707			
Awards	Academic	36	3.752	.627	60.002	83	.000
	Administrator	48	3.870	.551			
The whole questionnaire	Academic	36	3.425	.453	88.004	83	.000
	Administrator	48	3.274	.474			

Table 9 shows that there are statistical significance differences at the level (α ≤ 0.05) referred to the responder variable for all domains and the total score. The differences are favored to employee in 5 domains out of 8. Table 10 shows that there are statistical significance differences at the level (α ≤ 0.05) referred to the responder variable for all domains and the total score. The differences are favored to both academics (4 domains out of 8) and administrators (4 domains out of 8); however, in the total score the differences are slightly directed toward the academic task.

Table 11. ANOVA based on RANK variable of the degree of members' satisfaction level

Domain	*Source*	*Sum square*	*df*	*Mean square*	*F*	*Sig.*
Academic Supervision	Between groups	1.193	2	.597	.545	.580
	Inside groups	328.140	324	1.095		
	Total	329.603	326			
Student Services	Between groups	2.027	2	1.013	.917	.401
	Inside groups	331.701	324	1.106		
	Total	333.728	326			
Website	Between groups	1.488	2	.744	.734	.481
	Inside groups	304.191	324	1.014		
	Total	305.679	326			
Alumni Satisfaction	Between groups	.276	2	.133	.117	.890
	Inside groups	342.478	324	1.142		
	Total	342.745	326			
Activity	Between groups	1.089	2	.905	.697	.499
	Inside groups	389.519	324	1.298		
	Total	391.328	326			
Lecturer Satisfaction	Between groups	1.452	2	.726	.675	.510
	Inside groups	322.476	324	1.075		
	Total	323.927	326			
Job Satisfaction	Between groups	1.015	2	.507	.476	.622
	Inside groups	320.026	324	1.067		
	Total	321.041	326			
Awards	Between groups	1.002	2	.680	.422	.911
	Inside groups	301.679	324	.834		
	Total	302.681	326			
The whole questionnaire	Between groups	1.677	2	.811	.990	.548
	Inside groups	330.634	324	.705		
	Total	332.311	326			

Table 11 shows that there are no statistically significant differences at the level (α ≤ 0.05) referring to the independent variable "rank" in all domains and the total score. That is all sig. values obtained by the ANOVA are not less than 0.05.

5.4 Analysis Using Machine Learning Methods

This subsection presents the flow of experiments which starts at some preliminary experiments and ends at building the suitable model of data analysis.

5.4.1 Experimental Settings

This research conducted preliminary experiments to determine the suitable data mining techniques for a proper data analysis. The selected techniques are implemented with their default parameter settings. Then, their parameter values are carefully tuned to obtain the desired results. Experiments are conducted on a Windows 10 machine with Core i5 processor and 8 GB of RAM. The following table shows the parameter settings for the *WekaDeeplearning4j* (M1), *GenClust++* (M2), and *Apriori* (M3).

Table 12. Parameter Settings for Weka models

Model	*Hidden Layers*	*Learning Rate*	*Momentum*	*Epoch*	*Optimization*	*Mean Decay*	*Test Mode*	*Time (s)*
M1	Attributes + Classes	0.3	0.2	100	Gradient Descent	0.9	Split 66%	300
Model	*Distance Function*	*Initialization*	*Clusters*	*Seed*	*Optimization*		*Cluster Mode*	
M2	Euclidean	Random	5	10	Genetic Algorithm	-	Split 66%	120
Model	*Delta*	*Metric*	*Minimum metric*	*Rules*				
M3	0.05	Confidence	0.9	10	-	-	-	30

5.4.2 Experimental Results

In all 8 datasets, the attribute values vary between numeric and nominal data types. Tables 13 and 14 summarize the obtained results from each model.

I. Classification and Regression Results

Classification is a supervised learning which provides the ability of predicting the category of a labeled data. Based on an independent variable, it determines the class of a dependent variable. Regression is also a supervised learning which predicts a numerical value based on previously observed data. For predicting two categories (*satisfied, not satisfied*), the following table shows the classification model (*WekaDeeplearning4j*) for each dataset. For each dataset, the classifier has obtained results in a matter of 2-3 seconds.

