INTERNATIONAL JOURNAL OF CIRCUITS, SYSTEMS AND SIGNAL PROCESSING
DOI: 10.46300/9106.2023.17.12 Volume 17, 2023

Development of Predictive Models for Quality Assurance of Local Higher Education Institutions

Sharmaine Justyne Ramos Maglapuz,
Department of Computing and Engineering – Pamantasan ng Cabuyao,
Cabuyao City
Philippines

Luisito Lolong Lacatan
Department of Computing and Engineering – Pamantasan ng Cabuyao
Cabuyao City
Philippines

Received: August 21, 2022. Revised: January 24, 2023. Accepted: February 23, 2023. Published: March 6, 2023.

Abstract— Quality Assurance in local higher education institutions (LHEIs) to determine its performance based on set standards is necessary as to ensure that quality education is enforced holistically. However, due to the myriad of services that the institution is providing, this task could often be overlooked. However, with the availability of Information Technology systems, and Mathematics, the regular evaluation of the LHEIs can be managed and monitored consistently. This paper discusses the development of a basic framework to allow LHEIs monitor their performance across ten (10) areas to determine quality assurance of services in an institution. This study combines the application of Data Mining Models as well as Statistical Methods to develop a Predictive Model to determine the quality assurance levels of a local higher education institution. Moreover, it provides a model in which the institution can look into in determining whether it provides quality service to its students. The developed model was tested for accuracy using existing historical data.

Keywords— Educational Data Mining, Prediction Model, Quality Assurance in Education, Statistical Models

I. Introduction

The rise of technology in the current century has been continuously helping to improve the simplest to the most complex of tasks that are necessary for one's life to flourish and reach its maximum potential. Joining together Science, Technology, Engineering and Mathematics is very critical as it brings all disciplines together, defying boundaries and discovering new methods of turning basic forms of sources into useful ones.

Technology and Mathematics are among two disciplines that have been useful in the field of Information Technology and Computer Science. Among its many branches is Data Mining (DM) which basically combines machine learning, statistics, and visualization techniques to discover and extract knowledge to discover relationships and patterns that exist within the data - that can be interpreted by humans easily. (3) One of its applications is Educational Data Mining (EDM), specifically determining and interpreting underlying information in the data produced in educational institutions.

Higher Educational Institutions (HEIs) are one of the organizations that can benefit from Educational Data Mining. These advanced methods of producing information from raw data could aid in the goals of these institutions by providing new means of accurate information from a statistical and technological perspective. According to the report of the Global University Network for Innovation (GUNI), the aim of HEIs is to produce quality graduates to drive local and national socio-economic development as well as to address global challenges. Quality Education has also been addressed in the United Nations Development Program as one of its Sustainable Development Goals.

To be able to deliver quality education and ensure that student and faculty needs are constantly met, it is the institution's duty

to be consistent in maintaining a holistic approach in the institution, specifically addressing the four pillars in an academic institution - Curriculum, Instruction, Research and Community Extension. Managing and monitoring a local institution with different working groups and processes is difficult, and with it being an important concern - the most important problem that the institution faces is the evaluation of the quality assurance process. To overcome this, a system can be made to automatically identify data, determine the right processes to break down the given set of data, and in return generate meaningful reports and provide intervention and suggestions to help with administrative decision-making.
The main objective of this paper is to create models to correlate the perception of the students of a local higher education institution with the institution's actual performance across ten (10) areas to provide Quality Service based from the Assessment Instrument of the Association of Local Colleges and Universities. The ten (10) areas of quality assurance in an local higher education institution are as follows: Government & Administration, Academic Qualifications & Teaching Experiences, Curriculum & Instruction, Student Development & Services, Entrepreneurship & Employability, Community Extension Services, Research, Library, Laboratories and lastly, Physical Plant. These 10 areas are a well rounded scope of all the services provided by the institution to ensure that quality education and services is provided to its clientele.

The use of a model represents knowledge in precise and practical ways in a particular field. Additionally, a model facilitates the process of comprehending a field of because the information is presented visually, it makes learning easier [8]. This paper applies correlational statistical methods and various data mining techniques such as Classification and Association Rule Mining to clean academic datasets to extract the patterns underlying and interpret the relationships between the parameters in the study.

The objectives of the study are the following:

1. Determine the degree of relationship between the perception of the students and the actual compliance of the institution with the areas of quality assurance.

