Journal of Innovative Technology Convergence
Vol. 6, No. 2 June 2024, pp. 79-88
https://doi.org/10.69478/JITC2024v6n002a08

Cloud-Based Institutional Quality Assurance Model for a Local Higher Education Institution Using Data Mining Techniques

Sharmaine Justyne R. Maglapuz [1], Jenny Lyn V. Abamo [2]

Abstract: This study presents a Cloud-Based Institutional Quality Assurance Model for Local Higher Education Institutions (LHEIs). It utilizes advanced data mining techniques and custom data models that enhance the monitoring, assessment, and improvement of quality assurance processes through the utilization of cloud computing infrastructure in a developed system that offers scalability, flexibility, and accessibility, empowering policymakers to make data-driven decisions and to efficiently analyze vast amounts of data collected from various sources and be guided by algorithms. The findings advance quality assurance methodologies and offer valuable insights for practitioners, policymakers, and researchers in higher education.

Keywords: Data-Driven Decision Making, Data Models, Data Mining, Institutional Quality Assurance

1. Introduction

Academic data processing and analysis have been greatly impacted by the use of technology in education. Regular operations at educational establishments produce useful data for data-driven decision-making, which supports the quality of instructions. Decision-makers in higher education use techniques like data mining, analytics, and social network analysis to support difficult decisions that impact several stakeholders [1]. Finding patterns in huge datasets using database administration, artificial intelligence, and statistics is known as data mining, or knowledge discovery. The majority of these current information analysis tools are based on data mining, statistics, and social network analysis. They are focused on helping the general public, such as students and teachers [2].

Large-scale educational data is analyzed in the developing discipline of educational data mining (EDM) to gain a deeper understanding of learners and learning environments. Personalized instruction, enhanced programming, student retention, academic results, evidence-based decision-making, and tactical reactions to international trends are some of the advantages of EDM. In order to improve services

[1] School of Graduate Studies, AMA Computer University, Quezon City, Philippines
Email: sjmaglapuz@gmail.com
[2] School of Graduate Studies, AMA Computer University, Quezon City, Philippines
Email: jlvabamo@amaes.edu.ph

Received [December 14, 2023]; Revised [March 12, 2024]; Accepted [May 26, 2024]

Published by InnoCon Publishing
ISSN 2704-4440

and obtain insights, universities need to combine data mining techniques with operational data. Achieving these outcomes requires that the universities move beyond their operational data and integrate their operational data with various data mining methods to deeply understand trends and further improve their services to students [3].

In order to improve educational quality, quality assurance (QA) in education conducts systematic reviews of educational provisions. It consists of assessments of educators, school administrators, and students, in addition to external and internal evaluations. For the institution to support accessible, high-quality education, a strong quality assurance system is necessary.

Researchers and educational stakeholders concentrate on data-driven education development through learning analytics in the era of big data. Significant progress has been achieved in fields like self-regulated learning analytics, learner interaction visualization, and collaborative learning. Digital trails that support new research subjects, collaborative groups, and transdisciplinary capacity-building activities have been investigated, ranging from individual courses to large-scale academic analytics [4].

This study focuses on the Educational Quality Assurance (EQA) of an LHEI in Laguna (Philippines) by integrating IT concepts such as system development, cloud based technology, data mining, and algorithms to create a comprehensive system that addresses the need for data-driven decision-making.

The objectives of the study are as follows:

1. To apply data models to predict the outcomes of Quality Assessment of a Local Higher Education Institution;
2. To develop a student and faculty clustering and evaluation model for the Quality Assurance Model using Classification and Regression Tree (CART) algorithm;
3. To measure the performance of the student evaluation model using Accuracy, Precision, Recall, and F-score; and
4. To apply the prediction data models and learning clustering model to a Quality Assurance System.

2. Related Literature

The need for the application of machine learning on the educational frontier has become crucial. Most educational administrators and researchers are using various machine-learning methods to improve students' retention, predict students' performance, mitigate students' dropout rates, and position them to make better decisions in curriculum design and admission policy in their learning institutions.

Decision support systems, or DSS, are essential for effective leadership and operations in higher education institutions. Decision-making is aided by DSS in the domains of academics, administration, and management. A study examined the various applications of DSS and emphasized its importance in higher education. Students, faculty, administrative staff, and server and client end users were surveyed. Results show that DSS improves data analysis, decreases manual labor, expedites decision-making, and boosts output. Furthermore, DSS increases revenue and profitability in addition to improving customer and staff satisfaction [5].