With a total number of 327 instances in each dataset, it is shown in table 13 that the classifier has obtained highly accurate classifications for all 8 datasets. It has obtained an accurate classification for both datasets (namely: job satisfaction and lecturer satisfaction) with 100% of correctly classified instances, a value of 1 for kappa, precision, recall and F-measure. On the other hand, a relatively lower accuracy of the classification is obtained for the student activity dataset with a 94.3% of the instances is correctly classified.

Table 13. Accuracy of the classification model (M1)

Dataset	Cross-validation	Correctly classified	Incorrectly classified	Kappa	Mean abs. err.	Relative error	Precision	Recall	F-measure
Academic Supervision	10 folds	96.4%	3.6%	0.93	0.029	8.75%	0.964	0.964	0.964
Student Activity	10 folds	94.3%	5.7%	0.91	0.036	9.47%	0.943	0.943	0.943
Alumni Satisfaction	10 folds	97.6%	2.4%	0.94	0.025	5.67%	0.977	0.977	0.977
Staff Awards	10 folds	99.6%	0.4%	0.99	0.007	18.45%	0.991	0.991	0.989
Job Satisfaction	10 folds	100%	0%	1	0.085	1.83%	1	1	1
Lecturer Satisfaction	10 folds	100%	0%	1	0.095	19.30%	1	1	1
Student Services	10 folds	98.7%	1.3%	0.97	0.025	17.15%	0.972	1	0.986
Website Evaluation	10 folds	98.2	1.8%	0.95	0.025	5.41%	0.944	0.985	0.964

Based on the outcomes of the classifier, it can be described that the members of the FCSIT are satisfied with the services, activities, and working environment. This helps bridging the gap between the conducted strategic planning in the FCSIT (presented by the SWOT analysis) and the actual outcomes from the last three academic years. That is to measure the success of the conducted strategy in empowering the faculty's members and identifying the strengths and weaknesses or opportunities and threats.

Although the classifier has obtained highly accurate results, it is speculated that some issues remained in bridging the gap between students' satisfaction and lecturers' satisfaction. It is seen that lecturers are satisfied with their job and duty, which is confirmed by a 100% accurate classification of their satisfaction. But relatively it is not the case of students. This may be since a variety of students' needs and/or expectations are diverse from year to year during their study. So, it is hard to meet their needs at the right time. Add to that, the reality of labor market is indeed out of Alumni's perspectives and expectations which lead to an unsatisfied responder.

II. Clustering Results

Clustering is an unsupervised learning which provides the ability of grouping data in order to find the frequent patterns from our dataset. With clustering there is no class attributes in the data. That is clustering helps us determining the class attributes from our dataset. The processing speed for the obtained results ranges between 0.01 seconds and 0.06 seconds for each dataset while performing 9-11 iterations. The sum of squared errors within clusters about 27.8. The following table shows the outcomes of the *GenClust++* algorithm applied to the 8 datasets.

Considering attributes values, the *GenClust++* showed the percentage of respondents who belong to their cluster based on their answers (data types or scales). Based on Table 14, all the 111 items with all 8 datasets are clustered into 2 different clusters. A dataset with 50% split of its 2 clusters is not statistically relevant where no clear conclusion can be drawn from the responders to their satisfaction. For example, in the Alumni dataset its 2 clusters are distributed as cluster 0 with 50% and cluster 1 with 50%. The rest of datasets are clustered into 2 variants ranges from 41% to 87% for cluster 0, and from 13% to 59% for cluster 1. In brief, a high level of satisfaction is clearly seen in lecturer and job datasets, while the rest of datasets are favored to a level of satisfaction among students except the alumni dataset. Overall, the classification accuracy of the data is 99.4%.