2. Predict the compliance of the institution in the future assessments of the institution's quality service through curved fitting linear regression using historical data of the institution's evaluation to the compliance on the 10 standards on quality assurance.

3. Determine the accuracy of the developed predictive model

II. Background and Related work

Methods can provide an accurate means of viewing the correlation of data, any underlying factors to the relationship, and outliers.

In recent studies by Ashaduzzaman, students' data were evaluated and decision trees and associative rules were developed in order to gain a better understanding of the student's academic performance enabling the academic institution to make more informed decisions that will benefit students' future academic performance. Data mining methods such as Decision Tree, Naïve Bayes, and SVM were used to develop a prediction model that can suggest probable grades by analyzing parameters. Through this, key factors were identified that could gain insight into critical variables. Doing this would give a substantial impact on the outcome of the students by helping administrative personnel in making well-informed decisions in order to enhance student performance which will result in a decrease in student dropout rates. [4]

A study at how curriculum analytics was found to be of use to improve learning-centered curricula in a variety of higher education environments by Mansmann and Scholl (2007). Participating in evidence-based practices for curriculum evaluation and monitoring is critical to the success and sustainability of efforts to modify undergraduate and graduate programs. Emerging technology-enabled inquiry methodologies hold tremendous promise for informing evidence-based practice in complicated curriculum settings. For instance, curriculum analytics can be mined from various student learning systems and analyzed to support curriculum renewal initiatives and demonstrate impact at both the program and course level. Curriculum analytics can help develop an inquiry-based and scholarship-based culture of program improvement that is characterized by information sharing within and across disciplinary boundaries.

The study produces an innovative tool that leverages social network approaches to assess and visualize the integration and links between different courses that ultimately comprise a student's whole program of study. [5]

A method in which instructors use learning analytics to assess students' achievement of course objectives in order to facilitate student reflection, remediation, and faculty curriculum review was deemed effective in a study by Komenda, et.al (2015). The authors used learning analytics to improve advising, student reflection, remediation, and curriculum evaluation following the completion of a backward curriculum design process. To aid with the learning analytics process, the learning management system included a learning analytics application. The learning analytics were used throughout the academic year by the course teachers, student advisors, and students. Findings Instead of relying just on student grades and other proxy indicators of learning, the learning analytics application supplied real-time and direct data to various actors for advising, student reflection, student remediation, and course curriculum evaluation. It was, in their opinion, a worthwhile endeavor. It facilitated meaningful discussions about course learning objectives and gave precise information about each student. Additionally, the learning analytics technology provided specific insight into which areas of the program staff needed to enhance. [6]

III. Methodology

A. Pearson Correlation

The computational formula for the simple Pearson product-moment correlation coefficient between a variable labeled X and a variable labeled Y are as follows:

Where:
• *rxy* is the correlation coefficient between X and Y
• *n* is the size of the sample
• *X* is each individual's score on the X variable;
• *Y* is each individual's score on the Y variable;
•*XY* is the product of each X score times its corresponding Y score;
• X^2 is each individual's X score, squared; and
• Y^2 is each individual's Y score, squared.

In this study, Pearson Correlation was used to determine if there is a significant relationship between the evaluation/perception of the students on the institution's performance on the areas of quality assurance and the actual evaluation of the performance of the institution among the areas.

B. Curve Fitting or Linear Regression

Linear regression is a method of finding a linear correspondence between two data sets. The method is based on fitting a collection of data points by a simple linear form, $Y = A + BX$. The variable X is considered to be an explanatory variable, and the other, Y , is considered to be a dependent variable. Sometime, the variable A is called the intercept, and B is the slope of the line.

The method of curve fitting was used to determine the best line of fit to make prediction. The model was based from the computed r value and linear regression analysis of the 2018 student evaluation and quality assurance evaluation result.

IV. The Experiment

A. Data Sources and Data Set

The research heavily relies on 3 main data from a local higher education institution. The data was retrieved from a local higher education institution. Academic Performance of the Institution was evaluated starting from 1st Year Students of Semester of F.Y 2018-2019 to 1st Semester of F.Y 2021-2022 of the same students in their Fourth Year level, a total of 4 years worth of data. The actual institution's evaluation results of the two years (2018 & 2021) were gathered from the Quality Assurance Office of the institution and the results of the student surveys. The data were treated with ethical considerations and confidentiality and a model was only built to determine the effectiveness of the statistical and data mining method in the necessary process of data analysis. Due to the fact that the researcher needed real and actual data, policies were extremely stringent and ethical considerations were followed. Although the research is constrained by data security policies, the research lacks information about students' gender, age, and other sociodemographic characteristics.