Outcome-based academic programs are centered around Program Educational Objectives (PEOs) and Student Outcomes (SOs). The mapping and correlation between PEOs and SOs are crucial for program success and effective decision-making. This paper proposes a data mining-based approach to uncover hidden knowledge about PEOs, SOs, and their correlations in engineering programs. Using association rule mining techniques, the approach generates and analyzes association rules from PEOs-SOs mapping data, helping to discover valuable insights [2].

In summary, for the education sector to advance in terms of technological aspects, machine learning and decision support systems (DSS) may be integrated. Academicians and administrators are starting to apply machine learning techniques more often to improve student retention, forecast student performance, lower dropout rates, and guide admissions and curriculum design.

DSS also greatly assists decision-making in academics, administration, and management and is essential for efficient leadership and operations in higher education institutions. Research indicates that DSS enhances data analysis, lessens manual labor, and speeds up decision-making, resulting in more output, more income, and more satisfied staff and students. For outcome-based academic programs to be successful, it is also essential to comprehend the mapping and relationships between PEOs and SOs. By using data mining methods like association rule mining, programs can become more effective since important insights into these relationships can be found. When taken as a whole, these technologies highlight the revolutionary potential of data-driven strategies for reshaping education [6].

3. Methodology

3.1 Software Design

The software design of the Quality Assurance Framework is built on several layers, each of which displays the report visualization, report generation, data consolidation through tables, and expert support systems in both tabular and textual form. To determine these, institutional needs were considered when the software was designed. In selecting the platform to be used for the study, cloud computing was considered to ensure that the system's functionalities are available and reliable 24/7 without compromising the ability to deliver real-time data.

The proposed system aims to provide an institutional quality assurance framework by developing a Cloud-based system that would process data and provide the institution with evaluations, results, and interventional actions to serve as the platform to monitor academic quality.

Specifically, the system will cater to the administrative officials of the local institutions. The users will primarily be the institution's president, vice presidents and all members that are involved in decision-making.

3.2 Algorithms Used

The choice of data mining technique hinges on the application domain and data characteristics, guiding decision-makers toward appropriate methods. Techniques fall into two categories: descriptive, revealing interpretable patterns, and predictive, projecting future values. This study employs three techniques on given datasets: Association Rule for pattern identification, employing Correlational Analysis to gauge the strength of relationships in assessing academic quality by using two correlational statistical methods, Pearson and Spearman Rank Correlation, will be used to further analyze dataset patterns; Clustering Method for grouping similar data, utilizing the CART Algorithm; and Classification Data Mining, employing a decision tree approach to recommend supplementary actions for academic improvement.

3.2.1 Predictive Model for Quality Assurance

The developed system allows the prediction and analysis of the institution's quality assurance levels, which are measured through accreditation standard metrics. In a previous study by the author, a predictive model was developed that predicts the compliance of an institution and whether it would be able to meet the specific standards set by any accrediting body. It uses the correlation between students' perceptions of the services provided by the institution based on a set of criteria and the historical results of the institution. A regression model was developed based on the correlation of the results and is provided in this study [7].

3.2.2. Classification and Regression Tree (CART) Algorithm

This method will be used to cluster the students based on their academic performances into four groups according to their Grade Weighted Average (GWA). For both classification and regression issues, the Classification and Regression Trees (CART) algorithm is an effective machine learning tool. Its decision tree form is adaptable to a wide range of applications since it can handle continuous and categorical data with efficiency. Because of CART's flexibility and interpretability, it is usually recommended for educational datasets [8]. It helps in comprehending the fundamental reasons for performance variations by offering a clear visual representation of variable interactions. This clarity is essential for developing targeted interventions and improvement strategies in educational settings [9][10].

3.3 Conceptual Framework

Figure 1 illustrates the conceptual framework of the cloud-based quality assurance system, highlighting the integration of academic data, predictive analytics, and an Agile development approach to enhance decision-making and continuous improvement in the LHEIs.