Table 14. Statistics, Cluster Centroid and clustered instances using (M2)

Dataset	Number of clusters	Number of features	Number of data objects (categories)	Incorrectly clustered instances	*Best*	*Average*	*Std.*	*t-test*	*p-value*
Academic Supervision	2	4	327 (109, 218)	2.17%	97.90	102.4	0.06	1.425	0.004
Student Activity	2	3	327 (72, 255)	18.21%	94.14	110.2	1.42	3.719	0.063
Alumni Satisfaction	2	3	327 (46, 281)	49.12%	97.12	128.5	14.28	7.211	0.024
Staff Awards	2	3	327 (187, 140)	12.55%	99.01	114.3	1.04	2.627	0.013
Job Satisfaction	2	4	327 (268, 59)	3.52%	99.19	100.1	0.00	-	-
Lecturer Satisfaction	2	4	327 (201, 126)	3.98%	99.02	100.2	0.00	-	-
Student Services	2	8	327 (190, 137)	14.81%	97.02	111.3	6.09	2.204	0.094
Website Evaluation	2	6	327 (197, 130)	34.22%	97.34	121.9	3.91	5.106	0.008

III. Association Rules Results

Finally, the Apriori algorithm is implemented for more detailed relationship between instances as to understand the responders' answers. For example, that is to identify the instances (e.g., services) that impact students' satisfaction. All 8 datasets combined have 111 attributes and 2616 (327 responders x 8 datasets) instances. Attributes are scaled (1-5) reflecting the satisfaction level, and the nominal class attribute indicates weather the responders are satisfied or not. So, useful patterns could be found that may help predicting the attribute. By default, the algorithm stops after 10 rules learned from each dataset. Therefore, a set of rules are produced with a minimum support of 55% of the instances and a confidence greater than 95%. Hence, the best rules found are:

```
1. service=5 supervision=5 269 ==>satisfaction=5 214
   <conf:(0.99)> lift:(1.07) lev:(0.04) [5] conv:(3.5)

2. services=1 activity=5 196 ==>satisfaction=1 194
   <conf:(0.98)> lift:(1.04) lev:(0.02) [3] conv:(1.91)

3. supervison=5 alumni=1 161 ==>satisfaction=3 154
   <conf:(0.98)> lift:(1.04) lev:(0.02) [3] conv:(1.81)

4. activity=1 alumni=1 109 ==>satisfaction=1 102
   <conf:(0.97)> lift:(1.03) lev:(0.02) [3] conv:(1.52)

5. supervision=3 lecturer=5 175 ==>satisfaction=5 171
   <conf:(0.97)> lift:(1.03) lev:(0.02) [2] conv:(1.42)

6. service=5 awards=3 300 ==>satisfaction=4 264
```

```
   <conf:(0.97)> lift:(1.05) lev:(0.03) [4] conv:(1.83)

7. supervision=1 services=3 195 ==>satisfaction=3 191
   <conf:(0.96)> lift:(1.04) lev:(0.02) [3] conv:(1.51)

8. lecturer=5 awards=3 79 ==>satisfaction=5 65
   <conf:(0.96)> lift:(1.04) lev:(0.02) [3] conv:(1.41)

9. website=1 services=5 289 ==>satisfaction=3 278
   <conf:(0.95)> lift:(1.01) lev:(0.01) [1] conv:(1.04)

10. job=5 lecturer=5 84 ==>satisfaction=5 82
    <conf:(0.95)> lift:(1.03) lev:(0.02) [2] conv:(1.31)
```

For example, the 10th rule, an employee is 95% satisfied if he/she is totally satisfied with their lecturer and job roles. This too applies to the 1st rule regarding students. Overall, in 5 rules, a high level of satisfaction for both students and employees, and in 3 rules they are neutral, while in 2 rules only students are not satisfied.

By doing so, the association rule implementation has automated the buildup of the TOWS matrix to identify the association between factors in four perspectives to make decisions for a soundful environmental examination of all FCSIT aspects. It has also reflected the importance of the TOWS matrix in mapping relations between pairs of factors that are based on crucial Strength and Opportunity coefficients.

Hence, the usefulness of machine learning methods in quality management is demonstrated in high manner and expectations. It has the privilege of automation over the traditional means used for measuring and analyzing the perceived models of quality management. It also has the privilege of smoothly grouping critical factors, accurately classifying and/or predicting patterns of paired factors, and associatively building a rule set of factors.