Future analyses should include sociodemographic and socioeconomic data. However, the academic and administrative data gathered in this study is effective for analysing the insightful findings.

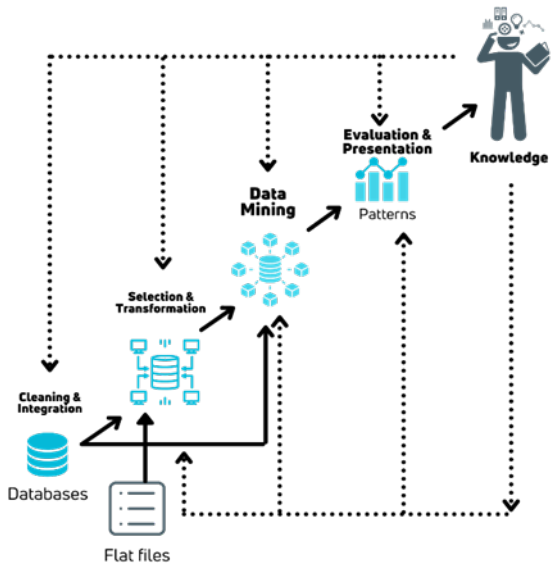

Figure 1. Knowledge Discovery in Databases

Discovery in Databases (KDD). This will be used to pre-process data for further analysis and model generation. The following are the essential processes in the data mining method:

B. Selection and Transformation

The data used mainly for this study are the institution's performance on the areas of quality assurance. Meanwhile, another set of data were gathered from the students to evaluate their perception on the performance of the institution on the areas of quality assurance. From this survey, the students were able to evaluate whether the institution meets the standards according to their own perception as the main clientele. The survey instrument was adapted from the ALCUCOA Accreditation Survey Tool. The internal consistency was established through a pilot testing to thirty (30) students and validated using Cronbach Alpha testing.

Table 1. Results of Instrument Reliability using Cronbach's Alpha

Quality Assurance Areas	Cronbach's Alpha	Reliability Level
1 – Government & Administration	.926	Excellent
2 – Academic Qualifications & Teaching Experiences	.844	Good
3 – Curriculum & Instruction	.940	Excellent
4 – Student Development & Services	.872	Good
5 – Entrepreneurship & Employability	.694	
6 – Community Extension Services	.897	Good
7 - Research	.859	Good
8- Library	.847	Good
9 - Laboratories	.943	Excellent
10 – Physical Plant	.924	Excellent

Legend: More than 0.90 - Excellent, 0.80-0.89 - Good, 0.70-0.79 Acceptable, 0.6-.69- Questionable, 0.5-0.59-Poor, Less than 0.59- Unacceptable

Cronbach's Alpha was used in validating the results of the usage of the instrument per area. Table 1 shows that the results in Areas 1,3,9 and 10 was Excellent and Areas 2,4,6,7 and 8 was Good. Further validating the use of the survey instrument in gathering the necessary data to identify the perception of the students on the compliance of the institution in the areas of quality assurance.

C. Data Pre-Processing

In order to process the huge chunk of data from the data sources, the datasets were organized. The necessary data were extracted and categorized. Missing values were identified and handled to avoid inconsistencies with the results. The dataset gathered were also divided into 80% training sets and 20% testing sets.

V. RESULTS & DISCUSSION

After the process of correlating the perception of the students on the evaluation of the institution's ten (10) areas of quality assurance, the model to predict the future results of the institution on the areas of quality assurance using the historical data in comparison with the present quality assurance results was developed.

5.1 Students' Perception on the areas of Quality Assurance

The first set of data needed to create the prediction model was processed to determine the actual perception of the students with the performance of the institution in the areas of quality assurance.