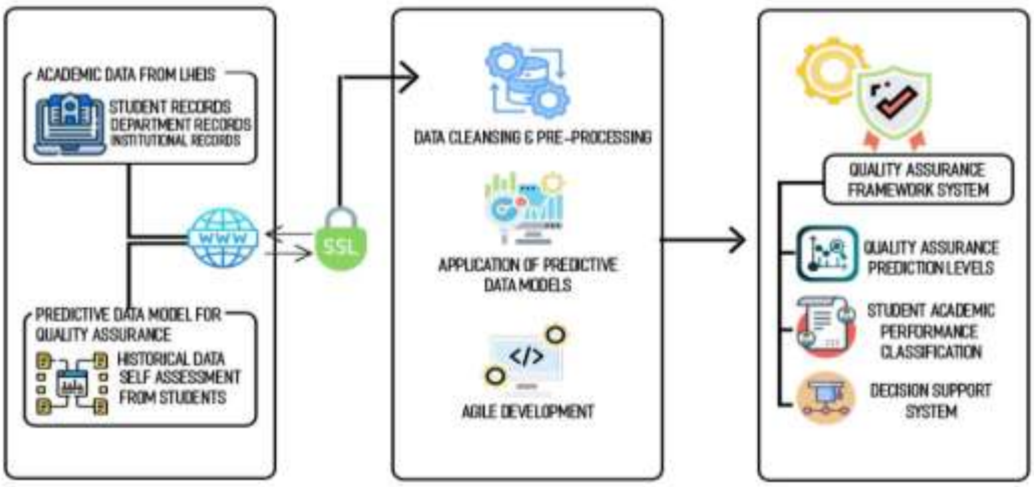

Figure 1. Conceptual Framework of a Cloud-Based Quality Assurance System for LHEIs

Academic information from LHEIs, including records for students, departments, and institutions, is integrated into the cloud-based quality assurance system. To guarantee accuracy and dependability, this extensive data is put through a thorough cleansing and pre-processing procedure. Advanced predictive models are loaded with student self-assessment data and historical data. By forecasting trends and identifying areas for improvement, these models examine the data to provide proactive quality assurance and data-driven decision-making.

The system was developed utilizing the Agile methodology, which places a strong emphasis on collaborative and iterative procedures to ensure constant improvement based on changing needs and continuing feedback. This strategy ensures that the system can adjust to evolving conditions in the field of education. The system produces a number of outputs, such as student performance classifications, prediction levels for quality assurance, and tools for decision support. These characteristics enable stakeholders to properly oversee quality assurance and encourage ongoing improvements in institutional dynamics and academic performance.

3.4 Evaluation of Data

The primary respondents of the study are its intended main users – the administrators and the academic heads of the institution: one university president, three vice presidents, six college deans, fourteen program chairs, and six information technology (IT) practitioners. A total of 31 respondents. The respondents were selected because they are actively engaged in managing and making academic decisions, especially during the accreditation period, during which the quality assurance of the college is assessed. It could also help their decision-making in determining which students to focus on and what programs to develop to support all the clusters of students generated by the reports of the system. The administrative officials will primarily benefit the system, as critical decision-making points will be backed up by the data provided by the quality assurance framework system.

Additionally, seven IT experts are included among the respondents to serve as guides in the technical aspects of the system, ensuring that all insights are properly discussed in the evaluation of the system. To evaluate the performance of the system, system criteria based on the software quality model of ISO 25010:2011 will be used, such as Functional Suitability, Performance Efficiency, Usability, Reliability, Security, and Maintainability.

Lastly, to determine the performance of the CART algorithm for clustering data and classifying the outcomes of the data, this study will use binary classifiers to classify instances as positive or negative. Positive instances are classified as a member of the class the classifier is trying to identify, and on the other hand, negative instances are classified as not being a member of the class being tried to identify. The basis of the evaluation model Precision, Recall, and F-1 score comes from the concepts of True Positive, True Negative, False Positive, and False Negative.

4. Results and Discussion

The outcomes of using data mining techniques on the cloud-based institutional quality assurance model created for this study are shown in this section. The importance of patterns and trends for quality improvement initiatives is clarified through analysis, offering a comprehensive grasp of the effectiveness of the model. The results are evaluated critically, with a focus on their significance, limitations, and relevance to the field of quality assurance in higher education.

With the results of the statistical analysis, a model was created to predict the outcomes of the quality assurance evaluation using the GWAs of the students and their perception evaluation results.