6 Conclusions

In the application at the FCSIT, the SWOT analysis was just one input to the planning process. An array of factors was generated which triggered a range of potential strategic initiatives. The high scoring factors had a bias towards opportunities and strengths, and the strategies proposed were also largely driven by those factors. Therefore, the FCSIT appeared to be pursuing a set of balanced strategies. The planning process itself yielded a balanced range of strategic initiatives covering most of the factors identified as being important, although a small number of factors needed further consideration. The strategies generated by the analysis were highly symmetrical with those in place, or subsequently adopted by the strategy committee. SWOT analysis is often presented as a method of rapidly moving towards an agreed strategy. It can certainly be an aid to generating new strategies, but a strategic development process also requires considerable analysis and testing of new strategies before adoption.

Hence, the required analysis and testing is conducted in this paper utilizing the strengths of 3 machine learning methods, namely, classification, clustering, and association rules. They proved to be handy in validating the conducted conventional SWOT analysis, and predicting its outcomes regarding FCSIT members. This is crucial in ensuring that significant weaknesses and threats are not overlooked, and that the potential of the FCSIT is fully realized.

Overall results, all 3 data mining tasks lead to the same conclusion from different perspectives. The classification task predicts the satisfaction level of FCSIT members, while clustering describes the common factors among FCSIT members by grouping them into areas of interests. The association rule task also describes the relationship between instances to further understanding of FCSIT members' needs.

The outcome of the study presents a high level of satisfaction regarding employee, and a vibrant level of satisfaction regarding students. Therefore, the developed quality assurance framework for the FCSIT is stable but needs more improvements to overcome the dissatisfaction of many students regarding services, supervision, awards and activities.

In future studies the author intends to assist the decision-making process of adopting appropriate strategies for the related factors, the author aims at testing and evaluating the iterative process of adopting strategies and models for perhaps a guaranteed successful implementation. This could be achieved by classifying a set of strategies that are most suitable for a certain factor in less time and effort.

References

[1] Sammut-Bonnici, T., Galea, D. (2015). SWOT analysis. Wiley Encyclopedia of Management, edited by Professor Sir Cary L. Cooper. John Wiley & Sons, Ltd.

[2] Habimana, T., Mutambuka, D., Habinshuti, P. (2018). The Contribution of SWOT Analysis in the Competitiveness of Business Enterprise in Rwanda. Journal of Economics, Business and Management, 6(2):56-60.

[3] Donaldson, S.I. (2019). Where do we stand? Recent AEA member views on professionalization. Journal of Evaluation and Program Planning, 72:152-161.

[4] Islam, M.S. (2019). An assessment of child protection in Bangladesh: How effective is NGO-led Child-Friendly Space? Journal of Evaluation and Program Planning, 72:8-15.

[5] Weihrich, H. (1993). Daimler-Benzs move towards the next century with the TOWS matrix. European Business Review, 95 (1): 4–11.

[6] Yildirim, O., Talo, M., Ay, B., Baloglu, U. B., Aydin, G. and Acharya, U. R. (2019). Automated detection of diabetic subject using pre-trained 2D-CNN models with frequency spectrum images extracted from heart rate signals, Computers in Biology and Medicine, Vol. 113: 103387.