Table 2. Students' Perception of the Institution's Response to the areas of Quality Assurance on 2018

Area	Mean	Interpretation
1 – Government & Administration	2.42	High Extent
2 – Academic Qualifications & Teaching Experiences	1.97	High Extent
3 – Curriculum & Instruction	2.03	High Extent
4 – Student Development & Services	2.01	High Extent
5 – Entrepreneurship & Employability	2.17	High Extent
6 – Community Extension Services	2.14	High Extent
7 - Research	2.13	High Extent
8- Library	2.03	High Extent
9 - Laboratories	2.48	High Extent
10 – Physical Plant	2.42	High Extent

Legend: 5 - Excellent(E), 4 - Very Good (VG), 3- Good (G). 2- Fair(F), 1- Poor (P), 0- No Provision (NP)

Table 3 depicts the perception of the students in the year 2018 with regards to the response of the institution in the 10 areas using the adapted questionnaire. The results of the perception of the students whether the institution is compliant with all the ten areas turned out with positive results, with all of the mean equivalent to High Extent. This means that based on the perception of the students, the institution has a high extent with regard to the standards set in all the criteria of quality assurance.

5.2. Quality Assurance Historical Results

The second set of data to used to create the prediction model is the actual results of the institution during the evaluation of Quality Assurance last 2018. The data are as follows:

Table 3. Actual Results of Quality Assurance Evaluation during 2018

Program	***Grand Mean***	***Descriptive Evaluation***
Program 1	2.50	Fair
Program 2	2.51	Fair
Program 3	2.58	Fair
Program 4	2.66	Fair
Program 5	2.59	Fair
Program 6	2.93	Fair

Table 4 shows the Grand Mean of the actual results from the 2018 Accreditation Visit per program. All 6 programs were able to get a grand mean that has a descriptive evaluation of Fair.

After gathering the necessary data, the initial stages of developing the model started with identifying which area has a significant value from the Students' Perception and the historical results of the Accreditation evaluation to determine which areas had significant data to develop models from. In order to do this, Pearson Correlation method was used. The result of the correlation are detailed in table 4.

Table 4. Significant Relationship Between the Factors Associated to the Areas of Quality Assurance result 2018

Variable	r-value	p-value	Decision	Interpretation
Quality Assurance Results and Area 1 Evaluation	.094	.291	Failed to Reject Ho	Not Significant
Quality Assurance Results and Area 2 Evaluation	.013	.888	Failed to Reject Ho	Not Significant
Quality Assurance Results and Area 3 Evaluation	.085	.338	Failed to Reject Ho	Not Significant
Quality Assurance Results and Area 4 Evaluation	.126	.155	Failed to Reject Ho	Not Significant
Quality Assurance Results and Area 5 Evaluation	.185	.036	Reject Ho	Significant
Quality Assurance Results and Area 6 Evaluation	.146	.098	Failed to Reject Ho	Not Significant
Quality Assurance Results and Area 7 Evaluation	.174	.049	Reject Ho	Significant
Quality Assurance Results and Area 8 Evaluation	.841	.103	Failed to Reject Ho	Not Significant
Quality Assurance Results and Area 9 Evaluation	.243	.006	Reject Ho	Significant
Quality Assurance Results Area 10 Evaluation	.247	.005	Reject Ho	Significant
Quality Assurance Results and GWA	.152	.086	Failed to Reject Ho	Not Significant

From the student's evaluation of the school's compliance to accreditation areas, 4 were found to have a significant relationship to the actual quality assurance evaluation result. The areas that were found to be significant are Area 5 Employability & Entrepreneurship, Area 7 Research, Area 9 Laboratory, and Area 10 Physical Plant. This findings mean that these four areas will be used as predictors to the school's performance the next time the institution will be evaluated for quality assurance.

5.3. Prediction Model

Due to the results of the statistical analysis, a model was created to predict the outcomes of the quality assurance evaluation using the GWAs of the student, and their perceived evaluation results.

Regression models was then developed based from the curve fitting linear regression analysis to predict the results of the quality assurance evaluation using the attributes that were perceived to have a significant correlation with the actual results.

Table 5. Factors that Significantly Predict the Quality Assurance Results

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	.753	1.382		.544	.587
Area 5	1.113	.525	.185	2.120	.036
(Constant)	.590	1.368		.431	.667
Area 7	1.034	.520	.174	1.989	.049
(Constant)	1.808	1.523		1.187	.238
Area 9	1.632	.578	.243	2.821	.006
(Constant)	1.808	1.523		1.187	.238
Area 10	1.632	.578	.243	2.821	.006

Table 6 depicts the values derived from the curve fitting linear regression among the four (4) areas of quality assurance evaluation based on the previous statistical method. Area 5 has a value of 1.113, Area 7 has a value of 1.034, Area 9 has a value of 1.632 and lastly, Area 10 has a value of 1.632.