Regression models were then developed based on 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. Regression models were then developed based on 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.

In order to develop the student evaluation model and faculty assurance capabilities of the study, the researcher developed a model applying the CART Decision Tree Model, as shown in the visualization in Figure 2.

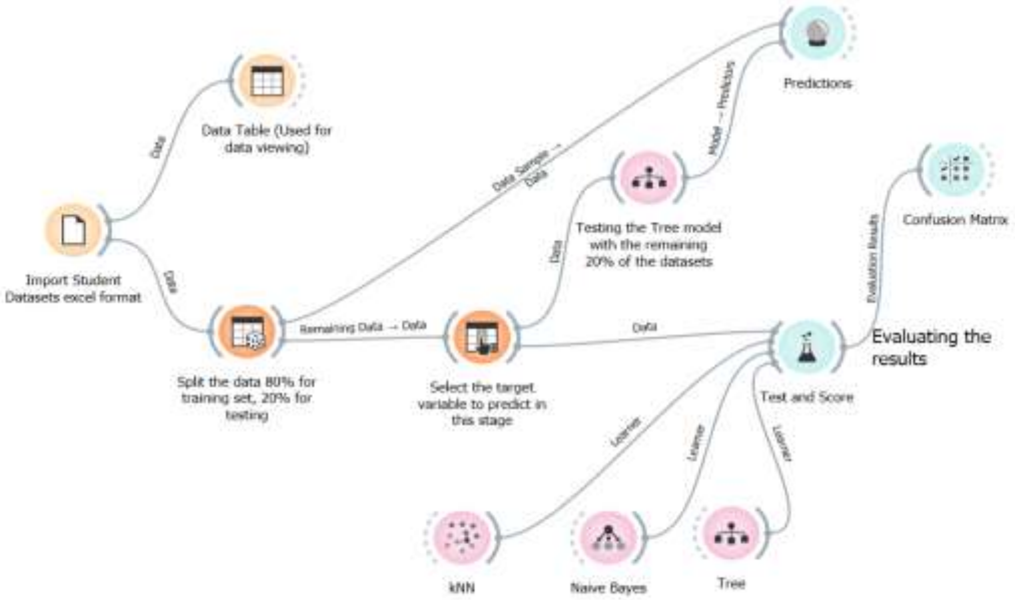

Figure 2. Student and Faculty Clustering Model using Cart Decision Tree

In order to develop the student evaluation model and faculty assurance capabilities of the study, the researcher developed a model applying the CART Decision Tree Model. The process starts by importing the raw academic data of the students as described in the previous portions of the study for pre-processing. For this specific process, open-source software called Orange Data Mining was used, from the cleansing of data up to generating the confusion matrix, to determine which algorithm suits the academic clustering processes. Undergoing the process of Data Cleansing, data redundancies were removed to ensure that there were no similar instances in the distribution of data in the testing and training of data. After this, the data were then split into the following: an 80% training set and a 20% testing set, using the process of stratified sampling to ensure the class distribution was preserved in both sets.

To determine the best algorithm that would match the needs of the system, three algorithms were applied in the testing: KNN (K-Nearest Neighbor), Naïve Bayes Algorithm, and the CART Decision Tree Model. These algorithms were chosen based on the related literature gathered in the study, and the majority of the research has mentioned that these three decision tree algorithms work well with academic performances and are appropriate for categorization tasks.

From this comparison of the performances of these algorithms, CART showed a significant advantage given the necessity of its use in the data set, with the highest F1 score of 86%.

Following this simple yet direct approach to identifying clusters based on the data of the students and faculty members, the system was able to provide a reliable and scalable module based on the highest performing algorithm.

To measure the performance of the student and faculty clustering evaluation model using Accuracy, Precision, Recall, and F1 score, the three algorithms compared were CART, kNn, and Naïve Bayes. In order to determine the best-fit algorithm for the specific uses of decision trees in this data, the models were tested based on their Accuracy, Precision, Recall, and F1 score. The findings of this testing are depicted in Table 1.