[7] Narin, A., Kaya, C., Pamuk, Z. (2020) 'Automatic Detection of Coronavirus Disease (COVID-19) Using X-ray Images and Deep Convolutional Neural Networks', arXiv: 2003.10849.
[8] Cohen, J.P., Hashir, M., Brooks, R. and Bertrand, H. (2020) 'On the limits of cross-domain generalization in automated X-ray prediction', Journal. arXiv preprint arXiv:2002.02497
[9] Alhawamdeh, M., Jaradat, G., Alhawamdeh, H. (2020). Deep learning techniques to diagnose COVID-19 big data based on features extracted from CT and x-ray images. International Journal of Applied Engineering Research 15 (8):873-883.
[10] Lee, K. L., & Lin, S. C. (2008). A fuzzy quantified SWOT procedure for environmental evaluation of an international distribution center. Information Sciences, 178(2), 531-549.
[11] Gao, C.-Y. and Peng, D.-H. (2011) Consolidating SWOT Analysis with Nonhomogeneous Uncertain Preference Information. Knowledge-Based Systems, 24, 796-808.
[12] LeCun, Y., Bengio, Y. and Hinton, G. (2015) 'Deep learning. Nature. Vol. 521, No. 7553:pp. 436–444.
[13] Agarwal, C.C. (2015). Data Mining: the textbook. Springer Cham Heidelberg, ISBN 978-3-319-14142-8, DOI 10.1007/978-3-319-14142-8.
[14] Goodfellow, I., Bengio, Y. and Courville, A. (2016) 'Deep learning', Cambridge (Mass).
[15] Ning, C., & You, F. (2019). Optimization under uncertainty in the era of big data and deep learning: When machine learning meets mathematical programming. Computers & Chemical Engineering, (125), 434-448.
[16] Han, J., Kamber, M. (2000). Data Mining: Concepts and Techniques. Morgan Kaufmann. ISBN: 1-55860-489-8.
[17] Fayyad U. M., Piatetsky-Shapiro G., and Smyth P. 1996 "Knowledge Discovery and Data Mining: Towards a Unifying Framework", Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining. Portland, Oregon, AAAI Press, Menlo Park, California, pp. 82 - 88.
[18] Olaru, C., Geurts, P., & Wehenkel, L. (1999). Data mining tools and applications in power system engineering. In Proc. of PSCC99, 1999.
[19] Shaleena, K.P., Paul, S. (2015). Data mining techniques for predicting student performance, 2015 IEEE International Conference on Engineering and Technology (ICETECH), Coimbatore, 2015, pp. 1-3, doi: 10.1109/ICETECH.2015.7275025.
[20] Wiemer, H., Drowatzky, L., Ihlenfeldt, S. (2019). Data Mining Methodology for Engineering Applications (DMME)—A Holistic Extension to the CRISP-DM Model. Appl. Sci. 2019, 9, 2407; doi:10.3390/app9122407.
[21] Lang, S., Bravo-Marquez, F., Beckham, C., Hall, M. and Frank, E. (2019). WekaDeeplearning4j: a Deep Learning Package for Weka based on DeepLearning4j, In *Knowledge-Based Systems*, Volume 178, 15 August 2019, Pages 48-50. DOI: 10.1016/j.knosys.2019.04.013

[22] Islam, M. Z., Estivill-Castro, V., Rahman, M. A. and Bossomaier, T. (2018). Combining K-Means and a Genetic Algorithm through a Novel Arrangement of Genetic Operators for High Quality Clustering. Expert Systems with Applications.

Notes on contributors

Ahmad Abu-Al-Aish *is an Assistant Professor in the Department of Computer Science, Faculty of Computer Science and Information Technology, at Jerash University, Jordan. He received his PhD in Computer Science from Brunel University, United Kingdom in 2014. His Research interests including electronic and Mobile learning, Algorithms, Combinatorial Optimization and Educational Data Mining Techniques.*

Ghaith M. Jaradat *is an Associate Professor in the Department of Computer Science, Faculty of Computer Science and Informatics, at Amman Arab University, Jordan. His research interests are mainly directed to Metaheuristics and Combinatorial Optimization Problems including Course and Exam Timetabling, Vehicle Routing, Travelling Salesman, Knapsack, and Nurse Rostering Problems. He received his bachelor's degree from the Computer Science Department at Jerash University in 2004. He received his master's degree from Intelligent Systems Department at Utara University in Malaysia in 2007. He received his Ph.D. in Intelligent Research algorithms - Computer Science from the National University of Malaysia in 2012. He has published a number of high-quality research papers in international journals and conferences.*

Mahmoud Al-Shugran *obtained his Ph.D. degree from UUM (Universiti Utara Malaysia) in 2014. He is now working as an assistant professor at the Department of computer science/networking in the Faculty of Computer Science and Information Technology, Jerash University. He has more than 6 years of experience in teaching and researching, in the field of networking. His research interests include Wireless and Mobile Networks, routing protocols design, Internet of Things.*

Iyas Alodat *obtained his Ph.D in 2015. He is now working as an Assistant Professor at the Department of computer science/networking in the Faculty of Computer Science and Information Technology, Jerash University. He has 5 years of experience in teaching and researching, in the field of networking. His research interests include Mobile Networks, Cybersecurity, Internet of Things and Deep learning.*