To develop the predictive model, the results of the 2018 perception were compared to the Historical data based on the Quality Assurance Evaluation last 2018.

Based on the factors that significantly predict the quality assurance evaluation, the linear models formed as per the constant and beta values are as follows:

Prediction from the Evaluation Results of Area 5

Formula:

$$y = 1.113a + .753$$

Where: y = Quality Assurance Evaluation Predicted Result

a = grand mean of Area 5 evaluation

Prediction from the Evaluation Results of Area 7

Formula:

$$y = 1.034b + .590$$

Where: y = Quality Assurance Evaluation Predicted Result

b = grand mean of Area 7 evaluation

Prediction from the Evaluation Results of Area 9:

Formula:

$$y = 1.113a + .753$$

Where: y = Quality Assurance Evaluation Predicted Result
a = grand mean of Area 9 evaluation

Prediction from the Evaluation Results of Area 10

Formula:

$$y = 1.448d + 1.353$$

Where: y = Quality Assurance Evaluation Predicted Result
d = grand mean of Area 10 evaluation

To use the predictive model, the y value will represent the actual predicted quality assurance evaluation results, and the value of a, b and c will be the grand mean of the area's evaluation based on the perception of the students on the compliance of the institution to areas of quality assurance.

The value of the grand mean of the area's evaluation will be extracted from the survey results of the students. If the model will be used in predicting the results of another quality assurance re-evaluation, a new survey shall be distributed to the students to provide accurate results for the future survey.

To determine the accuracy of the predictive model, the average prediction per area gathered across all four (4) models will be the projected quality assurance evaluation. In this study, the accuracy of the model is tested by comparing the results of the 2020 Quality Assurance Evaluation.

The results of the accuracy of the model are as follows:

Table 6. Factors that Significantly Predict the Quality Assurance Evaluation

Variables	P1	P2	P3	P4	P5	P6
Projected Results from the Evaluation of Area 5	2.664	2.608	2.964	2.005	2.608	2.522
Projected Results from the Evaluation of Area 7	2.100	2.238	2.509	1.702	2.169	2.276
Projected Results from the Evaluation of Area 9	4.468	4.419	4.928	3.848	4.884	4.595
Projected Results from the Evaluation of Area 10	3.795	3.968	4.575	3.179	3.870	3.986
Average Projected Results	3.257	3.308	3.744	2.684	3.383	3.345
Actual 2020 Results	3.860	3.860	3.590	3.900	3.900	3.760
N	20	30	25	8	33	13
Level of Accuracy	**84.37%**	**85.71 %**	**95.70%**	**68.81 %**	**86.74 %**	**88.96%**

As provided in Table 6, the accuracy of using the linear regression models produced the projected results of the accreditation with the following values:

The results of Program 1 based on the evaluation of Area 5 is 2.664, based on the evaluation of Area 7 is 2.100, based on the evaluation of Area 9 is 4.468, and based on the evaluation of Area 10 is 3.795. The average projected result for Program 1 is 3.257 which is 84.37% accurate as compared to the actual 2020 Accreditation results which is 3.860. For Program 2, the results of Area 5 is 2.608, based on the evaluation of Area 7 is 2.238, based on the evaluation of Area 9 is 4.419, based on the evaluation of Area 10 is 3.968. The average projected result for Program 2 is 3.308 which is 85.71% accurate as compared to the actual 2020 Accreditation results which is 3.860. The results for Program 3 shows that the results for Area 5 is 2.964, based on the evaluation of Area 7 is 2.509, based on the evaluation of Area 9 is 4.928, based on the evaluation of Area 10 is 4.575. The average projected result for Program 3 is 3.744 which is 95.70 accurate as compared to the actual 2020 Accreditation results which is 3.590. The results for Program 4 shows that the results for Area 5 is 2.004, based on the evaluation of Area 7 is 1.702, based on the evaluation of Area 9 is 3.848 and based on Area 10 is 3.179. The average projected result for Program 4 is 2.684 which is 68.81% accurate as compared to the actual 2020 Accreditation results which is 3.179. As for Program 6, the results for Area 5 is 2.608, based on the evaluation of Area 7 is 2.169, based on the evaluation of Area 9 is 4.884, and based on Area 10 is 3.870. The average projected result for Program 5 is 3.383 which is 86.74% accurate as compared to the actual 2020 Accreditation results which is 3.383. Lastly, the results of Program 6 based on the evaluation of Area 5 is 2.522, based on the evaluation of Area 7 is 2.276, based on the evaluation of Area 9 is 4.595 and based on the evaluation of Area 10 is 3.986. The average projected result of Program 6 is 3.345 which is 88.96% accurate as compared to the actual 2020 Accreditation results which is 3.345.