Table 1. Summary of Evaluation

Model	Accuracy	Precision	Recall	F1 Score
CART	0.904	0.854	0.875	0.863
kNn	0.908	0.811	0.900	0.900
Naïve Bayes	0.898	0.911	0.750	0.750

The accuracy of CART, kNn, and Naïve Bayes is 90.4%, 90.8%, and 89.8%, respectively. CART outperformed the other two algorithms in this category, which indicates that it is more effective in correctly classifying instances within the dataset used in the study. With the precision derived from the confusion matrix, CART demonstrated a precision score of 85.4%, which indicates a high proportion of correctly identified positive cases; Naïve Bayes got a precision of 81.1%, and lastly, Naïve Bayes outperformed the other two algorithms by having a precision of 91.1%. Despite Naïve Bayes getting high performance, the precision offered by CART is competitive and balanced.

In terms of Recall, CART also performed well with 87.5%, several points higher than Naïve Bayes with 75%. KNn also provides a high recall of 90%. Lastly, the F1 score, which combines precision and recall in a single metric, attained by CART was a solid 86.3%, while kNn and Naïve Bayes got 90% and 75%, respectively.

In summary, with CART's performance in the confusion matrix composed of Accuracy, Precision, Recall and F1 score, along with the related studies found while conducting this research, particularly looking for the best choice of algorithm for data sets in academic performances, the researchers ought to use this particular algorithm.

Since the use of a reliable algorithm plays a huge part in developing the system, its selection is crucial in this study. With the performance of the CART algorithm, the processes that are required for the specific use case of the study – in this case, clustering the grades of the students and assigning them the appropriate intervention or rewards, similar to the processes done in clustering the faculty performances – requires a lot of simplicity and versatility in the process. This system is intended to fit local higher education institutions; the processes may seem simple and direct, but in retrospect, the scale of data would be more complicated and flexible with the kind of policies that are implemented over time. As a fact, these policies may change and may differ per institution. Using a decision tree-based algorithm would be more efficient and direct, and it would avoid having biases as this type of model is non-parametric and does not make assumptions about the underlying distribution of the data or the relationship between features and the target variable. To address the need for the system to be scalable and to adapt to the expected growth of data per semester, the divide and conquer approach of CART allows it to efficiently process large amounts of data and construct decision trees with minimal computational overhead.

Decision trees using different approaches and under different models are available for use. However, upon close examination and considering the use of the confusion matrix, the related studies gathered in this research, and the intricacies applied when using the CART model, it was determined to best fit the needs of this study.

After the appropriate algorithms are identified and the data models are tested, these are applied to the developed Quality Assurance System (QAS). In order to achieve this objective, the researchers improved the QAS of an LHEI by using the CART algorithm and developing a predictive data model using regression models custom-fit to the existing records of the institution. The predictive data model was developed to forecast quality-related problems, specifically to determine whether the institution was qualified in terms of standards set by the Commission on Education.

The developed system includes features like student and faculty performance clustering and intervention for efficient decision-making. Furthermore, a predictive model using the institution's historical results and students' perceptions of accreditation was utilized to enhance the system's predictions that are proven accurate to determine the institution's accreditation passing rate. The predictive model and the CART algorithm were both applied in the developed system as a result of a credible systematic study, providing policymakers at the institution with the opportunity to make informed decisions for their institution. A sample screenshot of the interface is provided in Figure 3.