VI. CONCLUSIONS AND RECOMMENDATIONS

In this paper, models were developed to predict the overall results of a local higher education institution towards determining whether it is providing quality services to its clientele following the areas of quality assurance. This paper is multidisciplinary as it combines the data mining, as well as complex statistical methods to create a prediction model. The correlation between the perception of the students towards the institution's compliance to the 10 key areas of quality assurance, and from this process, it was known that four (4) areas – Area 5 Entrepreneurship & Employability, Area 7 Research, Area 9 Laboratories and Area 10 Physical Plant were highly significant to predict the future results of the institution's ability to deliver quality assurance. This discovery is an important step to determine which area to is beneficial in determining whether the institution is truly delivering quality services to the students and stakeholders through actual data driven evaluation tools.

However, it is important to note that the four (4) areas specified is only tailored fit to which institution the study was made. Once used in a different institution, the process of correlation shall be made to determine how the process could tailor fit other local higher institutions using the data that is provided by the study. The procedure done by this study serves as a framework that can be followed to predict quality assurance evaluation results in other local higher education institution since the local academe follows strict guidelines mandated by the government. This model can be used by other local institutions as well. The actual predictive models were developed through curve fitting linear regression. The accuracy of the results showed that all models fall under the Highly Effective percent range.

The model developed in this study can be of big help to local higher education institutions to determine fully if the institution provides quality services by reflecting on the data gathered from the students, as well as its past historical data.

This framework will be adapted to process academic data efficiently through a Decision Support System that may be used by school administrators and other institution heads to determine the performance of the school and have the ability to make data driven decisions.

REFERENCES

[1] M. E. Corbeil, j. R. Corbeil, and b. Khan, "A framework for identifying and analyzing major issues in implementing big data and data analytics in e-learning: introduction to special issue on big data and data analytics.," educ. Technol., vol. 57, no. 1, pp. 3–9, 2017.

[2] *Home*. NIU Newsroom. (2022, March 31). from https://newsroom.niu.edu/ Accessed September 18, 2022

[3] A. Aziz, N. H. Ismail, F. Ahmad, and H. Hassan, A framework for students' academic performance analysis using naïve bayes classifier, vol. 75. 2013.

[4] M. Ashaduzzaman, s. Zaman, and a. Pritom, An analysis on students' academic record using data mining techniques and identification of key factors to aid students' performance. 2020.

[5] S. Mansmann and m. H. Scholl, "Decision support system for managing educational capacity utilization," ieee trans. Educ., vol. 50, no. 2, pp. 143–150, 2007, doi: 10.1109/te.2007.893175.

[6] M. Komenda et al., "Curriculum mapping with academic analytics in medical and healthcare education," plos one, vol. 10, no. 12, pp. 1–18, 2015, doi: 10.1371/journal.pone.0143748.

[7] M. F. Musso, e. C. Cascallar, n. Bostani, and m. Crawford, "Identifying reliable predictors of educational outcomes through machine-learning predictive modeling ," *frontiers in education* , vol. 5. 2020, [online]. Available: https://www.frontiersin.org/articles/10.3389/feduc.2020.00104.

[8] A. . Salisu and e. Ransom, "The role of modeling towards impacting quality education," *int. Lett. Soc. Humanist. Sci.*, vol. 32, pp. 54–61, jun. 2014, doi: 10.18052/www.scipress.com/ilshs.32.54.

[9] O. Abu Khurma, Y. Alramamneh, T. Hassell, M. Hammadi, S. Zoelfakar, and Y. Fairouz, "Quality assurance model in enhancing the professional development content," *Int. J. Sci. Basic Appl. Res.*, vol. 54, pp. 49–68, Dec. 2020.

[10] G.-L. Curpănaru, "The model of quality assurance in education by implementing e-learning and blended learning," *Proc. 2021 Int. Conf. Mod. Manag. Educ. Res. (MMER 2021)*, vol. 577, no. Mmer, pp. 102–104, 2021, doi: 10.2991/assehr.k.210915.024.