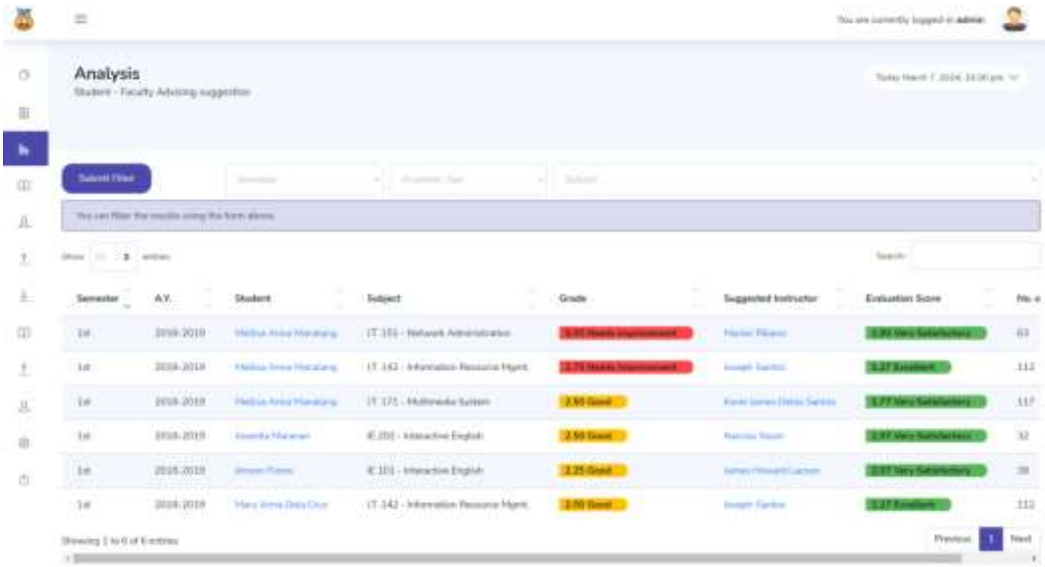

Figure 3. Student and Faculty Clustering and Suggestion Module Interface

Table 2. System Evaluation on the Quality Assurance System

Statements	Weighted Mean	Verbal Interpretation
Functional Suitability	4.80	Strongly Agree
Performance Efficiency	4.60	Strongly Agree
Usability	4.80	Strongly Agree
Reliability	4.40	Strongly Agree
Security	4.73	Strongly Agree
Maintainability	4.73	Strongly Agree
Total	4.67	Strongly Agree

After the development of the system was accomplished, it was evaluated by members of the academic community of local colleges and universities in the region of Laguna. The overall results of the evaluation of the system are depicted in Table 2.

All in all, the performance of the Quality Assurance System in terms of the criteria used, such as Functional Suitability, Performance Efficiency, Usability, Reliability, Security, and Maintainability, had an evaluation that had a verbal interpretation of Strongly Agree. Its overall weighted mean is 4.67, which proves that the system was able to perform all functions required to be accomplished based on its goals and objectives.

5. Conclusion and Recommendations

Summing up the main conclusions on cloud-based institutional quality assurance systems, this section integrates ideas from theoretical models and the actual findings of the research. Additionally, it benefits policymakers and upcoming researchers by offering feasible recommendations for utilizing cloud technology to improve quality assurance in higher education institutions.

Using an extensive approach, the clustering and predictive capabilities of the CART algorithm and the generated predictive data model were applied to the Quality Assurance System. By utilizing the CART algorithm, data related to both students and faculty could be efficiently clustered, enabling targeted measures and interventions. Similar to previous studies, decision support systems are utilized to help support LHEIs in decision-making through quantitative results measured by information within the organization [11][12].

In addition to this, a prediction model was used to predict institutional compliance with accreditation body standards. The system was able to evaluate an institution's compliance with accrediting standards due to this model, which was built on carefully established predictions from thorough research. Strong predictive capabilities were built by seamlessly integrating these approaches into the Quality Assurance System, enabling stakeholders to make well-informed decisions and lead proactive quality improvement toward the benefit of local higher education institutions.

Finally, through this study, the researcher was able to develop a Quality Assurance System for Local Higher Education Institutions using the student and faculty clustering model and prediction for quality assurance, allowing the users of the system to access its capabilities and maximize their data. Since the target users are decision-makers of local higher education institutions, they will be guided in making decisions for the college or university. The system was evaluated by different academic experts across the districts of the province of Laguna.

In conclusion, the researcher was able to develop a Quality Assurance System for Local Higher Education Instructions by integrating prediction models and algorithms to perform and process data that may be used by decision-makers in an academic organization to predict the quality of the institution as perceived by its stakeholders following accreditation standards and a system module that produces reports based on academic data of the institution. This further helps decision-makers determine the current performance of its students and faculty members. The system was evaluated by academic experts across Laguna.

After undergoing topic discovery, conceptualization, and study proper, the researchers have determined the future courses of action that can be taken to enhance the study. The prediction model for the accreditation of the institution can be added with varying factors such as faculty perception, the number of attempts per level of accreditation, the number of transitions or changes in leadership in the academic institution, and the population of the institution. To make the study more accurate, these may be considered and determined to be relevant to the prediction of the results.

An additional recommendation for the study is the integration of the accreditation instruments used by private higher education institutions. Since the accrediting bodies are all following CHED minimum requirements, they undeniably contain similar criteria, which can be seen as an opportunity to develop a model that can be used by these institutions, which eventually will allow the system to be used by other institutions regardless of their classification. Since the goal of the system is to make data analytics more accessible to educational institutions, this would expand the capabilities of the system, introduce data-driven decision-making to academic institutions and to its policymakers, leverage its use in this field, and further help institutions achieve a higher level of quality academic and non-academic services.

Lastly, to make the study efficient and usable by any type of local college or university, the integration of determining the predictors based on the framework used by the study as a function of the system could be added to ensure that its users would be able to perform this based on their institutional preferences, school history, and processes.

References

[1] A. Poth, Q. Beck, A. Riel, "*Artificial Intelligence Helps Making Quality Assurance Processes Leaner*," in *Communications in Computer and Information Science*, vol. 1060, A. Walker, R. O'Connor, R. Messnarz, (Eds.), Systems, Software and Services Process Improvement (EuroSPI 2019), August 2019, pp. 722-730, https://doi.org/10.1007/978-3-030-28005-5_56

[2] R. Shalabi, "*The Importance and Applications of Decision Support Systems (DSS) in Higher Education*," https://doi.org/10.6084/m9.figshare.12465599.v1.

[3] A. A. Yahya, A. Osman, "*A Data-Mining-Based Approach to Informed Decision-Making in Engineering Education*," Computer Applications in Engineering Education, vol. 27, no. 6, September 2019, pp. 1402-1418, https://doi.org/10.1002/cae.22158.

[4] Y. Nieto, V. García-Díaz, C. Montenegro, R. G. Crespo, "*Supporting Academic Decision Making at Higher Educational Institutions Using Machine Learning-Based Algorithms*," Soft Computing, vol. 23, June 2019, pp. 4145-4153, https://doi.org/10.1007/s00500-018-3064-6.

[5] V. P. Bresfelean, N. Ghisoiu, R. Lacurezeanu, D. A. Sitar-Taut, "*Towards the Development of Decision Support in Academic Environments*," in Proc. of the ITI 2009 31st International Conference on Information Technology Interfaces, June 22-25, 2009, Cavtat, Croatia, https://doi.org/10.1109/ITI.2009.5196106.

[6] A. Kayanda, L. Busagala, M. Tedre, "*User Perceptions on the Use of Academic Information Systems for Decision Making Support in the Context of Tanzanian Higher Education*," International Journal of Education and Development Using Information and Communication Technology, vol. 16, no. 1, 2020, pp. 72–87, ISSN: 1814-0556.

[7] S. Maglapuz, L. Lacatan, "*Development of Predictive Models for Quality Assurance of Local Higher Education Institutions*," International Journal of Circuits, Systems and Signal Processing, vol. 17, March 2023, pp. 100-106, https://doi.org/10.46300/9106.2023.17.12.

[8] L. Al Hallak, A. Pakštas, P. Oriogun, D. Novaković, "*Decision Support Systems for University Management Processes: An Approach towards Dynamic Simulation Model*," in Prof. of 2009 International Conference on Computer and Electrical Engineering, December 28-30, 2009, Dubai, United Arab Emirates, https://doi.org/10.1109/ICCEE.2009.264.

[9] M. Krishna, B. S. B. P. Rani, G. K. Chakravarthi, B. Madhavrao, S. M. B. Chowdary, "*Predicting Student Performance using Classification and Regression Trees Algorithm*," International Journal of Innovative Technology and Exploring Engineering, vol. 9, no. 3, January 2020, pp. 3349-3356, https://doi.org/10.35940/ijitee.C8964.019320.

[10] S. N. M. Nawai, S. Saharan, N. A. Hamzah, "*An Analysis of Students' Performance Using CART Approach*," AIP Conference Proceedings, vol. 2355, no. 1, May 2021, https://doi.org/10.1063/5.0053388.

[11] L. Surya, "*Machine Learning-Future of Quality Assurance*," International Journal of Emerging Technologies and Innovative Research, vol. 6, no. 5, May 2019, pp. 1078-1082, ISSN:2349-5162.

[12] J. Nouri, M. Saqr, U. Fors, "*Predicting Performance of Students in a Flipped Classroom Using Machine Learning: Towards Automated Data-driven Formative Feedback*," Journal of Systemics, Cybernetics and Informatics, vol. 17, no. 2, August 2019, pp. 17-21, ISSN: 1690-4524.